

Assessing Deviance, Crime and  
Prevention in Europe

L'évaluation des politiques de  
sécurité et de prévention en Europe

Philippe ROBERT

*Assessing Deviance, Crime and Prevention in Europe (CRIMPREV). Contrat n° 028300. Action de Coordination du 6e PCRD, financée par la Commission Européenne. Date de début : 1er juillet 2006. Durée : 36 mois. Projet coordonné par le Centre National de la Recherche Scientifique (CNRS).*

*www.crimprev.eu  
contact@crimprev.eu*

Imprimé dans la communauté Européenne  
sur les presses de l'imprimerie Flémal en avril 2009  
Mise en page JH Graphic.  
N° ISBN 978 2 917565 30 8  
Dépôt légal en avril 2009

**GROUPE EUROPÉEN DE RECHERCHE  
SUR LES NORMATIVITÉS**  
Immeuble Edison - 43, boulevard Vauban  
78280 - Guyancourt - France  
Tel : +33 (0)1 34 52 17 30 - Fax : +33 (0)1 34 52 17 32  
[www.gern-cnrs.com](http://www.gern-cnrs.com)



*Cette brochure a été réalisée avec le soutien du  
Forum Européen pour la Sécurité Urbaine  
[www.fesu.org](http://www.fesu.org) - [fesu@urbansecurity.org](mailto:fesu@urbansecurity.org)*

# SOMMAIRE

I - Introduction : le contexte	5
1 - L'axe Méthodologie et bonnes pratiques	5
2 - L'atelier consacré à l'évaluation des politiques de prévention et de sécurité	8
II - Des avantages et des inconvénients de l'évaluation	10
1 - Des raisons qui militent en faveur de l'évaluation	10
2 - Des résistances à l'évaluation	14
III - Un échantillon de conceptions de l'évaluation	17
1 - Une évaluation en peau de chagrin	17
2 - Une évaluation administrative	19
3 - Un exemple de méta-évaluation d'une politique de prévention et de sécurité	22
4 - Abondance d'évaluations et discussions de méthode	23
IV - Les apparences de l'évaluation sans ses risques	25
1 - Les charmes de l'auto-évaluation	25
2 - Neutraliser l'évaluateur externe	28
V - Les critères de l'évaluation	29
1 - Audit interne et évaluation	29
2 - L'évaluation et ses données	30
VI - Mesurer préalablement la consistance de l'action à évaluer	34
VII - Les savoir-faire évaluatifs	37
VIII - Conclusion	43
Références	46



# L'ÉVALUATION DES POLITIQUES DE SÉCURITÉ ET DE PRÉVENTION EN EUROPE

---

Philippe ROBERT<sup>1</sup>

## I - Introduction : le contexte

Dans le cadre du 6<sup>e</sup> PCRDT, la Commission européenne a financé une action coordonnée *Assessing Deviance, Crime and Prevention in Europe* (CRIMPREV).

Cette action coordonnée comprend six axes (*work-packages*) thématiques dont l'un est consacré à des problèmes de *Méthodologie et bonnes pratiques*.

### 1 - L'axe Méthodologie et bonnes pratiques

Il s'agit de recenser les mises en œuvre les plus significatives d'instruments de connaissance de la délinquance et les usages qu'on en fait.

On doit donc :

- dresser une cartographie de la situation en Europe,
- identifier les bonnes – mais aussi les mauvaises – pratiques,
- dégager les éléments de comparaison à l'intérieur de la zone européenne.

---

<sup>1</sup> Directeur de recherches émérite au Centre national de la recherche scientifique (CNRS), Centre de recherches sociologiques sur le droit et les institutions pénales (CESDIP, CNRS, ministère de la Justice, Université de Versailles Saint-Quentin).

Le dernier demi-siècle a vu apparaître de nouveaux et puissants instruments de connaissance de la délinquance. Leur particularité consiste à s'affranchir des données institutionnelles dans lesquelles l'étude du crime s'était traditionnellement enfermée.

Non seulement ces instruments ont renouvelé – au moins partiellement – la connaissance scientifique que l'on avait de la délinquance, mais encore ils peuvent constituer des outils importants d'aide à la décision.

Pour autant, leur introduction s'est faite de manière très variable dans les différents pays d'Europe. En outre, la maîtrise de ces instruments est assez inégalement répartie dans la mesure où le nombre de spécialistes confirmés est restreint ; en conséquence, les usages qu'on en fait sont plus ou moins pertinents. Enfin, les utilisateurs non scientifiques n'ont souvent qu'une connaissance limitée des potentialités de ces outils, mais aussi de leurs limites et des conditions de leur usage.

De la sorte, il y a place dans CRIMPREV pour un axe (*workpackage*) dédié à recenser les mises en œuvre les plus significatives de ces méthodes dans les principaux pays européens et l'usage qui en a été fait... dans l'espoir de mettre au point et de diffuser à la fois des états du savoir et des catalogues de bons usages.

La responsabilité de l'axe est partagée entre un pôle scientifique, le Centre de recherches sociologiques sur le droit et les institutions pénales (CESDIP), un enseignant de l'*Universidad central de Barcelona*, un réseau européen de collectivités locales en matière de sécurité, le Forum européen pour la sécurité urbaine (FESU), et un organisme régional de coordination des programmes locaux de sécurité, *Città sicure*, afin de bien le situer à l'interface du monde scientifique et de celui des utilisateurs.

Quatre méthodes ont été sélectionnées:

- les enquêtes en population générale sur la victimation et l'insécurité,
- les enquêtes en population générale sur la délinquance auto-reportée,
- la confrontation des données d'enquêtes et de celles provenant de sources institutionnelles comme les statistiques de police,
- l'évaluation des politiques de prévention et de sécurité.

Pour chacune de ces méthodes, on procède selon le même protocole en six phases :

- Phase 1 : élaboration d'une grille de rapport, choix d'un rapporteur général chargé de synthétiser les informations recueillies, et de rapporteurs chargés de dresser l'état des savoirs et des usages dans différents pays où la méthode est suffisamment développée. Bien entendu, on ne parvient pas à les couvrir tous compte tenu de la disponibilité des experts pressentis. On ne vise pas l'exhaustivité, mais plutôt à présenter une sélection raisonnablement représentative de ce qui se fait dans la zone européenne, notamment dans les principaux pays.
- Phase 2 : rédaction par chaque rapporteur d'un bilan concernant le pays ou la zone dont il est chargé, et circulation de ces documents.
- Phase 3 : présentation des rapports et discussion au cours d'un séminaire réunissant les promoteurs de l'axe, le rapporteur général et les rapporteurs.
- Phase 4 : rédaction par le rapporteur général d'une synthèse des rapports et des débats.
- Phase 5 : validation de ce document par les promoteurs de l'axe et diffusion sous forme d'un fascicule en anglais et en français.
- Phase 6 : publication de l'ensemble constitué par le rapport général et les apports nationaux en deux ouvrages, l'un en anglais aux *VUBPress*, l'autre en français chez L'Harmattan, dans la série *Déviante & Société*, sous la responsabilité de l'un des promoteurs de l'axe Méthodologie et bonnes pratiques.

Le présent rapport est consacré à l'évaluation des politiques de prévention et de sécurité.

## **2 - L'atelier consacré à l'évaluation des politiques de prévention et de sécurité**

Si, pour les autres ateliers – enquêtes sur la victimation et l'insécurité, enquête de délinquance autoreportée, comparaison entre enquêtes et statistiques officielles – il a été assez facile de trouver des rapporteurs du moins pour des pays européens où ces démarches étaient suffisamment pratiquées, nous nous sommes heurtés à beaucoup de difficultés pour constituer l'équipe chargée de travailler sur l'évaluation<sup>2</sup>. Il a donc été décidé de ne pas chercher cette fois à couvrir une partie significative des pays européens, mais de constituer une gamme de pays où l'emploi de l'expression 'évaluation des politiques de prévention et de sécurité' se réfère à des pratiques très contrastées. Encore n'étions-nous pas au bout de nos peines puisque les rapporteurs allemands qui avaient accepté de participer, se sont défaits à peine quelques jours avant le séminaire et n'ont plus donné ensuite signe de vie.

Finalement, l'équipe constituée pour cet atelier a compris:

- Sybille Smeets et Carrol Tange (U.L.Bruxelles) pour la Belgique,
- Tim Hope (Keele U.) pour l'Angleterre et Galles,
- Anne Wyvekens (Centre d'études et de recherches de sciences administratives et politiques – CERSA – U. Panthéon-Assas et CNRS) pour la France,
- Karin Wittebrood (Sociaal en Cultureel Planbureau, SCP) pour les Pays-Bas,
- pour le rapport général, Philippe Robert (Centre

---

<sup>2</sup> Ainsi les différents rapporteurs scandinaves qui ont été approchés n'ont pu finalement donner suite à notre offre de collaboration.



de recherches sociologiques sur le droit et les institutions pénales – CESDIP) en remplacement de Wolfgang Heinz (U. Konstanz) qui devait initialement assumer cette tâche,

- Renée Zauberman (Centre de recherches sociologiques sur le droit et les institutions pénales – CESDIP), Amadeu Recasens i Brunet (Universidad central de Barcelona), Rossella Selmini (Città sicure, Università degli studi di Macerata), Michel Marcus (Forum européen pour la sécurité urbaine – FESU) pour le groupe de pilotage.

