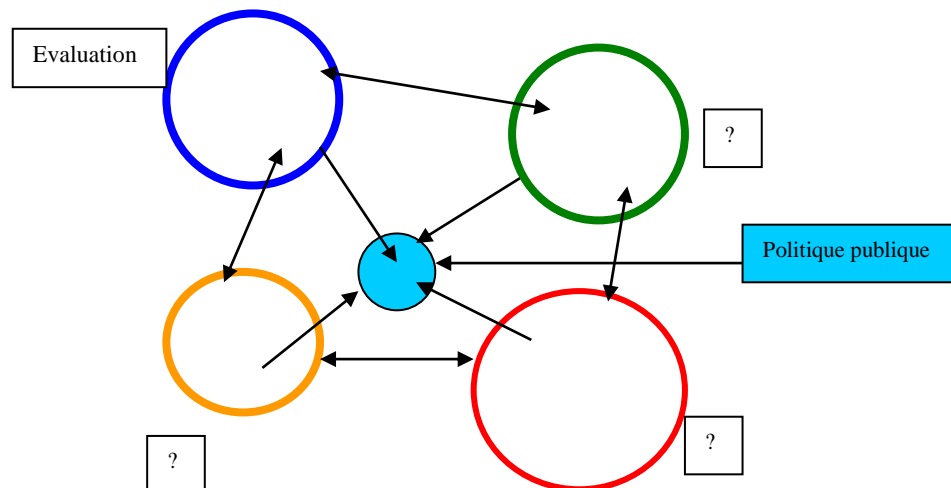


## A QUOI ET A QUI SERVENT LES EVALUATIONS ?

Une recherche sur la place des évaluations dans les processus de décision et  
d'apprentissage dans l'administration fédérale



### Jury

Prof. Katia HORBER PAPA ZIAN, rapporteur  
Prof. Luzius MADER, co-rapporteur  
Prof. Jean-Daniel DELLEY, expert

## RESUME

La présente étude donne des informations nouvelles et précieuses à la fois sur le développement de la pratique de l'évaluation en Suisse et sur les effets des évaluations effectuées dans le cadre de l'administration. Centrée sur la question de l'utilisation des évaluations, elle tente de définir les critères causant un usage direct ou une influence indirecte des évaluations et de pondérer entre eux ces deux types d'utilisation en fonction des publics utilisateurs.

Partant d'une base empirique de plus de 600 titres issus de l'administration fédérale, nous montrons que la moitié environ des études d'évaluation ont eu une forte influence et un usage direct dans le suivi de la politique en question.

Nous interrogeant ensuite sur les causes qui expliquent cette utilisation, nous soulignons l'importance de la participation aussi bien du Parlement que des responsables de l'administration, relativisant au passage le rôle présumé négatif des conflits. L'utilisation est liée à des facteurs plus importants que des éléments de type technique (longueur, simplicité...), comme la fonction dévolue à l'évaluation, le type de politique évaluée et le contenu. Le rôle de l'évaluateur et d'un bon système de valorisation ressortent également clairement des résultats de l'enquête, incluant une médiatisation des résultats du rapport.

En affinant et complétant notre technique de recherche, nous montrons ensuite que ces résultats impressionnants d'usage direct s'expliquent surtout par le fait que les besoins des utilisateurs directs ont été satisfaits, à savoir ceux des responsables de l'administration. L'importance des changements opérés à cause des évaluations dans la politique publique en question reste cependant assez modeste. Nous montrons ainsi que l'usage direct d'une évaluation est essentiellement possible au niveau de la mise en œuvre d'une politique publique et de son réglage, et nettement moins au niveau d'une redéfinition des objectifs centraux de la politique publique. A ce niveau là, l'influence de l'évaluation est en effet plus souvent indirecte, concurrencée qu'elle est par le jeu des acteurs, les intérêts et credos politiques. Les informations scientifiques fournies par les évaluations ne correspondent pas assez aux besoins des Parlementaires ou aux Chefs de Département. Toutefois, à ce niveau là aussi (Parlement, Commissions, Départements...), la participation et l'intérêt actif des personnes impliquées est la clé explicative principale de l'utilisation des résultats des évaluations menées.

Enfin, nous documentons au travers de la démarche même de cette étude l'évolution de l'institutionnalisation de l'évaluation dans le système politico-administratif suisse et soulignons les changements importants survenus ces 10 dernières années, mais aussi la nécessité de les poursuivre afin que chacun des groupes d'acteurs concernés (Parlement, administration etc.) puisse mieux utiliser conformément à ses besoins le potentiel d'apprentissage caché dans toute démarche évaluative.

## TABLE DES MATIERES

	Page
<b>RESUME</b>	2
<b>Table des matières</b>	3
<b>1. Introduction</b>	
1.1. Thème	5
1.2. Intérêt, définition du domaine et question de recherche	6
<b>2. Les données de départ</b>	
2.1. Les recensements de l'AGEVAL en 1989 et de l'OFJ de 1996	8
2.2. Les hypothèses de travail et le modèle causal	9
2.3. Méthodologie, techniques utilisées	14
<b>3. Eléments théoriques</b>	
3.1. Définitions : évaluation, utilisation	18
3.2. Science et pouvoir, ou évaluation vs. utilisation ?	23
3.3. Le cadre théorique général de l'utilisation des évaluations	24
3.4. Les mérites de M. Q. Patton	28
3.5. Les particularités du système suisse	29
<b>4. Déroulement et résultats de l'enquête</b>	
4.1. Le déroulement de l'enquête	30
4.2. La première étape : la définition de l'utilisation	31
4.3. Résultats des clauses évaluatives	34
4.4. Les causes explicatives de l'utilisation des évaluations	35
4.5. Valorisation des résultats	42

<b>5. L'interrogation critique</b>	
5.1. Les limites des deux questionnaires	45
5.2. Les limites des résultats	46
5.3. Les limites de l'approche	47
5.4. Analyse et hypothèses explicatives	48
<b>6. Hypothèse et approches complémentaires</b>	52
6.1. Eléments théoriques	53
6.1.1. Les trois autres grands facteurs explicatifs de l'utilisation	53
6.1.2. Les processus de décision. Le modèle des 4 « I » de C. H. Weiss	55
6.1.3. Les processus d'apprentissage	58
6.2. Hypothèse et indicateurs, techniques et démarche	60
6.3. Les entretiens qualitatifs et leurs résultats	63
6.3.1. Les responsables d'Office	63
6.3.2. Entretiens avec des parlementaires	70
6.3.3. Valorisation intermédiaire	74
6.4. Eléments d'autres études	75
6.5. Vérification des hypothèses	76
<b>7. Conclusions et valorisation des résultats</b>	77
7.1. Résultats de notre étude	77
7.1.1. Ce que la démarche quantitative a montré	77
7.1.2. Ce que la démarche qualitative a montré	78
7.1.3. Ce que le tout nous apprend	79
7.1.4. Et par rapport à la théorie scientifique ?	80
7.1.5. Nouveau modèle explicatif	80
7.2. L'évolution de l'évaluation dans l'administration	81
7.3. Conclusion	83
<b>7.4. POST SCRIPTUM POUR SOURIRE</b>	85
<b>8. Bibliographie</b>	86
<b>9. ANNEXES</b>	90

# 1. INTRODUCTION

## 1.1. THEME

« Les effets des évaluations (de l'Office) sont en partie décevants. Ainsi par exemple, une évaluation réalisée sur les décomptes individuels de chauffage avait montré qu'il s'agissait d'une mesure très efficace. Mais elle fut refusée par le politique. Une évaluation sur une autre mesure montrait que celle-ci avait des résultats très faibles. Le politique exigea qu'elle fut maintenue » (Widmer 2001 : 67).

A quoi et à qui servent les évaluations ? Des exemples comme celui-ci, les évaluateurs et les mandataires en connaissent beaucoup. A quoi bon mandater et payer des dizaines de personnes si aucun résultat tangible n'en découle ? Pourquoi tel rapport reste-t-il dans un tiroir, tel autre est-il refusé ou tel autre est-il immédiatement mis en pratique ? Les politiciens sont-ils les seuls responsables ? Les rapports d'évaluation sont-ils tous de qualité suffisante ?

L'évaluation est à la mode. Porté par la vague des « audits » et autres mesures de « controlling », le sujet des évaluations pénètre de plus en plus les cercles des décideurs politiques. Pourtant, de nombreuses évaluations commandées finissent manifestement dans des tiroirs, souvent sans être lues. D'autres ne sont pas ou que partiellement valorisées dans le sens d'une utilisation dans les processus de décisions politiques. Apparemment, une minorité d'entre elles sont véritablement suivies d'effets sensibles (Weiss 1986 : 54). On constate aussi (Bussmann 1996 : 2) que ce sujet semble plus facile à aborder dans certains domaines des politiques publiques que dans d'autres, dans certaines administrations ou départements que dans d'autres.

Il s'agit là d'un sujet difficile et conflictuel, car il s'inscrit dans le cadre des relations entre politiques et scientifiques, entre pouvoir et savoir. Les experts ont-ils toujours raison ? Les politiques ne s'intéressent-ils qu'à leur pouvoir et à leur réélection ? La science rationnelle et prudente doit-elle s'imposer, ou bien des idées, peut-être (encore) peu étayées ou démontrées scientifiquement, peuvent-elles provoquer l'émergence de réalités, par pure option politique, en repoussant à plus tard les (toujours légitimes) questions de l'objectivité et de la pertinence du besoin ?

Cette problématique, vieille comme le pouvoir et la science, nous est apparue d'autant plus intéressante et pertinente qu'elle s'inscrit aujourd'hui, en Suisse comme ailleurs, dans un

contexte politique, économique et social qu'on peut caractériser très sommairement par ces quelques caractéristiques<sup>1</sup> :

- La baisse des *budgets publics* et les difficultés croissantes à les financer, ce qui amène à une réflexion plus intense sur l'efficacité/efficience des moyens investis et l'importance accrue des contrôles
- L'accélération nécessaire des *processus de décisions* politiques, donc la nécessité d'une expertise technique de haut niveau et rapidement disponible
- Une forte *médiatisation* des enjeux politiques et des coûts engendrés par toute décision du gouvernement
- Une crise d'orientation politique dans le grand public et une tendance à la *polarisation* de la vie politique, donc un danger d'*instrumentalisation* de toute expertise, danger renforcé encore par le poids des médias et les techniques de communication modernes avant des votations
- La *complexité* croissante et l'interdépendance apparente des problèmes de société à résoudre
- Le besoin accru de *coordination* entre les divers services gouvernementaux (parlement, administration, gouvernement)
- Les propositions et projets divers visant à *réformer* les droits politiques, les procédures et fonctionnements des institutions politiques de ce pays, donc à faire le bilan du système ayant prévalu jusqu'ici
- Le début d'institutionnalisation de l'évaluation qu'on peut observer au sein de l'administration fédérale, avec en particulier l'introduction de plus en plus fréquente d'une *clause évaluative* dans les textes législatifs.

L'évaluation se retrouve donc comme un thème et un enjeu en filigrane et au croisement de nombreux sujets d'actualité.

## **1.2. INTERETS, DEFINITION DU DOMAINE ET QUESTION DE RECHERCHE**

- Notre motivation à aborder ce sujet est d'ordre professionnel. Dans le cadre de notre parcours à l'IDHEAP, le fil rouge de notre questionnement a été celui des instruments de gestion moderne, et en particulier celui des outils d'aide à la décision. Nous nous intéressons en particulier à ce qui renforce une démarche rationnelle et la plus objective possible pour prendre des décisions, sur la base de données quantifiables et vérifiables.

---

<sup>1</sup> Il ne s'agit ici que de quelques caractéristiques particulièrement en lien avec notre sujet.

Les progrès des techniques des sciences sociales et politiques rendent possible l'observation a posteriori des éléments ayant contribué à influencer une décision politique, permettant ainsi de vérifier ou d'infirmer les hypothèses les plus répandues sur les raisons qui ont influé sur le processus de prise de décision.

- Notre recherche se concentrera sur *les résultats des évaluations des politiques publiques au niveau fédéral* en Suisse.

Un autre domaine touché par notre sujet est celui des relations entre *le politique, l'administration et les scientifiques ou experts* amenés à les conseiller, et en particulier la question des processus d'apprentissage dans les organisations.

Nous nous intéresserons plus précisément aux *évaluations réalisées entre 1985 et 1995 des politiques publiques fédérales*. Nous retenons la dimension *fédérale* à cause de l'impact potentiellement plus grand des politiques publiques à ce niveau, de leur plus grande médiatisation et de l'importance des intérêts publics en jeu. Autant de facteurs qui augmentent les chances de trouver des évaluations. De plus, la croissance progressive d'une « culture d'évaluation » est pour l'instant limitée essentiellement au niveau fédéral. La *période choisie de 1985 à 1995* se justifie par l'existence de trois bases de données fédérales<sup>2</sup> sur les études d'évaluations réalisées pendant ces années. Nous compléterons nos résultats sur ces années par des informations plus récentes sur les mêmes sujets.

Notre étude entend donc prolonger le travail entrepris par W. Zimmermann en 1989 et surtout celui de W. Bussmann en 1996 sur les évaluations menées au sein de la Confédération. Ces travaux constituent la base de données primaires que nous avons utilisée pour notre questionnement sur l'utilisation qui a été faite de ces études. Celui-ci prend en compte en outre une base de données au sujet des clauses évaluatives dans les textes législatifs. Nous chercherons à savoir lesquelles parmi ces études ont été utilisées, ce que cela a signifié concrètement, et tenterons de dégager les raisons qui expliquent le sort réservé à ces évaluations dans le processus politico-administratif, que ce soit dans un sens positif ou négatif. Nous essayerons de dégager différents types de causes ou facteurs explicatifs.

Ainsi notre *question centrale* est-elle de savoir *pourquoi certaines évaluations sont utilisées dans les processus de décision politique et d'autres pas*.

---

<sup>2</sup> W. Zimmermann, Doris Kolly, Pascal Mahon : Die wissenschaftliche Evaluation von Bundespolitik, DFJP, Bern 1989. Werner Bussmann : Résumé des études exécutées, mandatées ou subventionnées par les services de la Confédération pendant la période législative 1991-1995 sur l'efficacité et l'effectivité de l'action étatique, DFJP, 1996. Office Fédéral de la Justice : Liste des clauses évaluatives. Site internet de l'OFJ, mai 1999.

Nous testerons à cet effet un modèle explicatif de travail sur les bases de données existantes et évaluerons sa pertinence et son bien-fondé.

## 2. LES DONNEES DE DEPART

### 2.1. LES RECENSEMENTS DE L'AGEVAL EN 1989 ET DE L'OFJ DE 1996

Notre démarche se construit sur trois recensements des évaluations effectuées au niveau fédéral. Tout d'abord celle effectuée sur mandat du Groupe de Travail d'Evaluation législative (AGEVAL) de l'Office fédéral de la Justice (OFJ) par *W. Zimmermann*<sup>3</sup> pour la période 1978-1989. Il avait rassemblé pour la période retenue 170 études d'évaluations de politique publique en Suisse et les avait analysées selon divers critères.

Puis, en 1996, donnant suite à un mandat de la section « Efficacité » de la Commission de Gestion (CdG<sup>4</sup>) du Conseil National, *l'Office Fédéral de la Justice (OFJ)* réalise une enquête devant lui permettre de fournir une information sur les démarches évaluatives en cours ou réalisées dans l'administration fédérale de 1990 à 1995. Un questionnaire est envoyé à tous les Offices, demandant de fournir des informations sur des études « mandatées, exécutées ou subventionnées » par les services de la Confédération sur la question de l'efficacité et de l'effectivité de l'action étatique. 51 des 58 Offices ont répondu. Les résultats de l'enquête de l'OFJ sont publiés dans le Bulletin no 2 / 1996 de la Société suisse d'évaluation (SEVAL).

L'évaluation de cette liste faite par *W. Bussmann* de l'OFJ est essentiellement quantitative. On y apprend que si les 7 Départements fédéraux ainsi que la Chancellerie fédérale ont livré des informations, on constate de grandes différences, par exemple entre le Département Fédéral des Affaires Etrangères (DFAE), où la Direction pour le Développement et la Coopération (DDC) fournit plus de 200 titres, et le Département de la Défense, de la Protection de la population et des Sports (DDPS) par exemple, avec une seule étude. Les titres de deux Départements (DFI et DFAE) représentent plus de 80 % du tout. A eux deux, les Offices DDC/DFAE et OFSP/DFI forment près de 50% de l'ensemble. La disparité est énorme. Les titres des 8 premiers Offices (sur 25) constituent 80% de tous les titres. Il y avait à l'époque 58 Offices fédéraux. L'évaluation n'est donc connue et/ou pratiquée que dans une minorité d'Offices. L'analyse montre en outre une nette et rapide progression du nombre d'évaluations au fil des ans. Mais elle ne permet d'en tirer aucune conclusion en ce qui concerne la *méthodologie* ou la *qualité* de ces études. Cependant, il faut noter ici qu'il s'agit

<sup>3</sup> Cf. Bibliographie. La démarche est présentée également dans Horber :1990.

<sup>4</sup> Cf. Liste des principales abréviations en Annexe I

d'un recensement historique en ce qu'il est la première compilation donnant un aperçu à peu près exhaustif de l'état de l'évaluation en Suisse à cette époque. Il nous informe indirectement aussi sur quelques particularités du système suisse comme la structure décentralisée de l'Administration, l'importance des Offices, etc. Nous reviendrons plus bas sur ce point.<sup>5</sup>

Enfin, l'OFJ toujours avait recensé sur son site Internet la liste de tous les textes législatifs connus en mai 1999 et munis d'une clause évaluative. Il y avait 30 lois ou ordonnances recensées, datant pour la plupart des années 80 et 90.

La base de données recueillie par l'OFJ contient 532 titres. Ajoutés aux 170 titres signalés par W. Zimmermann et aux trente lois disposant d'une clause d'évaluation, nous avons donc une liste potentielle (sous réserve de doublets) de 732 titres d'évaluations effectuées entre 1978 et 1995<sup>6</sup>.

## **2.2. LES HYPOTHESES DE TRAVAIL ET LE MODELE CAUSAL**

Quel sort a été réservé aux évaluations en question ? Et surtout : peut-on identifier les raisons qui expliquent l'usage qui a été fait de ces études ? Quels sont les principaux facteurs explicatifs ? Il ne saurait évidemment n'y avoir qu'une seule réponse à ces questions. Ainsi peut-on imaginer, en toute naïveté, que le facteur temps joue par exemple un rôle important : la plupart des rapports sont assez longs et complexes, nos politiciens plutôt pressés et le personnel de l'administration de plus en plus sollicité. Ou alors le rapport arrive tout simplement trop tard et n'a plus d'actualité politique. Peut-être un conflit de personne a-t-il introduit un grain de sable dans les rouages du processus de décision ? Ou bien certains sujets sont trop brûlants et les découvertes d'experts trop « explosives » politiquement ?

La diversité des explications possibles est grande. Nous avons donc émis plusieurs hypothèses dans cinq domaines explicatifs, allant du plus large au plus pointu.

### **1. Les circonstances dans lesquelles la politique publique à évaluer a été élaborée**

Une première série de variables explicatives concerne non pas l'évaluation elle-même, mais la politique publique qui a fait l'objet de l'évaluation, en particulier la manière dont elle a été définie et décidée. Nous émettons l'hypothèse qu'une politique publique qui naît dans de grands efforts et après de nombreuses procédures démocratiques a plus de chances de se faire évaluer qu'une politique publique reposant sur un grand consensus et ne provoquant aucun débat. Par analogie, nous en déduisons qu'une évaluation d'une telle politique publique a plus de chances d'être utilisée et d'avoir des effets.

---

<sup>5</sup> Cf. § 3.5.

<sup>6</sup> Une présentation consolidée des 3 listes de base se trouve en Annexe 2.

Les hypothèses sont alors les suivantes :

- *la longueur des processus d'élaboration de la politique publique et la préparation de celle-ci augmentent les chances pour une évaluation d'être utilisée.*
- *les conflits lors de l'élaboration et la préparation de la politique publique augmentent les chances pour une évaluation d'être utilisée.*

Comme indicateurs, nous retenons l'existence et la qualité de débats parlementaires et de travaux de Commissions parlementaires ou extraparlimentaires sur le sujet, la durée de ces processus et l'existence de conflits publics d'opinion sur le sujet.

## **2. Les circonstances dans lesquelles l'évaluation elle-même est élaborée, organisée et réalisée**

Le deuxième groupe de variables explicatives a trait cette fois à l'évaluation elle-même, soit les conditions de sa naissance, sa genèse.

Les causes sont ici à chercher dans l'attitude et les compétences de l'administration face à l'évaluation décidée. Puisque l'administration est chargée en général de la mise en œuvre des politiques publiques, une évaluation qui se fait sans le concours et/ou l'appui de celle-ci a peu de chances d'être utilisée.

Nos hypothèses sont donc les suivantes :

- *les chances pour une évaluation d'être utilisée augmentent si l'évaluation fait partie des tâches habituelles (planifiées, préparées...) de l'administration. Une évaluation imposée par une décision extérieure aura moins de chance.*
- *plus les acteurs de l'administration en charge des évaluations sont informés et formés à cette discipline, plus l'évaluation aura de chances d'être utilisée.*
- *plus les acteurs de l'administration responsables de l'évaluation participent à la démarche conduisant et réalisant l'évaluation, plus celle-ci a de chances d'être utilisée.*

Les indicateurs retenus sont l'origine et la réalisation internes ou externes à l'administration de l'évaluation, la présence de clauses évaluatives, l'application de démarches participatives, ainsi que la familiarité des acteurs de l'administration avec l'évaluation. Nous nous intéresserons en particulier au niveau et au type de formation des responsables de l'administration.

### 3. Les circonstances dans lesquelles l'évaluation et ses résultats sont reçus

Un troisième groupe de causes concerne le moment où le rapport d'évaluation est rendu. Là aussi, le contexte général, les constellations politique, médiatique et sociale sont susceptibles de jouer un rôle. Les causes sont à chercher dans les jeux d'alliance et d'intérêts entre divers groupes, en fonction de l'opportunité. Si le thème n'est plus d'actualité, ou si les médias ne s'y intéressent pas, il est peu probable qu'elle ait beaucoup d'impact.

Par ailleurs la fonction attribuée à l'évaluation joue un rôle : si l'évaluation sert à justifier, défendre et légitimer les choix faits, elle sera certainement utilisée. De même si les résultats de l'évaluation laissent beaucoup de place à l'interprétation politicienne.

Nous formulons par conséquent nos hypothèses de la manière suivante :

- *plus la pression médiatique sur le thème de l'évaluation est grande, plus elle sera utilisée.*
- *le type de fonction attribué à l'évaluation influencera son utilisation.*

Les indicateurs retenus : l'intensité médiatique et l'actualité politique des sujets d'évaluation, le degré d'urgence du thème, la fonction dévolue à l'évaluation.

### 4. Les caractéristiques du sujet d'évaluation

Le groupe suivant dirige notre regard sur le contenu même du sujet de l'évaluation. Ce n'est plus le contexte général qui politise ici l'évaluation, mais son objet même. Il n'est certainement pas indifférent pour expliquer l'utilisation qui en est faite. Un domaine nouveau aura a priori plus de chances qu'une vieille politique. De même, une évaluation axée plus sur les questions de rendement, ou de contrôle des coûts aura plus de chances qu'une étude générale sur les effets à long terme et le taux de réalisation des objectifs politiques.

Nous retenons donc les hypothèses suivantes :

- *plus le sujet est pionnier et traite d'un nouveau domaine, plus l'évaluation sera utilisée ;*
- *plus l'évaluation s'intéresse aux questions de rendement, d'efficacité, de maîtrise des coûts, plus elle sera utilisée ;*
- *plus l'évaluation s'intéressera à l'atteinte du groupe cible ou à la réalisation des objectifs politiques, moins elle sera utilisée.*

Les indicateurs sont dans ce cas sont la nouveauté du sujet, le type de questions sur lesquelles l'évaluation est orientée, en particulier les questions de contrôle d'efficacité de la mise en œuvre ou d'atteinte de résultats politiques, la relation entre mesures prises et résultats, etc.

## 5. Les aspects formels du rapport d'évaluation

Enfin, tous ces facteurs se cristallisent dans le rapport d'évaluation lui-même, sa forme et ses caractéristiques. Il est vraisemblable qu'un rapport rapidement mené, aux conclusions très concrètes, écrites de manière concise et dans un langage compréhensible de chacun sera plus facilement utilisé. Le respect de certaines règles de base de déontologie aide également (convention écrite, respect du droit et des intérêts des parties, publication, transparence, gestion des conflits...). De même, la présence d'une stratégie de valorisation, avec par exemple un résumé pour journalistes, un communiqué de presse etc., augmente les chances d'utilisation. Il s'agit sans doute du groupe le plus important de facteurs.

Nos hypothèses seront donc les suivantes :

- *La présence de recommandations précises et simples renforce les chances d'utilisation.*
- *Le respect des règles déontologiques augmente les chances d'utilisation des évaluations.*
- *Les rapports brefs et écrits dans un langage simple et aussi peu technique que possible seront plus utilisés.*
- *La présence de résumé et de traduction facilite l'utilisation.*
- *Plus les évaluations sont réalisées rapidement, plus elles seront utilisées.*
- *Plus les évaluations ont une stratégie de valorisation élaborée (y compris au niveau budgétaire), plus elles seront utilisées.*

Les indicateurs retenus sont alors l'existence et l'importance de recommandations, leur qualité, le respect de standards déontologiques, la présence d'une stratégie de diffusion des résultats et de leur valorisation. Mais aussi la technicité du rapport, ses dimensions, la rapidité avec laquelle le mandat a été rempli.

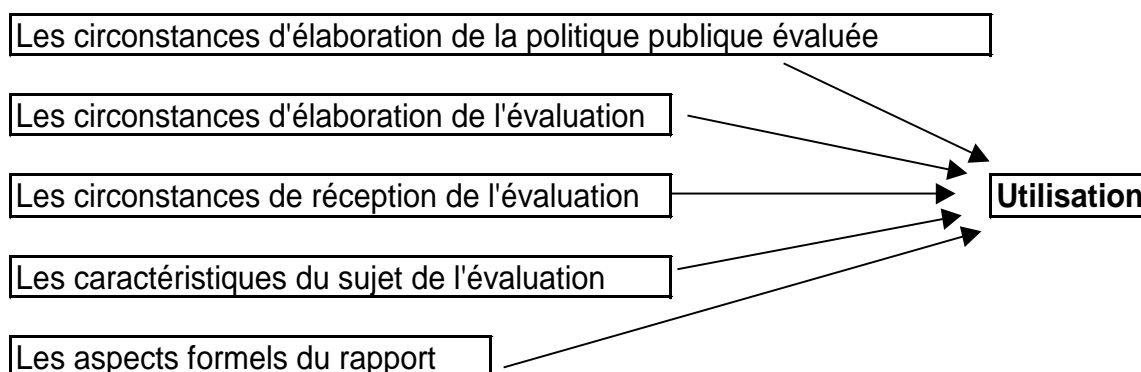
En résumé, nous retenons donc le système suivant d'hypothèses :

<b>Hypothèse</b>	<b>Variable explicative</b>	<b>Indicateurs</b>
1. Les circonstances dans lesquelles la politique publique à évaluer a été élaborée	- processus de préparation de la politique publique - longueur du processus - conflits	- débats et travail parlementaires - durée du processus - navettes entre les Conseils, prises de position conflictuelles
2. Les circonstances dans lesquelles l'évaluation elle-même est élaborée, organisée et réalisée	- tâche intégrée au travail de l'administration - collaborateurs (in)formés - participation de l'administr.	- origine de la demande d'évaluation - familiarité avec la technique - technique participative, présence de clauses évaluatives

3. Les circonstances dans lesquelles l'évaluation et ses résultats sont reçus	-actualité médiatique du sujet du rapport -attente de l'utilisateur par rapport à l'évaluation	-taux de médiatisation (articles, éditos..), urgence du thème -fonctions de l'utilisation de l'évaluation
4. Les caractéristiques du sujet d'évaluation	- type de politique évaluée - contenu de l'évaluation	- sujet nouveau ou traditionnel - orientation sur les questions d'efficacité, ou d'efficacité, ou d'impact, etc.
5. Les aspects formels du rapport d'évaluation	- importance des recommandations - déontologie - simplicité, résumé, traduction - rapidité d'exécution - stratégie de valorisation	- importance des recommandations - respect des règles - langage simple, résumé, traduction - délai - poste au budget et techniques de valorisation

Notre modèle causal de travail que nous appliquerons à notre base de données pour expliquer pour quelles raisons des évaluations sont utilisées ou non sera donc le suivant :

### Le modèle causal de travail



### **2.3. METHODOLOGIE ET TECHNIQUES UTILISEES**

L'importance de la base de données nous invitait à tenter une *démarche essentiellement quantitative*.