Les rapporteurs nationaux ont reçu du groupe de pilotage de l'axe une lettre de commande en octobre 2007 et remis leur rapport en juin 2008. Puis un séminaire a réuni, à l'Université de Bologne, promoteurs, rapporteur général et rapporteurs nationaux pendant trois jours en juillet 2008 afin de présenter les rapports, de les discuter et d'en tirer des enseignements transversaux.

Le rapporteur général a ensuite préparé le présent rapport de synthèse.

\*\*\*\*

L'évaluation constitue une matière paradoxale : chacun chante ses louanges, mais, en réalité, tout le monde s'en méfie.

Du côté des décideurs – politiques et administratifs – on serait charmé d'entendre démontrer (avec l'aura de la science) que *ça marche*, mais on craint toujours, comme le dit Tim Hope (2009a), que la politique sur laquelle on a fondé son succès et sa réputation ne soit pas jugée aussi bonne qu'on l'affirme.

Du côté des scientifiques, on redoute la difficulté de la démarche, on craint toujours le ridicule d'avoir dé-

claré efficace un programme qui s'avèrera finalement contre-productif. Surtout, on appréhende de se faire sans cesse *tordre le bras* par des commanditaires qui n'admettent qu'une évaluation louangeuse. Ainsi Sybille Smeets et Carrol Tange (2009) montrent que les universitaires belges ont accepté de participer à des appels d'offres d'évaluation parce qu'ils y trouvaient le moyen de pénétrer des domaines jusque-là fermés à leur investigation plutôt que par enthousiasme pour la recherche évaluative.

En somme, chacun des protagonistes, décideur ou chercheur, a peut-être beaucoup à gagner potentiellement dans le recours à l'évaluation, mais il risque aussi d'y perdre énormément. Et c'est peut-être ce qui explique les réticences concrètes à travailler sur un thème dont pourtant tout le monde parle.

Dans une première partie, nous allons d'abord détailler les avantages et les inconvénients de l'évaluation.

Nous montrerons ensuite les différents usages qui en sont faits dans les pays retenus comme exemples typiques.

Ensuite, nous montrerons comment on peut chercher à bénéficier des apparences de l'évaluation sans en assumer les risques.

C'est seulement après que nous passerons en revue les critères de l'évaluation, ses préalables et ses savoir-faire.

## **II - Des avantages et des inconvénients de l'évaluation<sup>3</sup>**

### **1 - Des raisons qui militent en faveur de l'évaluation**

Barder les politiques que l'on lance de dispositifs

---

<sup>3</sup> Voir par exemple Robert, 2003, 2009.

d'évaluation peut être un moyen efficace de prévenir les critiques ou d'y répondre. Qui ne se préoccupe pas de faire évaluer ce qu'il entreprend s'expose à rester sans voix et sans argument le jour où se lèveront les critiques. L'action dont l'impact n'est pas mesuré finit toujours par attirer son Martinson (1974) qui vient déclarer *Nothing works*. Et quand bien même la critique serait faiblement fondée, médiocrement menée, voire même de mauvaise foi, ses résultats peuvent être dévastateurs si l'on n'a pas pris les devants en commanditant régulièrement des études d'impact. Des politiques qui pouvaient cependant mobiliser de bons arguments en leur faveur ont ainsi été balayées tout simplement parce que leurs responsables n'avaient pas prévu à temps les moyens de prouver leur utilité.

Évaluer peut donc permettre d'abord de résister aux critiques mal fondées ; mais toute contestation n'est pas dénuée de raison : quand une politique ou un programme s'est longtemps dispensé de toute évaluation, il peut avoir dérivé à bas bruit si massivement qu'on finira par lui découvrir des nocivités ou des inutilités si fortes et si anciennes qu'il paraîtra alors trop tard pour redresser la barre : tout abandonner semblera préférable ou changer radicalement de cap<sup>4</sup>. Sherman *et al.* (1997) en donnent un bel exemple avec l'histoire de la *Law Enforcement Administration Agency*. À ce moment, on aura l'impression d'avoir dépensé pour rien beaucoup d'énergie et beaucoup de moyens. Disposer d'évaluations peut aider à prévenir une issue aussi fatale.

Au delà de ces fonctions défensives, mesurer l'impact de ce que l'on entreprend peut permettre d'évoluer, d'ajuster son action. Les politiques de sécurité et de prévention recouvrent un domaine tellement vaste que leur mise en œuvre est fatalement instable et qu'il faut constamment redresser la barre. Elles recouvrent, en effet, tous les moyens possibles pour rendre moins

---

<sup>4</sup> Ex. *in* Robert, 1994.

fréquents des comportements incriminés. Les vertus dissuasives prêtées à la répression pénale sont loin de les épuiser, loin de là : elles n'occupent qu'un chapitre dans le rapport de Lawrence Sherman (1997<sup>5</sup>) au Congrès américain sur la prévention de la délinquance<sup>6</sup> ; les autres sont consacrés au voisinage, à la famille à l'école, au travail, aux lieux criminogènes, à la police enfin dont les actions ne sont pas uniquement répressives. On peut en imaginer une très longue liste : tout le jeu des relations sociales peut être mis à contribution. Chacune de ces formules a ses faiblesses. Celles qui sont très générales risquent toujours de manquer la cible, de se disperser sans aucun effet préventif notable : c'est le reproche fait dans les années 1960 aux grands programmes urbains sursaturants<sup>7</sup>. Elles risquent aussi d'entrer en contradiction avec des politiques répressives décidées par ailleurs sans aucun souci de synergie tant il est peu aisé de mettre en phase le social et le pénal : ainsi reproche-t-on souvent aux activités policières de répression de l'immigration irrégulière et du commerce de produits prohibés de ruiner tous les efforts pour éviter la ghettoïsation des zones urbaines en difficulté en créant dans ces quartiers un climat d'affrontement permanent entre habitants et institutions. Quant aux préventions spéciales, elles peuvent manquer d'adossement dans un environnement trop défavorable : ainsi Sherman (1997) remarque-t-il que les programmes préventifs n'arrivent pas à lutter à eux seuls contre les effets ravageurs de la ségrégation urbaine et sociale dans les hyperghettos urbains. En ce cas, les actions spécialisées peuvent céder à une exaspération sécuritaire qui les rend surtout productrices d'effets pervers. Devant ces périls, les politiques de sécurité restent rarement – ou du moins rarement longtemps – abonnées à un seul modèle. Plus généralement, en tout cas tôt ou tard,

---

<sup>5</sup> Rééditions élargies : Sherman *et al.*, 2002, 2006.

<sup>6</sup> Un bilan qui revisite quelques 500 évaluations étalées sur quarante ans.

<sup>7</sup> Voy. par ex. bilan *in* Robert, Lascoumes, 1974.

elles les mélangent dans des proportions variables. La prévention oscille sans cesse entre le trop général et le trop spécifique. C'est une politique publique instable par nature dont il faut sans cesse corriger la trajectoire. Un accompagnement évaluatif constitue alors un équipement utile pour le décideur qui veut échapper à la tyrannie des modes, des effets d'opinion ou des pressions intéressées des praticiens les plus influents.

Le discours évaluateur peut surtout aider à rendre compte de l'action (*accountability*) non seulement aux instances politiques qui décident et financent, mais aussi aux populations concernées ou à celles qui vivent à proximité. On néglige souvent cet usage, peut-être parce qu'il demande quelque effort : il y faut évidemment un langage adapté pour être compréhensible. Il peut être pourtant fort utile pour prévenir les hostilités et susciter les participations. Souvent, les responsables locaux se plaignent de la difficulté à impliquer les populations concernées ; il faut bien dire qu'on n'en prend pas souvent les moyens : il y faut au moins un débat sur la base d'un compte rendu sérieux des impacts (positifs et négatifs) des actions ou de ceux qui sont attendus des projets au vu d'évaluations antérieures. Cet usage de l'évaluation est également susceptible de mettre en lumière les attentes de sécurité : les responsables ont naturellement tendance à penser que leur production s'adapte naturellement aux attentes sociales. C'est l'occasion pour eux de vérifier si les impacts de leurs actions répondent à des attentes. À vrai dire, le problème se pose en termes plus compliqués. Dans le domaine de la sécurité publique comme dans d'autres matières, les attentes sont généralement diverses et souvent contradictoires : les groupes de jeunes des quartiers à l'abandon n'ont pas souvent les mêmes que les commerçants du centre-ville ou les retraités des banlieues résidentielles. L'art du politique consiste justement à combiner des attentes divergentes dans toute la mesure

du possible<sup>8</sup> et à opérer des arbitrages entre ce qui est incompatible en se référant au compromis social qui fonde dans une société donnée la légitimité de l'action publique. Rendre compte de l'impact des programmes en termes compréhensibles à des publics divers peut rendre plus aisée la clarification des choix qui sont à la base des politiques et faire apparaître d'éventuels effets pervers qui resteraient autrement inaperçus.

## 2 - Des résistances à l'évaluation

Pourquoi donc, avec tous ces mérites, l'évaluation se heurte-t-elle à d'aussi fortes résistances ?

Il faut prendre en compte les réticences, peu avouées, de responsables, nationaux ou locaux, qui répugnent à ce qui leur apparaît comme un contrôle. Ce n'est pas que tous rechignent devant l'obligation de rendre compte, mais s'ils en admettent la nécessité, c'est souvent au seul profit du débat politique, celui qui leur est familier. Moins volontiers, admettent-ils une démarche qui leur paraît technocratique. Dans certains pays, on soupçonne même l'évaluation de participer d'une conception gestionnaire<sup>9</sup> qui évacue la spécificité de l'action publique au profit d'un emprunt aux modes de raisonnement et d'action de l'entreprise. Il y a du vrai dans cette objection, toutefois l'emprise gestionnaire plaide plutôt en faveur d'un contrôle de programme (*monitoring*) ou d'un audit (*auditing*) qui se préoccupent surtout de vérifier l'adéquation des mises en œuvre aux objectifs

---

<sup>8</sup> Anne Wyvekens (2008) donne d'intéressants exemples d'opérations de prévention visant à faire apprendre à des publics aux attentes hétérogènes à coexister dans des lieux publics parisiens comme le parc de la Villette ou des grandes gares de chemin de fer.

<sup>9</sup> Managériale, en français.

proclamés ou leur rentabilité financière<sup>10</sup>, ce qui ne constitue qu'une caricature de recherche évaluative.

Quelquefois, les réticences des responsables viennent d'un souci moins noble que celui de préserver le champ du débat politique. Par exemple la volonté de conserver une marge de manœuvre : ainsi dans les formules de prévention générale, les responsables locaux peuvent trouver dans les programmes de financement des moyens complaisants pour abonder une action sociale sans aucun ciblage préventif. Ou, tout bonnement, la répugnance à rendre des comptes, la conviction que le responsable sait ce qui est efficace et bon pour la société, que toute vérification est superflue.

L'évaluation peut aussi manquer non pas parce que les responsables y répugnent pour une raison ou pour une autre, mais parce que leurs attentes sont trop divergentes pour laisser à l'évaluateur une marge suffisante : écartelé entre les forts présumés des forces politiques nationales, les exigences contradictoires des Finances, de la police, des travailleurs sociaux, des enseignants, des médecins, des procureurs, des juges, des élus locaux, des associations... il ne parvient pas à ménager l'autonomie nécessaire pour un travail sérieux. Finalement, on n'évalue rien et les arbitrages s'opèrent seulement au rapport de forces entre les différents protagonistes.

Certaines mesures résistent davantage que d'autres à l'évaluation. D'abord, on ne peut, à proprement parler, évaluer toute une politique que par une sorte de méta-évaluation qui fait la somme des évaluations des différents programmes qui la composent. Pour évaluer, après les critiques de la Cour d'Audit, la politique

---

<sup>10</sup> Exemples dans Leigh (2000) – à propos de l'obligation d'une amélioration de l'*efficiency* de 2% par an dans le *Crime and Disorder Act* - ou Braga *et al.* (1999) avec l'importance de l'estimation *Value for Money* du programme de réforme de la police de New Jersey City.

*Naar een veiliger samenleving* inaugurée en 2002 par le premier gouvernement Balkenende, L. Van Noije et Karin Wittebrood (2008) ont, ainsi, dû passer en revue 149 évaluations. Par ailleurs, on a souvent soutenu que les mesures de prévention sociale sont plus malaisées à évaluer que celles de prévention situationnelle. Ainsi Karin Wittebrood (2009) déclare-t-elle inévaluables les programmes qui prétendent s'attaquer à des causes supposées de la délinquance comme l'école buissonnière, la qualité éducative des parents ou l'alcoolisme dans la mesure où cette relation causale demeure hypothétique<sup>11</sup>. La difficulté vient probablement du fait que les politiques sociales ne s'attachent qu'indirectement à réduire la délinquance ou l'insécurité, plus exactement elles cherchent à contribuer à de tels effets par l'intermédiaire de leur impact direct sur les conditions de vie des populations-cibles. Enfin, il est surprenant de constater que le domaine répressif est beaucoup moins soumis à évaluation que le préventif<sup>12</sup> ; il en a pourtant encore plus besoin tant il repose massivement sur des assertions improuvées<sup>13</sup> et l'évaluation s'y heurte peut-être à de moindres difficultés car il implique une participation individuelle et obligatoire des populations-cibles tandis que la prévention cherche à mobiliser une participation qui reste collective et volontaire<sup>14</sup>.

Devant une situation aussi contradictoire, on a donc choisi de commencer par passer en revue les pratiques

---

<sup>11</sup> Pour autant, Hope (2009a) rappelle que le *New Deal for Communities Programme (NDC)* qui poursuit la réduction du chômage, celle de la délinquance, l'amélioration de la santé et des emplois, la rénovation de l'habitat et de son environnement sur 39 sites particulièrement défavorisés a fait l'objet d'un programme national d'évaluation.

<sup>12</sup> L'un des mérites de la méta-évaluation néerlandaise relatée par Wittebrood (2009) consiste justement à ne pas faire l'impasse sur le domaine pénal. Voy. Cpd Mackenzie, 2006.

<sup>13</sup> Comme Lawrence Sherman le rappelait en 1997 à propos du recours massif à l'emprisonnement.

<sup>14</sup> Comme le fait remarquer Tim Hope (2009a).



de l'évaluation des politiques de prévention et de sécurité dans quelques pays européens.

### **III - Un échantillon de conceptions de l'évaluation**

Plutôt que de couvrir un nombre important de pays européens, on a constitué une sorte d'échantillon de pratiques et même de conceptions très différentes de l'évaluation des politiques de prévention et de sécurité.

En Belgique l'effort évaluatif – qui est lié aux contrats de sécurité entre l'État fédéral et des communes – semble plutôt perdre de la vigueur dans les dernières années.

La France présente le cas d'une évaluation 'officielle' prise en charge par l'administration.

Les Pays-Bas offrent l'exemple d'une méta-évaluation qui toise l'ensemble des matériaux nationaux disponibles à l'aune d'un canon scientifique 'quasi-expérimental' d'origine anglo-saxonne.

Finalement, l'Angleterre et le Pays de Galles présentent le cas d'une pratique évaluative ancienne et systématique qui couvre tout le champ des politiques de prévention et permet d'aborder des problèmes méthodologiques complexes

#### **1 - Une évaluation en peau de chagrin<sup>15</sup>**

Les contrats de sécurité passés entre le ministère fédéral de l'Intérieur et des communes se sont accompagnés d'évaluations destinées à décider de la poursuite, de l'extension ou de la réduction des financements

---

<sup>15</sup> D'après le rapport présenté sur le cas belge par Sybille Smeets et Carrol Tange (2009 et réf. cit.) lors du séminaire de Bologne.

consentis par l'État central. L'ensemble de ces 'évaluations' sont dites 'sommatives' en ce sens qu'elles ont pour but de dresser un bilan de contrats entre l'État et une commune pour décider de sa poursuite.

Certaines évaluations – dites 'administratives' – se présentent sous forme de tableaux synthétisant le nombre de contrats et de localités concernées, les allocations budgétaires, les dispositifs retenus, avec quelquefois des statistiques policières ou des données tirées du Moniteur de sécurité<sup>16</sup>.

D'autres – dites internes – doivent être réalisées par des agents rémunérés par les communes bénéficiant d'un contrat de sécurité. Mais, depuis 2005, ces évaluations se sont transformées en diagnostics (préalables) de sécurité – obligatoires depuis 2007 – réalisés sur la base d'un guide méthodologique mis au point par le ministère de l'Intérieur.

À côté, un programme de recherche a été initié – toujours par le ministère de l'Intérieur – dès 1985 d'abord sur la police. À partir du début de la décennie suivante, il a été élargi aux politiques de prévention et de sécurité en vue de déterminer les faiblesses des programmes et de contribuer à leur amélioration. On parlait alors d'évaluations 'formatives'. Pendant toute la dernière décennie du XX<sup>e</sup> siècle, des équipes universitaires ont ainsi étudié les processus et les fonctionnements de programmes portant sur la police communale, la police de proximité, le rapprochement entre police et citoyens, l'accueil des victimes, la réinsertion professionnelle, le hooliganisme, le décrochage scolaire, la coordination des projets... Il s'est agi de recherches surtout qualitatives qui ne se proposaient pas expressé-

---

<sup>16</sup> Un questionnaire préparé par la police et administré par téléphone par une firme privée à un échantillon aléatoire de quelque 40 000 enquêtés. Les questions portent sur les problèmes dans le quartier, la victimation, l'insécurité, les contacts avec la police et le sentiment sur son fonctionnement. Il a été administré 6 fois depuis 1997.

ment de mesurer l'impact sur la délinquance et l'insécurité.