#### *Critères d'épuration des listes de base*

Nous avons tout d'abord défini une *limite temporelle*, et retenu la période de 1985 à 1995, afin d'avoir une période assez longue. La distance temporelle jusqu'en 1978 nous semblait trop risquée par rapport à l'objectif de retrouver relativement facilement trace des auteurs et des effets des études. Il y avait par ailleurs relativement peu d'études entre 1978 et 1985. Nous avons en outre retranché un certain nombre de doublets et d'autres titres de la liste de W. Zimmermann, car leur *mandataire* n'était pas clairement la Confédération. Des 170 titres, cette liste s'est réduite finalement à 70. Grâce à la liste de l'OFJ de 1996 de 532 titres, nous avons finalement à notre disposition une liste nette de 602 titres entrant dans nos critères.

#### *Présence de clauses évaluatives*

En plus de cette liste, nous disposions d'une liste de 30 textes législatifs munis d'une clause évaluative. Nous les avons également retenus. Il était intéressant de faire une brève incursion dans ce domaine et d'observer l'utilisation des évaluations rédigées sur cette base à des fins de comparaison avec les autres titres. *Notre hypothèse implicite était ici que la présence de clauses évaluatives augmentait les chances d'utilisation des études*. Nous avons traité ce fichier en parallèle du grand fichier vu la disparité des situations de départ des deux listes. Notre base de données se monte donc finalement à 632 titres.

#### *Dispositif de recherche*

Pour mettre en œuvre notre modèle causal, nous avons retenu l'approche par questionnaires, adressés aux mandataires, soit les Offices de l'administration fédérale. Pour ce faire, nous avons élaboré un *premier questionnaire* destiné à définir le terme « utilisation ». Nous avons proposé aux Offices une définition de l'utilisation basée sur un modèle *d'intensité progressive* permettant de classer facilement les études en 6 catégories selon qu'elles avaient été utilisées très fortement, fortement, moyennement, un peu, faiblement ou pas du tout. Les Offices étaient invités à se conformer à cette définition, ce qui ne fit apparemment aucune difficulté.

Pour le groupe des clauses *évaluatives*, le questionnaire recensait plus simplement 4 variantes d'utilisation : faible, moyen, fort, très fort. Nous avons supprimé les rubriques « pas d'effet » et « je ne sais pas - influence non nette » car ces questions ne nous semblaient pas pertinentes dans ce cadre, ces évaluations devant par nature avoir un impact. Par contre, il est utile dans ce cas d'avoir une information sur le degré d'amélioration de l'efficacité de la Loi obtenue

grâce à ces clauses. Nous avons maintenu ce groupe distinct, afin de pouvoir par la suite comparer les résultats d'utilisation de ce groupe avec les évaluations « classiques ».

Sur la base des réponses fournies par les Offices, nous avons élaboré *un second questionnaire à orientation causale*, structuré en fonction de nos hypothèses. Afin de vérifier les réponses des Offices sur les causes d'utilisation, nous avons choisi de travailler avec un *groupe de contrôle*. Puisque nous cherchions les causes de l'utilisation, les critères de choix permettant de constituer les échantillons ne devaient pas être aléatoires, mais au contraire dictés par l'intensité d'utilisation des évaluations communiquée par les Offices. Le *groupe principal* a donc été constitué par les études dont l'utilisation par les Offices avait été déclarée très forte. Le *groupe de contrôle*, lui, a été constitué avec les études qui n'avaient pas été utilisées. Un questionnaire ad hoc par étude<sup>7</sup>, établi sur un logiciel adapté, interrogeait les responsables d'Office. Il était structuré en 5 sections, suivant notre modèle causal et visant à obtenir des informations pertinentes pour nos indicateurs. Ses 60 questions reprenaient toutes sortes de caractéristiques propres à l'évaluation, à sa genèse, à son sujet, à son organisation, etc. Il s'agissait bien sûr d'un seul et unique questionnaire, envoyé aussi bien au groupe principal qu'au groupe de contrôle. Rien ne permet formellement de distinguer les deux groupes d'études. Les Offices ne sont pas informés de l'existence de ces deux groupes, et ne savent donc pas pourquoi des questions sont posées sur cette étude en particulier. Nous avons veillé à avoir deux groupes de taille comparable.

Ce dispositif incluant deux questionnaires successifs et l'emploi d'un groupe de contrôle permet d'atteindre trois objectifs :

1. Solliciter la réflexion de l'administration sur l'usage qu'elle fait des études qu'elle mandate et obtenir d'elle le jugement sur la valeur de chaque étude.
2. Dégager les caractéristiques principales des études ayant été fortement utilisées grâce aux pourcentages de réponses recueillies pour chaque variable explicative.
3. Vérifier l'importance de ces variables en comparant ces résultats à ceux obtenus sur le groupe de contrôle.

*Ainsi c'est la différence entre les réponses obtenues (dans les deux groupes) sur l'importance que chaque variable explicative a eue pour chacun des deux groupes qui nous donnera les causes de l'utilisation ou non de l'évaluation (différence causale).*

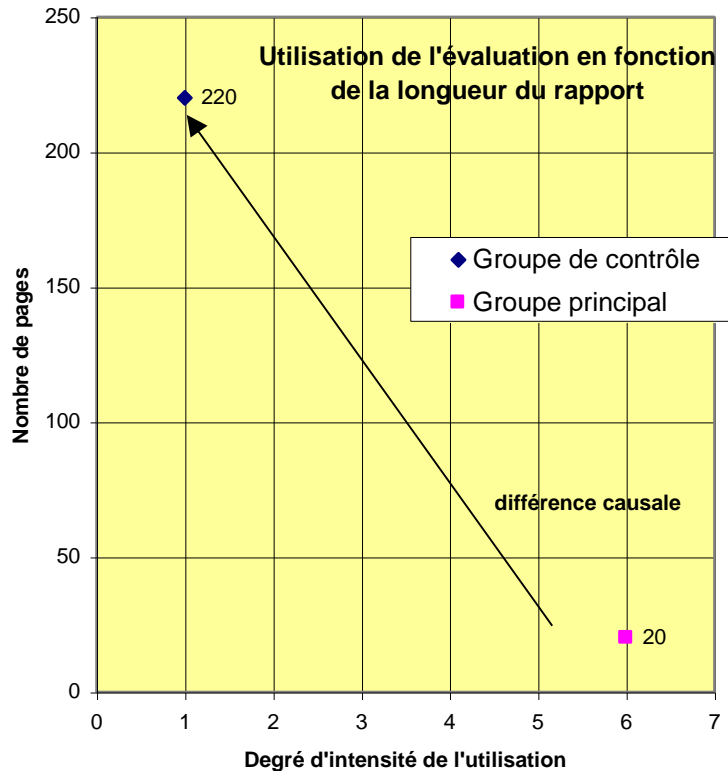
Si, par exemple, à la question sur le nombre de pages qu'avait l'évaluation, nous constatons que les évaluations du groupe de contrôle se trouvaient en général dans la catégorie « plus de

---

<sup>7</sup> Cf. Annexe 10

200 pages » tandis que la majorité des études du groupe principal se trouvait dans la catégorie « moins de 20 pages », nous aurions une preuve que l'hypothèse selon laquelle les évaluations longues ne sont pas utilisées est correcte. Et ainsi de suite pour chaque variable.

### Schéma



La cause sera d'autant plus claire que l'inclinaison de la flèche sera forte.

Nous envisageons ensuite de compléter et vérifier cette base de données avec des informations qualitatives plus récentes par quelques interviews ciblés d'acteurs importants dans le domaine de l'évaluation dans l'administration fédérale.

### Constitution des groupes

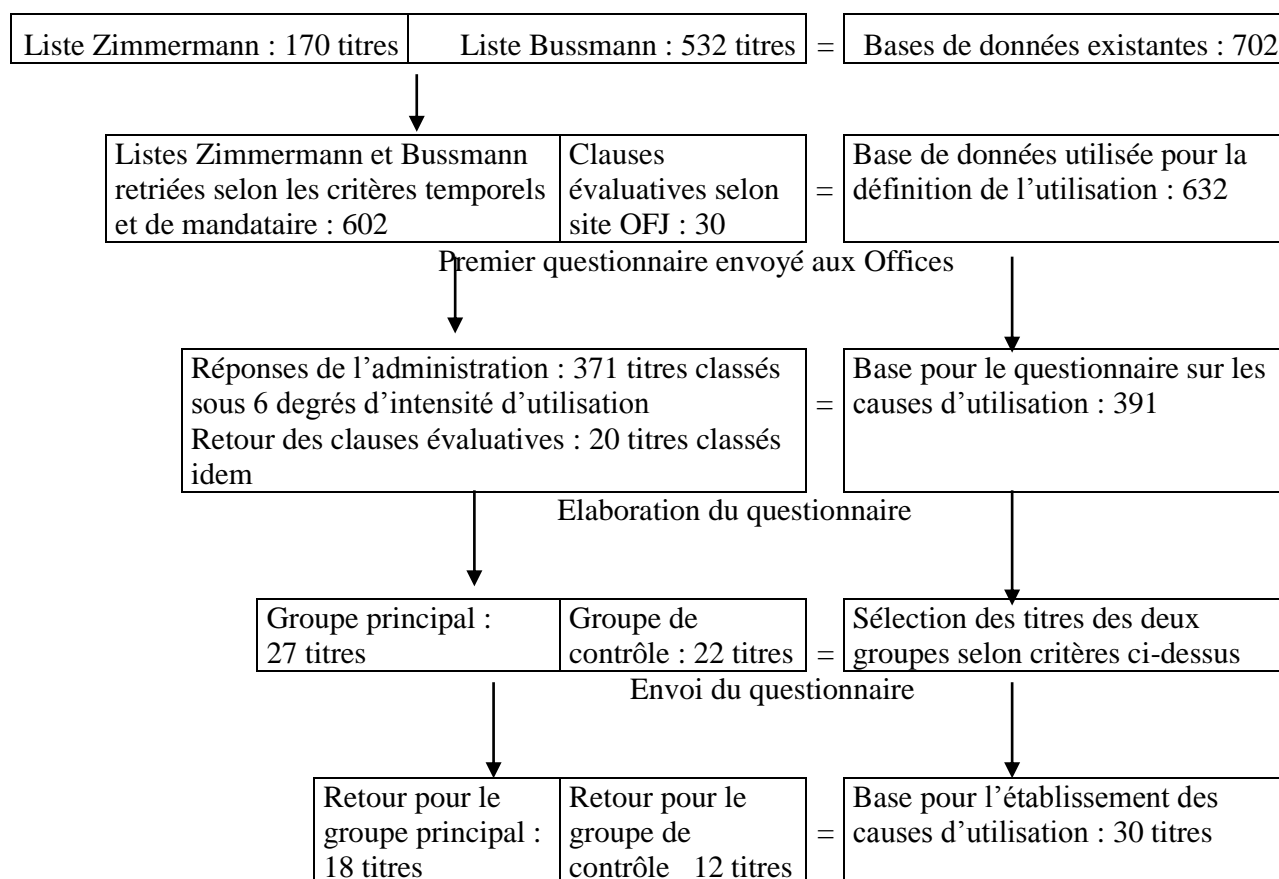
Les critères suivants ont été appliqués pour sélectionner les études destinées à constituer le groupe principal et le groupe de contrôle :

- Faire partie des études classées soit sous « très forte utilisation » - « usage direct » (intensité 6), soit sous « aucune utilisation » (intensité 1) par les Offices concernés. Pour

le groupe de contrôle, nous avons dû prendre en plus quelques études classées sous « faible influence » (intensité 2) afin d’avoir un groupe suffisamment important<sup>8</sup>.

- Proximité temporelle la plus grande possible.
- Degré de difficulté probable à retrouver le cas échéant l’auteur ou identifier clairement le mandataire.
- Représenter le plus grand nombre possible de domaines et de départements
- La plus grande ampleur possible du sujet par rapport à la politique publique (sur la base des informations disponibles dans le titre).

Notre dispositif de recherche et le resserrement progressif de la base de données peut donc être représenté de la manière suivante :



<sup>8</sup> Nous avons volontairement renoncé à prendre en compte dans les deux groupes les études utilisées avec l’intensité 5 (« forte ») et l’essentiel de celles utilisées avec l’intensité 2 (« faible ») afin de garantir la meilleure fiabilité possible de notre démarche.

### 3. ELEMENTS THEORIQUES

Qu'entend-on par *utilisation* ? Comment peut-on la définir ? Comment peut-on retrouver des traces sûres qu'un texte ait eu un impact ou une suite dans un processus ? Et pendant combien de temps faut-il chercher ? Telles sont les questions de base qui ont guidé notre utilisation des bases de données. Aborder le thème des évaluations par la porte de leur utilisation semble assez logique. L'intérêt des évaluations est de savoir à quoi elles ont servi, si elles ont pu changer positivement quelque chose. Elles font partie du processus de décision politique, et n'ont donc de sens en principe que si elles sont utilisées.

Mais avant de parler d'utilisation, il nous faut nous arrêter sur le concept d'*évaluation* lui-même. En effet, vu notre grande base de données contenant 3 listes d'auteurs et de dates différentes, il y avait de fortes chances de trouver des conceptions variées de ce qu'on mettait sous ce mot.

#### 3.1. DEFINITIONS : EVALUATION, UTILISATION

*Qu'est-ce qu'une évaluation ?*

La discussion dans la littérature est très fournie et ancienne (Patton 1986 :103ss. ; Rangeon 1993 : 14ss.), aussi nous contenterons-nous de commenter les éléments classiques en fonction de points plus particulièrement intéressants dans l'optique de notre recherche<sup>9</sup>.

1. Un premier élément rappelle qu'une évaluation implique l'existence *d'objectifs et de buts à atteindre* et donc à mesurer. L'évaluation est donc une démarche de mesure de l'atteinte d'objectifs.

La question qui se pose ici est de savoir ce qu'on peut faire de toutes les évaluations qui ne sont pas centrées sur des objectifs à mesurer. Une autre question est de savoir qui fixe ces objectifs, respectivement si tous les acteurs de l'évaluation ont vraiment les mêmes objectifs et attendent les mêmes résultats en ce qui concerne leur degré de couverture.

2. Une autre aspect est la rigueur *scientifique et méthodologique* de l'évaluation. C'est une approche des sciences sociales, basée sur des expérimentations et des mesures empiriques. Il s'agira ici de s'interroger sur cette rigueur et les biais de méthodes et leur relation à l'utilisation. Y a-t-il un lien entre la qualité scientifique d'une étude et celle de son utilisation ?

Ces deux premiers éléments se retrouvent explicitement dans les démarches effectuées par W. Zimmermann et W. Bussmann.

---

<sup>9</sup> Nous situons l'évaluation comme un élément à l'intérieur du cycle des politiques publiques (Bussmann 1998). Cf. schéma classique reproduit en Annexe 3.

3. Un troisième éclairage souligne l'importance de la *nature comparative* de l'évaluation. Elle s'est profilée en particulier face à l'étroitesse de l'approche axée purement sur l'atteinte d'objectifs, en essayant de mettre en perspective des résultats de politiques similaires ou comparables. La comparaison donne certes un étalon de mesure, mais ne répond pas implicitement à la question du choix des entités comparées entre elles.
4. On peut mettre en avant aussi l'importance du *jugement* inhérent à une évaluation. Celle-ci est un processus visant à juger la valeur d'un programme. La question cruciale ici est celle du processus de définition des critères qui permettent ensuite de poser un jugement valable sur un programme.
5. Plusieurs théoriciens ont mis en évidence le fait que les évaluations constituent un *système d'information* pour les décideurs (Weiss 1999). L'évaluation est un processus de résolution d'un problème. Dans cette approche, l'évaluateur devient partie prenante dans le processus politique de résolution du problème.
6. Enfin, signalons que l'évaluation peut aussi être considérée elle-même comme un *processus d'apprentissage*, comme un changement d'état de l'apprenant, qui n'est jamais le même après qu'avant l'évaluation, quel que soit par ailleurs le statut réservé à l'étude elle-même.

L'évaluation se distingue en particulier des *audits*, dont l'objectif est principalement de faire une photographie de l'organisation ou d'un programme, souvent sur le plan financier seulement, et aussi du *controlling*, intéressé d'abord à vérifier le degré de mise en œuvre de décisions prises, mais sans poser la question du sens. Avec Patton (cité par Horber 1999), nous dirons que l'évaluation « *permet de collecter et d'analyser de façon systématique de l'information pour les utilisateurs spécifiques afin de réduire l'incertitude, d'améliorer l'efficacité, l'efficience et l'aide à la décision* ».

### *Les types d'évaluation*

Classiquement, on distingue entre les évaluations *summatives* (ou récapitulatives) d'une part, destinées à fonder une décision pour la suite d'un programme en montrant la convergence ou non des résultats avec les objectifs (Palumbo in Palumbo 1986 : 15), et les évaluations *formatives* (ou endoformatives) d'autre part, orientées plus sur le développement du programme, son amélioration. Palumbo (Palumbo in Palumbo 1986 : 39) utilise ces deux concepts en fonction du moment où l'évaluation se déroule, la *formative* ayant lieu au moment de la mise en œuvre, tandis que la *summative* se situe soit avant la décision d'une politique publique, soit au moment de l'impact de la politique, là où l'on peut faire un bilan et prendre une option pour la suite. Toutefois Monnier (Monnier 1992 :112ss.) fait remarquer que dans les deux cas des décisions sont prises, et il propose de distinguer les deux types plutôt en fonction des protagonistes concernés. Un type *summatif* intéressera d'abord des

personnes extérieures au programme, en particulier les décideurs politiques ou publics, tandis que le type *formatif* concernera en premier lieu les personnes participant au programme. Cette distinction nous paraît plus pertinente pour notre cas d'espèce, puisque nous sommes toujours dans une situation « ex post », mais nous nous intéressons en particulier aux utilisateurs.

De même, on peut distinguer également entre des évaluations à vocation *instrumentale*, c'est-à-dire dont il est attendu que les résultats puissent entrer rapidement dans les processus de gestion administrative, et celles à vocation *conceptuelle ou symbolique*, où l'attente est plus de l'ordre de la connaissance, de l'« enlightenment ». Dans notre cas, il est clair que nous cherchons les traces d'usage instrumental direct.

### *Les fonctions de l'évaluation*

Une évaluation peut remplir différentes fonctions. La littérature (Rossi & Freeman 1993 : 60ss. ; Chelimsky 1986 : 74ss.) en cite le plus souvent 5 possibles :

1. *La légitimation*. L'étude sert à défendre une position préétablie, à justifier les choix et les décisions prises, à renforcer le rôle et le statut du mandataire.
2. *Le contrôle*. La décision prise à l'époque a-t-elle été appliquée, l'argent investi a-t-il été correctement dépensé, quels dérapages ou imprévus ont-ils eu lieu ?
3. *L'information*. Où en est l'avancement du programme, de quoi s'agit-il avec ce projet, quelles sont les variantes possibles pour la suite ?
4. *Le pilotage*. L'étude sert à boucler la boucle du cycle d'une politique publique, à le maintenir en mouvement, à donner les éléments permettant une correction, que ce soit au niveau des définitions de départ (objectifs, groupes cibles, problème à résoudre), de sa mise en œuvre, de ses résultats, etc.
5. *L'argumentation*. L'évaluation donne des arguments supplémentaires pour (con)vaincre l'adversaire politique ou le supérieur, et augmente le pouvoir et la crédibilité du mandataire.

### *Qu'est-ce que l'utilisation d'une évaluation ?*

Nous partons d'une définition assez large du mot, incluant une influence *indirecte*. Néanmoins, nous délimitons l'usage du mot de manière positive, c'est-à-dire que la simple lecture d'un rapport n'est pas encore une utilisation. Nous entendons principalement que des éléments provenant de l'étude ont été perçus par les mandataires comme utiles et repris comme arguments ou sources qui ont conduit à une décision sur le sujet évalué.

Une première distinction fondamentale est à opérer entre *l'usage direct* et *l'influence*. (Cf. Rossi & Freeman, 1993 : 60-61 ; Patton, 1986 : 100-143). *L'usage direct* implique une action et relation immédiate et matérielle de l'étude sur le programme et la politique. C'est le cas

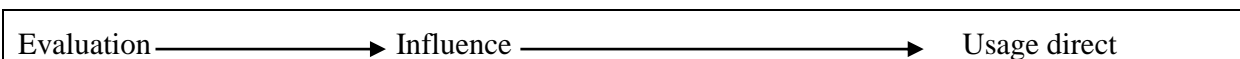
idéal où le changement est observable directement et où l'évaluation détermine les décisions. *L'influence* consiste plutôt dans une relation indirecte ou pas directement ou immédiatement observable sur le programme. L'influence se fait plutôt sur l'esprit et la réflexion des acteurs que sur les programmes.

Utilisation	Agit sur	Genre	Type	L'utilisation s'observe au niveau de :
<b>USAGE DIRECT</b>	Programme	Instrumental	Direct, immédiat	Décision
<b>INFLUENCE</b>	Acteurs, organisation	Conceptuel	Indirect, médiat	Réflexion

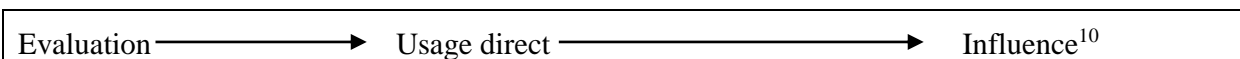
Il y a clairement une tension entre ces deux modèles. Mais c'est un peu aussi une question de type « l'œuf ou la poule ». Il est possible que des utilisations au sens d'« usage direct » puissent inclure l'utilisation dans le sens d'« influence », dans la mesure où cela peut à terme induire en plus une habitude ou un début de culture évaluative. Mais cela n'est pas certain, car on peut aussi imaginer une évolution de type « contrôle », peu propice à un apprentissage. A l'inverse, il semble a priori moins facile d'inclure le type « usage direct » à l'intérieur d'une approche « influence ». Néanmoins, on ne saurait exclure qu'une influence répétée sur la réflexion des acteurs ne débouche à terme et de manière soudaine et très objective sur un changement d'attitude et de décision.

Schématiquement on pourrait représenter cela ainsi :

### *Types d'utilisation d'une évaluation*



Ou bien :



L'utilisation de type « influence » n'est contestée par personne. Mais elle reste par définition extrêmement difficile à quantifier. Seuls les acteurs directs peuvent l'affirmer ou en témoigner. La littérature, et notre étude aussi, s'est plus orientée sur la notion d'« usage

<sup>10</sup> Une question restant ouverte est de savoir si la définition de l'influence reste la même dans les deux options décrites ici. Il est possible que l'influence résultant d'une utilisation directe soit plus importante et forte que celle qui la précède.

direct». Elle est l'espoir de tout évaluateur, et se prête mieux à une approche matérielle et quantitative telle que nous l'avons esquissée. Sur l'ensemble de notre base de données, nous pensons réaliste d'espérer trouver des cas clairs d'usage direct et nous pensons que la quantité d'études recensées et la technique utilisée devraient nous permettre d'aboutir à des conclusions probantes<sup>11</sup>.

Avant de présenter la définition de l'utilisation que nous avons pratiquée dans notre enquête, un mot encore pour relever le parallèle possible entre la définition de *l'évaluation* et celle de *l'utilisation* de l'évaluation. La même tension entre une vision restrictive ou large de la définition d'une évaluation se retrouve pour celle de l'utilisation. Une définition de l'évaluation axée sur l'atteinte chiffrée de résultats par rapport à des objectifs fixés par des dirigeants induit naturellement l'attente d'une utilisation elle aussi très matérielle, de type «usage direct». A l'inverse, une vision de l'évaluation comme d'un processus d'apprentissage des protagonistes rendra plus attentif à tous les éléments informels ou inattendus qui entrent dans l'utilisation faite de l'évaluation. Suivant le courant principal de la recherche, nous avons opté pour une définition matérielle de l'utilisation.

Sur la base de ces éléments, nous pouvons mieux positionner notre *définition de l'utilisation*. Il ne nous paraissait pas possible de partir d'une alternative « oui ou non », « tout ou rien ». Nous avons plutôt essayé d'introduire des nuances de type « oui, oui », « plutôt oui », etc. Nous avons assimilé l'absence d'information sur l'effet d'une étude ou l'impossibilité à le retrouver à une influence faible. Si personne ne s'en souvient ou est capable d'en dire quelque chose 5 ans après, il est peu probable que cette étude ait apporté des éléments cruciaux. Nous avons privilégié l'approche « usage direct » à « influence », en essayant de distinguer des étapes *progressives* où cette influence est devenue matériellement de plus en plus visible et traçable. Les chiffres de 1 à 6 représentent autant d'échelons dans l'intensité de l'utilisation. Enfin, nous avons tenté de définir une fonction qui correspondait à chaque degré et qui pouvait s'appliquer dans un processus de prise de décision dans un Office.

---

<sup>11</sup> Il est peut-être utile ici de faire une remarque d'ordre historique. En effet la question de l'utilisation des évaluations s'est particulièrement développée dans les années 70, marquées entre autres par des convictions rationalistes : l'on s'attendait à des impacts clairs et rapides des évaluations sur les politiques (Freiburghaus/Zimmermann 1985 ; Weiss 1999 : 470). Les attentes des scientifiques de cette époque par rapport à l'usage que les gouvernements feraient de leurs travaux étaient très claires : il s'agissait d'aider les gouvernements à décider s'il fallait continuer ou arrêter certaines initiatives politiques, d'étendre et institutionnaliser les programmes qui réussissent et stopper ceux qui ne marchent pas. Ce faisant, l'évaluation participerait à l'amélioration des politiques et de l'humanité entière. Tel ne fut pas vraiment le sort réservé aux sciences sociales. La recherche sur l'utilisation s'est développée sur la base de ce constat un peu frustré, cherchant en particulier les éléments non rationnels jouant un rôle important.

Voici le libellé des *définitions*<sup>12</sup> tel que nous l'avons soumis aux Offices.

Degré	Genre	Définition	FONCTION
1	« Pas d'effet ».	Travail enregistré, mais pas de suite sensible donnée, ou résultat non intégré, pour toutes sortes de raison	PAS DU TOUT
2	« Je ne sais pas », influence non nette	Le travail fournit une information utile, mais son influence reste indirecte, difficile à saisir, plutôt de l'ordre conceptuel pour les lecteurs avertis. <b>Ou bien</b> : aucune information n'a été retrouvée sur cette étude	FAIBLE
3	« Discussion dans un groupe de travail »,	Le travail a fait l'objet d'une discussion interne, peut-être aussi avec l'auteur. L'influence reste faible	UN PEU
4	« Discussion générale ».	Le travail a été présenté à un groupe plus large et a été l'objet d'une discussion plus approfondie	MOYEN
5	« Effet localisé »	Le travail a été explicitement, quoique partiellement, intégré dans le travail ultérieur de l'office sur le sujet	FORT
6	« Usage direct »	L'essentiel du travail a été repris et a modifié le cours du travail ultérieur sur le sujet dans l'Office/Service	TRES FORT

### 3.2. SCIENCE ET POUVOIR OU EVALUATION VS. UTILISATION

Le fameux dilemme entre science et pouvoir est au cœur de la problématique de l'évaluation. Celle-ci met en présence des acteurs et des besoins issus de ces deux mondes. D'un côté le politique, que ce soient les employés de l'administration ou les parlementaires, intéressés à la régulation de la société, de l'autre des scientifiques, experts et évaluateurs, consacrés à l'objectivité de la recherche scientifique.