À partir de 1999, les appels d'offres ont répondu à un souci de devenir plus opérationnels. Ils ont fait appel à des organismes gouvernementaux ou des structures privées. Le montant des contrats de recherche et leur durée se sont sensiblement réduits. Loin de chercher à évaluer l'impact de programmes ou de politiques, ces études se présentent surtout comme des inventaires de moyens ou des propositions de méthodes de gestion, de planification des politiques ou encore d'analyse des données. Malgré tout, les résultats de ces travaux sont très peu pris en compte par leurs commanditaires sauf dans une perspective de 'mieux se connaître pour mieux se justifier'.

Jusque-là, les politiques publiques tentaient des compromis entre une tendance libérale plutôt favorable à un renforcement des moyens répressifs, et une autre social-démocrate faisant plutôt confiance aux contrats locaux de sécurité. À leur image, les évaluations réalisaient une sorte de compromis entre la mesure de l'impact des programmes et une étude qualitative des discours et du vécu des acteurs. Dans les années récentes, l'accent est mis sur la rationalisation des moyens dédiés à la sécurité et la concentration des financements d'évaluation sur des études directement gestionnaires où le monde scientifique tient de moins en moins de place.

## **2 - Une évaluation administrative<sup>17</sup>**

Depuis le gouvernement Rocard de la fin des années 1980, l'évaluation des politiques publiques constitue une activité administrative tout à fait officielle. Un texte réglementaire lui a même été consacré une dizaine d'années plus tard : le décret 98-1048 du 18 novem-

---

<sup>17</sup> D'après le rapport présenté au séminaire de Bologne sur l'exemple français par Anne Wyvekens (2009 et réf. cit.).

bre 1998 relatif à l'évaluation des politiques publiques. Cette activité semble généralement prise en charge par l'administration elle-même, le plus souvent par le truchement des corps d'inspection qui existent dans chaque branche de l'administration. Le plus prestigieux d'entre eux, l'inspection générale des Finances (IGF), paraît tenir une place prééminente, généralement en association avec l'inspection propre au ministère impliqué. La Cour des Comptes a également, parmi ses tâches permanentes, une fonction d'évaluation des politiques publiques. La place de ces deux organismes souligne la coloration très 'financière' de la conception officielle de l'évaluation à la française : il s'agit avant tout d'une vérification de la bonne utilisation des deniers publics.

En ce qui concerne les politiques de prévention et de sécurité, toutefois, on ne trouve qu'un nombre limité d'exemples à citer. En 1993, le conseiller d'État Jean-Michel Belorgey avait présidé une mission d'évaluation de la politique dite de la Ville<sup>18</sup> mais sans faire une place très considérable aux programmes de prévention qui ont été intégrés dans cet ensemble de politiques publiques à partir de la fin des années 1980. Ce rapport faisait suite à un bilan des contrats locaux de Développement social des quartiers – l'ancêtre de la politique de la Ville – mené à bien en 1988 sous la direction de François Lévy, ingénieur général des Ponts et Chaussées.

Ce quasi-monopole administratif de l'évaluation des politiques, notamment de celles de prévention et de sécurité, semble s'accompagner d'une solide méfiance des décideurs politiques et administratifs à l'égard du monde scientifique, soupçonné de nourrir trop facilement des pensées critiques envers l'action publique.

---

<sup>18</sup> Il faut préciser que cette mission n'était pas composée seulement d'administrateurs puisque l'essentiel du travail préalable a été réalisé par des scientifiques, notamment Jacques Donzelot qui s'est appuyé sur cette expérience pour publier ensuite un ouvrage avec Philippe Estèbe (1993).

De leur côté, les milieux universitaires paraissent peu mobilisés par cette activité évaluatrice. Seuls des politologues spécialisés dans l'analyse des politiques publiques semblent lui accorder un certain intérêt<sup>19</sup>, mais pas tous cependant : cette optique ne retient guère l'attention de ceux qui étudient les politiques publiques par le bas, sous l'angle des interactions entre l'agent administratif de base et le client, ou encore sous le point de vue de ce dernier. Il existe une sorte de consensus scientifique pour juger que ce qui est mis en œuvre par l'administration sous le terme 'évaluation' constitue seulement de l'audit interne et qu'il n'existe pratiquement pas en France de véritable évaluation des politiques publiques, notamment de celles de prévention et de sécurité.

Toutefois, l'Institut national de la statistique et des études économiques (INSEE) a procédé à une étude de l'évolution de la situation relative des quartiers faisant l'objet de la politique de la Ville entre les recensements de 1990 et 1999. Ce travail de statisticiens publics est ce qui semble se rapprocher le plus d'une étude évaluative.

Pour la période la plus récente, parmi les multiples observatoires nationaux opérant dans le champ des politiques de prévention et de sécurité, celui des zones urbaines sensibles et surtout celui des drogues et toxicomanies ont publié des rapports qui se rapprochent clairement d'une perspective évaluative.

On peut encore citer quelques travaux particuliers, par exemple sur le recours à la vidéosurveillance dans les lycées de la région d'Île-de-France, mais c'est pour relever que leur diffusion est sérieusement entravée par les consignes des commanditaires.

---

<sup>19</sup> Voy. par ex. Lascoumes, Setbon, 1995 ; Kessler, Lascoumes, Setbon, Thoenig, 1998.

Finalement, il faut mentionner un certain nombre de travaux à tournure évaluative concernant le devenir de condamnés à différentes sortes de peines.

### **3 - Un exemple de méta-évaluation d'une politique de prévention et de sécurité<sup>20</sup>**

Les interrogations de la Cour néerlandaise d'Audit sur les performances d'un programme de prévention et de sécurité lancé en 2002 ont conduit à la réalisation, par l'Institut néerlandais pour la recherche sociale (*Sociaal en Cultureel Planbureau, SCP*), d'une méta-évaluation en trois phases confiée à Van Noije et Wittebrood.

Ces auteurs ont d'abord réuni quelque cent cinquante évaluations disponibles sur tel ou tel aspect de cette politique.

Ils ont ensuite sélectionné celles d'entre elles qui semblaient conformes à un canon évaluatif défini par rapport à l'exercice classique réalisé sous la direction de Lawrence Sherman pour le Congrès des États-Unis, également aux critères du *Campbell Collaboration Crime and Justice Group* destiné à favoriser le développement d'une évaluation quasi-expérimentale dans le domaine de la délinquance et de la justice pénale.

Ils ont enfin utilisé les travaux retenus pour passer en revue les différents chapitres de la politique gouvernementale sous examen – l'application de la loi, la prévention 'développementale'<sup>21</sup>, la prévention situationnelle, enfin les mesures systémiques – tout en soulignant qu'on ne disposait pas toujours des travaux nécessaires pour porter un jugement sur plusieurs aspects du programme inauguré en 2002 et qu'il était nécessai-

---

<sup>20</sup> D'après le rapport présenté au séminaire de Bologne sur le cas néerlandais par Karin Wittebrood (2009 et réf. cit.).

<sup>21</sup> *Developmental prevention* regroupe l'ensemble des programmes destinés à prévenir la délinquance des jeunes.

re d'encourager un développement plus systématique encore de l'évaluation quasi-expérimentale.

Ce n'est pas la première fois que le gouvernement néerlandais fait ainsi procéder à une revue systématique des évaluations disponibles<sup>22</sup>. La matière première en est fournie grâce à l'habitude ancienne de réserver à l'évaluation 10% du montant des programmes de prévention et de sécurité. Pareille situation tend à montrer une certaine habitude de la méta-évaluation, même si, à chaque fois, les réviseurs déplorent la qualité trop inégale des évaluations qui ont été pratiquées.

#### **4 - Abondance d'évaluations et discussions de méthode<sup>23</sup>**

En Angleterre et au Pays de Galles, depuis les années 1980, l'évaluation accompagne systématiquement les programmes de prévention et de sécurité. Qui plus est, depuis son retour au pouvoir en 1997, le *Labour* a couplé sa politique de réduction de la délinquance et des désordres avec une *Evidence-based policy and practice (EBPP)* qui repose avant tout sur l'accumulation d'un savoir-expert construit à partir de l'examen systématique d'évaluations. Dans la réalisation de ces dernières, sont devenus de plus en plus influents les critères de standardisation développés dans la foulée des travaux de Sherman à l'Université de Maryland et de ceux du groupe qui poursuit la réflexion de Donald Campbell.

On peut ainsi dénombrer les évaluations de :

- \* *Home Office Crime Prevention Feasibility Study* basée sur une recherche action utilisant une méthode d'étude de cas ;

---

<sup>22</sup> Voy. Junger-Tas, 1989 ; Polder, Van Vlaardingen, 1992 ; Soetenhorst de Savornin Lohman, 1992 ; Van Dijk, 1993.

<sup>23</sup> D'après le rapport présenté au séminaire de Bologne sur l'exemple anglo-gallois par Tim Hope (2009a et réf. cit.).

- \* *Neighbourhood Watch Evaluation* qui met en œuvre une démarche quasi-expérimentale à partir de données d'enquêtes ;
- \* *Community Policing Fear Reduction* également de tournure quasi-expérimentale ;
- \* *Priority Estates Project* elle aussi quasi-expérimentale ;
- \* *Kirkholt Burglary Reduction Project* focalisées sur l'appréciation de l'efficacité de programmes axés sur le cambriolage à répétition ;
- \* *Safer Cities programme* parmi lesquelles une utilisation sophistiquée de méthodes micro-économétriques par Ekblom *et al.* (1996), mais aussi une autre recourant à des techniques qualitatives (Tilley, 1993) ;
- \* *Design Improvement Controlled Experiment (DICE)* au moyen d'une batterie de démarches incluant des données d'enquêtes ;
- \* *Communities that Care (CtC) Demonstrations Projects* pour la Joseph Rowntree Foundation par une équipe de l'Université de Sheffield ;
- \* *New Deal for Communities Programme (NDC)* recourant à une batterie d'indicateurs et de démarches analytiques ;
- \* *Crime Reduction Programme (CRP)* regroupant quelques 80 études évaluatives focalisées sur la réduction du cambriolage et la surveillance à distance ;
- \* *Community Policing Reassurance.*

Cette abondance de travaux utilisant des démarches et des données diverses permet une discussion approfondie de la méthodologie évaluative. C'est ainsi que le rapporteur britannique peut passer en revue les insuffisances d'une évaluation se contentant d'un modèle de 'boîte noire', discuter des mérites de l'évaluation 'd'expérimentation réaliste' de Pawson & Tilley (1997), montrer comment la démarche 'quasi-expérimentale' elle-même (Farrington, 2003) se heurte à des difficultés pour la solution desquelles il propose d'emprunter des enrichissements à la micro-économétrie.



L'implication systématique de l'évaluation dans le pilotage des politiques de prévention et de sécurité met, par ailleurs, en mesure de débattre des retombées d'une si intense mobilisation des ressources scientifiques et de leur arrimage à des finalités de politiques publiques. À ce propos, Hope (2009a, b) avance qu'une telle situation pourrait conduire paradoxalement, non à une 'scientification' de la politique, mais à une politisation de la science, si elle se combinait avec une irrésistible aversion pour les résultats critiques. Il ne peut y avoir d'apport de la science à la politique que si chacune conserve son autonomie par rapport à l'autre, et donc ses vertus propres.

Que l'évaluation des politiques de prévention et de sécurité renvoie à une gamme aussi contrastée de pratiques dans des pays européens par ailleurs assez proches montre la difficulté de la synthèse. Il faut pouvoir traiter des problèmes suscités par le déficit d'évaluation mais aussi de ceux suscités par son excès.

#### **IV - Les apparences de l'évaluation sans ses risques**

La tentation est grande de capturer les bénéfices de l'évaluation sans se soumettre à ses risques. Il y a deux manières d'y parvenir : la première consiste à s'évaluer soi-même, la seconde à contrôler tellement bien l'évaluateur externe qu'il est pratiquement contraint à des conclusions positives.

##### **1 - Les charmes de l'auto-évaluation**

La manière la plus simple consiste à s'évaluer soi-même.

Ainsi la réforme de l'organisation policière belge a permis, d'après Smeets et Tange (2009), de réserver les

volets policiers des contrats de sécurité<sup>24</sup> à une évaluation interne à cette administration. Wyvekens (2009) en donne l'exemple encore plus significatif de 'l'évaluation des politiques publiques' en France, une activité administrative régie par l'autorité réglementaire<sup>25</sup> et prise en charge par les corps d'inspection des différentes administrations. En réalité, le vérificateur est trop proche du dispositif d'action pour produire une véritable évaluation et ces prétendues 'évaluations' sont, au mieux, des audits internes. Parfois, elles n'atteignent même pas ce niveau : Roché (2005) cite ainsi 'l'évaluation' de la mise en place d'une police de proximité par des inspecteurs de l'administration et de la police : il n'y découvre qu'une *longue liste de jugements subjectifs*.

Sur la condamnation de l'auto-évaluation, il y a consensus des spécialistes, quelquefois même en termes assez rudes<sup>26</sup>.

De l'intérieur, le risque est trop grand de confondre le produit d'un programme avec ses effets sur une population ou un territoire, pour toute une série de mauvaises raisons. De cette position, on a si facilement accès aux données internes qu'on s'en satisfera volontiers<sup>27</sup> sans examiner avec beaucoup de rigueur si elles mesurent ce que l'on a fait ou l'influence qu'a pu avoir cette action<sup>28</sup>.

De l'intérieur, il est difficile d'échapper à la logique de l'action : si les objectifs n'ont pas été mis en œuvre

---

<sup>24</sup> Ensuite devenus plans stratégiques de sécurité et de prévention.

<sup>25</sup> Décret n° 98-1048 du 18 novembre 1998 relatif à l'évaluation des politiques publiques.

<sup>26</sup> Voy. par ex. Brodeur, 1998b.

<sup>27</sup> Ainsi, selon Smeets et Tange (2009), le ministère fédéral belge de l'Intérieur impose aux titulaires de contrats de sécurité d'utiliser les statistiques internes à l'administration 'pour des raisons d'objectivité' (*sic*).

<sup>28</sup> Éventuellement, on pourra même parfois les présenter de façon à donner l'impression d'un résultat positif.

ou s'ils l'ont été infidèlement, on le verra facilement ; mais s'ils l'ont été, il sera difficile d'imaginer que l'effet sur la cible n'est pas nécessairement atteint ou qu'il diffère de ce qu'on escomptait. En tout cas, s'il parvient à apercevoir les effets escomptés, l'auditeur interne est dans la plus mauvaise position possible pour voir ceux qui sont inattendus, latéraux ou de déplacement.

Concrètement, cette exigence produit deux conséquences. L'évaluation ne peut être demandée ni à un organe chargé de concevoir l'action ou de la mettre en œuvre, ni même à un organe chargé de contrôler l'action publique, par exemple à un corps d'inspection ou d'audit<sup>29</sup>.

S'évaluer soi-même ne procure que des résultats illusoire, des apparences d'évaluation. Dans sa révision des évaluations disponibles aux Pays-Bas, Karin Wittebrood (2009) conclut qu'aucune de celles réalisées en interne par les promoteurs de programmes de prévention et de sécurité n'atteint le niveau minimal de validité.

Pour tenter de mieux sauver les apparences, un substitut consiste à faire appel à des cabinets privés. Smeets et Tange (2009) tout comme Wyvekens (2009) donnent de multiples exemples de la faveur dans laquelle est tenue cette solution tant en Belgique qu'en France : le plus souvent affranchis des canons de la recherche évaluative, ils sont aussi rapides que leur commanditaire le souhaite et leurs résultats ne sont pas contrariants. Aux Pays-Bas, Wittebrood (2009) les juge d'un niveau technique parfois acceptable, parfois insuffisant, mais en tous cas jamais critiques.

---

<sup>29</sup> Ainsi le dispositif d'évaluation de la réforme policière à Chicago a été confié à un consortium de départements universitaires locaux commandité et financé par l'État de l'Illinois. Il se situait tout à fait en dehors de la municipalité et du corps policier qui ont la charge de concevoir et de mettre en œuvre le programme CAPS (Skogan, Hartnett, 1997a, b).

## 2 - Neutraliser l'évaluateur externe

On comprend aisément que l'auto-évaluation (ou le recours à des cabinets privés complaisants) permette de s'approprier l'apparence de l'évaluation sans en courir les risques. Discerner un risque finalement comparable dans un 'excès' d'évaluation paraît plus paradoxal.

Pourtant l'intégration systématique de l'évaluation dans tout programme de prévention et de sécurité peut produire une sorte de domestication de l'évaluateur externe qui devient abonné aux contrats de l'administration au risque d'y perdre une partie de son indépendance, un peu comme le journaliste *embedded* dans une unité militaire en opération.

Il n'est pas absolument certain que la standardisation même stricte des opérations de recherche évaluative suffise à garantir contre ce péril. La technique méthodologique peut constituer une défense illusoire que l'on brandit d'autant plus volontiers que l'on sacrifie les autres critères de la production scientifique, le jugement des pairs et la publication des démarches et des résultats dans un format qui permette une réplique.

Dans ce cas, loin de parvenir à cette 'scientification' de l'action politique qu'appelait de ses vœux l'initiateur de la démarche, Donald T. Campbell (1969) comme critère d'une bonne société 'ouverte', on parviendrait plutôt à une 'politisation' de l'activité scientifique.

C'est probablement ce risque, plus encore, que les difficultés de méthode, qui retiennent beaucoup de scientifiques de s'impliquer dans l'évaluation.

Mais, dans ce cas-là, le recours à l'évaluation n'apporterait plus réellement de valeur ajoutée qui suppose d'accepter le risque d'une conclusion négative de l'évaluateur et que ce dernier, lui-même, prenne toujours les moyens de pouvoir conclure négativement aussi bien que positivement.

C'est ainsi que des évaluateurs confirmés (ainsi, Hope, 2009b) insistent pour un maintien sourcilleux de l'autonomie et de la déontologie propres au monde scientifique dans une matière dont l'enjeu politique est si élevé parce que l'insécurité est devenue moins un problème à résoudre qu'un levier de gouvernement (*governing through crime*).