Dans la mesure où l'évaluation cherche à mesurer les effets de l'action étatique, voire à mesurer les résultats par rapport à des objectifs, l'évaluateur n'est pas seulement un expert au service de l'exercice d'un pouvoir, mais intervient directement dans la forme et l'exercice de celui-ci. Et le sort réservé à son étude est le lieu par excellence où cette relation entre science et pouvoir s'exprime finalement, le lieu où le pouvoir montre ce qu'il entend faire des

<sup>12</sup> Nous nous sommes inspirés des distinctions proposées par Willy Zimmermann *Un inventaire des évaluations au niveau fédéral*, in K. Horber (éd.), *Evaluations des politiques publiques en Suisse*, 1990, p.86

connaissances apportées par l'évaluateur. En nous penchant sur l'utilisation des évaluations, nous nous intéressons donc à la résolution d'une *relation de pouvoir* entre le politique, qui a commandité ou financé l'étude, et le scientifique qui l'a réalisée avec une vision sectorielle.

On peut même se demander s'il n'y pas un antagonisme irréductible entre les deux termes du titre de notre paragraphe. En effet, si l'on suit certains sociologues des organisations, le pouvoir y est défini (Crozier&Friedberg 1977 : 59ss) comme l'art de préserver et maîtriser une marge de manoeuvre par rapport aux autres acteurs. Cette marge de manoeuvre se construit essentiellement sur la zone et la quantité d'*incertitude* et d'*imprévisibilité* qu'un acteur peut maintenir face à un autre. Or le but général de toute démarche évaluative est de diminuer, voire de supprimer les éléments d'incertitude dans les programmes, de montrer les logiques parfaitement prévisibles et démontrables dans la conceptualisation et la mise en œuvre d'une pensée politique, de diffuser un savoir supplémentaire sur eux et/ou ces incertitudes de manière assez large. Ainsi le concept d'utilisation fait-il directement référence au monde des *décideurs*, de leur contexte et de leurs contraintes. Tandis que l'évaluation implique une *réflexion* la plus objective possible et hors de toute pression.

### **3.3. LE CADRE THEORIQUE GENERAL DE L'UTILISATION DES EVALUATIONS**

Quels sont les facteurs qui influencent l'utilisation d'une évaluation ? Partant du fait que les études recensées dans notre base de données entrent dans un processus de prise de décision, nous devons être attentifs aux éléments pouvant entrer en ligne de compte dans ce genre de situation (cf. Rossi&Freeman 1993 : 459ss., Patton 1986 : 107ss., Chelimsky 1986 : 74ss.).

#### **3.3.1. Eléments généraux**

- Les *types* d'évaluation jouent un rôle dans ce domaine. En reprenant la distinction entre une évaluation sommative ou formative, on peut imaginer intuitivement que les évaluations formatives seront plus marquées par la raison et la démonstration scientifique, tandis que les évaluations de type sommative, même si elles ont la même qualité scientifique, sont susceptibles de se prêter mieux à une approche plus émotionnelle, ou politique. En fonction du type de décision ou de groupes d'acteurs chargés de prendre ces décisions, cela peut jouer un rôle.
- De la même manière, on peut émettre l'hypothèse que la *fonction* attribuée à l'évaluation aura une influence sur l'utilisation qui en sera faite<sup>13</sup>:

---

<sup>13</sup> Pour les définitions, cf. p. 20.

- La légitimation: il est dans ce cas raisonnable de supposer un fort souhait ou besoin de pouvoir utiliser l'étude.
  - Le contrôle : là également, l'attente pour une forte utilisation de l'étude sera haute, en particulier venant de milieux parlementaires.
  - L'information : cela peut être aussi une première étape menant à l'une des deux autres fonctions mentionnées ci-dessus. Ici il est difficile d'estimer de manière générale l'intensité de l'utilisation qu'on peut attendre d'une étude menée avec cette fonction.
  - Le pilotage : l'évaluation ici est clairement un outil de gestion d'une politique et sert à ajuster tel ou tel élément. Son utilisation est donc très directement d'ordre politique et non seulement d'ordre administratif ou vaguement policier comme dans les cas cités plus haut. Une évaluation qui aurait cette fonction est théoriquement la plus orientée sur une utilisation, mais elle implique une volonté et une pratique de suivi politique du sujet et peut remettre en question certains choix opérés. Elle semble donc à priori être plus difficilement prévisible et régulière en ce qui concerne son utilisation.
  - L'argumentation : bien évidemment, dans la mesure où les conclusions de l'étude confortent l'opinion du mandataire, avec le danger de l'instrumentalisation, ou des études de complaisance.
- D'autres facteurs entrent en outre en ligne de compte pour influencer directement l'utilisation qui sera faite de l'évaluation. Par exemple *le type de politique* choisie. C. Weiss (1986) a déjà montré que des évaluations qui ne s'intéressent pas à la structure du pouvoir politique ont plus de chances d'être bien reçues que celles qui le remettent en question. De même, s'intéresser à des vieilles politiques bien établies aura moins de chances de succès que d'étudier un phénomène nouveau.

### 3.3.2. Eléments structurels (cf. Patton 1986 :107ss.)

- La question de l'utilisation pose bien sûr la question de *qui va l'utiliser*. Les intérêts quant à l'utilisation ne sont pas du tout les mêmes entre les décideurs du programme, les réalisateurs, les bénéficiaires et l'évaluateur lui-même, pour ne prendre que ces acteurs. Il n'est pas possible pour une évaluation de répondre de manière égale aux besoins de tous les acteurs. Quelqu'un qui ne participerait au processus qu'au moment de la réception des résultats est certainement moins intéressé à les utiliser que celui qui a fait activement toute la démarche.
- Sur le plan temporel, remarquons qu'il est difficile de déterminer *quand* l'évaluation sera utilisée. Il peut se passer pas mal de temps jusqu'à ce que ses décisions ou découvertes

soient mises en œuvre. Les processus de décision sont plus qu'on le croit de l'ordre incrémental et progressif, et l'immédiateté de l'exploitation de résultats est plutôt l'exception.

- *Comment l'évaluation est-elle utilisée ?* Bien sûr il y a les effets planifiés et formels. Mais qu'en est-il des effets indirects, involontaires, informels, lors de discussions de groupes, par le bouche à oreille ou des articles de journaux ?
- Un autre cercle de questions tourne autour de *ce qui va être utilisé* dans l'évaluation. On imagine a priori que ce qui est d'abord utile, ce sont les recommandations et les résultats. Mais la recherche a montré que dans un nombre important de cas, d'autres éléments ont été très utiles, et en particulier le processus lui-même.
- Une autre variable intervient encore dans le processus d'utilisation, à savoir *où* l'évaluation est utilisée. Les attentes et intérêts sont différents au niveau local, régional ou national, pour un programme à l'étranger ou en Suisse. Une évaluation ne peut pas satisfaire pleinement aux attentes de tous les niveaux.
- Enfin, last but not least, on ne saurait passer sous silence le *rôle de l'évaluateur lui-même* dans tout le processus. Au-delà de la valeur scientifique de son travail, il joue souvent un rôle déterminant dans la valorisation de ses découvertes.

### 3.3.3. Eléments techniques<sup>14</sup>

- La *punctualité et l'opportunité* ne sont pas à négliger. Un rapport arrivant en retard ou traitant d'un phénomène déjà dépassé a peu de chances d'échapper au placard.
- La *relevance*. Les résultats d'évaluation ne sont utilisables que s'ils répondent aux objectifs et aux questions du mandataire et parlent de sujets intéressants. Les chercheurs laissent parfois parler leur hobby ou leur intérêt particulier ou personnel.
- *Les buts, les intentions*. L'utilisation d'une évaluation peut être tout à fait indépendante du souci d'informations supplémentaires, mais servir les intérêts personnels des politiciens ou des mandataires, par exemple pour se faire apprécier ou reconnaître. Ces attentes, explicites ou implicites, jouent un rôle parfois sous estimé.

---

<sup>14</sup> Ces éléments reprennent-quelques uns des standards d'évaluation de la SEVAL particulièrement pertinents pour notre sujet. Cf. bibliographie.

- *La communication.* Pour augmenter les chances d'utilisation, celle-ci doit être permanente entre l'évaluateur et ses mandataires, et ceci bien avant que la moindre recommandation ne soit écrite.
- *La participation.* Tous les acteurs concernés par l'évaluation (du mandataire au bénéficiaire) devraient être impliqués dans la construction de celle-ci si l'on veut augmenter ses chances d'utilisation. Les demandes, réserves et sujets de désaccord doivent être abordés de manière conjointe.
- *Les résultats.* Leur compréhensibilité tout comme leur qualité sont bien sûr déterminantes. Mais aussi la manière de les présenter, leur simplicité. Cela doit être particulièrement soigné si ces résultats sont négatifs ou vont à l'encontre des positions ou attentes du mandataire principal.
- *La méthodologie* scientifique doit être irréprochable et les diverses techniques bien maîtrisées qu'il s'agisse de méthodes quantitatives ou qualitatives. A priori, les méthodes quantitatives tirées des sciences exactes sont plus susceptibles de convaincre. Mais il est logique également de voir dans une autoévaluation la meilleure technique pour utiliser des résultats. Une approche qualitative a pour elle le grand avantage de faire participer les gens à la démarche, ce qui est un plus pour l'utilisation future.
- *La crédibilité de l'évaluateur* est certainement un facteur décisif. Cela ne concerne pas seulement sa réputation scientifique, mais aussi son honnêteté intellectuelle, sa capacité à l'autocritique.
- *Une stratégie de valorisation.* Le souci de l'utilisation doit aussi se manifester par des pas concrets et planifiés dès le départ, aussi au niveau du budget, en vue d'une utilisation. Les démarches participatives sont ici certainement les mieux placées.
- *Les recommandations.* Celles-ci auront intérêt à être précises et bien mises en évidence pour être facilement utilisables.
- *Le traitement de l'information.* Il s'agit de la manière dont les informations issues de l'évaluation sont travaillées en termes de communication pour être transmises aux commanditaires. Il peut être important par exemple d'être attentif au potentiel de conflit qu'elles contiennent.

### 3.4. LES MERITES DE PATTON

Il n'est pas possible de terminer cette introduction sur les conditions cadres de l'utilisation des évaluations sans mentionner brièvement le travail de M. Q. Patton, qui a fortement influé sur la recherche et la pratique depuis une trentaine d'années, et dont l'approche est aussi à l'origine de notre étude. Il a poussé le plus loin la réflexion sur ce qu'il a appelé « utilisation-focused evaluation ». Pour lui (Patton 1997), une évaluation n'a pas de sens et ne devrait en définitive pas être lancée s'il n'y a pas dès le départ un potentiel élevé et réaliste de chances qu'elle soit utilisée. L'utilisation d'une évaluation est la seule aune à laquelle mesurer la qualité de celle-ci. Le sens premier et le plus important de l'utilisation est « intended use by intended users ». Dans son esprit, les utilisateurs principaux des évaluations à privilégier et à cibler sont bien les bénéficiaires des programmes et pas seulement - mais aussi - les employés de l'administration.

On peut regrouper les présupposés à la base du modèle de Patton de la manière suivante.

1. Le souci de l'utilisation doit être le moteur de l'évaluation. Chaque fois que l'évaluateur a une décision à prendre, il doit se demander quel effet cela aura sur l'utilisation de l'étude.
2. Le souci de l'utilisation doit accompagner le travail pendant toute sa durée et pas seulement à la fin.
3. L'évaluation doit être orientée sur l'utilisateur, son identité, ses besoins spécifiques. Il faut donc identifier avec soin les utilisateurs premiers de l'évaluation et analyser leur position par rapport à des décisions qui découleront de l'étude.
4. Les utilisateurs potentiels (stakeholders) doivent être intégrés dans les processus décisionnels autour de l'évaluation.
5. Se soucier de l'utilisation implique des coûts supplémentaires non négligeables.
6. Tout une gamme de facteurs vont influencer l'utilisation qu'il s'agit d'analyser, comme par ex. la nature de l'étude, la crédibilité de l'évaluateur, etc.

Patton s'intéresse bien sûr à l'usage direct des évaluations et développe une série de techniques et de conseils pour accroître leur impact. Mais il ne néglige ni ne méprise l'usage d'influence. Son évolution biographique semblerait même indiquer que ce type d'usage est devenu de plus en plus important pour lui au fil des ans<sup>15</sup>. Néanmoins, le fait de concentrer son effort sur les protagonistes directement impliqués dans l'évaluation (les « primary users ») renforce cette attente de démonstration d'usage direct. Ses remarques et ses recommandations sont donc bien en phase avec le cadre et les objectifs de notre recherche.

---

<sup>15</sup> Cf. son intervention lors de la Conférence de l'European Evaluation Society (EES) de Lausanne en octobre 2000 ([www.europeanevaluation.org](http://www.europeanevaluation.org))

Armé de toutes ces informations et considérations théoriques, nous pouvions lancer notre propre enquête sur le plan suisse. Il nous fallait toutefois être attentif à quelques-unes des spécificités du système suisse.

### **3.5. LES PARTICULARITES DU SYSTEME SUISSE**

Certaines particularités du système politique suisse et de ses institutions doivent encore être mentionnées dans ce contexte. En effet, elles peuvent à la fois compliquer et nuancer certains des aspects signalés ci-dessus, inspirés essentiellement des pratiques anglo-saxonnes et allemandes. Nous choisissons de placer ces informations à ce moment de notre démarche, car elles ne concernent pas seulement la définition de l'évaluation, mais aussi directement la pratique et l'*utilisation* des évaluations dans le contexte suisse. Elles font partie de notre approche théorique des bases de données.

- *La démocratie directe* peut être à la fois un frein et un catalyseur pour les évaluations. Comme le rappelle W. Bussmann (1998 :14 et 31), lancer une évaluation sur un sujet qui vient d'être l'objet d'une votation, c'est la condamner à n'avoir qu'un effet très limité. A l'inverse, publier des évaluations peu avant des votations sur les sujets en jeu peut renforcer leur impact si on sait utiliser la « fenêtre d'opportunité politique », au risque d'une instrumentalisation. De manière générale, on peut considérer que la fréquence à laquelle les citoyens suisses sont appelés à donner leur opinion sur toutes sortes de sujets rend l'utilisation des évaluations plus difficiles, à cause de cet effet « veto » de chaque votation, qui est perçue un peu comme une évaluation. L'utilisation de toute étude sera influencée par cet élément. En outre le système des processus de consultation organisés pour chaque loi fait effet de feed back permanent diminuant les chances d'évaluation externes.
- *Le système fédéraliste* est un autre obstacle. D'une part, parce que le niveau de pertinence de chaque étude (communal, cantonal ou fédéral) limite d'autant son utilisation à un autre échelon du système politique suisse. D'autre part, parce que les législations cantonales sur le même sujet sont souvent différentes et rendent l'exercice d'une évaluation comparative plus difficile, sans parler de l'emploi de ses conclusions dans un autre canton. Enfin, le fédéralisme d'exécution des lois rend souvent l'évaluation de mesures fédérales pénible, vu la variété des applications et adaptations des cantons. S'interroger sur les causes expliquant l'utilisation d'une évaluation devra le cas échéant tenir compte de cette composante.
- *La faiblesse de l'administration centrale.* Ce point n'est pas sans lien avec le précédent. De manière générale, l'administration centrale suisse est peu équipée et sous dotée en

ressources en comparaison avec les pays voisins. L'histoire du pays et de la construction de son système politique en sont les principales causes. Le domaine des statistiques sur le plan fédéral est un exemple souvent cité. L'administration fédérale dispose de peu de moyens pour effectuer elle-même des évaluations. Elle n'a encore que relativement peu d'expériences dans la réflexion sur l'évaluation et l'utilisation, peu ou pas de banques de données centralisées et de forum institutionnalisé d'échanges scientifiques. Cela influence également l'utilisation.

## **4. DEROULEMENT ET RESULTATS DE L'ENQUETE**

### **4.1. LE DEROULEMENT DE L'ENQUETE**

Les données de base de notre enquête étaient les titres des évaluations effectuées et recensées entre 1985 et 1995. Cette liste n'existant pas (plus ?) sous forme électronique, nous avons dû commencer par la reconstituer sur un format électronique, en distinguant les titres issus de l'enquête de M. Bussmann de ceux de M. Zimmermann. Pour les clauses évaluatives, il suffisait de consulter le site de l'OFJ.

L'enquête proprement dite s'est déroulée en deux étapes, entre mai 2000 et février 2001. La *première étape* a consisté à écrire aux 25 Offices répertoriés dans notre banque de données en leur demandant de bien vouloir classer les titres d'évaluation qui les concernaient en fonction des 6 catégories d'utilisation que nous avons définies (cf. p.22), allant de « pas d'effet » à « usage direct ». A cause du grand nombre d'études de la DDC et de l'OFSP, nous avons établi pour ces Offices, et d'entente avec eux, des listes par Section ou unité plus petite. La lettre d'accompagnement qui présentait l'objectif et le déroulement de notre démarche soulignait l'intérêt de l>IDHEAP à cette enquête. De manière générale, l'accueil à ce questionnaire a été très positif. Les questionnaires ont été renvoyés en majorité dans les délais. Si ce n'était pas le cas, nous avons été en général informés des retards prévus et de leurs causes. Des recherches souvent longues ont été entreprises dans les Offices pour pouvoir apporter les réponses. Plusieurs entretiens téléphoniques de clarification et des lettres avec des commentaires précis et des remarques pertinentes ont accompagné le retour des questionnaires.

Une variante de ce questionnaire a été utilisée pour aborder la question des *clauses évaluatives*. Le listing présent sur le site Internet de l'OFJ ne précisait pas les Offices concernés par la mise en œuvre de ces clauses. Après avoir attribué chacune à un Office, nous

avons adressé un petit questionnaire aux Offices répondant pour les 30 articles recensés. Cela concernait 13 Offices ou Services.

#### **4.2. LA PREMIERE ETAPE : LA DEFINITION DE L'UTILISATION**

Le taux de réponse au premier questionnaire fut très bon. Plus de 60 % des titres (371 sur 602) recensés au départ étaient classifiés selon notre typologie. Nous en avons retranscrit les résultats détaillés en *annexe 4*. Voici les principaux résultats obtenus :

##### *Résultats significatifs*

- *Plus d'un quart des études (27%) d'évaluation ont eu une influence directe (catégorie 6), un impact immédiat sur la suite de la gestion de la politique en question. En tenant compte des deux degrés d'utilisation les plus élevés (5 et 6), on obtient que 60 % des études ont été importantes ou très importantes pour la suite.*
- *Seules 5% des évaluations sont déclarées avoir été inutiles. En additionnant les résultats des deux catégories les plus basses, les évaluations ayant un impact faible ou nul s'élèvent à 12%.*
- *Env. 28% des études ont eu une influence moyenne, donnant lieu à des discussions plus ou moins importantes, mais où l'influence reste limitée.*

##### *Biais possibles*

Une lecture attentive du tableau des réponses appelle encore les remarques complémentaires suivantes :

- *Le type de listes fourni au départ par les Offices n'est pas homogène. Ainsi, l'Office fédéral de la statistique n'a pas été en mesure de répondre à notre questionnaire, car il ne pouvait donner aucune information sur l'usage fait dans les Offices concernés ou le Parlement des résultats des études menées. De même les études du FNRS ou de l'OFIAMT, dissous entre temps. Si l'on enlève les listes de ces Offices du total des réponses, on arrive à un *taux de retour de 81,6%* .*
- *14 Offices sur 25 ont répondu, soit 56 %. Si l'on fait la même réflexion que ci-dessus, on arrive à 63,6%. Tous ces éléments nous incitent à penser qu'il ne s'agit pas de réponses de complaisance, et que les résultats sont dans l'ensemble fiables. Néanmoins, beaucoup de cas sont restés sans réponses (par exemple à l'OFSP), principalement à cause du taux de rotation du personnel et l'absence de mémoire institutionnelle organisée. Nous avons pu obtenir quelques réponses supplémentaires grâce à des contacts personnels.*
- *Un biais important est celui de la propension naturelle à exagérer ou optimiser les résultats. Cela d'autant plus que quelques années avaient passé depuis.*

- Quelques résultats laissent apparaître des besoins d'informations supplémentaires. Ce sont des résultats qui sont tous favorables à l'hypothèse d'une forte utilisation et font donc augmenter le taux d'utilisation : ainsi le Conseil des EPF a estimé que tous les rapports cités avaient eu une utilisation très forte. L'OFSP pour sa part n'a évalué l'impact d'aucun rapport au dessous de l'appréciation « a provoqué une discussion générale » (note 4).

A cause de cela, mais aussi du nombre élevé d'études recensées à la DDC et à l'OFSP, et d'une méthodologie de l'évaluation très élaborée dans ces Offices, il nous a semblé intéressant de refaire les calculs sans ces 3 Offices. Les résultats détaillés sont également joints en annexe 4, ainsi que les résultats propres à la DDC, le plus grand fournisseur d'évaluations.

### *Résultats significatifs*

- Avec ce nouveau calcul, nous constatons quelques nuances importantes : d'une part, le *taux de réponses net descend à 52%*. D'autre part, les pourcentages par catégorie d'utilisation varient sensiblement, surtout dans les taux de forte utilisation. Ainsi le pourcentage des évaluations fortement et très fortement utilisées baisse de *60 à 47%*. C'est surtout le taux des évaluations très fortement utilisées (catégorie 6) qui baisse, passant de *27 à 16 %*. Le taux des évaluations inutiles reste à peu près stable et celui des évaluations à utilisation moyenne augmente proportionnellement.

### *Résultats provisoires*

Des renseignements fournis par ce premier questionnaire, nous tirons les résultats provisoires suivants :

1. La pratique de l'évaluation est relativement bien connue dans l'administration fédérale et en progression certaine. Sur l'ensemble des Offices, il n'y en a cependant encore qu'une minorité qui la pratique occasionnellement ou régulièrement. Et parmi les Offices qui la pratiquent, il y a encore de très grandes disparités aussi bien en termes d'habitude qu'en ce qui concerne l'intérêt ou l'infrastructure à disposition pour effectuer ces tâches, ou même l'intérêt pour ce domaine, celui-ci allant de la plus ouverte des collaborations avec l'enquêteur à l'absence totale de réaction. Les domaines où l'évaluation est la plus pratiquée sont ceux de la Coopération au développement (80 à 100 études par an), de la santé publique (20-30 par an), de l'énergie et de l'environnement et de l'agriculture.
2. La très grande majorité de ces évaluations (env. 80%) concerne des projets précis, ou des programmes assez pointus et limités dans le temps, souvent intégrés dans une stratégie d'action par étapes successives. *Il y a relativement peu d'évaluations larges ou thématiques.*

3. La très grande majorité des évaluations connaît une forte valorisation dans le processus de la gestion de la politique en question. *Si on peut estimer qu'environ 10 % des études n'ont pas d'impact réel ou significatif, nous constatons qu'environ 20%<sup>16</sup> ont joué un rôle très important pour la suite.*
4. 80 % des évaluations ont un degré d'utilisation plus fort que le degré moyen.
5. Néanmoins, il serait faux d'en déduire que l'utilisation de l'évaluation est le seul élément qui influence la suite de la politique en question. Cette enquête permet d'affirmer que *les évaluations ne sont un élément décisif pour définir la suite de la politique ou du programme que dans un cas sur deux environ<sup>17</sup>*. Est-ce beaucoup ou est-ce peu ? A priori cela semble énorme. Cela indique très clairement qu'il y a un suivi bien organisé des études commanditées par l'administration, et que rares sont celles qui ne servent à rien. Mais il s'agira d'identifier avec plus de précision le niveau sur lequel le changement s'est opéré dans la politique en question.
6. Il vaut la peine de comparer ces résultats avec des recherches similaires faites dans le contexte suisse récemment<sup>18</sup>.
  - Ainsi W. Zimmermann (Zimmermann in Horber 1990 : 86) enregistre-t-il des évaluateurs l'impression que 20% de leurs études ont eu un effet direct alors que 50% n'en ont apparemment pas eu. Il s'agit d'un résultat assez similaire en ce qui concerne la forte utilisation.
  - Widmer (Widmer 2001) transcrit l'opinion d'un responsable d'Office du DETEC estimant que plus de 50 % des recommandations des évaluations ont été appliquées. Mais on est bien ici au niveau de la mise en œuvre de recommandations techniques. Il est difficile d'en déduire l'importance des études pour l'évolution de la politique publique, ou d'extrapoler sur d'autres Offices. Nous ne voyons pas là non plus de contradictions fondamentales avec nos résultats.
  - Janett (Janett 2000) dans sa recension à l'Organe parlementaire de contrôle de l'administration (OPCA) arrive à 80-85% d'application des recommandations de la CdG par le Conseil Fédéral (CF). Mais il n'y a pas de chiffres disponibles pour évaluer le pourcentage de conclusions de l'OPCA reprises dans les recommandations de la CdG au CF. D'autre part, l'obligation faite au CF de donner suite à ces recommandations explique peut-être aussi ce taux très élevé. Il a été obtenu en outre sur la base d'un échantillon assez limité. Là non plus, rien n'est dit sur le niveau d'importance des changements opérés.

---

<sup>16</sup> Moyenne entre les 16% obtenus sans compter la DDC ni l'OFSP et les 27% avec ces deux Offices

<sup>17</sup> Idem. Moyenne entre les 60 % avec DDC et OFSP et 47% sans ces Offices.

<sup>18</sup> Pour ce qui est du contexte international, on lira avec intérêt les chiffres pas très différents contenus dans les études de John Mayne et de Frans L. Leeuw et Piet J. Rozendal, publiées dans : Leeuw Frans L. Rist Ray C. et Sonnichsen Richard C. (ed.), *Can governments learn ?*, Transaction publishers, 1993, p. 17 –43, 67-85.

### 4.3. RESULTATS DES CLAUSES EVALUATIVES

Le questionnaire visait à savoir le degré d'utilisation des rapports reçus sur la base de ces clauses ainsi qu'une appréciation sur le taux d'amélioration de l'efficacité de la Loi ainsi obtenu. Nous supposons que le taux d'utilisation serait plus élevé. Le taux de réponses fut de 73% (22 sur 30).

#### *Résultats*

La clause évaluative avait déclenché une évaluation dans 72% des cas. En ce qui concerne les 28% restants, environ la moitié des Offices signalait que la chose était sur le point d'être lancée. Les résultats plus détaillés sont disponibles en *annexe 5*.

#### *Degré d'utilisation de l'évaluation*

Faible (discussion en groupe)	0%
Moyen (discussion générale) :	8%
Fort (influence localisée):	69%
Très fort (influence directe) :	23%

#### *Commentaire*

- Même une clause évaluative ne suffit pas à déclencher une évaluation dans près de 15% des situations. Les raisons les plus souvent évoquées sont la complexité de la matière à évaluer.
- Contrairement aux études mentionnées plus haut, nous avons affaire ici à des évaluations assez larges et complexes.
- Comme pour le premier questionnaire, nous constatons que le taux de très forte utilisation se situe autour de 20%. En y ajoutant le taux d'utilisation « forte », nous atteignons déjà plus de 90% des études. C'est tout à fait considérable.
- Il semble donc possible d'en déduire provisoirement qu'une *clause évaluative renforce clairement l'utilisation* par rapport à une situation sans clause évaluative. Dans tous les cas, cela semble éviter l'absence complète d'utilisation.

En ce qui concerne l'amélioration de l'efficacité de la Loi grâce à l'évaluation, les résultats montrent ceci :

Aucune :	0%
Faible :	0%
Moyenne (discussion en groupe):	21%
Forte (discussion générale) :	68%
Très forte (influence directe) :	10%

### *Commentaire*

Près de 80 % de forte amélioration de l'efficacité de la Loi est là aussi très spectaculaire. Notons cependant qu'il s'agit là de l'appréciation des Offices chargés de la mise en œuvre, et non par exemple des groupes cibles. Nous n'avons pas d'information non plus sur ce qui fonde ce jugement.