## V - Les critères de l'évaluation

La faiblesse de tous ces raccourcis destinés à obtenir les bénéfices d'une évaluation sans en courir les risques vient toujours d'une méprise sur ce qu'est l'évaluation.

### 1 - Audit interne et évaluation

Comparer les objectifs et leur mise en œuvre, chercher à déterminer des coûts et à les comparer à des prévisions initiales, décrire des produits, les rapporter aux objectifs, toutes ces démarches s'apparentent à l'audit (*auditing*), au calcul coût/efficacité (*value for money*) ou au contrôle de programme (*monitoring*), mais elles ne constituent pas une évaluation<sup>30</sup>.

Se limiter à mesurer ce qui a été fait n'est pas évaluer. L'évaluation ne commence qu'avec la détermination d'un impact. Un exemple<sup>31</sup> permettra de faire sentir la différence : l'installation de vingt caméras dans les rues d'une ville constitue un produit (*output*) pas un impact (*outcome*) : c'est ce que l'on a fait. En revanche, obtenir une baisse de 20% de la délinquance dans la rue ou une atténuation de la peur constituent bien des

---

<sup>30</sup> Voy. par ex. Sherman (*et al.*, 1997) : la description de la mise en œuvre, son contrôle (*monitoring*), son audit ne valent pas évaluation ; voy. aussi Brodeur, 1998b ou McKenzie, 2006, 342.

<sup>31</sup> Emprunté au manuel édité par Hough, Tilley (1998) pour le Home Office afin de guider les autorités locales dans la mise en œuvre du *Crime and Disorder Act de 1998*.

résultats. Vérifier que l'on a bien fait ce que l'on avait prévu, ce n'est pas encore une évaluation même si ce préliminaire lui est nécessaire. *L'évaluation commence quand on cherche, non pas ce qui a été fait, mais la conséquence qu'a eu cette action sur une cible.* Souvent cependant, on croit abusivement avoir évalué son action alors que la vérification n'a pas été poussée jusque là, qu'elle s'est bornée à enregistrer les produits pas les impacts.

On peut agir ainsi sous l'effet d'une mode gestionnaire ('managériale') qui insiste sur l'adéquation de la mise en œuvre aux objectifs ou sur la capacité à contrôler ses dépenses, et qui offre différentes techniques de vérification. Paradoxalement, cette posture peut très bien découler aussi d'une forte tradition administrative où le 'centre' dispose de moyens d'inspection qu'il mobilise pour vérifier la fidélité des mises en œuvre par rapport aux intentions des politiques.

## 2 - L'évaluation et ses données

Cette clarification de la nature même de l'exercice a des conséquences sur ses matériaux : règle générale, les données internes au dispositif ne suffisent pas à évaluer<sup>32</sup>.

La raison en est simple : tous ces comptages ne saisissent les événements qu'à travers les prises en charge institutionnelles dont ils font l'objet. Qu'un vol ou qu'une agression figure dans les statistiques policières dépend, non seulement de la survenance d'une victimation, mais encore de la propension de la victime à déposer une plainte et de celle de la police à enregistrer l'affaire sous une forme qui en permette le comptage.

Un exemple permettra de saisir la difficulté : soit une augmentation des enregistrements policiers<sup>32</sup> de la

---

<sup>32</sup> Par exemple Brodeur, 1998b.

délinquance. Doit-on conclure à l'aggravation de cette dernière ? Pas nécessairement, des enquêtes peuvent révéler, au contraire, une baisse des victimations. C'est la propension des victimes à se plaindre qui a crû<sup>33</sup> ou la manière d'enregistrer les plaintes qui est devenue plus systématique<sup>34</sup>. La seule consultation des données officielles aurait induit en erreur.

Par ailleurs, des comptages de fait ne permettent pas toujours d'apprécier la prévalence d'une victimation : soit une augmentation du nombre de délits enregistrés, elle peut provenir d'un élargissement de la population touchée ou bien, au contraire, d'une aggravation de l'incidence – il n'y a pas davantage de victimes mais chacune l'est plus souvent. Pour apprécier l'impact d'un programme de prévention, il faut absolument distinguer ces deux hypothèses, victimation diluée ou victimation concentrée. Souvent, seule une enquête le permet.

Enfin, le relief d'une victimation dépend beaucoup de la manière dont elle est vécue. Dans une enquête<sup>35</sup>, nous étions en présence de victimes peu alarmées – leur mésaventure tenait à leur style de vie et il leur était facile de le modifier légèrement pour éviter une répétition, par exemple d'éviter de se promener dans certains quartiers à certaines heures – tandis que d'autres vivaient très mal des incidents guère plus graves – leur victimation était liée à leurs conditions de vie, il leur aurait fallu déménager pour se mettre à l'abri et ils n'en avaient pas les moyens. Évaluer l'impact des programmes de prévention suppose de distinguer ces deux sortes de vécus. Là encore, de simples données institutionnelles ne suffisent pas, une enquête s'impose.

---

<sup>33</sup> Cet exemple est rapporté par Wittebrood (2009) à propos de programmes néerlandais, comme il l'était par Sherman (1997) à partir d'expériences américaines.

<sup>34</sup> Exemple *in* Hough, Mayhew, 2004.

<sup>35</sup> Zauberman, Robert, Pottier, 2000.

Tilley (1995) fait aussi remarquer que les données institutionnelles sont susceptibles – c’est bien compréhensible, mais néanmoins trompeur – d’être présentées d’une manière tendancieuse pour donner à croire à l’efficacité de l’action entreprise, même si ses résultats sont incertains.

Seulement, se contenter de données institutionnelles préexistantes est généralement bon marché et (relativement) rapide, tandis que faire enquête coûte cher et demande du temps<sup>36</sup>. Que faire alors ? D’abord, on peut mobiliser d’autres données préexistantes appartenant à d’autres secteurs<sup>37</sup> et qui permettent des recoupements : les impayés des logements sociaux, les statistiques d’incidents scolaires, d’école buissonnière... Il est ensuite possible de pré-positionner des enquêtes, par exemple en réalisant dans le cadre d’un observatoire régional des enquêtes régulières sur la victimation, l’insécurité, la délinquance auto-reportée. Même s’il faut grossir certains sous-échantillons pour travailler sur des zones précises, le coût sera moindre que celui d’une enquête montée à partir de rien et surtout la réalisation sera plus rapide. Enfin, mieux vaut évaluer seulement certains programmes et y mettre le prix en réalisant les enquêtes utiles plutôt que de prétendre tout évaluer sans données suffisantes. C’est ici que s’avère utile la disposition réservant, comme c’est l’usage aux Pays-Bas, à l’évaluation un pourcentage des crédits affectés aux politiques de prévention ; Lawrence Sherman (1997) le dit explicitement : ils doivent permettre de mettre en œuvre les enquêtes coûteuses qui sont nécessaires à une évaluation sérieuse.

Finalement, il ne faut pas attendre que le programme soit mis en œuvre ou, pis, achevé pour se préoccuper

---

<sup>36</sup> Ekblom, 1998 ; Hough, Tilley, 1998.

<sup>37</sup> Par ex. Hough, Tilley, 1998 ; cpr. Lagrange, Sina, 1998, encore qu’ils traitent du diagnostic préalable et estiment que l’évaluation doit faire appel à d’autres données.



des données qui seront utiles à son évaluation : certaines doivent être recueillies avant son démarrage afin de pouvoir ensuite comparer les situations *ex ante* et *ex post*.

Au chapitre des données, on doit encore se demander si celles qui sont quantitatives sont seules éligibles à la recherche évaluative. *A priori*, rien n'empêche de mobiliser dans une évaluation des données qualitatives. En concluant récemment, au Centre Marc Bloch à Berlin, une conférence de politologues spécialisés dans l'examen des politiques publiques au niveau des interactions entre 'l'agent de guichet' et ses clients, je remarquais que certaines de ces observations ethnographiques auraient pu fournir un bon matériau évaluatif. On sait, par ailleurs, que Tilley (1993a, b) a mobilisé des techniques qualitatives dans l'évaluation de portions de *Safer Cities*. Pawson & Tilley (1994, 1997) ont plaidé pour une évaluation 'réaliste' qui recourt au qualitatif, ce qui a déclenché une controverse avec les promoteurs du *Campbell Collaboration Crime and Justice Group* et notamment David Farrington (2003). À vrai dire, la discussion dépasse le seul statut des données ; elle porte plutôt sur la critique d'une méthode quasi-expérimentale à laquelle Pawson & Tilley reprochent de rechercher ce qui marche. Pour eux, en effet, il n'est possible d'évaluer que ce qui a marché dans un contexte donné et compte tenu de ce contexte... au moyen d'observations plutôt ethnographiques. Quel que soit le sort de cette querelle, il n'est pas certain qu'il porte condamnation du recours au qualitatif dans l'évaluation<sup>38</sup>.

En tout cas, mesurer l'impact d'un programme demande de ne pas s'arrêter à un simple audit ou à un contrôle de gestion, de séparer l'évaluation du pilotage de l'action, enfin de se donner les moyens de générer des données extérieures au dispositif sous examen.

---

<sup>38</sup> Toute l'œuvre de M.Q. Patton (voir notamment 1980) constitue un énergique plaidoyer en faveur de démarches qualitatives d'évaluation. Il définit cette dernière : déterminer comment un programme affecte les gens (328).

## VI - Mesurer préalablement la consistance de l'action à évaluer

Pour autant, si la mesure de ce qui a été fait ne peut tenir lieu d'évaluation, elle en constitue pourtant le préalable obligatoire, afin d'éviter de mesurer l'impact de ce qui n'a pas vraiment eu lieu.

Il faut donc commencer par déterminer la consistance de ce dont on va rechercher l'impact. Peu de domaines de l'action publique sont davantage sujets aux effets d'annonce que celui de la criminalité : on parle et l'on promet, mais ces affichages sont loin de constituer des descriptions fidèles de ce que l'on va réellement entreprendre. Il faudra ensuite être capable de distinguer l'échec dû à l'inefficacité du programme de ceux imputables tout simplement à une absence de mise en œuvre ou à son incomplétude. On admet généralement que la phase préliminaire de l'évaluation doit rendre compte des *objectifs*, des *moyens (inputs)*, de la *mise en œuvre*, des *produits (outputs)*.

Ce sont les premiers qui posent le plus de difficulté : généralement, on charge la barque et l'on annonce une multitude d'objectifs, parfois sans aucune relation compréhensible avec les mesures prises<sup>39</sup>. En cas de partenariat multiple, qui est le plus souvent la règle, ces finalités peuvent différer d'un protagoniste à l'autre et quelquefois se contredire<sup>40</sup>. Enfin, certains jugent né-

---

<sup>39</sup> Ainsi Brodeur (2000) note la distance qui existe entre la prolifération du discours sur le *Community Policing* et la pauvreté relative des démarches concrètes pour sa mise en œuvre.

<sup>40</sup> Au point de défier toute évaluation qui voudrait s'attacher à cette multiplicité contradictoire : Crawford, 1998b. Ainsi Breen (2002) et ses associés, travaillant sur l'évaluation de la justice, constatent l'absence de consensus sur les objectifs ; ils entreprennent alors d'en déterminer eux-mêmes pour se voir finalement désavoués par leurs commanditaires qui 'oublie' dans la mise en œuvre ceux d'entre ces objectifs qui pourraient s'avérer gênants, comme l'impartialité ou l'égalité (Beauvallet, 2009).

cessaires de distinguer les buts affichés de ceux qui restent latents. On risque bien de se lancer dans une course épuisante à la détermination des intentions. La tâche serait-elle plus aisée quand on descend de la politique globale au programme particulier ? Je suis arrivé à conclure qu'il n'était pas souhaitable de trop s'attarder à ce premier point : l'auditeur ou le contrôleur de gestion s'y attacheront eux fortement ; pour l'évaluateur, c'est le contenu de l'action qui compte bien plus que les objectifs de départ. Cependant, leur prise en compte peut peser, dans une certaine mesure sur le choix des critères en évitant de choisir ceux qui correspondraient à un but étranger à la logique de l'action entreprise. Ainsi la formule de *police communautaire* ou de *Community Policing*<sup>41</sup> implique le choix d'une démarche préventive ou aprioriste plutôt que répressive ou a posterioriste. On va exercer une surveillance propice à dissuader plutôt que d'attendre la commission du délit pour intervenir. L'évaluateur serait alors à côté de la plaque s'il choisissait comme instrument de mesure des comptages de l'activité répressive, par exemple des délits enregistrés : il jaugerait l'action à l'aune d'un but qu'elle ne poursuit pas. La difficulté est cruciale pour les programmes de prévention sociale qui ne poursuivent la réduction de la criminalité ou de l'insécurité qu'indirectement à travers d'autres objectifs, comme l'accès à l'emploi stable ou la diminution de la pauvreté. En les évaluant à l'aune d'objectifs qu'ils ne poursuivent pas directement, on se prive de la possibilité de vérifier s'ils atteignent leurs objectifs immédiats. Reste ensuite la difficulté consistant à déterminer si le succès de l'objectif immédiat a ensuite ricoché favorablement sur l'objectif indirect de la réduction du crime ou de l'insécurité.

La détermination des moyens importe surtout aux contrôleurs de gestion qui veulent, dans une optique

---

<sup>41</sup> Voy. par ex. Skogan, 1990 ; Fielding, 1995 ; Rosenbaum, 1994 ; Skogan, Hartnett, 1997a, b ; Brodeur, 1998a, 2000.

‘managériale’, mesurer le rapport entre les coûts exposés et les produits obtenus<sup>42</sup>. Elle est moins centrale, me semble-t-il, pour l’évaluateur. En tout état de cause, il s’agit – ce n’est pas toujours tâche aisée – de distinguer ce qui est moyen nouveau ou additionnel spécifiquement lié à l’action entreprise. Souvent, les programmes s’opèrent par redéploiement de dotations existantes ou à moyens constants ou relativement constants. Dans une telle situation, le surcoût spécifique demeure marginal, du moins dans sa partie chiffrable car sa mise en œuvre peut exiger des personnels des dépenses d’énergie, de volontarisme et d’imagination d’autant plus élevés qu’on leur fournit peu de moyens additionnels<sup>43</sup>.

Dans l’évaluation, déterminer ce qui a effectivement été mis en œuvre constitue, me semble-t-il, le point crucial de la phase préliminaire<sup>44</sup>. C’est ainsi qu’on pourra éviter d’imputer au programme les échecs qui proviennent des défauts de sa réalisation ou, tout bonnement, de son absence. Tant il est vrai qu’en notre domaine on dit beaucoup plus qu’on ne fait, l’annonce tient souvent lieu d’action. Pour déterminer sans trop de difficulté quelle intervention a effectivement eu lieu, il est vraiment utile d’avoir décidé de l’évaluation avant même le démarrage du programme ; il est aussi indispensable que l’évaluateur soit indépendant de ceux qui déterminent le programme comme de ceux qui le mettent en œuvre ; il faut cependant qu’il puisse sans difficulté observer ce qui se fait. Cette contradiction n’est pas toujours facile à gérer : que le responsable national de la politique publique – législatif ou exécutif – ait prévu l’évaluation et provisionné son financement, qu’un pôle central sélectionne les projets à évaluer et cautionne les évaluateurs constituent alors des précautions propres à mettre de l’huile dans les rouages.

---

<sup>42</sup> Voy. par ex. Braga *et al.*, 1999.

<sup>43</sup> Dans ce cas, le lancement du nouveau programme peut entraîner – l’évaluateur devra en tenir compte – un certain délaissement des tâches habituelles.

<sup>44</sup> Par ex. Brodeur, 1998b.

Enfin, la phase préliminaire de l'évaluation s'achève par la mesure des produits (*outputs*). Mais cet élément est surtout crucial pour le contrôleur de gestion ou l'auditeur. À l'évaluateur, sa détermination sert surtout à le distinguer de l'impact (*outcome*) qui constitue, lui, le vrai critère de l'évaluation<sup>45</sup>. Et c'est pourquoi l'évaluateur, contrairement à l'auditeur ou au contrôleur de gestion, ne peut se contenter de données internes ; il lui faut faire enquête pour constituer des données externes au dispositif et donc susceptibles d'en mesurer l'impact. Sinon, selon l'expression de Sherman (1997), on évalue les efforts pas les résultats.

C'est à ce stade que les données internes sont surtout utiles : pour rendre compte de la mise en œuvre et de ses produits. Encore faut-il qu'on ne réduise pas l'action à quelques indicateurs qui la caricatureraient : seule une analyse compréhensive permettra ensuite d'estimer l'impact des interventions<sup>46</sup>. Au surplus, décrire les opérations réalisées ne suffirait pas, encore faut-il les remettre dans leur contexte pour juger ensuite de la transposabilité des résultats observés.

Hope (2009a) va plus loin et conseille de prendre en compte, non seulement l'existence d'une action, mais aussi son intensité<sup>47</sup> en estimant, grâce à des emprunts méthodologiques à l'économétrie, un score d'intensité<sup>48</sup>, une manière élégante de mesurer la consistance globale de l'action que l'on va évaluer.

---

<sup>45</sup> P. ex. Crawford, 1998b ; Hough, Tilley, 1998.

<sup>46</sup> Voy. Pawson, Tilley, 1994.

<sup>47</sup> En ce sens également McKenzie, 2006, 342.

<sup>48</sup> *Time series analysis of programme impact*.

## VII - Les savoir-faire évaluatifs

Il s'est installé une sorte de standard minimal comportant une comparaison avant/après, la prise en compte de groupes ou de zones de contrôle, enfin l'examen de la relation entre l'action et l'impact.

C'est d'ailleurs à cette aune que L. Van Noije et Karin Wittebrood (2008) ont sélectionné 69 évaluations valides sur les 149 qu'ils ont pu réunir aux Pays-Bas. Ce faisant, ils se réfèrent explicitement aux précédents de Sherman *et al.* (2002), du *Campbell Collaboration Crime and Justice Group* (Farrington, Petrosino, 2001) et du *Scientific Model Scale* ou SMS (Farrington & *al.*, 2001 ; Farrington, 2003).