#### **4.4. LES CAUSES EXPLICATIVES DE L'UTILISATION DES EVALUATIONS**

La deuxième étape de notre enquête a démarré après avoir répertorié et commencé à interpréter toutes les réponses reçues en fonction des 6 catégories possibles<sup>19</sup>. Nous avons constitué les deux groupes d'études d'évaluation : l'un avec des études dont les Offices nous avaient dit qu'elles avaient été « très utilisées » et avaient eu un effet direct, l'autre avec des études au sujet desquelles l'information était qu'elles n'ont pas été utilisées et sont restées sans influence. Nous n'avons retenu dans le groupe principal aucune étude provenant du sous-groupe des clauses évaluatives, car une partie de notre questionnaire, en particulier sur les circonstances qui ont vu naître la politique publique et la demande d'évaluation, était répondu par avance : par définition il y avait eu un travail de type parlementaire et une intervention de la direction de l'Office. Cela risquait de biaiser les résultats. Les études retenues dans le groupe principal ont été sélectionnées en fonction de l'importance du sujet évalué. Nous avons limité au maximum les études de type trop technique et favorisé les études plus larges, thématiques sectorielles ou stratégiques.

Nous avons en tout 49 titres, répartis ainsi : 27 pour le *groupe principal* et 22 pour le *groupe de contrôle*. Un questionnaire sur les facteurs ayant influencé l'utilisation été envoyé à chacun des mandataires des 49 évaluations. Ces questionnaires étaient munis d'un code de reconnaissance inconnu de nos interlocuteurs.

L'objectif du deuxième questionnaire était de réunir des informations permettant de vérifier ou d'invalider les hypothèses émises sur les causes qui expliquent pourquoi 16 ou 27% des évaluations étaient très utilisées et pourquoi 10% ne l'étaient quasiment pas du tout.

Le taux de retour fut de 80 % (40 études). Mais parfois avec un retard considérable et après plusieurs recherches d'information téléphoniques. Certaines questions ont été régulièrement laissées de côté, et ont fait l'objet de remarques critiques, en particulier celles sur la personnalité, la formation et l'appartenance politique de l'interviewé. Plusieurs questionnaires

---

<sup>19</sup> Raison pour laquelle nous avons préféré rendre compte des résultats de la première étape avant de présenter la seconde.

n'étaient pas remplis de manière suffisante pour pouvoir être utilisés. Quelques-uns ont été remplis par les consultants. Finalement, nous avons disposé de 30 titres pour établir des résultats, soit un taux de retour net de 61%<sup>20</sup>. Les titres étaient répartis entre 8 Offices de 6 Départements.

<b>Département / Groupe</b>	<b>Groupe principal</b> Nombre d'études	<b>Groupe de contrôle</b> Nombre d'études
DFAE	7	7
DFE	1	
DETEC	3	3
DFI	3	
DFJP	3	2
DDPS	1	
<b>Total</b>	<b>18</b>	<b>12</b>

Deux des études comprises dans ces listes provenaient de la liste de W. Zimmermann, toutes les autres de celle de W. Bussmann. Nous avons reproduit en *annexe 7 les résultats en %* pour chaque question et chaque groupe. Les deux groupes d'étude ont été enregistrés sur deux fichiers informatiques distincts ainsi que leurs résultats.

Pour présenter les résultats obtenus, nous reprendrons le libellé principal de chaque hypothèse et chaque variable explicative retenue, et nous comparerons les résultats obtenus dans le groupe principal des études très utilisées avec ceux du groupe de contrôle. Une différence significative des résultats constitue le cœur de la cause explicative recherchée.

### *Discussion des hypothèses*

#### **Hypothèse 1. Les circonstances dans lesquelles la politique publique à évaluer a été élaborée**

- Le principal résultat ici est la confirmation que *la politisation du sujet de la politique publique renforce sérieusement les chances pour une évaluation d'être utilisée*. Par « politisation », nous entendons l'existence d'un débat public sur la politique publique en question, que ce soit sous la forme d'un débat parlementaire, d'un travail de Commission parlementaire ou extraparlamentaire. Qui dit débat dit aussi en général conflit d'opinion. Plus de la moitié des évaluations très utilisées correspondent à des politiques publiques « politisées », contre seulement 9% pour le groupe de contrôle. Une évaluation faite sur un sujet qui n'a pas été discuté publiquement a moins de chances d'être utilisée. Il y a une certaine logique dans ce résultat, dans la mesure où le passage d'un sujet devant le

<sup>20</sup> Cf. liste en Annexe 6.

Parlement augmente le nombre des acteurs informés et intéressés à en connaître le suivi. Vu la composition de notre groupe principal, cela signifie également que les évaluations de type et d'importance plus grande que des évaluations de simples projets sont également très utilisées par l'administration, même si elles sont plus rares.

- Lorsqu'est évoquée *la raison* pour laquelle la politique publique a été évaluée, près de 30% des études très utilisées nomment la « politisation du sujet », contre 0% au groupe de contrôle. Le besoin d'informations est également cité dans les deux groupes comme un des facteurs déclencheurs principaux. Notons enfin que des *déclencheurs* comme une « mise en œuvre difficile » ou la « pression de lobbys/médias » est légèrement plus représentée dans le groupe des évaluations très utilisées. La différence entre les deux groupes est cependant trop peu importante pour en faire une cause explicative sans autres recherches.
- La *longueur du processus* d'élaboration est aussi un facteur séparant les deux groupes : dans le cas du groupe principal, plus de 40% des cas ont connu une durée d'élaboration supérieure à deux ans, contre 16 % au groupe de contrôle. Cela va de pair avec l'existence de débats au Parlement.
- Les cas où la politique publique était sujette à un *conflit* concernent plus du tiers des études du groupe principal, soit *deux fois plus* que dans le groupe de contrôle. Le processus de consultation est en particulier le lieu où ce conflit s'est exprimé. Le conflit ne semble donc pas nécessairement être un problème pour qu'une évaluation soit utilisée. Il faudra encore voir sur quoi porte ce conflit, ou ce qui le caractérise. En effet, dans la majorité des cas des deux groupes, le sujet n'était pas conflictuel.

## **Hypothèse 2. Les circonstances dans lesquelles l'évaluation elle-même est élaborée, organisée et réalisée**

- Il ressort clairement ici que la participation active, voire *l'initiative de l'administration* est absolument indispensable pour qu'une évaluation soit utilisée. Cela concerne 85% des études très utilisées, contre 65 % dans le groupe de contrôle.
- A noter également, et cela rejoint le point précédent, que 12% des études très utilisées sont *demandées par le Parlement*, contre 0% au groupe de contrôle. L'origine interne est donc un facteur important, que cela soit la Direction de l'Office, ses collaborateurs ou le vœu du Parlement pour qu'une étude ait un fort impact. Cela conforte également l'aspect « politisation » évoqué.
- A contrario des éléments relevés ci-dessus, il est utile d'observer que cela n'est qu'un facteur parmi d'autres. En effet, l'origine de l'évaluation est aussi très majoritairement interne dans le groupe de contrôle où les études n'ont eu aucun effet.

- On remarque aussi que plus les *contacts* entre l'évaluateur et l'Office sont nombreux, plus les évaluations sont utilisées (30 % contre 16%).
- Par contre, le fait de confier l'évaluation à *quelqu'un d'externe* à l'administration ne semble pas avoir d'impact direct sur son utilisation. La différence n'est pas significative.
- L'utilisation est plus fréquente lors d'évaluations *ex ante ou concomitantes*. C'est compréhensible dans une situation ex ante, car le but de l'évaluation est souvent de décider si on va réaliser le programme ou non. La différence au sujet des études concomitantes est explicable certainement pour les mêmes raisons. Par contre, une approche ex post est la moins à même de garantir une utilisation. Est-ce à dire qu'un jugement a posteriori a moins de chances de concerner beaucoup de gens ?
- Un résultat surprenant concerne le *degré de familiarité ou de formation* du personnel concerné de l'Office dans le domaine de l'évaluation. Certes, dans les deux groupes, il y avait du personnel formé à l'évaluation dans la majorité des cas. Toutefois, le pourcentage est sensiblement plus faible dans le groupe principal (60%) que dans le groupe de contrôle (85%).
- Par contre, une *longue expérience* dans l'administration semble être un atout pour valoriser les résultats de l'évaluation. Dans 65% des cas des études utilisées, le responsable d'Office travaillait depuis 10 ans ou plus dans l'administration, contre 25% pour le groupe de contrôle.
- Nous n'avons pas réussi à avoir des informations suffisantes sur la question de l'influence de l'appartenance à un parti politique des responsables de l'administration. Outre le motif du respect de la sphère privée, le taux de rotation du personnel fédéral responsable du dossier nous a empêché d'obtenir des réponses valides. Sur la base des réponses reçues, ce que l'on peut dire est que le fait d'appartenir à un parti politique ne semble pas représenter une variable explicative fiable.

### **Hypothèse 3. Les circonstances dans lesquelles l'évaluation et ses résultats sont reçus**

- La *médiatisation* du conflit cité plus haut ne semble pas jouer de rôle très important, ni son urgence. Les différences ne sont pas significatives. Par contre l'*actualité* du sujet a joué un rôle important (40% dans le groupe principal contre 8,3% pour le groupe de contrôle). Il y a là une information intéressante sur la nature du conflit sur laquelle nous reviendrons. Mais la décision sur l'utilisation de l'évaluation en général précède le conflit plutôt qu'elle ne le suit. Idem en ce qui concerne l'urgence du sujet.
- Par contre l'utilisation varie considérablement selon les *fonctions* attribuées à l'évaluation : une attente de légitimation et d'information garantissent une bonne utilisation. Les études très utilisées avouent une fonction d'*information forte à plus de 90%*, contre 40% pour le groupe de contrôle. Cela rejoint le besoin d'information comme

déclencheur important des évaluations cité plus haut. Le même constat touche de manière déjà moins nette la fonction de *légitimation*, forte pour 70% des cas très utilisés, contre 50% pour le groupe de contrôle. C'est nettement moins clair pour la fonction de contrôle, où la différence entre les deux groupes n'est pas significative. Enfin, la fonction d'aide au *pilotage* d'une politique publique ne garantit qu'un avantage modeste quant à son utilisation. Le groupe principal reconnaît la forte importance de cette fonction dans 65% des cas, contre 50% dans le groupe de contrôle.

#### **Hypothèse 4. Les caractéristiques du sujet de l'évaluation**

- Probablement lié à la dernière remarque, nous observons que la *nouveauté* d'un sujet implique plus de chances pour un rapport d'évaluation d'être utilisé. La majorité des titres très utilisés (53%) concerne une nouvelle politique, contre 17% pour le groupe de contrôle.
- En ce qui concerne *le contenu* qu'on cherchait à évaluer avec cette étude, le groupe des études très utilisées se distingue du groupe de contrôle surtout sur deux éléments, à savoir d'une part *l'atteinte de l'impact* (groupe-cible), 53% contre 33% des cas, et d'autre part *l'efficacité* politique (résolution du problème). La différence est toutefois ici déjà plus mince, soit 35% pour le groupe principal contre 25% au groupe de contrôle. Une autre différence significative concerne la question de *l'efficacité*. Alors que le groupe de contrôle contient beaucoup plus d'études orientées sur *l'efficacité du programme* (rapport coût-résultat) que le groupe des études très utilisées (42% contre 18%), celui-ci contient plus d'études orientées sur *l'efficacité administrative*, (rapport entre coût de la mise en œuvre et la résolution du problème). La plupart des études signalaient cependant plusieurs buts. Manifestement, être renseigné sur l'efficacité d'un programme n'est pas suffisant pour utiliser plus avant une étude. Par contre, vérifier que le problème ait été résolu et à quel prix est une information plus recherchée et utile.

#### **Hypothèse 5. Les aspects formels du rapport d'évaluation**

- Quelques surprises sont à enregistrer ici. Ainsi, il n'y a pas de différence vraiment significative entre les deux groupes en ce qui concerne *l'importance de recommandations précises, l'importance de leur place ou leur degré de précision*. Dans les deux cas, à plus de 75% les recommandations étaient faites, précises et bien mises en valeur. Ce n'est donc pas cet élément qui a fait la différence entre les deux types d'utilisation.
- Le même constat vaut pour la *rapidité* avec laquelle le rapport est rendu et le respect de *règles déontologiques*.

- La *simplicité* du rapport joue certes un rôle, mais ne constitue apparemment pas une clé décisive de succès. La majorité des études utilisées étaient d'un «*abord complexe ou très complexe* », contre 40% au groupe de contrôle. La *brièveté* non plus: 65% des études très utilisées ont autour de 100 pages, contre seulement 25% pour le groupe de contrôle.
- La *nouveauté de la méthodologie* utilisée semble pouvoir être aussi bien un avantage qu'un inconvénient. Dans les deux groupes la très grande majorité des cas utilisent une méthodologie connue ou bien connue. Néanmoins, le groupe des études très utilisées est le seul à signaler l'emploi de *méthodes nouvelles ou très nouvelles* dans près de 20% des cas.
- La technique *qualitative* semble plus à même de provoquer une utilisation forte. Cette technique a en tout cas été utilisée deux fois plus souvent dans des études très utilisées que lors des études négligées. La majorité des cas utilise toutefois une technique mixte. Ce qui fait la différence entre les deux groupes ici est que les études utilisées ont plutôt *choisi clairement entre l'une ou l'autre technique*. Près de 85% des études négligées ont systématiquement employé une technique mixte.
- La présence *d'une traduction* ne joue pas de rôle important. La différence n'est pas très significative.
- La présence *d'un résumé* donne un avantage, mais celui-ci reste léger, car la majorité des études des deux groupes en contenait un.
- L'importance de *l'implication de l'évaluateur* apparaît clairement : près de 60% des études très utilisées le mentionnent, contre 25% pour le groupe de contrôle.
- De même, la *médiatisation du rapport d'évaluation* aide manifestement à le prendre sérieusement en considération. Dans 35% des cas, les études très utilisées ont été fortement ou très fortement médiatisées, contre 8% au groupe de contrôle. La différence est d'autant plus forte si cette diffusion relève d'une *décision de l'Office* mandataire (47% contre 8%).
- La présence d'un poste et d'une stratégie de *valorisation* dans le budget de l'étude en est bien sûr une des conditions préalables. 35 % des études très utilisées en avaient une contre seulement 8% pour le groupe de contrôle. Notons toutefois que près de 47 % des études très utilisées n'en avaient pas ! Le pourcentage du budget réservé à cette valorisation allait dans près d'un tiers des cas jusqu'à 10% de l'ensemble du budget de l'étude.
- Enfin, l'utilisation varie également sensiblement en fonction du *degré de diffusion* des résultats à l'intérieur de l'Office et à l'extérieur de l'administration. Nous constatons une diffusion forte ou très forte dans 70% des cas du groupe principal, contre 50% pour le groupe de contrôle. Nous n'avons pas d'information sur le taux de diffusion au delà du strict Office, dans les niveaux hiérarchiques supérieurs.

Afin d'essayer d'interpréter ces résultats, nous avons voulu effectuer quelques tris croisés entre quelques variables à l'intérieur de chaque groupe.

### *Les corrélations*

Afin de vérifier nos commentaires, nous avons soumis certains résultats obtenus au test de corrélation dit du  $X^2$ , en particulier en ce qui concerne le *groupe principal*. Nous voulions en particulier revenir sur les résultats étonnants ou peu clairs. Cela nous permet de vérifier (sous réserve<sup>21</sup>) ou infirmer une corrélation des réponses pour les résultats suivants :

- Il y a une corrélation significative, voire très significative entre toutes les questions liées à la variable « politisation », y compris par rapport aux conflits et aux raisons qui ont fait se réaliser l'évaluation (Hypothèse 1). Cette corrélation n'est confirmée que pour le groupe des études utilisées, et pas pour le groupe de contrôle. Nous avons là une confirmation que l'utilisation d'une évaluation est liée au processus de discussion politique du sujet de la politique publique en question.
- Il n'y a pas de corrélation entre la politisation et la fonction d'utilisation de l'évaluation. Autrement dit, le processus de politisation d'une politique publique n'induit pas de fonction particulière de l'évaluation lorsqu'elle sera réalisée. Cela est valable pour les deux groupes. On aurait pu imaginer qu'une politique publique très débattue ait induit une attente de contrôle ou de légitimation particulièrement forte.
- En ce qui concerne les facteurs affectant favorablement l'utilisation, nous observons qu'il n'y a pas de corrélation entre la conflictualité potentielle de la politique publique évaluée et son actualité médiatique. Ce sont d'autres raisons que l'actualité qui font qu'une politique publique se voit politisée, comme par exemple son contenu. Et c'est de cela dont dépend son utilisation. Ce résultat n'est valable que pour le groupe des études très utilisées. En effet, pour le groupe de contrôle, il y a au contraire une corrélation entre la conflictualité du sujet et l'actualité médiatique. En cas de conflictualité et de médiatisation du sujet, nous avons donc un indice que cela portera préjudice à l'utilisation de l'évaluation.
- Il y a une corrélation significative entre la nouveauté du sujet et la politisation. Plus la politique en question est nouvelle, plus il est logique que des débats clarifient le sujet. Tout cela renforce les chances d'utilisation de l'évaluation.
- Par contre, la nouveauté d'une politique n'est pas liée à la fonction qu'on attribue à son évaluation. Evaluer une nouvelle politique ne présage en rien encore de l'usage qu'on voudra faire du rapport.

---

<sup>21</sup> Nous ne mentionnons ici que les résultats « significatifs » ou « très significatifs » et « l'absence de corrélation » dans les cas où on aurait pu en attendre une. La réserve provient du fait que les conditions pour réaliser le chi carré ne sont pas toujours données, car il n'y a pas toujours au moins 5 réponses par catégorie de réponse possible. Nous les mentionnons tout de même, car la probabilité de la confirmation des résultats avec une masse critique de réponses suffisante pour valider le chi carré est très forte.

- Le même constat peut être fait en ce qui concerne l'absence de lien entre la nouveauté et le contenu de l'évaluation. Evaluer une nouvelle politique n'implique pas qu'on cherche plus à savoir par exemple si l'on atteint ou non le groupe cible que dans une politique traditionnelle.
- Il y a par contre une corrélation très significative entre le contenu de l'évaluation et la fonction de pilotage dévolue à celle-ci. Ce qu'on cherche à évaluer dans une politique publique est lié à la volonté de la piloter. C'est dans ces cas-là que l'utilisation est la plus forte. Cette corrélation est donnée uniquement pour cette fonction et pas pour les autres (légitimation etc.). Elle ne concerne que les études très utilisées et pas celles du groupe de contrôle.

#### **4.4. VALORISATION DES RESULTATS**

Notre double enquête nous a permis d'accumuler une grande quantité d'informations, aussi bien sur le genre d'utilisation des évaluations que sur les causes de cette utilisation. Nous aimerions en particulier souligner les résultats suivants qui sont autant de réponses à nos hypothèses :

En ce qui concerne *le taux d'utilisation* des évaluations :

- que l'on parle des 20% de très forte utilisation ou des 50% de forte utilisation, nous avons à faire à un *bon, voire très bon résultat*. Il rejoint, voire dépasse les résultats obtenus dans d'autres enquêtes en Suisse et à l'étranger<sup>22</sup>.
- Ce résultat témoigne du fait que dans les Offices qui la pratiquent, l'évaluation est un *outil précieux et bien intégré* dans les processus de travail et de décision.
- Ces résultats montrent aussi que nous avons à faire à un type d'évaluation majoritairement centré sur des *projets et programmes précis*, mandaté pour l'essentiel pour répondre aux besoins de travail des Offices *à ce niveau-là*. La pratique de l'évaluation sur de plus grands objets est encore relativement faible pour la période observée. Toutefois, la composition de notre groupe principal ne laisse pas supposer que le taux d'utilisation soit radicalement différent. Nous n'avons que peu d'informations sur l'utilisation des évaluations par d'autres acteurs que l'administration.
- Nous observons également que la présence de *clauses d'évaluation* agit favorablement sur l'utilisation des études et contribue aussi bien à l'institutionnalisation de l'évaluation qu'à l'émergence d'une culture d'évaluation.

---

<sup>22</sup> (Cf. Zimmermann in Horber : 1990, Janett : 2000 ; Cf. aussi note 16, p. 32)

En ce qui concerne les *causes explicatives de l'utilisation* des évaluations :

- Notre première hypothèse prétendait que les processus de type politique précédant l'adoption d'une politique publique augmentaient les chances pour une évaluation d'être utilisée. *Nos résultats confirment que plus ce processus est élaboré, public et long, plus les évaluations futures seront utilisées. Le côté conflictuel de tout débat de politique publique ne gâche en rien les chances d'utilisation, au contraire, il semble légèrement les favoriser, à condition que ce conflit ne soit pas dépendant ou causé par une médiatisation forte, mais reste lié au contenu des décisions à prendre, aux options de base des courants politiques. Cet élément vient confirmer l'essentiel de la théorie sur le sujet (Janett 2000 :9). Quant au processus de politisation lui-même, nous n'avons pas trouvé de discussion théorique sur cet aspect et le considérons provisoirement comme un élément propre à notre recherche. Il est probable que les particularités du système démocratique suisse jouent un rôle sur ce point.*
  
- Notre seconde hypothèse liait le sort dévolu à une évaluation à l'intégration de celle-ci dans les processus internes de travail et de décision de l'administration. *Nos résultats confirment l'importance décisive de l'origine interne (y compris le Parlement) aussi bien de l'initiative de l'évaluation que de celle de sa gestion. Les évaluations imposées depuis l'extérieur n'ont que peu de chances de se voir valorisées, de même que celles qui ne font pas activement participer l'administration au projet d'évaluation. Cet élément structurel semble plus important que la présence en soi de collaborateurs bien formés en évaluation, même si celle-ci est bien sûr souhaitable et un atout. Mais l'élément décisif est que ce projet d'évaluation soit une affaire gérée par l'Office en tant que tel. Quoi qu'il en soit, nous avons ici une confirmation de l'importance de toute démarche participative telle que relevée par la littérature (Cf. par exemple : Rossi&Freemann 1993 : 61).*
  
- Notre troisième hypothèse mettait l'accent sur les conditions dans lesquelles le rapport d'évaluation était reçu. Nous partions en particulier du principe que l'attente existante par rapport à l'évaluation jouait un rôle, ainsi que l'influence de l'actualité et des médias au moment de la remise du rapport. *Nos résultats confirment que ce qui est surtout attendu et utilisé dans une évaluation est ce qui permet à l'administration d'être bien informée de l'évolution et du déroulement de la politique publique, ainsi que de la légitimer. La seule fonction de contrôle ne suffit pas, probablement car elle est aussi de la compétence d'autres acteurs du système politico-administratif. L'importance de la fonction de pilotage est plus difficile à cerner. La différence entre les deux groupes est un peu moins significative. Mais en même temps,*

*elle est la seule fonction clairement corrélée au contenu de l'évaluation. Cela souligne le rôle reconnu de l'administration dans la gestion du programme à évaluer. L'influence des médias en particulier joue à ce moment-là un rôle non négligeable, qui est souvent intégré dans les résultats de l'évaluation elle-même (stratégie de valorisation). Par rapport à la littérature (Rossi & Freemann 1993 : 61), nous observons ici un renforcement de l'idée de « dissémination des résultats » souvent évoquée. L'idée de médiatisation est plus forte et rend compte de l'évolution de nos sociétés ces dernières années.*

- Notre quatrième hypothèse voulait voir un facteur d'influence de l'utilisation dans le sujet et le contenu même de l'évaluation.

*Nos résultats ne confirment que partiellement cette hypothèse. Certes, en ce qui concerne le sujet, il est démontré que l'évaluation d'une nouvelle politique publique a bien plus de chances d'être utilisée que dans le cas d'une politique traditionnelle (Weiss 1986 : 58). Par contre, en ce qui concerne le contenu, nous n'avons pas trouvé de confirmation que les questions d'efficience étaient plus les bienvenues que des questions d'efficacité ou d'impact. Au contraire, les résultats obtenus indiquent que l'intérêt pour l'impact et l'atteinte des objectifs jouent un plus grand rôle que la recherche d'informations sur le coût et les moyens de la mise en œuvre. C'est au niveau du contenu de l'évaluation que l'on sent le plus l'attente par rapport à la fonction de pilotage.*

- Enfin notre dernière hypothèse, la plus largement détaillée, voulait voir dans les caractéristiques quasi techniques du rapport d'évaluation une cause principale de son utilisation.

*Nos résultats ne permettent pas vraiment de confirmer cette hypothèse globalement. Certains éléments semblent jouer un rôle beaucoup plus important que prévu, d'autres beaucoup moins. Décisive est la participation de l'évaluateur à la valorisation de ses résultats, ainsi que la présence d'une stratégie de communication (résumé) et de valorisation (cf. Rossi & Freemann 1993 : 455ss.). Mais les autres éléments techniques imaginés, comme la simplicité, la brièveté, la rapidité avec laquelle les rapports sont rendus ne jouent pas un rôle fondamental. Ou plutôt faudrait-il dire que ces aspects semblent assez garantis en général pour qu'ils ne soient pas (plus) aussi déterminants. En effet, ces éléments sont également présents et importants dans les deux groupes d'évaluations étudiés. Cela ne saurait en tout cas remettre en cause l'importance de ces facteurs. Mais cela nous invite à comprendre qu'une évaluation est très rarement ignorée pour ce genre de motifs uniquement, mais plutôt que dans ces cas, cela s'ajoute à d'autres éléments.*

Lesquels ? c'est là l'objet de la suite de notre recherche.

## 5. L'INTERROGATION CRITIQUE

Le déroulement de cette enquête, la réception et l'analyse des résultats résumés ci-dessus, partiellement étonnants, nous ont amené à nous interroger sur certains aspects de la démarche et des résultats. Nous nous attendions en fait à des résultats plus nets en particulier sur les causes d'utilisation et restions partiellement sur notre faim concernant l'efficacité de la technique quantitative retenue. Cela nous a amené à porter un regard rétrospectif sur notre démarche et à en questionner quelques éléments.

### 5.1. LES LIMITES DES QUESTIONNAIRES

Nous avons été un peu étonnés par quelques détails en cours de démarche :

#### *La liste établie par W. Zimmermann en 1989*

Cette première compilation se caractérise par l'hétérogénéité des informations recueillies. Elle a été constituée sur la base de 17 bases de données, d'informations recueillies auprès d'Offices et de dizaines d'évaluateurs. Le critère de sélection était l'approche scientifique, mais la définition de celle-ci n'était pas prédonnée. Comment vérifier ? W. Zimmermann s'est surtout attaché à recenser les méthodes utilisées ainsi que les effets de ces études, via un questionnaire adressé aux auteurs (et non aux mandants) des évaluations. Cela représentait un biais possible.

#### *La base de données de l'OFJ de 1996*

Le premier étonnement avec cette liste fut de constater qu'en 1996, il n'existait pas (plus) de base informatique de ces titres, mais uniquement des listes papier. Était-ce un signe de l'état de la culture d'évaluation et de son importance dans l'administration ? La liste n'était pas exhaustive, puisque tous les Offices n'avaient pas répondu. Cela pouvait-il changer quelque chose ?

Nous avons déjà signalé la grande diversité de la situation de l'évaluation entre les Départements, certains fournissant plus de cent titres, d'autres quasiment aucun. La même disparité se retrouve à l'intérieur de chaque Département, entre les Offices. Pourquoi ? Qu'est-ce qui justifiait une pareille différence ?

La question méritait d'autant plus d'être posée que quelques Offices nous firent remarquer que notre liste leur paraissait incomplète, pour ne pas dire incompréhensible quant aux critères qui avaient présidé à sa constitution.

La lecture de cette base de données montrait ensuite très clairement que cette diversité entre Départements cachait aussi une hétérogénéité considérable en ce qui concerne le genre de

titres retenus. On y trouve de tout, du rapport de voyage au sondage d'opinion, en passant par la culture du haricot dans la région des Grands Lacs, l'analyse d'un journal d'information et de prévention de la santé, le rapport du Conseil Fédéral sur l'agriculture et l'application de la Loi sur l'égalité des salaires, etc. Il y avait manifestement une très grande variété dans la définition du mot « évaluation », et un certain nombre de titres ne méritait que difficilement ce nom. Quel allait être l'impact de cette variété de compréhension sur les réponses ? Qu'est-ce que ce flou sémantique et conceptuel nous disait sur la situation de l'évaluation dans l'administration ? Plusieurs interlocuteurs, par exemple à l'OFAG, la DDC ou l'OFE nous ont rendu attentif en particulier à la distinction difficile à faire parfois entre « évaluation » et « recherche » ou « évaluation » et « rapport »<sup>23</sup>. Cela pouvait même se sentir à l'intérieur du même Office. Ainsi la DDC nous signalait-elle d'emblée des difficultés de terminologie dans les titres que nous lui avons envoyés : review, autoévaluation, étude, bilan... Plusieurs de ces questionnaires ont finalement été remplis par les auteurs des évaluations.

A cela s'ajoutait l'information obtenue ici ou là de l'existence d'autres études de la même époque qui n'apparaissaient pas dans la liste, y compris dans d'autres Offices du DFAE ou du DFI. La méthode de sélection des titres de cette base de données n'était pas très transparente, ni apparemment toujours très logique.

Le refus de répondre de certains Offices, la manière d'optimiser les réponses d'autres, ainsi que des lettres de responsables d'Office nous interrogèrent aussi sur la fiabilité de certaines réponses. Ceci d'autant plus que la distance temporelle transformait parfois la quête de l'informateur pertinent en véritable chassé croisé et faisait fonctionner une mémoire institutionnelle plus toujours entièrement fiable.

A cela s'ajoutèrent des interrogations sur les réponses aux questionnaires.

## **5.2. LES LIMITES DES RESULTATS**

Nous avons été également surpris du côté peu tranché et clair de nombreux résultats.

- Plusieurs Offices ont eu quelques difficultés à donner des réponses claires ou utilisables. Il y avait parfois plusieurs réponses assez différentes au sujet du degré d'utilisation d'une même étude.
- Si le premier envoi sur l'utilisation donnait au final un réservoir de plus de 370 titres, la base de données pour le questionnaire sur les causes était relativement modeste. Mener une étude quantitative sur un groupe de 22 et un autre de 27 évaluations était relativement fiable. Pour le *groupe principal*, nous pouvions, au vu du nombre élevé d'études très

---

<sup>23</sup> Cf. aussi Widmer, 2001, p.9

utilisées, faire facilement une sélection en privilégiant des études qui semblaient de plus grande importance. Par contre, le faible pourcentage d'études pas du tout utilisées nous interdisait quasiment cette sélection et/ou nous obligeait à aller puiser quelques études dans la catégorie d'utilisation supérieure. Ce problème s'est encore accentué avec le taux de retour. Les efforts déployés pour faire rentrer les réponses au deuxième questionnaire nous ont interpellés, et si le taux de retour brut fut finalement plutôt bon, le nombre de questionnaires utilisables n'était que de trente en définitive. Pourquoi des questions ont-elles été impossibles à répondre pour autant de questionnaires ?

- Sur l'ensemble des 60 questions du questionnaire, la majorité des réponses peinait à différencier clairement les deux groupes. *Comme si la réponse à la question de l'utilisation ou non utilisation ne se trouvait pas dans l'évaluation elle-même, ou que dans certains aspects seulement. Qu'est-ce que cela voulait dire ?*
- D'autre part, malgré le soin apporté au libellé des questions, nous obtenions parfois des réponses que nous ne pouvions classer et interpréter avec certitude. Ainsi par exemple, à la question de savoir si l'évaluation utilisée était plus orientée sur les questions d'efficacité, d'efficacité ou d'impact, qui sont des éléments importants de nos hypothèses.
- Le croisement de certains résultats s'avérait plus difficile que prévu et pas non plus toujours très probant. Il débouchait souvent sur l'absence de corrélation que sur un signal même faible de dépendance. Pourquoi ?

### **5.3. LES LIMITES DE L'APPROCHE**

Ces éléments nous ont fait nous interroger sur la pertinence de l'approche choisie. Etait-il raisonnable dans un tel contexte d'évaluation dans l'administration suisse (manifestement une situation de démarrage) et avec de telles données de base, de pouvoir espérer dégager à coup sûr des résultats probants simplement par des méthodes de type statistique ou mathématique ? Reprenant les éléments théoriques élaborés en particulier par Patton pour expliquer l'évaluation centrée sur son utilisation, nous avons été frappé a posteriori des très hautes exigences qu'il pose au départ pour réaliser des évaluations. Mentionnons par exemple (Patton 1997 : 32ss.) :

- toutes les questions sont-elles toujours uniquement de type empirique ?
- la réponse n'est-elle vraiment jamais prédéterminée par la question ?
- les utilisateurs primaires ne s'intéressent-ils vraiment qu'aux réponses à leurs questions ?
- les utilisateurs primaires veulent-ils des réponses uniquement pour eux-mêmes ?
- les utilisateurs primaires peuvent-ils toujours indiquer comment ils vont utiliser les réponses ?

Peut-on vraiment toujours garantir la très haute motivation des évalués ? Comment tenir compte des *inévitables résistances* que l'on rencontre à quelque étape que ce soit d'une évaluation ? Les personnes concernées, les bénéficiaires ou utilisateurs sont-ils vraiment toujours parfaitement conscients de leurs besoins ou peuvent-ils estimer avec précision les effets de telle résultat. Bref, l'approche choisie n'était-elle pas *un peu idéaliste* ?

En mettant l'accent ou la valeur de l'évaluation uniquement sur son utilisation, et si possible directement visible, ne manque-t-on pas une partie importante du bénéfice ou de l'effet d'une évaluation ? Si tout est fait pour satisfaire les utilisateurs primaires, *qu'en est-il des autres acteurs* ?

Cette focalisation sur les utilisateurs laisse peu de place à d'autres facteurs possibles, comme la politique ou l'expériences des gens sur la manière dont les décisions se prennent. N'y a-t-il pas une légère *surinterprétation du rôle de l'évaluation* elle-même ?

En définitive, cela nous amène à nous demander si la prémisse de base selon laquelle *le sens de toute évaluation est son utilisation* est correcte. Henry (2000 : 2) rend attentif au fait que que l'objectif final d'une évaluation est plutôt l'amélioration sociale générale que l'utilisation. L'amélioration peut avoir lieu bien sûr grâce à l'utilisation des évaluations, mais l'utilisation en tant que telle ne garantit pas encore que l'amélioration sociale soit réelle. « The evaluation entreprise is no less vulnerable to goal displacement than the interventions that evaluators study » (Henry 2000 : 3). En déclarant que l'objectif de tout évaluation est son utilisation, on court le risque de déplacement d'objectifs de l'évaluation elle-même. *Pourchasser l'utilisation induit que l'on produise toujours plus d'évaluations sur l'amélioration de programmes ou d'organisation*. On risque ainsi de négliger l'aspect de *développement de la connaissance* (Henry 2000 : 8) également fondamental dans toute évaluation.

#### **5.4. ANALYSE, HYPOTHESES EXPLICATIVES**

Notre réflexion s'est orientée en fonction de deux pré-hypothèses. La première était que nos résultats étaient ce qu'ils étaient d'abord parce qu'ils répondaient aux besoins des utilisateurs primaires, à savoir l'administration, et *donc moins aux autres utilisateurs potentiels*. Ceci expliquerait en partie le haut taux de réponses et le pourcentage très respectable, voire excellent d'utilisation décisive pour la suite du travail. Cette approche était bonne pour comprendre la manière de travailler de l'administration et ses besoins, mais ne décrivait les causes explicatives que pour ce public, et pas nécessairement dans la même mesure pour d'autres acteurs.

Deuxièmement, la variété des situations de l'évaluation entre les Départements et les Offices renvoie certainement aussi à des contextes et des fonctions différentes pour l'évaluation dans les Offices concernés<sup>24</sup>. Il est probable que les résultats soient à interpréter de manière différenciée d'Office à Office. Une évaluation pour mesurer l'effet de la Loi d'aide aux victimes ne saurait s'inscrire exactement dans la même dynamique qu'une étude sur la culture de sauterelles au Pakistan.

### *Les questions liées à l'approche choisie*

En confrontant nos résultats aux éléments théoriques résumés au point 3.3., nous faisons en outre les constats suivants qui nous semblent corroborer ces hypothèses :

- beaucoup des *éléments d'ordre technique* ne se vérifient pas, ou que très partiellement. N'est-ce pas lié au fait que par exemple la DDC et l'OFSP travaillent depuis longtemps avec des évaluateurs, et que les questions de crédibilité, de ponctualité ou de recommandations par exemple sont réglées depuis longtemps ?
- l'importance des *fonctions d'évaluation*. La fonction de contrôle ne semble pas importante dans notre cas. Certes, il y a l'existence parallèle du Contrôle Fédéral des Finances par exemple pour l'expliquer. Mais aussi peut-être de la longue pratique de la DDC, où l'évaluation est très régulière et prévue dès le départ. Dans ce cas c'est plutôt la régularité et la relation à long terme qui garantissent que la fonction de contrôle soit respectée plutôt que l'évaluation elle-même.
- La question de *l'utilisateur primaire* est particulièrement révélatrice : le deuxième questionnaire a montré que plus de 75 % de toutes nos études sont commanditées par les Offices eux-mêmes<sup>25</sup>. L'utilisateur est donc d'abord le collaborateur ou le Directeur d'un Office. Ses intérêts d'information se situent certainement plus du côté de la mise en œuvre d'une mesure que de son orientation, plus du côté de la légitimation de son travail que du côté de son achèvement. Si le taux d'utilisation peut sembler assez élevé au premier regard, c'est d'abord l'indice que ces études correspondent bien aux besoins des collaborateurs de l'administration, plus peut-être que des parlementaires ou des chefs de Départements.
- La question du *temps* nécessaire à l'utilisation des évaluations. Notre étude est probablement très bien placée ici pour atteindre des chiffres éloquentes d'utilisation, dans la mesure où certaines études datent de près de 10 ans. Que l'usage ait été instrumental ou conceptuel n'est plus important, car après 10 ans, il y a eu assez de temps pour pouvoir répondre que les deux usages ont eu lieu. La même remarque vaut pour la question de

<sup>24</sup> Cf. Widmer 2001, p.88

<sup>25</sup> Cf. résultats similaires chez Widmer 2001, p.12

savoir comment l'évaluation a été utilisée. La distance dans le temps a tendance à optimiser les résultats d'utilisation.

- Notre enquête ne nous donne pas non plus assez de données probantes par rapport aux éléments théoriques structurels mentionnés (p. 25) par exemple sur la question du « quand », du « comment » et de « ce qui » est utilisé dans l'évaluation.

Si nous reprenons en outre quelques-uns des résultats obtenus statistiquement, nous constatons les choses suivantes :

- l'importance de *la politisation du sujet* : les résultats obtenus renvoient à une réalité qui ne pouvait être que très partiellement l'affaire de l'administration à qui nous avons demandé de donner les réponses. Même si ces éléments ont joué un rôle important, les *acteurs responsables* de cette politisation ne sont que partiellement présents dans l'administration. Il faut plutôt les chercher à l'extérieur de celle-ci. Idem pour la question du caractère conflictuel du sujet.
- La mention d'une cause importante d'utilisation comme étant celle de l'initiative du projet par *le Parlement* va dans le même sens.
- L'importance des fonctions de légitimation et d'information, et, dans une moindre mesure, du pilotage : cette importance correspond très bien aux besoins des collaborateurs des Offices<sup>26</sup>. Mais il n'est pas sûr que les besoins des Chefs de Département ou des parlementaires soient entièrement les mêmes.
- Les réponses sur la *diffusion* du rapport et sa discussion à l'interne montrent clairement que l'Office, voire la Section sont les lieux où ces rapports sont reçus et utilisés. Il manque par contre des informations sur le sort réservé à ces études au niveau d'un ensemble d'Offices, voire d'un Département, sans parler du Parlement.
- Nous avons veillé au niveau de l'échantillonnage des deux groupes d'études à avoir des études pertinentes provenant d'un maximum d'Offices. Malgré cela, le poids des études de projets de la DDC et de l'OFSP reste très fort sur l'ensemble. On peut se demander s'il n'aurait pas fallu enregistrer les résultats une fois avec et une fois sans les études de ces deux Offices. La faible fiabilité de l'échantillonnage dont nous disposerions dans ce cas (15 et 22 études de base ; en tenant compte d'un taux de retour de 60%, nous aurions finalement entre 10 et 15 cas par groupe) nous a empêché de faire ce contrôle, mais la question doit en tout cas rester présente pour notre analyse.

### *L'efficacité ou l'output, ou l'efficacité de l'output ?*

Une des questions à laquelle ces résultats ne répondent que partiellement est celle du niveau de la politique publique sur lequel l'utilisation obtient son impact. Ou la relation qu'il y a

---

<sup>26</sup> Cf. résultats analogues chez Widmer 2001, p. 84

entre le « pourquoi » une évaluation est utilisée et le « pour quoi » on l'utilise. D'un côté, nous avons des résultats indiquant que les évaluations utilisées cherchaient surtout à mesurer l'impact et l'efficacité du programme. Donc nous aurions une influence et un changement de la politique publique à un niveau important. Mais par ailleurs, nous avons vu que les évaluations ont surtout comme fonction de légitimer et d'informer le mandataire. Donc l'utilisation débouche apparemment surtout sur une confirmation du statu quo. Surtout, nous avons l'indication que la nature de ces évaluations est majoritairement de type « projet ». Nous sommes par conséquent à un niveau relativement de détail de la mise en œuvre de la politique publique en question. Enfin, nous avons vu que la très grande majorité de ces évaluations sont non seulement mises en œuvre par l'administration, mais aussi demandées par elle. C'est elle qui est chargée de la mise en œuvre des politiques publiques. Elle situera donc naturellement le cadre de son action au niveau de la mise en œuvre plus qu'au niveau des objectifs supérieurs. C'est sur ce niveau-là qu'elle est intéressée à avoir des informations sur les effets de son action et à faire des propositions d'amélioration (d'où le fort taux d'utilisation), si possible ni trop controversées ni trop chères (Weiss 1999 : 472). Donc, l'importance accordée à la recherche du degré d'atteinte des objectifs et de l'impact est certainement correcte, oui, mais elle se passe à un niveau relativement modeste du programme administratif. L'intérêt de l'administration n'est pas prioritairement de mettre le doigt sur les gros problèmes de la politique publique en question, mais plutôt d'en améliorer coûte que coûte sa mise en œuvre. Cette interprétation nous semble renforcée par le constat fait jusqu'ici que les évaluations sont pour l'essentiel situées et utilisées au niveau des Offices et relativement peu plus haut.

Tout cela nous amène à nous demander si nos hypothèses sont suffisamment vérifiées avec les techniques choisies. Même si les résultats obtenus semblent tout à fait compatibles avec d'autres études ou l'expérience de spécialistes, les questions sur la fiabilité de la base de données de départ et d'une partie des réponses obtenues rendent en tout cas une généralisation des résultats obtenus un peu téméraire.

Les questions suivantes nous semblent en l'état n'avoir pas reçu de réponse satisfaisante :

- Nous avons globalement un échantillon *d'évaluations centrées sur des projets et programmes précis*. Mais nous n'avons que peu d'informations sur l'utilisation faite d'évaluations plus larges, plus stratégiques. De même, nous avons pas mal d'informations sur l'usage des évaluations au niveau de la *mise en œuvre* de programmes, mais nous manquons d'informations sur ce que ces évaluations ont modifié dans le cycle de la politique publique au delà de ce niveau, à un échelon supérieur.

- Notre approche est centrée sur les *besoins des utilisateurs*. Partant du constat de diversité et d'hétérogénéité de la définition et de la pratique de l'évaluation, nous devrions pousser plus loin la recherche et mieux identifier les besoins des utilisateurs *par Office* ou par Département.
- Les hypothèses de départ qui n'ont pas été vérifiées méritent un complément d'enquête. Nous pensons en particulier aux *facteurs techniques* expliquant l'utilisation faite, mais aussi au facteur lié au *contenu* de l'évaluation (output, ou résultat, ou impact...) ainsi que la question des *fonctions* de l'évaluation.
- Il serait bon également de revenir sur les hypothèses trop facilement vérifiées. Nous pensons ici en particulier au facteur de « *politisation* », qu'il faut mieux comprendre. Cela nous renvoie à d'autres acteurs que nous avons jusqu'ici peu sollicités.
- Enfin, il nous faut garder un œil sur certains biais possibles repérés, comme par exemple l'effet distordant du temps, la tentation *d'optimiser les résultats*. Il nous manque par conséquent des informations plus actuelles sur l'utilisation de l'évaluation telle qu'elle se pratique et se vit aujourd'hui.
- Certains résultats nous renvoie directement à l'importance d'autres cofacteurs dans le processus d'utilisation, comme par exemple *les contextes* du Parlementaire et de l'administration. Il serait utile d'élargir l'approche et le dispositif de recherche.
- Enfin, un des résultats principaux du premier questionnaire, à savoir que l'évaluation n'est vraiment déterminante pour un changement d'orientation de la politique/du programme que dans un cas sur deux environ nous invite à élargir un peu le champ de réflexion et d'investigation et nous interroger sur *ce qui se passe avec l'évaluation* dans l'autre moitié des cas. Y a-t-il un effet d'un autre ordre ?

Nous somme donc amenés à chercher d'autres éléments d'informations pour corroborer nos résultats

## 6. HYPOTHESE ET APPROCHES COMPLEMENTAIRES

Sur la base de l'analyse des résultats, notre hypothèse de départ complémentaire serait que l'utilisation d'une évaluation *dépend encore d'autres facteurs* pas suffisamment pris en compte dans notre questionnaire, des facteurs difficiles à identifier par des statistiques et renvoyant *plus au rôle des acteurs et du contexte* dans lequel intervient le rapport d'évaluation.

## 6.1. ELEMENTS THEORIQUES

Il nous faut dans un premier temps interroger les facteurs indépendants de l'évaluation mais qui influencent son utilisation, à quelque niveau que ce soit. Ensuite, nous devons nous pencher sur l'usage informel des évaluations, les processus indirects qu'elles déclenchent et qui permettraient d'observer si l'utilisation d'une évaluation en est renforcée, et comment, le cas échéant, elle dépasse le niveau de la mise en œuvre de projets où nous l'avons surtout rencontrée.

### 6.1.1. Les trois autres grands facteurs explicatifs de l'utilisation : les acteurs, le contexte et l'organisation

Mis à part les éléments mentionnés dans notre première approche théorique<sup>27</sup>, quels peuvent être les facteurs influençant fortement l'utilisation d'une évaluation ? Notre enquête nous a renvoyés en tout cas à deux autres groupes de causes : *le contexte* et *le facteur humain*<sup>28</sup>. Il nous semble opportun d'y ajouter aussi l'importance de *l'organisation*.

Par *facteurs humains*, on entend surtout les caractéristiques de l'évaluateur et de ou des utilisateur(s). En ce qui concerne l'évaluateur, son enthousiasme, sa créativité, sa capacité de communication et de conviction jouent certainement un rôle aussi important que sa méthode. De même sa participation active à la diffusion des connaissances acquises, son intérêt à faire connaître et reconnaître ses travaux. En ce qui concerne l'utilisateur, son attitude et son intérêt personnel pour le programme, ou pour l'évaluation en tant que discipline, l'arrière-fond personnel, quasiment biographique, la position dans l'organisation, le degré d'expérience professionnelle jouent potentiellement un rôle.

Par *facteurs contextuels*, il faut comprendre toutes les exigences, opportunités et contraintes qui interviennent dans le processus de l'évaluation et de son utilisation, comme les contraintes fiscales ou budgétaires, les relations entre le programme à évaluer et d'autres acteurs de l'organisation ou de l'environnement. Il faut mentionner aussi l'actualité politique, médiatique, parlementaire, le contexte social, financier ou public dans lequel l'étude s'effectue, etc. Notre enquête a mis le doigt en particulier sur l'aspect « politisateur » de ce

---

<sup>27</sup> Cf. § point 3.3, p.24.

<sup>28</sup> James Burry (1984) regroupe les facteurs en 3 catégories : les facteurs humains, contextuels et liés directement à l'évaluation elle-même. Par facteurs liés à *l'évaluation*, on comprend surtout la manière dont l'étude a été conduite, la méthodologie, les procédures utilisées, la qualité de l'information fournie. Nous avons déjà suffisamment abordé cet aspect. La qualité des informations fournies et de la forme de l'évaluation ne semblent que très partiellement responsables de l'utilisation ou non de ses résultats, si l'on fait exception du rôle de l'évaluateur lui-même. Mais cela nous renvoie justement à la dimension jouée par un des acteurs.

contexte. Un sujet peut tout à coup être instrumentalisé dans une optique politique. La question des « fenêtres d'opportunité politique » joue manifestement un grand rôle. Janett (2000 : 8) dans son étude sur l'évaluation parlementaire a relevé l'importance de cette « timeliness ».

Le facteur de *l'organisation* nous semble pouvoir être ajouté ici. Nous entendons en particulier les caractéristiques de la structure et la culture de l'organisation. Ainsi la question de la manière dont l'information est organisée et traitée dans l'administration, les processus de décision, l'importance, la structure et la place des divers éléments (dont l'évaluation) qui entrent en jeu lors d'une décision. En plus des éléments de structures et d'organigrammes (et leur réalité concrète pratique), il faut souligner la question de la culture de l'organisation. On pense ici bien sûr à la culture bien particulière d'une administration, à ses tendances immobilistes tout comme à sa capacité de changement, ou sa résistance aux changements. L'administration est en principe plutôt une organisation de type mécanique, où l'information circule principalement en suivant l'échelle hiérarchique. Mais des mécanismes de contre-pouvoir sont également bien connus et répandus, qu'ils soient liés à des facteurs de pression politique ou au jeu formel ou informel d'acteurs précis. Des éléments comme le courage, la peur ou l'intérêt personnel ne sont pas à oublier. Il n'est pas inutile de rappeler que l'utilisation d'une évaluation est une question de répartition du pouvoir.

#### *Les remarques de M. Q. Patton*

Patton (1987 : 113) a cherché à pondérer entre eux les facteurs. Y a-t-il une prédominance ou un hiérarchie clairement établie entre eux, ou bien reste-t-on dans le domaine du hasard ? Il est difficile de trouver des études explicites sur ce point, et les données de base sont souvent sujettes à de multiples interprétations. Néanmoins, selon Burry et Patton (Patton 1986 : 118), il semble évident que le facteur le plus déterminant est le facteur humain. Parmi les 3 facteurs, c'est clairement celui qui se trouve « au milieu », qui a à gérer à la fois la matière concrète de l'évaluation, le contexte dans lequel celle-ci « tombe » et l'avenir possible des résultats de l'étude à travers un chemin de piste politico-administratif où des intérêts et des attitudes divers cohabitent. Le facteur humain joue un rôle dans la recherche d'information, la qualité de celles recueillies pour rédiger le rapport, la qualité formelle du rapport. Il est aussi déterminant dans l'adéquation des résultats du rapport aux besoins des utilisateurs, dans l'approche diplomatique et politique des utilisateurs, etc. Enfin le facteur humain représente bien cette tension entre l'aspect « objectif-scientifique » et l'aspect « subjectif-politique » propre au sujet qui nous occupe, aussi bien au niveau de l'évaluateur lui-même qu'à celui de l'utilisateur. Notre enquête a montré le rôle favorable que semble jouer une longue expérience du contexte et des autres acteurs de l'administration pour valoriser les évaluations.

### 6.1.2. Les processus de décision, le modèle des quatre « I » de C. H. Weiss

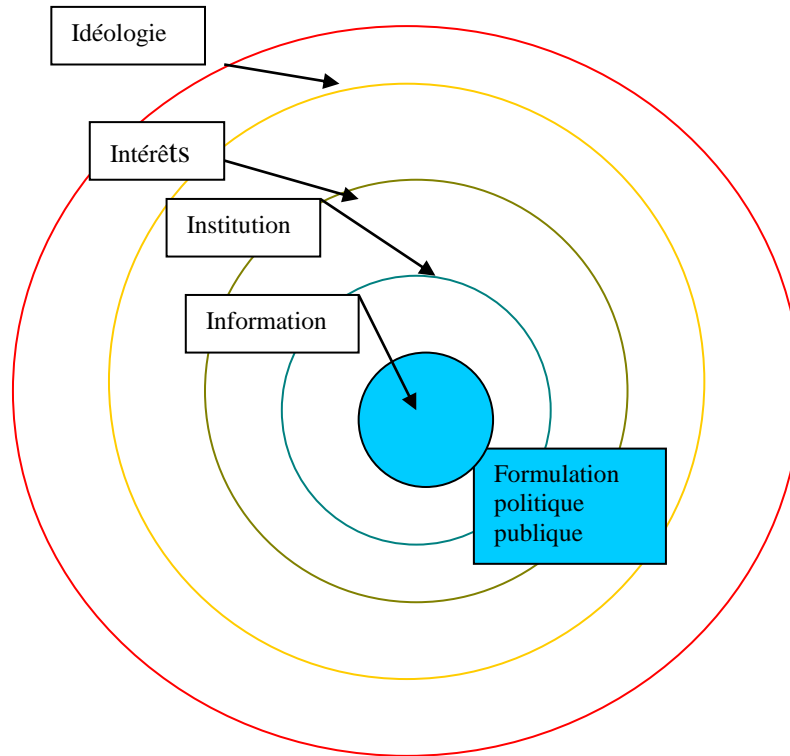
L'évaluation, et en particulier la question de l'usage qu'on va en faire, se situe au cœur de la question du pouvoir, et de la relation entre science et gouvernement. Par conséquent, il nous semble pertinent d'élargir la réflexion sur la question de l'utilisation des évaluations à la problématique des processus de prise de décision. Notre question sera de définir la place que prend un élément comme l'évaluation dans un processus de prise de décision. Au lieu de partir d'« en bas » comme nous l'avons fait avec nos questionnaires en cherchant à repérer les éléments qui ont aidé l'évaluation à être utilisée pour elle-même, nous partirons ici d'« en haut », c'est-à-dire des mécanismes qui interviennent en général dans tout processus de décision, qu'il y ait eu une évaluation ou non, et tenterons de définir la place que prend l'évaluation dans ce système.

« Evaluation, contexte, acteurs et organisation » sont les facteurs déterminant le genre d'utilisation des évaluations. Ce sont des facteurs peu homogènes, certains très étroits (évaluation), d'autres très larges (contexte). En élargissant un peu le cadre théorique, C.H. Weiss nous propose une autre manière de définir ces facteurs.

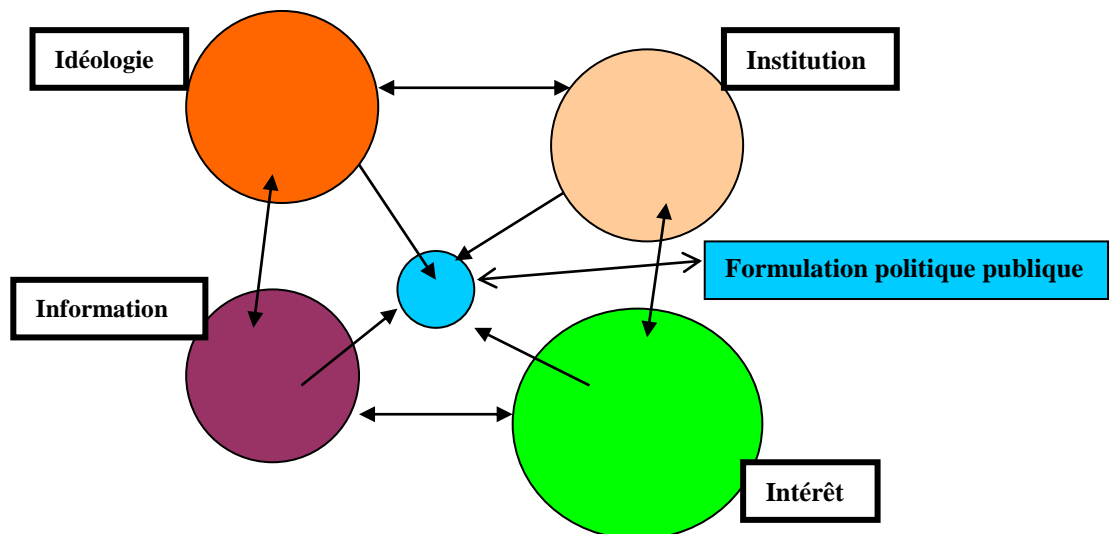
C.H. Weiss (1983 et 1999) a élaboré un modèle dynamique représentant l'interaction entre *l'Idéologie*, *l'Intérêt*, *l'Institution* et *l'Information* (les quatre « I ») dans un processus de prise de décision. L'évaluation elle-même fait partie pour Weiss du domaine « information ». Comment s'articulent les relations entre ces 4 éléments ?

Selon Weiss (1983 : 228 ; 1999 : 477), il est important de prendre conscience du modèle théorique idéal que nous véhiculons tous et de le distinguer du modèle que l'on rencontre de manière pragmatique dans la réalité quotidienne de tout chercheur. Selon le *modèle idéal*, l'élément le plus éloigné influant sur un processus de décision comme par exemple le processus de fabrication d'une politique est *l'Idéologie*, représentant les valeurs et croyances de base. Plus proche et importante vient ensuite la dimension de *l'Intérêt*, qu'il soit professionnel, privé ou politique. Les contraintes de *l'Institution*, qu'elles soient de type culturel, hiérarchique ou budgétaire donnent la marge de manœuvre réelle. Certaines options sont exclues d'avance. Et la partie dure qui détermine l'élaboration de la politique est *l'Information*. Weiss montre que dans la réalité, nous n'avons pas à faire un modèle de type cercles concentriques mais plutôt à quatre cercles adjacents, comme des électrons autour du « noyau » que représente l'élaboration d'une politique.

### Le schéma idéal théorique



### Le modèle réaliste de C.H. Weiss



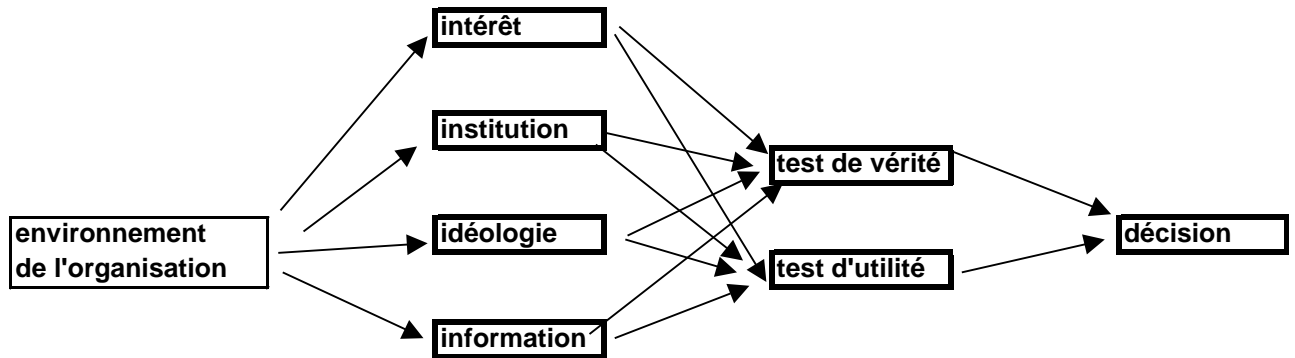
L'évaluation fait partie du cercle « Information », mais il n'en constitue qu'une partie. Les responsables (politiques ou administratifs) ont des connaissances préalables formelles ou informelles, ils comparent les éléments fournis par l'évaluation à d'autres informations qu'ils ont sur le sujet, ils disposent de plusieurs sources, en particulier celles fournies par leur expérience. La dernière information est aussi toujours la meilleure. Nous devons tenir compte du fait que l'information fournie par l'évaluation n'est pas seule dans un désert, mais qu'au contraire elle doit se frayer un chemin dans la jungle des informations de toutes sortes qui submergent les décideurs. Un premier facteur décisif est par conséquent la relation, la *compatibilité entre les informations fournies* par l'évaluation et les autres informations à disposition.

La même remarque peut être faite en ce qui concerne les relations entre l'information et l'idéologie, et l'information et l'intérêt. Préjuger d'un comportement complètement rationnel du décideur (qui changerait d'avis uniquement sur la base d'arguments rationnels) est naïf. La *compétition d'influence* entre ces 4 cercles est permanente.

Plus encore, la *taille* de ces quatre cercles peut fortement varier d'un cas de politique publique à l'autre. Certaines politiques sont très perméables à des arguments rationnels (la sécurité routière par exemple), d'autres relèvent du conflit d'idéologies (la politique des étrangers par exemple). Des alliances et des constellations de pouvoir s'opèrent également sans cesse entre elles en fonction des sujets pour aboutir à la formulation de la politique. Ainsi, il est très invraisemblable que dans une constellation où le facteur idéologique et celui de l'intérêt prédominent, l'élément d'information puisse prétendre jouer un rôle prépondérant. Par contre, sur un sujet où l'idéologie est faible ou confuse et les intérêts divisés, l'évaluation a plus de chance d'être entendue. La question est toujours de voir quelle va être la coalition dominante. *A priori, la meilleure constellation pour la promotion de l'importance de l'évaluation est celle où l'idéologie et les intérêts sont en conflit ouvert* (l'avortement par exemple ?).

Le modèle du processus de décision ne prévoit pas seulement que ces quatre facteurs jouent un rôle combiné. Il rend attentif au fait qu'ils sont repris par les décideurs qui les passent au crible de leur valeur de *vérité*, c'est-à-dire un jugement sur la qualité des informations fournies, sur leur adéquation à ce qu'il leur était demandé de fournir comme éléments et leur positionnement par rapport aux informations disponibles jusque-là sur ce sujet, ainsi qu'à leur valeur d'*utilité*, c'est-à-dire un jugement sur leurs effets, leur implication pour l'action future, le niveau de défi, de nouveauté et de risque qu'elles représentent. Si le test de vérité est encore relativement lié à une démarche de type scientifique, celui de l'utilité est très clairement d'ordre politique. Quoi qu'il en soit, pour Weiss (1999 : 220), poser la question de la manière

dont l'évaluation peut accroître son influence sans prendre en considération la compétition entre différentes sources d'influence est très certainement non productif.



**Le modèle des processus de décision selon Weiss**

Ailleurs, Weiss (1988 : 175) décrit ainsi la manière dont les évaluations influencent les processus de décision politique : « We have to recognize that it is not the single decision maker, the benevolent despot, who uses evaluation to improve the program. Evaluation findings - and generalizations from these findings - come into currency among many groups and interests involved in a policy domain. They infiltrate the 'policy system'...evaluation is the most useful when it facilitates negotiation. Evaluation that opens new facets of a situation or that discloses possible grounds for agreement can have widespread influence ».

Ceci nous amène directement au point suivant.

### 6.1.3. Les processus d'apprentissage

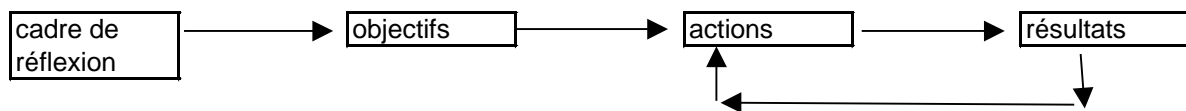
Nous avons fait par notre enquête un double constat. D'une part que les évaluations étaient très souvent très utilisées, mais d'autre part que cela concernait en majorité des projets et programmes assez limités. Nous n'avons pas ou peu d'informations sur les effets supérieurs des évaluations, sur l'influence de ces évaluations de programmes sur l'ensemble de la politique publique. Notre enquête touche ainsi la question du processus d'apprentissage déclenché par l'évaluation. Il vaut la peine pour notre propos de faire un bref détour vers les modèles d'apprentissage. Nous n'avons pas la possibilité d'entrer dans les détails de ce domaine et renvoyons pour l'essentiel à la littérature<sup>29</sup>. Mais nous sommes ici très clairement dans le domaine où l'évaluation n'est plus observée pour son usage direct, mais bien pour son

<sup>29</sup> Cf. en particulier les ouvrages de F. Leeuw (et alii) ainsi que P. Urio (et alii) cités dans la bibliographie. L'ouvrage de F. Leeuw (et alii) fait en particulier référence aux travaux de C. Argyris : *Reasoning, Learning and Action*, San Francisco, Jossey-Bass Publishers, 1982

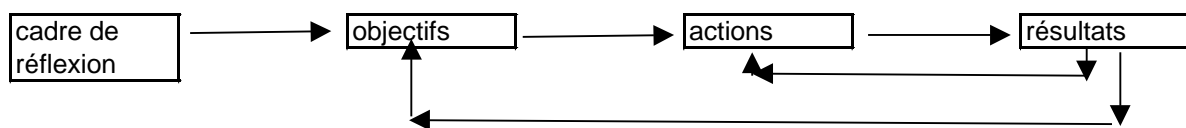
influence indirecte. Lors d'un processus d'apprentissage tel que par exemple celui déclenché chez les utilisateurs par une évaluation, nous avons à faire à la question du degré de profondeur de l'acquisition de la connaissance nouvelle. Qu'est-ce que la nouvelle information va changer dans le système cognitif de l'apprenant ? Jusqu'où dans son système le changement va-t-il s'opérer ? Suivant la terminologie anglaise ou française, on parlera de « single or double loop learning » (Leeuw et alii 1993 : 7), ou de la question de l'apprentissage par « adaptation ou reconstruction » (Urio et alii 1996 : 42). La plupart du temps, l'apprentissage se fait par « adaptation ». Cela correspond à une adaptation des *procédures* en vigueur aux objectifs fixés sans remettre rien d'autre en question dans le programme ou l'action. La « reconstruction » correspond à une *nouvelle théorie* d'action, remettant en cause les objectifs et normes établies. Le cadre de référence institutionnel est redéfini. Il existe encore un processus plus intense et fondamental, appelé « réflexion sociale », qui correspond au processus d'apprentissage de l'apprentissage, l'amélioration des capacités d'apprentissage. Ceci remet en cause le cadre même de réflexion de l'apprenant, exigeant un changement profond de système.

On peut représenter schématiquement cela ainsi :

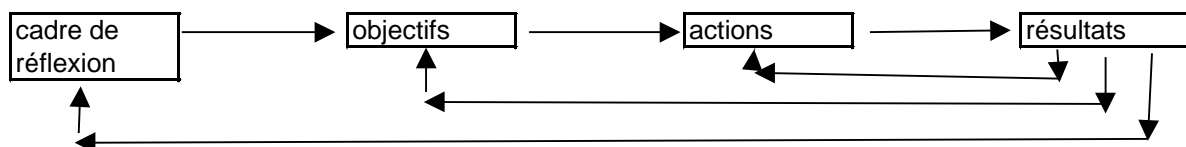
#### Correction et apprentissage par adaptation (single loop)



#### Correction et apprentissage par reconstruction (double loop)



#### Correction et apprentissage par réflexion sociale (deutero learning)



Nous aimerions relever ici deux éléments pour notre propos :

- notre enquête jusqu'ici a traité pour l'essentiel de cas de « single loop ». D'une part parce que beaucoup d'études concernaient des programmes bien précis. D'autre part parce nous avons vu que la forte utilisation démontrée n'impliquait pas clairement une réorientation complète de la politique, mais son pilotage administratif, voire sa légitimation. Ainsi les résultats d'une évaluation peuvent-ils être à la fois très utilisés mais garder une influence modeste eu égard aux objectifs de la politique.
- Une utilisation de type « reconstruction » est évidemment plus ambitieuse et difficile car elle implique une nouvelle discussion des objectifs. On en reste donc souvent au niveau d'un « single loop » parce que les mécanismes de rétroaction sur la définition des objectifs n'est pas faite, ou reste incomplète, inefficace, non prévue ou non souhaitée. Ainsi, il n'est pas certain que les responsables de l'administration aient toujours la compétence de procéder à une double boucle de rétroaction. Leur utilisation ira plutôt dans le sens de favoriser une nouvelle discussion des objectifs au niveau d'autorité approprié, donc d'utiliser les résultats de l'évaluation de manière indirecte.

Tous ces éléments nous invitent à repositionner la place et l'importance de l'évaluation en tant qu'instrument scientifique dans les processus de prise de décision. La question de l'utilisation ne saurait se résoudre à celle de l'usage direct d'une étude. Il faut également tenir compte de l'effet d'influence indirecte, et mieux cerner la fonction d'apprentissage liée à l'évaluation, y compris pour les acteurs extérieurs à l'administration. En effet, par rapport à la constellation d'influences des 4 « I », l'évaluation n'a apparemment qu'un espoir d'usage direct au niveau du chapitre « information ». Pour le processus de décision en tant que tel, elle doit plutôt compter sur son pouvoir d'influence et sur les processus d'apprentissage qu'elle peut aider à provoquer. L'utilisation dépend donc aussi du type d'apprentissage en jeu.

## **6.2. HYPOTHESE ET INDICATEURS, TECHNIQUES ET DEMARCHE**

Après ce détour théorique, nous sommes en mesure de formuler notre hypothèse explicative complémentaire, qui comporte 2 volets :

Notre hypothèse sera donc premièrement que *l'utilisation d'une évaluation sera pour l'essentiel proportionnelle à l'équilibre des forces entre les pôles « information », « idéologie », « institution » et « information » dans le processus de décision sur la politique publique ou le programme concerné.*

Une sous - hypothèse pronostiquerait que *le facteur « information » comme variable explicative d'une utilisation perd de son importance au fur et à mesure que l'on s'éloigne du domaine strict du suivi technique ou administratif d'une politique et que l'on s'oriente vers le domaine des objectifs politiques, que ce soit au niveau du Département ou du Parlement.*

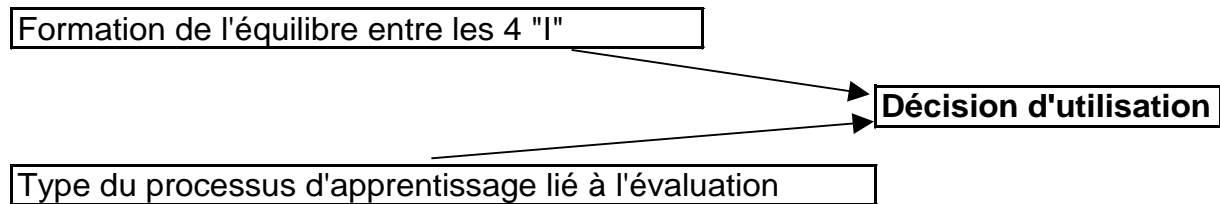
Par rapport à la question du pouvoir de changement des évaluations, nous en déduisons deuxièmement l'hypothèse que *son influence totale (à tous les niveaux du système politico-administratif) dépendra du type de processus d'apprentissage en jeu, que ce soit pour les acteurs dans les Offices, ou au niveau du Parlement.*

Tous ces éléments ont une influence sur l'utilisation en ce qu'ils renforcent l'influence indirecte de l'évaluation au détriment de l'usage direct.

Comme indicateurs, nous retiendrons d'abord la présence *d'éléments idéologiques, institutionnels et relatifs aux intérêts des acteurs* dans les causes de l'utilisation des évaluations. Mais encore, pour la seconde partie, nous serons attentifs au *niveau administratif* où les évaluations que nous avons observées sont le plus utilisées. Troisièmement, nous essayerons d'identifier les facteurs explicites *d'apprentissage* relevés par les acteurs au sujet ou en relation avec l'évaluation.

Nous avons donc formulé le modèle causal suivant.

#### *Modèle causal*



#### *Techniques*

Pour essayer de vérifier notre hypothèse, nous avons eu recours à une démarche qualitative. Nous avons cherché à rencontrer surtout des acteurs représentant les domaines « intérêt », « institution » et « idéologie », c'est-à-dire des parlementaires mais aussi des responsables d'Office. Par ailleurs, il nous a semblé utile de rencontrer des acteurs pouvant compléter notre information sur la pratique et l'évolution de la pratique en ce qui concerne le domaine « information », c'est-à-dire la pratique et l'organisation de l'évaluation depuis 1995 jusqu'à aujourd'hui. Il nous semblait important de tenir compte de l'évolution du dossier de l'évaluation dans l'administration ces dernières années, en particulier à cause du facteur d'apprentissage mentionné. De même, nous avons cherché des documents récents sur l'usage des évaluations en Suisse.

Afin de rester centré sur les domaines touchés par notre analyse quantitative, nous avons concentré nos efforts sur les *Offices* responsables de la politique de développement (DDC), de la politique de santé (OFSP), la politique de l'énergie (OFE), l'agriculture (OFAG) et les

transports (OFT). Nous avons également retenu les services du Parlement et l'Organe parlementaire de contrôle de l'administration (OPCA), à cause de leur rôle d'interface entre l'administration et le Parlement.

En ce qui concerne les *Parlementaires*, nous avons également privilégié les membres des Commissions couvrant les politiques publiques au centre de nos questionnaires. Nous avons tenté d'équilibrer le choix des interlocuteurs en fonction de l'importance des partis et de la représentation cantonale, mais également en fonction de la durée depuis laquelle le/la Parlementaire suit ces politiques publiques. Contrairement à notre souhait, nous n'avons finalement rencontré que des Parlementaires francophones. Nous avons élaboré un questionnaire d'entretien<sup>30</sup> orienté sur nos hypothèses, que nous avons adapté en fonction de chaque situation. Nous avons ensuite analysé et reformulé les réponses les plus fréquentes, tout en donnant place aux exceptions significatives.

Par ailleurs, nous avons consulté et utilisé des *études récentes* parues en Suisse sur le statut de l'évaluation.

### *Déroulement de la démarche*

Nos entretiens et la recherche de documentation ont eu lieu entre juin et juillet 2001.

A partir du moment où nous avons pu provoquer un contact direct, nous n'avons pas eu de difficultés à rencontrer des parlementaires fédéraux, si ce n'est celle de fixer des rendez-vous entre la fin d'une session parlementaire et des vacances. Six parlementaires ont aimablement répondu à notre demande<sup>31</sup>.

En ce qui concerne les Offices, nous avons pu obtenir la collaboration de M. Th. Widmer qui avait réalisé de nombreux entretiens avec plusieurs des interlocuteurs pressentis quelques mois auparavant, en partie sur les mêmes sujets d'intérêt que nous. Après quelques hésitations et retards (qui n'étaient pas le dû des responsables d'Office), nous avons finalement eu le droit d'accès aux procès-verbaux d'entretien que M. Widmer et son équipe avaient réalisés et qui étaient résumés dans son étude déjà citée. Dans trois cas, nous avons souhaité en plus rencontrer directement les personnes déjà interviewées, les procès-verbaux laissant trop de questions encore ouvertes. Là non plus, nous n'avons connu aucune difficulté à agender des rendez-vous. Six Offices et leurs responsables ont été interviewés<sup>32</sup>.

---

<sup>30</sup> Cf. Annexe 11

<sup>31</sup> Cf. liste en annexe 8

<sup>32</sup> Cf. liste en annexe 9

### 6.3. LES ENTRETIENS QUALITATIFS ET LEURS RESULTATS

Notre objectif était triple :

- d'une part mesurer l'importance que ces interlocuteurs accordent aujourd'hui aux divers facteurs expliquant l'utilisation des évaluations (donc obtenir des informations permettant de *confronter les informations et résultats obtenus* par notre démarche jusqu'ici), dans l'optique de la vérification ou de l'invalidation de nos hypothèses.
- d'autre part vérifier les *intérêts et attentes des divers types d'acteurs* face aux évaluations de manière à pouvoir essayer d'élaborer quelques pistes pour la conclusion de notre réflexion.
- Enfin, il s'agissait de s'informer sur *l'état actuel de l'évolution du dossier « évaluation »* dans l'administration et le Parlement fédéral.

#### 6.3.1. Entretien avec les responsables d'Office. Résultats

##### 6.3.1.1. L'Office Fédéral de l'Energie (OFE)

L'OFE a développé en particulier suite aux clauses évaluatives de la Loi sur l'énergie toute une série d'évaluations au cours des années 90.

Les éléments suivants nous semblent particulièrement intéressants à relever pour notre propos<sup>33</sup> :

- L'initiative pour de grands programmes d'évaluation comme celui lié à « Energie 2000 » n'est venue ni du Parlement ni du Gouvernement mais de l'interne. Malgré sa nouveauté, elle ne rencontra pas de résistance notable dans la hiérarchie.

*Nous observons ici une confirmation d'une part pour l'origine toujours interne de l'initiative de l'évaluation, et d'autre part de l'intérêt et du besoin des divers acteurs à évaluer une politique nouvelle.*

- La plupart des 58 évaluations réalisées ont été très utiles au niveau de l'Office. Mais elles n'étaient pas spectaculaires et donc inintéressantes pour le politique. L'écho a été assez faible, y compris au Parlement. Sauf lorsque le sujet était politiquement l'objet d'un conflit, comme le cas des décomptes individuels de chauffage. Mais en général, l'utilisation des évaluations a souffert du fait que l'énergie n'est pas un sujet de souci prioritaire dans la population.

*Nous observons ici que la plupart des évaluations restent au niveau de l'Office, et que le faible intérêt politico-idéologique du sujet explique peut-être qu'elles n'ont pas beaucoup*

---

<sup>33</sup> Nous nous limiterons à relever les éléments confirmant peu ou prou nos hypothèses et laisserons les informations neutres de côté. Nous n'avons pas rencontré ou reconnu d'éléments qui iraient entièrement à l'encontre de nos hypothèses. *Nous soulignerons par le caractère italique des confirmations directes de nos hypothèses ou résultats obtenus jusqu'ici et /ou des commentaires les explicitant.*

*été lues « plus haut ». Notons au passage le grand nombre d'études réalisées dans ce programme.*

- Les évaluations représentent un élément important pour la conduite d'une équipe. Leur portée est donc beaucoup plus forte « vers en bas » que « vers en haut » de la hiérarchie. *L'évaluation est aussi un instrument de direction et de formation d'une équipe.*

- Le contrôle de la mise en œuvre des recommandations des évaluations doit être amélioré. Les rapports sont publiés et diffusés, mais seuls les dossiers politiquement les plus importants sont envoyés au Secrétariat général. Le reste reste au niveau de l'Office. On estime que plus de la moitié des recommandations sont appliquées.

*Nous entendons ici un indice confirmant l'hypothèse du niveau relativement bas dans la structure administrative sur lequel l'évaluation a un impact direct.*

- Les facteurs favorisant l'utilisation sont le professionnalisme de l'évaluateur (aussi par rapport au domaine évalué), la scientificité et sa capacité à communiquer. Une forte résistance du personnel provenait de la crainte des « contrôleurs de l'extérieur ».

*Les éléments techniques prédominent. Confirmation de l'importance de la rigueur scientifique et du rôle de l'évaluateur, de l'inefficacité des pressions externes.*

- Les effets des études sur le plan légal furent décevants. Cf. l'exemple des décomptes individuels de chauffage. Une mauvaise expérience avec une évaluation fut un cas provoqué directement par une intervention parlementaire. Il y a aussi des domaines à l'intérieur de l'OFE qui restent bizarrement épargnés par les évaluations. Il est supposé que les enjeux et les acteurs politiques concernés sont trop gros et sensibles, tandis qu'il était aisé de faire évaluer des domaines plus insignifiants économiquement et politiquement.

*Cela confirme la préférence pour l'origine interne, la prédominance de coalition entre l'idéologie et l'intérêt au détriment de l'information, ainsi que le côté peu dangereux car trop technique des 58 évaluations réalisées.*

- Les évaluations ont été très utiles pour réaliser le *pilotage* du programme.
- Les évaluations ne sont pas seulement des études de causalité. Leur plus-value est dans le *processus d'apprentissage* et les connaissances nouvelles qu'elles génèrent, sur le sujet lui-même mais aussi dans d'autres domaines. Il importe à l'avenir d'augmenter la capacité d'apprentissage de tous les acteurs. La distinction entre évaluation et recherche est donc nécessairement fluctuante et souple.

## **Commentaire**

*Nous aimerions relever encore deux éléments qui corroborent nos hypothèses :*

- *d'abord la confirmation de cette tension entre d'un côté l'effet direct et appliqué de nombreuses petites études et recommandations et de l'autre la difficulté à faire monter le*

*principe de l'évaluation au niveau supérieur pour des questions politiques et de rapports de force.*

- *d'autre part la tension entre l'évaluation comme une science causale à impact précis et comme un processus d'apprentissage beaucoup plus flou à cerner.*

### **6.3.1.2. L'Office Fédéral de l'Agriculture (OFAG)**

L'évaluation est relativement récente dans ce domaine. Une partie importante consiste en des clauses évaluatives. Nous retiendrons ici les points suivants.

- L'évaluation est comprise de manière large, incluant aussi bien un contrôle des résultats qu'une forme de recherche appliquée. Cela est dû aux particularités du domaine de l'agriculture, où recherche, contrôle et évaluation sont liés.  
*Cette remarque donne un exemple de la variété des approches et définitions de l'évaluation, mais aussi du fait que celles-ci sont liées aux besoins spécifiques des utilisateurs primaires de l'administration.*
- L'utilisation politique des évaluations est toujours plus facile si les résultats sont positifs. Elles sont donc en général plutôt *légitimatrices* ou donnant les bases d'*information* pour une décision.
- L'organisation décentralisée, les nombreux domaines d'activité et l'histoire de l'OFAG font que la supervision et la coordination du tout sont difficiles.  
*Une indication qui confirme l'importance du contexte de l'institution et de l'organisation dans la manière d'utiliser les résultats.*
- Les études sont diffusées en général jusqu'au niveau du chef de section. Elles sont publiées assez largement. Si les résultats sont un peu trop « chauds », on attend un moment propice pour les divulguer. Leur utilisation dépend beaucoup du chef de projet et du sujet traité.  
*Cela confirme l'importance de l'idéologie et des intérêts des acteurs dans l'utilisation.*
- En ce qui concerne des améliorations possibles, on signale aussi que l'évaluation n'est peut-être pas très bien placée au niveau de l'organigramme.  
*Une remarque allant dans le même sens que celles de l'OFE et soulignant l'importance du facteur organisation.*

### **Commentaire**

*Nous observons ici en particulier une confirmation que les évaluations sont surtout utilisées à un niveau plutôt bas du domaine, que l'usage terminologique est large, que la fonction d'évaluation est plutôt sous évaluée. Mais aussi que ce domaine est assez sensible politiquement (paiements directs, agriculture durable...).*

### 6.3.1.3. L'Office Fédéral de la Santé Publique (OFSP)

- La pratique de l'évaluation à l'OFSP date de la fin des années 80. Ses *domaines d'action se sont élargis au fil des ans*. Une nouvelle réorganisation en cours doit prochainement encore *renforcer le statut de l'évaluation* en le plaçant au niveau de la Direction.
- La compréhension de l'évaluation est clairement celle d'un *processus d'apprentissage* et absolument pas d'un *instrument de contrôle* (« not to prove but to improve »). La distinction entre évaluation et recherche est par conséquent parfois difficile à faire.
- L'évaluation peut remettre en question telle ou telle activité ou prestation, elle nomme des problèmes, mais n'est pas le *seul facteur influant sur la décision*, même si son importance gagne progressivement en croissance.
- La *valorisation des rapports* d'évaluation est une des caractéristiques de l'Office. L'utilisation y est définie comme la dissémination des résultats, que ce soit à l'intérieur ou à l'extérieur de l'Office. C'est une stratégie qui est due aussi au genre de thèmes traités (SIDA, etc.). Une partie substantielle *du budget* finançant l'OFSP est affectée à cette tâche.
- L'utilisation des résultats des évaluations est essentiellement du ressort de l'acheteur et / ou de la section impliquée. *Une utilisation au niveau de la Direction a lieu très rarement*.
- Un faible nombre d'évaluations de *type stratégique* ont lieu pendant la législature, pour vérifier, si possible confirmer ou alors corriger les options de politique de santé choisies.
- L'usage direct des évaluations sur des programmes précis est visible et traçable dans chaque cas. Par contre pour *les évaluations stratégiques, il est plutôt question d'influence*.
- L'élément déterminant pour expliquer une utilisation est *l'intérêt et la participation* de l'acheteur ou du partenaire du projet. En outre, la *qualité* de l'étude et son caractère *provoquant la discussion* et la réflexion, allant parfois jusqu'au *conflit*, jouent un rôle important. Une discussion animée sur des résultats étonnants voire dérangeants est une excellente utilisation.
- La *culture d'évaluation* est d'abord une mentalité analytique, toujours à la recherche du bon critère pour juger, comme dans la vie privée.

### 6.3.1.4. La Direction pour le Développement et la Coopération (DDC)

La DDC pratique l'évaluation de projets et programmes depuis près de 20 ans et a développé une méthodologie adaptée à cet effet, appelée PSER dans le jargon de la branche : Planification, Suivi, Evaluation, Réalisation. Comme pour l'OFSP, il n'est pas exagéré de parler d'une culture d'évaluation dans cet Office. L'importance de l'échantillon d'études dans nos questionnaires nous invitait à lui réserver une attention particulière. Nous retenons les faits saillants suivants :

- L'évaluation est intégrée dans un cycle de programme, un peu à l'instar du cycle d'une politique publique. L'utilisation est donc prévue dès le départ. L'objectif est l'amélioration de la qualité du travail. On attache de l'importance aussi bien aux mesures concrètes destinées à *améliorer la qualité du programme* qu'au *processus d'apprentissage* des acteurs concernés. La participation des acteurs locaux et des bénéficiaires sont des conditions de succès reconnues à la réalisation du mandat confié à la DDC.
- La DDC pratique aussi bien l'autoévaluation que l'évaluation externe. Les bureaux locaux et les partenaires sur le terrain sont *parties prenantes* à tous les moments de l'évaluation. Ceci augmente les *chances d'application* des mesures préconisées.
- La valorisation des évaluations est en général l'affaire des sections, et varie passablement d'une section à l'autre. Un comité particulier gère le contrôle de l'application des recommandations. Il y a là encore une marge d'amélioration importante à réaliser. Les évaluations plus importantes sont discutées au niveau des Directeurs. Mais cela reste encore l'exception.

*Une remarque particulièrement importante au vu du nombre d'études de la DDC dans nos échantillons. L'impact recherché par ces études se limite bien dans leur grande majorité au niveau du projet / de la section. La valorisation politique reste limitée, voire absente.*

- L'utilisation des évaluations dépend en premier lieu de *l'intérêt* que l'on aura réussi à éveiller chez les parties concernées. Sans *participation active*, il n'y a aucun effet sérieux ou durable. En outre, *le choix et la crédibilité de l'évaluateur, la planification et la préparation de l'évaluation sont fondamentaux*.
- Les évaluations répondent en premier lieu *aux besoins des partenaires du terrain et au management du projet*. *Qu'elles répondent aux besoins extérieurs (Parlement, population...)* reste une exception, même si cet aspect gagne et gagnera encore en importance à l'avenir.
- Outre les fonctions classiques énoncées, les évaluations servent à développer le savoir, la responsabilité (accountability), le développement de l'organisation.
- L'OCDE a récemment critiqué cette technique privilégiant des acteurs en définitive internes au programme pour l'évaluer. De même, une évaluation du fonctionnement interne de la DDC en 2000 a montré qu'il y aurait avantage à mieux distinguer les évaluations courantes de projets d'évaluations plus importantes, sectorielles, thématiques ou politiques.

### **Commentaire**

*L'intérêt ici pour notre propos est cette double approche, qui conjugue à la fois l'évaluation technique de projets précis qui a des impacts directs et retraçables, et l'apprentissage des*

*acteurs comme facteur d'influence indirecte sur les programmes. Il y a la volonté de privilégier à la fois l'usage direct et l'influence.*

### **6.3.1.5. L'Office Fédéral de la Justice (OFJ)**

L'OFJ joue un rôle important dans la discussion et la promotion de l'évaluation en Suisse. C'est en son sein que s'était constitué un groupe de travail pour l'évaluation (AGEVAL). C'est lui qui est à l'origine du PNR 27 du FNRS dont nous avons utilisé les résultats. C'est aussi l'OFJ qui avait réalisé il y a 6 ans l'enquête de base, dont les résultats constituent la base matérielle de notre étude. Son action concerne bien évidemment d'abord l'évaluation législative et les clauses évaluatives, ce qui distingue l'OFJ d'autres Offices rencontrés, plus orientés sur l'évaluation de projets. Nous aimerions relever en particulier les points forts suivants :

- L'évaluation se définit ici surtout comme un élément de la *méthode* de la démarche qui aboutit à l'élaboration de législations. Elle intervient donc beaucoup au niveau de la préparation des lois afin de savoir si la loi prévue pourra développer ses effets. L'évaluation est utile ensuite au niveau du suivi des conséquences de la loi. La Confédération a souvent la compétence du contrôle de l'application des lois au niveau cantonal ou communal.  
*Outre l'insistance sur la qualité méthodologique et scientifique, soulignons ici les particularités des besoins de l'Office en termes d'évaluation.*
- L'évaluation à l'OFJ n'est pas pratiquée de manière systématique mais ponctuelle, plus ou moins régulière, et dans certains domaines seulement, comme par exemple la Loi sur l'aide aux victimes. Des évaluations régulières ont lieu, qui servent de *base très importante pour les corrections apportées* par la suite à la Loi. Mais généraliser les évaluations dans tous les domaines n'a pas beaucoup de sens, car il importe qu'on puisse reconstituer un modèle causal. Or il y a plusieurs domaines dans le droit où cet élément ne joue pas de rôle (par exemple le Droit international, la réforme des institutions). *L'OFJ opte clairement pour un usage instrumental de l'évaluation.*
- La condition principale pour l'utilisation d'une évaluation est qu'elle *intervienne au bon moment dans le travail de préparation législatif, c'est-à-dire le plus tôt possible*. Faire une évaluation au moment où une loi est en Commission ou devant le Conseil Fédéral est souvent trop tard. *Elle risquera à ce moment-là d'être utilisée comme argumentation pour ou contre la Loi.*
- *L'institutionnalisation* de la fonction d'évaluation est à soutenir, que ce soit sous la forme des clauses évaluatives et/ou la création de ressources supplémentaires. Mais le danger existe de voir ces clauses se transformer en routine formelle, voire en alibi.

- Les difficultés rencontrées pour l'utilisation des évaluations en général dans la Confédération reposent sur les facteurs suivants : grande variété de styles d'évaluation et manque de systématique, l'ancrage de l'évaluation au niveau de l'Office et non du Département, le manque de formation, l'insuffisance d'expertises vraiment externes, la difficulté à coordonner les évaluations avec le niveau politique, en particulier le niveau parlementaire.

*Ces remarques générales d'analyse sur la situation de l'évaluation vont dans le même sens que plusieurs de nos résultats ainsi que des remarques d'autres Offices.*

### **Commentaire**

*Nous aimerions relever ici la confirmation du fait que non seulement les besoins en évaluation diffèrent d'un Office à l'autre, mais aussi les besoins quant au moment où l'évaluation est le plus facilement utilisable. Nous notons aussi la politisation possible des évaluations lorsqu'elles sont communiquées dans les Commissions.*

#### **6.3.1.6. Les Services du Parlement : le Secrétariat de la Commission de Gestion (CdG) et l'Organe parlementaire de contrôle de l'administration (OPCA)**

Situé à l'interface entre le Parlement et l'Administration, l'OPCA est l'instrument mandaté par la Commission de Gestion (CdG) des Chambres pour effectuer ou mandater des évaluations. L'OPCA doit donc servir les besoins des Parlementaires mais s'inspirer des compétences et connaissances acquises dans l'administration ou le monde académique. En outre la demande de la CdG d'annexer au rapport annuel du CF dès 2000 une liste des évaluations réalisées et les discussions autour de la loi devant mettre en œuvre l'Art. 170 de la Constitution fédérale placent cet organe dans une position intéressante, idéale pour faire la transition avec les milieux parlementaires.

Nous aimerions souligner les éléments suivants tirés des divers entretiens :

- L'évaluation est certes un outil scientifique, mais dont les *résultats doivent être politiquement utilisables* immédiatement. Pour un parlementaire par contre, une réponse à une interpellation peut être déjà une évaluation.
- Les évaluations réalisées par l'OPCA servent de base pour une action ou décision de la CdG, mais *elles ne sont pas la seule source d'information* ou de jugement à disposition.
- Un parlementaire *n'utilisera une évaluation que si ses résultats l'intéressent immédiatement*. Pour le reste, l'intérêt pour les évaluations, et en général pour les sciences sociales, n'est pas très élevé chez les parlementaires, qui sont majoritairement des juristes.
- Les *parlementaires font rarement usage* des évaluations réalisées par l'administration. Ils font faire plutôt leurs propres recherches ou commanditent une étude à l'OPCA.

- *Un usage aussi bien direct qu'indirect des évaluations* a lieu avec les études de l'OPCA, mais il n'est pas possible de pondérer les chiffres de l'une ou l'autre variante. Le groupe qui en profite le plus varie avec la fonction qu'on lui attribue. Une fonction de légitimation renforcera les partisans du statu quo, une fonction d'amélioration aidera les réformateurs, et ceci à travers tous les partis ou les niveaux de l'administration.
- Les facteurs décisifs qui expliquent que des parlementaires utilisent les évaluations sont : *être là au bon moment, avoir une base scientifique solide, être simple et compréhensible, être appuyé par des relais de communication (médias), être politiquement facile à utiliser, la capacité au changement de l'Office...* Non seulement les utilisateurs sont importants, mais aussi les mandataires.
- *Les changements opérés par les recommandations de la CdG sont certes élevés en pourcentage<sup>34</sup>, mais ils ne concernent qu'exceptionnellement le cœur de la politique en question. En général, il s'agit d'adaptations et de correctifs de mise en œuvre.*
- En ce qui concerne les processus de décision, *l'influence reste indirecte et faible, il y a beaucoup d'autres facteurs plus importants qui jouent un rôle.*

### 6.3.1.7. Valorisation intermédiaire

Après ces diverses interviews dans l'administration, nous aimerions souligner cinq résultats saillants :

- d'une part, les nombreuses confirmations d'éléments de nos résultats obtenus jusqu'ici, en particulier en ce qui concerne les facteurs explicatifs de l'utilisation ;
- d'autre part les nombreuses remarques faites au sujet de l'importance du *processus d'apprentissage inhérent à toute démarche d'évaluation* et de sa prise en compte lorsqu'on parle d'utilisation ;
- troisièmement, la confirmation de *l'importance relative des recommandations d'évaluation* lors de prise de décision de type « stratégie politique » ;
- ensuite, l'information confirmée selon laquelle le niveau d'impact des résultats de l'évaluation se situe plutôt au *niveau de la mise en œuvre* d'une politique plutôt qu'au niveau de sa redéfinition, de sa réorientation ;
- enfin, plusieurs éléments importants sur *l'évolution du statut de l'évaluation*, manifestement en phase de consolidation, avec des idées pour améliorer l'intégration de ses résultats dans les processus politiques ;

---

<sup>34</sup> 80-85%, selon l'étude de Janett (2000)

### 6.3.2. Entretiens avec des parlementaires

#### *Introduction*

Tout comme pour les Offices, nous avons privilégié dans notre choix de parlementaires ceux qui siègent dans des Commissions correspondant aux principaux domaines d'action étudiés jusqu'ici. Notre sélection de 6 membres du Parlement panachait les cantons, les partis et le degré d'expérience en tant que parlementaire.

#### *Résultats et commentaires synthétisés*

Notre intérêt n'est pas de personnaliser les propos mais de rendre compte de l'approche et de l'utilisation que fait le Parlement des évaluations. Par conséquent, nous regrouperons nos remarques sous cinq thèmes.

##### *1. La connaissance du domaine de l'évaluation*

L'évaluation en tant que science sociale n'est pas vraiment connue. Les thèmes et les contenus le sont bien sûr, mais les définitions varient entre les parlementaires encore une fois autant qu'entre les Offices, sans toujours se recouper d'ailleurs. Elles semblent liées aux domaines de travail des Commissions dont sont membres les parlementaires. Les notions de « surveillance » ou de « contrôle » reviennent le plus souvent, ainsi que d' « informations » arrivant au bon moment. Un seul utilisera l'expression d' « examens de l'application des objectifs ». Seul un parlementaire y voyait un élément de culture et d'apprentissage permanent. L'évaluation est essentiellement une information pour « apprécier la situation », pondérer ce qui est important de ce qui ne l'est pas.

##### *2. L'emploi et l'utilisation des études*

Fondamentalement, nous avons constaté un fossé assez large entre le travail d'évaluation que réalise l'administration et le degré d'information des parlementaires sur ce travail. Les sources d'information des parlementaires ne semblent être qu'exceptionnellement les évaluations fournies par l'administration. Un seul parlementaire trouvait cela problématique et souhaitait plus de transparence. Surtout, les Commissions tiennent à garder la compétence du jugement politique sur les informations qu'elles ont reçues et elles ont tendance à distinguer entre les rapports provenant d'experts ou de l'administration, qui fournissent une information « neutre », et le jugement politique qui est du ressort de la Commission. La fonction que la CdG attribue à des études, rapports ou évaluations varie en fonction du sujet. En général, les aspects « contrôle » et « pilotage » dominent. Mais en tant que parlementaire généraliste, l' « information » à transformer en outil d' « argumentation » est aussi utile. Tous les rapports *commandés directement par les Commissions* sont très utilisés et débouchent en général sur une action, en particulier vis-à-vis du Conseil Fédéral.

Signalons pour illustrer des exemples tirés de deux Commissions :

- *Expériences dans le domaine de la Commission de Gestion (CdG)*

Deux cas d'étude sont mentionnés comme ayant été particulièrement bienvenus.

L'un, établi sur mandat de la CdG, porte sur la situation du personnel fédéral. Ses résultats ont coulé ensuite avec succès dans le processus qui a abouti à la nouvelle Loi sur le personnel de la Confédération (LPers.). L'autre est la demande au Conseil Fédéral d'avoir en annexe du rapport annuel une liste des principales évaluations réalisées pendant l'année. La première liste a été fournie avec le rapport 2000. La CdG, tout en se félicitant de cela, regrette qu'il n'y ait aucun commentaire ni conclusions politiques, en particulier sur l'usage qui en a été fait (!), et constate qu'il y a un déséquilibre entre les sujets abordés<sup>35</sup> qui n'est pas satisfaisant ni entièrement compréhensible. La CdG en conclut que les raisons des choix opérés et les critères utilisés ne sont pas très clairs et devraient être mieux définis et discutés à l'avenir. Notons au passage que plusieurs parlementaires ignoraient l'existence de cette annexe, y compris des membres de la Commission de Gestion.

- *Expériences avec la Politique extérieure*

Plusieurs cas sont signalés ici :

- le rapport Voyame sur le Ruanda après le génocide de 1994. Ce rapport a été très discuté en Commission, mais son impact sur le suivi de la politique suisse reste vague. L'influence de la Commission sur la pratique de la DDC reste très marginale.
- La prochaine votation sur l'adhésion à l'ONU fait ressortir des tiroirs des rapports rédigés récemment.
- L'art. 47 de la Loi sur les rapports entre les Conseils (LREC) a été révisé suite à plusieurs rapports d'un groupe d'étude de la CPE. Il donne plus de poids à la CPE dans la conduite des affaires extérieures, et introduit le droit de consultation dans certains cas.
- Le travail concret de la DDC ne constitue qu'un élément parmi d'autres du travail de la CPE, souvent limité aux questions autour du crédit-cadre. Les gros rapports d'évaluation internes sont reçus par la Commission, mais cela n'est pas fréquent et ne provoque pas de grands débats. En cas de visite de parlementaires à l'étranger ou de délégations étrangères, des rapports ad hoc sur ces pays sont commandés à la DDC ou au service compétent.
- Les séances de la CPE comportent toujours un point appelé « actualités du DFAE » où le ministre compétent et les chefs de service concernés interviennent *oralement*. Mais la majorité des sujets traités touche à l'actualité ou aux grands thèmes de politique internationale du moment, et ne consiste pas à faire l'évaluation de l'action de la Suisse ici ou là.

---

<sup>35</sup> L'annexe rapporte un grand nombre d'études par exemple sur la LAMal.

### *3. Les besoins des parlementaires*

Fondamentalement, les parlementaires rencontrés ont souligné la nécessité de disposer de documents courts, nerveux et aisément transformables en action politique. Ils en regrettent l'aspect trop technique et voient leur rôle et leur compétence plus au niveau des grands principes qu'à celui du détail de tous les objectifs secondaires. Tout document est susceptible d'être compris comme une évaluation, quelle que soit par ailleurs son origine. Pour l'un, une interview d'un responsable de l'administration dans un journal de boulevard peut aussi bien servir à formuler les bonnes questions sur le sujet en question qu'un lourd rapport. « Un bon rapport est celui qui est là au bon moment ».

Les Commissions travaillent plus avec leurs propres « rapports », qu'elles réalisent et/ou commanditent elles-mêmes, auxquels elles adjoignent souvent des rapports complémentaires demandés à un expert et/ou à l'administration. Mais on parle bien de « rapports » et pas nécessairement d'évaluations. Les choix du parlementaire dans la masse d'informations reçues sont dictés par son engagement professionnel et de Commission. *C'est aussi cet intérêt et cet engagement personnel qui déciderent de l'utilisation ou non d'un rapport.*

En général, tous les parlementaires nous ont renvoyés au rapport de la Commission de Gestion (CdG) comme étant LE rapport d'évaluation du Parlement. Un seul parlementaire nous a dit que sa tâche était de s'assurer surtout que la méthode choisie pour faire le rapport était sérieuse, puis de lire le résumé.

### *4. La fonction et l'importance des rapports dans la formation d'une opinion et d'une décision*

Aucun interlocuteur n'a pu se souvenir d'un cas où un rapport l'aurait fait changer d'avis. Tout au plus cela a-t-il permis de pondérer différemment les arguments que l'on prévoyait d'utiliser. Aucun n'attend de lire un rapport pour avoir son opinion politique faite sur un sujet. L'impression est que les évaluations et les rapports en général n'ont que peu d'impact direct. Cela est dû au fait que les Commissions souvent commanditent un rapport quand il y a un gros problème ou une affaire. Lorsque le rapport est livré, le problème a souvent disparu ou s'est transformé. Quelques rapports ont certes eu un effet immédiat, mais c'était surtout lors « d'affaires », et l'on ne peut utiliser le terme d'évaluation pour cela. Le facteur principal qui explique pourquoi une étude a un effet ou non est son intérêt politique, souvent lié à l'importance du thème. Ainsi, la question pour un parlementaire n'est pas de savoir ce que dit tel ou tel rapport, mais de trouver l'information dont il a besoin au moment opportun pour son intérêt propre. Un seul parlementaire a souligné le danger de cette attention permanente à l'opportunité politique. Mais les Parlementaires se sentent en situation de faiblesse par rapport à la presse ou aux fuites organisées par tel collaborateur d'Office. Ils sont toujours en train de réagir et n'ont pas suffisamment d'outils pour gérer les dossiers.

### *5. Les relations avec l'administration et l'avenir de l'évaluation : chances et handicaps*

L'administration est encore ressentie comme assez opaque et toute puissante. Plusieurs entretiens ont fait référence très clairement aux luttes de pouvoir entre Parlement et administration, ou entre Parlement et Conseil fédéral, ou à l'intérieur du Conseil fédéral. A une exception près, les parlementaires rencontrés se sont tous exprimés de manière favorable à l'intensification du travail d'évaluation, ayant même de la peine à identifier des secteurs précis où celle-ci ne se justifierait pas. Une des questions difficiles reste celle de la « Federführung » dans certains dossiers, ainsi que la question de savoir où et comment renforcer cette fonction. Plusieurs interlocuteurs n'avaient pas (plus) conscience de l'existence de l'art. 170 de la nouvelle Constitution fédérale. Mais nous avons aussi perçu la crainte de voir gonfler des budgets, des Offices et des séries de publications, sans que cela ne réussisse à améliorer la position et la profession de parlementaire ni à garantir une meilleure gestion.

Parmi les handicaps freinant cette évolution, les constats suivants ont été faits :

- Le fédéralisme, et en particulier le fédéralisme d'exécution rend tout très (trop) compliqué pour avoir une information solide.
- La complexité des sujets et la lourdeur du système fait perdre de vue l'ensemble et rend le recul difficile.
- Le système de Parlement de milice ne permet pas (plus) de répondre à cette tâche sous la forme actuelle. Que ce soit par le biais d'une professionnalisation et/ou d'un collaborateur personnel, les parlementaires ont besoin de ressources supplémentaires pour gérer et trier la quantité d'informations qu'ils reçoivent. Mais leur capacité à s'organiser est aussi en cause.
- Les objectifs sont souvent absents, insuffisants, ou établis sur la base d'un tel compromis qu'ils n'en sont plus des objectifs. Le manque de suivi des dossiers dans les Commissions et d'une législature à l'autre est gênant.

Parmi les éléments favorisant le développement de l'évaluation, citons en particulier l'intérêt des parlementaires pour ces questions, ainsi que la lente pénétration d'habitudes et de techniques issues du monde professionnel, de l'économie privée, liées directement ou non à la « nouvelle gestion publique ».

### **6.3.3. Valorisation intermédiaire**

L'image reçue ici est très claire, et confirme nos hypothèses :

- *La connaissance et l'expérience que quelques parlementaires ont de l'évaluation en tant que branche scientifique est très faible.*

- *Les parlementaires utilisent pour former leur jugement essentiellement leur information sectorielle propre et leur jugement (flair) et sensibilité politique.*
- *Ils utilisent toutes sortes de documents internes ou externes à l'administration essentiellement dans un but argumentatif et pour se procurer un avantage à ce niveau.*
- *Les besoins des parlementaires en matière d'information pour juger d'une politique ne sont pas (ou seulement exceptionnellement) satisfaits par les évaluations. Même le résumé, s'il n'est pas formulé de manière politique, n'est pas très utile. Les besoins des parlementaires sont essentiellement liés à l'opportunité politique, à la simplicité d'utilisation de l'information reçue et à sa faisabilité politique. L'opinion du parlementaire se forge d'abord au niveau de ses convictions, de son expérience et des documents utilisés dans les discussions au niveau des Commissions ou des groupes parlementaires.*
- *La vision généralement rencontrée de la fonction d'évaluation va essentiellement dans le sens d'une haute surveillance de ce que font le Conseil fédéral et l'administration et peu dans le sens du pilotage.*
- *Les parlementaires n'ont que rarement connaissance des évaluations stratégiques faites par l'administration, et jamais des autres. Ils n'utilisent fondamentalement que celles qu'ils commanditent eux-mêmes.*
- *Les parlementaires rencontrés sont en général ouverts et intéressés au renforcement de cette fonction d'information et d'analyse que représente l'évaluation.*

#### **6.4. ELEMENTS D'AUTRES ETUDES**

Parmi les documents consultés pour notre étude, nous aimerions relever en particulier deux documents récents parus sur la question des évaluations en Suisse, que nous avons déjà cités à plusieurs reprises, à savoir les études de Th. Widmer et de D. Janett (cf. bibliographie). Nous aimerions relever rapidement encore un élément de chacune de ces études, qui chacun à sa manière est une réponse à nos hypothèses.

- Outre les interviews réalisées au niveau des Offices, Widmer (2001 : 75-77) s'est également entretenu avec des responsables des Secrétariats généraux des Départements, c'est-à-dire la charnière entre le niveau des Offices (où se déroulent et sont utilisées les évaluations) et la tête du Département et les fonctions hautement politiques. Il y a constaté en général un intérêt plutôt modéré pour les évaluations, qui ne sont qu'une forme parmi d'autres de l'information que les SG reçoivent en retour sur leur action (et pas la plus importante), ainsi que l'information selon laquelle les évaluations que les Secrétariats généraux reçoivent sont rarement utilisées ou discutées. Outre la surcharge de travail et le manque de personnel pour expliquer ce fait, on fait référence au fait que les résultats

d'évaluation sont rarement formulés ou présentés de telle manière qu'une utilisation dans les affaires courantes et actuelles du Département soit aisément possible. « *Les exceptions à cette règle sont en général les questions qui sont soit très politisées dans l'actualité immédiate, soit touchent à un intérêt personnel sur un thème* » (Widmer 2001 : 75)

- Janett s'est quant à lui penché sur les études menées par l'OPCA pour le compte de la Commission de Gestion (CdG). Sur la base d'interviews en profondeur, il aboutit à la conclusion que le *moment opportun*, *l'engagement des membres* des Commissions et la *vulgarisation politique* des résultats sont parmi les facteurs clés expliquant qu'un rapport est repris par la CdG. Il constate néanmoins (2000 : 7) que « *les rapports ont rarement changé le cœur de la politique, mais ont souvent touché des aspects secondaires des structures de mise en œuvre* »

## 6.5. VERIFICATION DES HYPOTHESES

Cette recherche qualitative nous a permis de récolter un lot important d'informations qui nous confortent dans nos hypothèses d'explication :

- *nous voyons tout d'abord confirmé en général le modèle élaboré par C. H. Weiss. L'information scientifique contenue dans une évaluation doit, pour être valorisée dans un programme d'action politico-administratif, se montrer plus influente que les facteurs « institution », « intérêt » et « idéologie », respectivement doit entrer dans une coalition d'influence avec l'un et/ou l'autre de ces facteurs.* En fonction des sujets, mais peut-être aussi du moment, ces 4 sphères sont plus ou moins volumineuses et le jeu des relations entre elles peut varier du modèle « roulement à billes » jusqu'au modèle « billard ».
- *Nous avons reçu plusieurs confirmations du fait que le type des évaluations que nous avons majoritairement dans notre échantillonnage de départ (projet/programme) étaient bien adaptées aux besoins d'information des Offices. Leur fort taux d'utilisation à ce niveau ne présage cependant pas beaucoup de leur impact au delà de la mise en œuvre du programme en question, par exemple au niveau de la correction des objectifs supérieurs. Dans la mesure où ce genre de décision implique la participation d'acteurs situés à des étages plus sensibles aux 3 autres domaines d'influence, l'impact de ces études d'évaluation à un haut niveau du cycle d'une politique publique est toujours plus difficile à réaliser et à vérifier.*

*Les exemples apportés par Janett laissent même entendre que les recommandations de la CdG ne modifient que rarement la direction des politiques choisies. D'une part parce que la CdG elle-même n'a pas de moyens suffisants pour réaliser des études vraiment globales, d'autre part parce que le système intégré des niveaux de prise de décision du système suisse, avec en particulier le principe de concordance et de compromis, font que*

*les intérêts des parties intéressées à changer cela ne peuvent qu'exceptionnellement s'imposer. Ce principe du système politique s'avère être à terme un des principaux obstacles et handicaps à l'élargissement de la fonction d'évaluation pour la conduite de politiques publiques.*

- *Il est néanmoins étonnant et réjouissant de constater que la culture d'évaluation continue à se propager dans l'administration. Nous avons constaté qu'à ce niveau-là, au contraire du niveau parlementaire, des instruments structurels se mettent en place pour favoriser une évaluation stratégique, sectorielle etc. plus fréquente et mieux qualifiée. Cette institutionnalisation « par en bas », plus celle « par en haut » de l'Art. 170 de la CF représentent des facteurs favorisant l'apprentissage, donc dessinent des éléments des conditions cadres propres au système suisse favorisant l'apprentissage par « reconstruction ».*

## **7. CONCLUSIONS ET VALORISATION DES RESULTATS**

Arrivés au terme de notre parcours, il est nécessaire de reprendre les diverses leçons à tirer des résultats et de la démarche même de notre étude. Nous voulons en particulier voir en quoi ces résultats obtenus modifient le schéma explicatif de départ (p.13) et celui élaboré par la suite (p. 58).

### **7.1. LES RESULTATS DE NOTRE ETUDE**

#### **7.1.1. Ce que la première démarche (quantitative) a montré**

1. L'évaluation reste une activité minoritaire dans les pratiques de l'administration fédérale. Elle est particulièrement pratiquée dans des domaines travaillant essentiellement par des projets et programmes d'action bien délimités, ne reposant pas sur des bases légales contraignantes, mais plutôt sur des nécessités de mesurer les résultats et si possible les succès obtenus par ces mesures (coopération au développement, prévention de la santé...).
2. La définition de l'évaluation est encore imprécise et sa qualité, en l'absence de standards communément connus et reconnus, variable.
3. Entre 20 et 50% des évaluations commanditées par l'administration fédérale sont fortement ou très fortement utilisées dans les processus de décision. Les évaluations jouent donc un rôle décisif dans un cas sur deux environ.
4. Seules 5 –10% des études finissent à la poubelle ou restent sans aucun effet dans un fond de tiroir.

5. L'évaluation en Suisse fait des progrès quantitatifs et qualitatifs depuis le début des années 90. Elle s'étend progressivement à toujours plus de domaines dans le travail de l'administration.
6. Cette amélioration se montre en particulier au niveau technique. Les standards des rapports d'évaluation sont d'ores et déjà assez élevés, au point que l'utilisation d'une évaluation est certes liée au respect de quelques éléments techniques, mais moins qu'on aurait pu l'attendre.
7. Les clauses évaluatives participent fortement à une utilisation intense des évaluations.
8. Les éléments déterminants suivants expliquant l'utilisation ont été identifiés ou confirmés :
  - l'intégration de la politique publique évaluée dans l'ordre du jour du débat politique (y compris conflictuel) du Parlement ;
  - l'intégration de l'évaluation dans un programme et une planification de travail de l'Office concerné ; sa participation, voire son initiative dans le processus ;
  - l'importance de la collaboration entre Office et évaluateur, y compris la participation de celui-ci au travail de valorisation de ses résultats dans le travail de l'Office ;
  - les autres circonstances dans lesquelles le rapport est reçu jouent également un rôle important, comme la médiatisation ;
  - les études cherchant à légitimer ou donner des informations permettant l'argumentation sont favorites à l'utilisation, ainsi que la volonté de piloter un programme ;
  - les études orientées sur la mesure de l'atteinte de l'impact et de l'objectif sont plus prisées que celles axées uniquement sur le contrôle ;
  - au niveau technique, les facteurs comme la simplicité ou la brièveté, l'emploi de méthodologies nouvelles ne sont pas toujours des aides facilitant l'utilisation ;
  - par contre l'approche qualitative, l'existence d'un résumé, d'une stratégie de communication et de valorisation sont décisives.

### **7.1.2. Ce que la démarche complémentaire (qualitative) a montré**

1. Les résultats quantitatifs obtenus sont encore relativement muets sur le niveau d'impact de l'utilisation.
2. Ils correspondent essentiellement aux besoins des Offices. Ces besoins peuvent varier d'un Office à l'autre. Ceux-ci se trouvant essentiellement chargés de la mise en œuvre, l'impact de ces études ne dépasse que rarement le niveau de l'Office et la rétroaction, l'accroissement de la connaissance susceptibles de remettre en question ou de modifier sensiblement l'orientation de la politique n'ont lieu que rarement.

3. Nous n'avons pas constaté de différences fondamentales au sujet de l'utilisation entre la situation observée via notre banque de données et celle se basant sur des informations plus récentes. Les résultats de l'étude de Widmer (2001) et Janett (2000) pour l'essentiel confirment les résultats obtenus par notre enquête. Tout au plus avons-nous rencontré plus de cas et identifié un intérêt apparemment plus grand pour des évaluations un peu plus larges et stratégiques, ainsi que le souci d'améliorer l'ancrage de la fonction d'évaluation dans les Offices et Départements.
4. Cette utilisation s'organise dans le cadre des interactions entre les 4 « I » : Idéologie, Intérêt, Institution et Information.
5. Nous observons une confirmation de notre hypothèse concernant la « politisation ». Les évaluations concernant des domaines discutés au niveau parlementaire sont plus facilement utilisées. Ceci d'autant plus que les Commissions parlementaires elles-mêmes commanditent toute une série de rapports et d'évaluations qu'ils utilisent selon leurs besoins.
6. L'utilisation se fait aussi bien au niveau d'un impact direct qu'à celui d'une influence indirecte sur les niveaux de la politique publique. L'impact direct est plus reconnaissable au niveau de la mise en œuvre des projets - programmes d'action administratifs. C'est à ce niveau que l'importance du « I » d'Information est la plus importante et traçable. C'est à ce niveau-là aussi que nous avons surtout repéré le souci de pilotage. L'influence sur les autres niveaux reste plus vague et aléatoire.
7. Afin de renforcer la composante Information dans le schéma des 4 « I », il est certes nécessaire de garantir un très bon niveau technique et scientifique des évaluations. Toutefois, il apparaît que le timing, l'engagement personnel de l'évaluateur et des commanditaires, la puissance médiatique des résultats et des recommandations jouent un rôle nettement plus important, surtout pour le niveau parlementaire.
8. Les niveaux administratifs supérieurs sont peu informés des résultats des évaluations importantes réalisées. Ils en déclenchent eux-mêmes également peu.
9. Les besoins en information des niveaux administratifs aptes à changer l'orientation de politiques qui s'avérerait inefficaces (SG, Dépts. et Parlement) ne sont pas satisfaits par les études d'évaluation réalisées habituellement.
10. Le Parlement n'est que très partiellement armé pour faire du pilotage de politiques publiques. Il intervient surtout au niveau de la surveillance.

### **7.1.3. Ce que le tout nous apprend**

1. Nous partions du principe que l'utilisation directe était la raison d'être de l'évaluation. Nous avons constaté que forcer sur l'utilisation des évaluations à tout prix c'est prendre le

risque de renforcer une vision très « projet-programme » de toute politique publique, au détriment d'une vision d'objectifs politiques à rediscuter. De même, c'est favoriser un apprentissage de type « single loop » au détriment d'un apprentissage plus profond.

2. Les facteurs favorisant une meilleure utilisation passent d'une part par une plus forte institutionnalisation (harmonisation des définitions, rigueur de la démarche, lien et importance hiérarchique plus forts, bonne communication...) et d'autre part par le développement intellectuel et de la connaissance, de l'apprentissage permanent, que ce soit dans les Offices ou dans le Parlement.

#### **7.1.4. Et par rapport à la théorie scientifique ?**

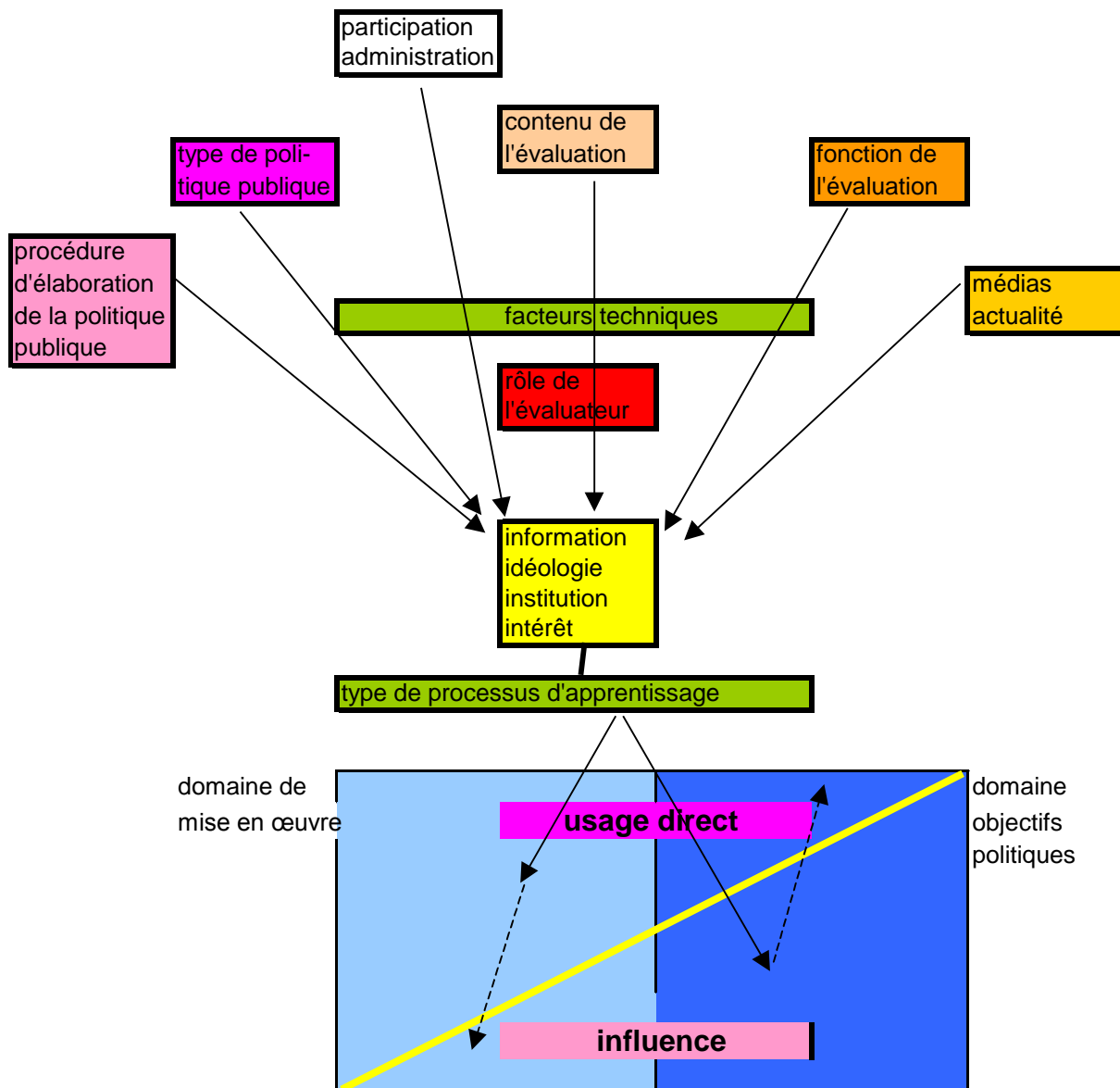
1. Sans le chercher ni le deviner, nous avons obtenu une confirmation du côté « frères ennemis » de Weiss et Patton, chacun insistant avec raison sur l'autre axe du pôle de facteurs décisifs, à savoir les utilisateurs (et bénéficiaires) primaires et les commanditaires ou décideurs.
2. Nous observons une confirmation de l'approche de Patton dans l'importance primordiale des « primary users » et de leurs besoins pour rendre l'évaluation plus utilisée.
3. La suite de la démarche souligne également combien son intérêt est centré sur ce groupe, au détriment des autres groupes potentiels, comme les politiciens par exemple.
4. Mais notre travail confirme également la finesse de la perception de Weiss, qui montre bien pourquoi tous ces facteurs importants pour l'administration ou les bénéficiaires n'en sont pas pour les politiciens, qui ont eux besoin d'une traduction supplémentaire des résultats d'évaluation pour les rendre efficaces dans leur travail.

Tout ceci nous amène à revenir sur notre modèle explicatif de départ et à le modifier.

#### **7.1.5. Nouveau modèle explicatif. Les facteurs et le processus d'utilisation d'une évaluation**

Sur la base des informations recueillies pendant cette recherche, nous pouvons modifier notre modèle causal de départ, et l'enrichir de la manière suivante :

## Les facteurs et le processus d'utilisation des évaluations



Cette représentation ne cherche qu'à reprendre l'essentiel de nos découvertes et essaye de les organiser graphiquement. Nous renonçons à l'expliquer plus avant.

### **7.2. L'EVOLUTION DE L'EVALUATION DANS L'ADMINISTRATION ET DANS LE PARLEMENT**

Dans la mesure où les progrès de l'évaluation et de son utilisation en Suisse vont dépendre en particulier des processus d'apprentissage générés par l'expérience déjà acquise en la matière depuis 10 ans, il nous semble utile de regrouper pour conclure quelques-unes des informations

recueillies sur ce sujet dans la seconde partie de notre démarche afin de les mettre en perspective. De la même manière que notre étude a été provoquée en quelque sorte par l'existence de bases de données fédérales qui donnaient indirectement une image de la culture d'évaluation il y a 10 ans, nous voulons souligner pour terminer l'évolution de ce dossier et observer la direction qu'il prend aujourd'hui.

Nous pouvons mentionner les éléments suivants qui témoignent de l'avancement du thème de l'évaluation en Suisse.

- Il y a bien sûr d'abord l'Art. 170 de la Constitution Fédérale. Cet article existe maintenant. Même s'il est difficile de savoir ce que la Loi qui doit le mettre en œuvre dira exactement et sous quelle forme cet article se réalisera, il n'est déjà pas anodin que cette fonction de vérifier l'efficacité des mesures étatiques ait finalement été reconnue jusque dans le plus haut texte normatif de Suisse. Les discussions et préparations parlementaires sont en cours. Le lieu institutionnel où sera ancrée cette fonction n'est pas indifférent<sup>36</sup>. Faut-il créer un service interdépartemental, ou se contenter d'un réseau professionnel interne, faut-il renforcer la Chancellerie, et en particulier les Services du Parlement, ou au contraire confier la tâche à chaque Département et Office de manière directe, ou à chaque Commission parlementaire ? Toutes les options sont encore en l'air, et nul doute que le schéma des 4 « I » va trouver là prochainement un nouvel exemple de sa pertinence pour expliquer la décision qui sera prise. Cet article constitue sans nul doute un premier élément de conditions cadres nécessaires au développement de la culture d'évaluation en Suisse, et en particulier pour le Parlement. Si l'on tient compte de l'importance du facteur d'influence indirecte des évaluations et de la nécessité de répondre aux besoins d'informations des premiers utilisateurs, on pourrait imaginer une solution très profitable pour les parlementaires si les Services du Parlement avaient une fonction importante dans le programme de la mise en œuvre de l'Art. 170.
- La publication décidée des principales évaluations en annexe du rapport du Conseil fédéral est une première mesure pour familiariser les parlementaires avec cette fonction. Reste à trouver le moyen de la rendre lisible et attractive.
- L'accroissement des textes de loi avec clauses d'évaluation est sensible ces dernières années. Nous avons vu qu'ils sont jugés comme particulièrement efficaces pour augmenter l'efficacité de la Loi. C'est un premier modèle d'institutionnalisation.
- Quelques Offices ont développé d'autres modèles ou sont sur le point de le faire: ainsi l'OFE a-t-il, sur la base de l'évaluation du programme Energie 2000, suggéré de rattacher l'évaluation non plus au niveau des sections, mais directement au niveau du Secrétariat

---

<sup>36</sup> Cf. Widmer (2001. 90ss.)

général (SG), dans un groupe stratégique. Cela devrait améliorer le relais politique des recommandations des rapports. Un modèle analogue est sur le point d'être inauguré à l'OFSP, où le Service d'évaluation va quitter le niveau de service quasi technique d'une ou deux sections et dépendre directement de la Direction de l'Office. A la DDC également, des réformes sont en cours, suite à l'évaluation de l'OCDE (un bon exemple d'utilisation d'évaluation !). Les évaluations seront séparées en 3 catégories d'importance diverse. La plus importante d'entre elles, seule appelée «évaluation», développera, surtout avec des consultants externes, des évaluations larges, sectorielles, régionales ou thématiques, et servira de base à l'orientation stratégique et politique des options de la DDC. Dans les trois cas, nous voyons clairement cette tendance à rattacher l'évaluation plus haut dans l'organigramme, si possible au niveau des décisions du Département. Cela permettrait également de mieux réserver le mot évaluation à des questions d'importance et contribuerait à la clarification sémantique du concept.

### **7.3. CONCLUSION**

Construite sur une hypothèse relativement « quantitativiste », notre recherche nous a amenés à comprendre que celle-ci n'était pas entièrement vérifiable avec les moyens et les ressources dont nous disposions. Les premiers résultats, même s'ils étaient correctement obtenus, ont donné naissance à la question de savoir si une telle approche avait des chances de « toucher juste », et si ses prémisses n'étaient peut-être pas assez nuancées. L'ajout d'un complément de type qualitatif a permis d'une part de vérifier cela, en apportant des compléments importants à la compréhension du processus d'utilisation des évaluations dans plusieurs Offices. D'autre part cependant, il a permis indirectement aussi de vérifier et valider aussi certaines des hypothèses émises au départ, mais en restreignant leur champ de validité pour l'essentiel au domaine interne de l'administration. Le principal acquis et apprentissage de cette étude aura été de considérer que l'évaluation est un, et un seul, des éléments qui entrent en ligne de compte dans le processus de décision qui aboutit à une modification d'une politique publique (Weiss 1986). Les trois autres domaines doivent trouver un équilibre particulier avec celui de l'information pour que l'évaluation puisse jouer pleinement son rôle de pilotage.

Néanmoins, notre recherche complémentaire a montré que, de la même manière que l'administration est à l'origine de la quasi totalité des évaluations commandées, elle est aussi la première intéressée et concernée par le développement ou non de cette science dans le travail pour assurer l'amélioration de l'efficacité de l'action étatique.

La pertinence des remarques ci-dessus sur l'avenir et le statut de l'évaluation dans l'administration suisse est étroitement liée à celles sur les conditions qui facilitent les processus d'apprentissage dans une organisation. Une évaluation se situe en effet toujours dans un processus qui, même s'il n'est pas conçu comme tel au départ, génère du savoir supplémentaire. Par conséquent, même si l'effet immédiat de l'évaluation X ou Y n'est pas objectivement visible dans telle ou telle décision, elle participe d'une augmentation de connaissance chez les acteurs concernés, et en tant que telle, est un actrice de changement. A moins qu'il ne lui arrive un sort plus sombre, comme dans cette histoire racontée par M.Q. Patton<sup>37</sup>.

*« Au commencement, Dieu créa les cieux et la terre.... Et il vit que cela était très bon. C'était le sixième jour » (Dieu 0000, in : Bible, Genèse 1).*

*Peu après, ce soir-là, réuni avec ses archanges, Dieu se vit cependant fermement interpellé par Lucifer, qui lui dit : « Comment pouvez-vous dire une chose pareille ? Sur quels résultats vous basez-vous ? Quels étaient vos objectifs au départ ? Et selon quels critères portez-vous un tel jugement ? Par rapport à quoi le comparez-vous ? Quels sont vos indicateurs de succès ? Quel est votre benchmark et qui l'a fixé ? Avez-vous interrogé les utilisateurs ?... » Lassé, Dieu l'interrompit et lui promit de réfléchir à ses questions lors de sa journée annoncée de repos. Il lui donna rendez-vous le dimanche soir.*

*Le lendemain soir venu, tous les archanges étaient à nouveau rassemblés pour entendre la suite de la dispute avec Lucifer et la réponse de Dieu. Celui-ci dit alors :*

*« Lucifer, ... go to hell !! »*

Serge Fornerod

Lausanne, le 31 juillet 2001

---

<sup>37</sup> Transcription libre d'une intervention lors de la Conférence de la European Evaluation Society à Lausanne, en octobre 2000.

#### **7.4. POST SCRIPTUM POUR SOURIRE**

##### *Quelques recommandations pour apprenti évaluateur*

Si vous voulez que vos études soient lues et participent directement à l'amélioration de l'état du monde, y compris de la situation sociale en Suisse, prenez bien note des conseils gratuits suivants:

1. Faites de la politique et participez comme citoyen à la définition des politiques publiques.
2. Assurez-vous de bons contacts dans les parlements et appréciez l'importance de leurs membres à leur juste valeur.
3. Faites des contacts tout aussi bons avec l'Administration, à tous les échelons, et valorisez son travail.
4. Ne communiquez jamais à un de ces deux groupes que vous fréquentez aussi volontiers l'autre.
5. Choisissez d'évaluer plutôt quelque chose de nouveau, si possible avant ou alors tout au début de sa réalisation.
6. Intéressez-vous à l'évaluation d'objectifs ambitieux et de thèmes importants plutôt que des aspects secondaires, mais ne refusez pas si on vous propose le contraire.
7. Il va de soi que vous avez toujours raison dans vos analyses, mais évitez quand même de montrer que les autres ont jusqu'ici toujours fait tout faux.
8. Obtenez que le plus haut chef possible soit dans votre groupe d'accompagnement.
9. Constituez aussi un groupe de suivi des conclusions de votre rapport dans l'Administration, afin de vous assurer qu'on exécute ce que vous avez si brillamment démontré.
10. Parlez de votre travail partout, écrivez des lettres de lecteurs anonymes sur vos découvertes.
11. Organisez bien les fuites et les relais sur les conclusions de votre rapport grâce aux conseils 1 à 3 ci-dessus.
12. Faites un effort pour que votre rapport intervienne juste avant le bon moment et que ses recommandations soient applicables facilement.
13. Si votre rapport n'est pas utilisé directement, consolez-vous en vous disant que vous avez tout de même considérablement augmenté la connaissance universelle sur le sujet, que votre carnet d'adresses s'est enrichi et que vous connaissez maintenant grâce à cette étude des nouveaux coins sympas pour des bons business lunches.

## 8. BIBLIOGRAPHIE

AGEVAL (Arbeitsgruppe Gesetzevaluation/groupe de travail « Evaluation législative »), *Mieux connaître les effets de l'action étatique, problèmes, possibilités, propositions*. Rapport final au DFJP, Berne, OFCIM, 1991.

BALTHASAR ANDREAS Dr., *Zusammenfassung Energie 2000. Programmwirkungen und Folgerungen aus der Evaluation*, INTERFACE, Institut für Politikstudien, Luzern, 1999

BALTHASAR ANDREAS, BÄCHTIGER CHRISTINE, *Evaluationskonzept für die schweizerische Verkehrspolitik, Bestandesaufnahme bisheriger Ex-Post-Evaluationen und Eckpfeiler für die Weiterentwicklung*, Bericht D14 des PNR 41 Verkehr und Umwelt, Bern, EDMZ 2000

BURRY JAMES, *Synthesis of evaluation Use Literature*. Evaluation Productivity Project, Center for the Study of Evaluation, UCLA, 1984

BUSSMANN WERNER, KLÖTI ULRICH, KNOEPFEL PETER (éd.), *Politiques publiques, Evaluation*. Economica, Paris, 1998

BUSSMANN WERNER, *Accompagner et mettre à profit avec succès les évaluations des mesures étatiques*, Genève, 1995

BUSSMANN WERNER, *Résumé des études exécutées, mandatées ou subventionnées par les services de la Confédération pendant la période législative 1991-1995 sur l'efficience et l'effectivité de l'action étatique*, DFJP, 1996.

BUSSMANN WERNER, *Evaluationsneugigkeiten aus der Schweiz*, in Bulletin SEVAL No2, novembre 1996

DIRECTION DU DEVELOPPEMENT ET DE LA COOPERATION (DDC), *Controlling, Evaluation externe. Faisons-nous ce qu'il faut ? Comme il le faut ?*, juin 2000, Berne

CHELIMSKY ELEANOR, *Linking programm evaluation to user needs*, in Palumbo Dennis (ed), *The politics of programm evaluation*, Sage, London, 1986, op. cit. p. 72-99

CROZIER MICHEL, FRIEDBERG EMMANUEL, *L'acteur et le système*, Paris, Seuil, 1977

DE COSTER MICHEL, *La logique du pouvoir dans les organisations*, in : De Coster Michel, *Sociologie du travail et gestion du personnel*, Labor, Bruxelles, 1987, p.125-150

FREIBURGHAUS DIETER, ZIMMERMANN WILLI, *Wie wird Forschung politisch relevant ?*, *Erfahrungen in und mit den Schweizerischen Nationalen Forschungsprogrammen*, Haupt, Bern, 1985

FREIBURGHAUS DIETER, *De l'efficacité des évaluations. Un discours en trois étapes*, Cahiers de l'IDHEAP 98, Chavannes, 1992

HENRY GARY T., *Using Evaluation findings for Policy : a realist perspective*. Workshop paper presented at EES Conference in Lausanne, October 2000.

HORBER PAPA ZIAN KATIA (éd.), *Evaluation des politiques publiques en Suisse*, 1990

HORBER PAPA ZIAN KATIA, *L'évaluation des politiques publiques*, Matériel du cours donné à l'IDHEAP, semestre hiver 1999-2000, Chavannes 1999

JANETT DANIEL, *Parliamentary Evaluation in Switzerland and its effects on legislative and administrative Action*, Workshop paper presented at EES Conference in Lausanne, October 2000

KLÖTI ULRICH, SCHNEIDER GERALD, *Die Informationsbeschaffung des Gesetzgebers*, Rüegger Verlag, 1989

LEE UW FRANS L., RIST RAY C., SONNICHSEN RICHARD C. (ed.), *Can governments learn ?*, *Comparative perspectives on Evaluation&Organizational learning*, Transaction Publishers, New Brunswick and London, 1993

MONNIER ERIC, *Evaluation de l'action des pouvoirs publiques*, Economica, Paris, 1992

OFFICE FEDERAL DE LA SANTE PUBLIQUE (OFSP), Unité d'évaluation, *Guide pour la planification de l'évaluation de projets ou programmes de santé*, Berne, 1997

PATTON MICHAEL QUINN, *Utilization-focused evaluation*, Thousand Oaks, Sage, London, 1997<sup>3</sup>

PATTON MICHAEL QUINN, *Evaluation's political inherency : practical implications for design and use*, in : Dennis J. Palumbo (Ed.), *The politics of programm evaluation*, Sage, London, 1986

PALUMBO DENNIS J. (Ed.), *The politics of programm evaluation*, Sage, London, 1986,

PALUMBO DENNIS, *Politics and evaluation*, in Palumbo Dennis (ed.), *The politics of programm evaluation*, Sage, London, 1986, p. 12-46

RANGEON FRANÇOIS, *La notion d'évaluation*, in : *L'évaluation dans l'administration*, PUF, Paris, 1993

ROSSI PETER, *Practice, Method and Theory in evaluating Social-Action Programms*, , in : J.L. Sundquist (ed.) *On Fighting poverty : perspectives from experience*, New York, Basic Books, 1969, pp. 217-234

ROSSI PETER, FREEMANN, *Evaluation, a systematic approach*, Sage, London, 1983

STOCKBAUER UTA, *Was macht Evaluationen nützlich ? Grundlagen und empirische Untersuchungen zum Thema Verwertung und Verwertbarkeit von Evaluationen*, Dissertation, Universität Salzburg, 2000

URIO PAOLO, BARI JEAN-PAUL, GLAUSER CHRISTOPHE, *L'évaluation comme instrument d'apprentissage et de changements organisationnels, l'exemple des programmes cantonaux de promotion de la santé*, PNR 27 / FNRS, Rapports finals, Dépt. Science politique, Université de Genève, Berne, 1996

WEISS CAROL H, *Where Politics and Evaluation research meet*, in Palumbo Dennis (ed.), *The politics of programm evaluation*, Sage, London 1986, p.47-69

WEISS CAROL H., *Ideology, Interests and Information, The basis of policy positions*, in : Callahan D., Jennings B. (ed.), *Ethics, the social sciences and policy analysis*, Plenum Press, New York and London 1983, p. 213-245.

WEISS CAROL H., *Evaluation for decision. Is anybody there ? Does anybody care ?* in : *Evaluation Practice* 9 (1), 1988

WEISS CAROL H., *The Interface between Evaluation and Public Policy*, in : Evaluation, Volume 5, Nr. 4, pp. 468-485, Sage Publications, 1999

WIDMER, THOMAS, LANDERT CHARLES, BACHMANN NICOLE, *Standards d'évaluation, recommandés par la SEVAL, 1999* ([www.ofj.admin.ch/f/index/html](http://www.ofj.admin.ch/f/index/html))

WIDMER THOMAS, RÜEGG ERWIN, NEUENSCHWANDER PETER, *Stand und Aussichten der Evaluation beim Bund*, Schlussbericht zuhanden der Schweizerischen Bundeskanzlei vom 9. März 2001, Universität Zürich, Institut für Politikwissenschaft, Zürich 2001 (y compris les procès-verbaux d'entretien avec les responsables de l'OFE, OFAG, OPCA, DDC, OFSP)

ZIMMERMANN WILLY, KOLLY DORIS, MAHON PASCAL, *Die wissenschaftliche Evaluation von Bundespolitik*, DFJP, Groupe de travail Evaluation législative, 1989

## **A QUOI ET A QUI SERVENT LES EVALUATIONS ?**

# **ANNEXES**

## LISTE DES ANNEXES

1. Abréviations utilisées	1
2. Listes de base	2
3. Schéma du cycle d'une politique publique	3
4. Résultats bruts du questionnaire sur la définition de l'utilisation, résultats nets (sans DDC ni OFSP), résultats de la DDC seulement.	4
5. Liste des clauses évaluatives	5
6. Liste des études ayant répondu au questionnaire sur les causes explicatives de leur utilisation	6
7. Résultats du dépouillement du questionnaire en %	7
8. Liste des parlementaires rencontrés	8
9. Listes des responsables d'Offices rencontrés	9
10. Questionnaire du logiciel SPHINX	10
11. Questionnaires d'interview	11