Créé et longtemps présidé par David Farrington, professeur de psychologie criminelle à l'Université de Cambridge, le *Campbell Collaboration Crime and Justice Group* se propose de réaliser une revue systématique des études d'évaluation de l'efficacité de l'intervention criminologique et de la rendre largement accessible. Il entend ainsi poursuivre l'œuvre *princeps* de Donald T. Campbell (1969) qui est généralement considéré comme le promoteur de la recherche évaluative. On notera que ces deux auteurs essentiels sont tous deux des psychologues expérimentalistes : l'évaluation quasi-expérimentale qu'ils proposent des politiques de prévention et de sécurité doit beaucoup au mode de raisonnement et aux méthodes de leur discipline d'origine.

Farrington (2003) retient cinq critères de qualité méthodologique des évaluations ; il attribue les quatre premiers à Campbell :

- la validité statistique concerne le caractère significatif de la liaison entre l'intervention et l'effet supposé (*outcome*) ;
- la validité interne concerne l'examen du lien de causalité entre intervention et effet ;
- la validité du construit concerne l'adéquation

de l'effet avec la définition opérationnelle et la mesure des construits théoriques qui sous-tendent l'intervention ;

- la validité externe concerne la possibilité de généraliser la relation causale entre intervention et effet ;
- enfin la validité descriptive concerne l'adéquation de la présentation, dans un rapport de recherche, des traits cruciaux d'une évaluation.

Parmi toutes les échelles de qualité méthodologique des évaluations, la plus célèbre est la *Maryland Scientific Method Scale* (SMS) mise au point à partir du bilan mené à bien pour le Congrès des États-Unis par Lawrence Sherman et ses collaborateurs.

Elle possède cinq niveaux :

- niveau 1 : simple corrélation entre un programme et une mesure de la délinquance ;
- niveau 2 : mesure de délinquance avant et après la mise en œuvre du programme ;
- niveau 3 : mesure de la délinquance avant et après instauration d'un groupe de contrôle ;
- niveau 4 : mesure de la délinquance avant et après et contrôles multiples avec prise en compte des autres variables pouvant influencer la délinquance ;
- niveau 5 : assignation au hasard dans le groupe soumis au programme et celui dédié au contrôle.

Mais Farrington (2003) lui reproche de ne pas couvrir toutes les dimensions de validité retenues par le *Campbell Group*.

La comparaison avant/après est évidemment fondamentale : sans elle, pas d'évaluation, tout simplement. Quatre points au moins méritent l'attention. D'abord, mieux vaut prévoir l'évaluation avant le commencement de l'action<sup>49</sup> : on sera plus à l'aise pour observer la situation *ex ante*, plutôt que d'avoir à la reconstituer

---

<sup>49</sup> Par ex. Sherman *et al.*, 1997 ; Brodeur, 1998b.

péniblement après coup. Ensuite, mieux vaut retenir un nombre suffisant de critères pour cette comparaison avant/après afin de ne pas manquer des effets auxquels on n'avait pas pensé. Pareille précaution permet de découvrir plus facilement les effets pervers<sup>50</sup> : un harcèlement des *dealers* diminue en effet l'impact de la drogue dans un quartier, mais les méthodes d'intervention policière exaspèrent tellement les jeunes que les violences augmentent. Que cette mesure d'avant/après déborde le cadre de la zone d'intervention du programme permet d'observer d'éventuels effets de déplacement<sup>51</sup> de la délinquance – elle diminue là où s'applique le programme mais se transporte à côté – mais aussi des effets de contagion vertueuse<sup>52</sup> – l'investissement préventif est assez puissant pour rayonner aux alentours de sa zone propre d'intervention<sup>53</sup>.

Les spécialistes critiquent beaucoup les mesures d'avant/après qui ne s'accompagnent pas de l'observation de zones ou de populations de contrôle où le programme à évaluer ne s'applique pas. On insiste même sur l'utilité de disposer d'un panier de zones ou de population de contrôle afin de neutraliser l'effet d'une crise soudaine dans l'une d'entre elles<sup>54</sup>. Ainsi ambitionne-t-on de passer du modèle assez primitif de la 'boîte noire'<sup>55</sup> à un dispositif que l'on qualifie de quasi-expérimental. Pour autant, la mise en œuvre de cette procédure n'est pas sans poser de redoutables problèmes. Passe encore si l'on divise en deux un même quartier

---

<sup>50</sup> Par ex. Sherman *et al.*, 1997, mais aussi Hope (2009a) à propos du programme Kirkholt.

<sup>51</sup> Hough, Tilley, 1998.

<sup>52</sup> On pense fréquemment aux effets de déplacement, moins souvent à ceux d'élargissement (Hope, 2009a).

<sup>53</sup> Par ex. Ekblom, 1996b.

<sup>54</sup> Voy. par ex. Bennett, 1991.

<sup>55</sup> Se dit d'une observation qui considère les intrants et les extrants mais pas les mécanismes intermédiaires.



ou une population pour réserver une moitié au contrôle. Mais s'il faut chercher ailleurs une ou des zones ou populations de contrôle, deux difficultés apparaissent: i) il est difficile de parvenir à un appariement convaincant – aussi proches que paraissent deux quartiers par leurs populations et leurs problèmes, leurs histoires peuvent différer profondément et donc leurs capacités d'autocontrôle – et ii) une zone ou population de contrôle peut faire l'objet d'un autre programme produisant des effets semblables à celui que l'on veut tester. Pour souhaitable qu'elle soit, la comparabilité est souvent malaisée à assurer. Disposer de zones ou de populations de contrôle a pour but de trancher l'interrogation : le changement observé peut-il être attribué au programme évalué ou se serait-il passé quand même en son absence ? Mais le fonctionnement de ce contrôle n'est pas automatique. Ainsi l'effet nul d'un programme de surveillance du voisinage peut simplement traduire l'identité de 'capital social' de la zone d'application et de celle de contrôle (Hope, 2009a). Ce type de difficulté disparaît évidemment avec l'évaluation 'réaliste' de Tilley. Puisqu'il ne s'agit plus de déterminer 'ce qui marche', mais ce qui a marché dans un contexte donné compte tenu de ce contexte, il n'est évidemment plus utile de rechercher des zones ou populations de contrôle, ni de s'exposer aux affres de cette quête. Dans une telle perspective, ce n'est plus le programme que l'on teste mais la théorie qui sous-tend sa mise en œuvre concrète.

Plus généralement, l'application mécanique d'une démarche quasi-expérimentale ne permet pas d'éviter tous les biais de sélection sur lesquels Hope (2009a) a particulièrement attiré l'attention. Ainsi on peut choisir des terrains parce qu'ils semblent bien disposés envers le programme que l'on veut mettre en œuvre, mais l'effet observé peut être dû autant au 'capital social' dont dispose ce terrain qu'au programme qu'on y

déploie<sup>56</sup>. Mais l'on peut aussi sélectionner des zones ou des groupes particulièrement défavorisés et n'observer alors que leur retour à la moyenne. C'est pourquoi Hope suggère de modéliser les effets de sélection en s'inspirant de la micro-économétrie.

En tous cas, l'ensemble de ces difficultés donne une importance particulière à la troisième phase de l'évaluation : avant de conclure à l'existence d'un impact, il faut pouvoir i) envisager et rejeter des explications alternatives<sup>57</sup>, ii) expliquer au contraire comment les actions concrètement réalisées ont pu parvenir au résultat observé<sup>58</sup> et iii) décider de la vraisemblance de ce processus. C'est là un point crucial où se déploient le savoir-faire de l'évaluateur et son expérience.

Reste enfin l'aval de l'évaluation, son utilisation. On peut distinguer l'accompagnement de l'action et sa transposabilité.

Simple jugement après-coup, l'évaluation n'a pour le programme sous examen qu'un intérêt historique qui peut ne pas passionner énormément les responsables et les praticiens. C'est tout l'avantage d'une étude d'impact mise en place dès le début qui est donc à même de produire des informations en cours de route permettant d'ajuster les dispositifs<sup>59</sup>.

---

<sup>56</sup> Encore qu'un programme peut avoir au moins l'effet positif de mobiliser les ressources latentes du groupe ou de la zone où il s'applique. Il devient alors la variable intermédiaire qui produit l'impact (*outcome*), alors que l'on ne parvient pourtant pas à observer de relation directe entre sa mise en œuvre et l'impact.

<sup>57</sup> Rosenbaum, 1986 ; entre autres, il faudra disposer de données suffisamment longues sur la situation antérieure pour déceler d'éventuelles tendances qui rendaient compte du changement avant/après même si le programme n'avait pas existé (Hough, Tilley, 1998 ; Hope, 2000). Il faut aussi s'assurer que les changements quantitatifs observés sont suffisamment significatifs (Sherman *et al.*, 1997).

<sup>58</sup> Voy. par ex. Tilley, 1993 ; Hough, Tilley, 1998 ; Brodeur, 1998b.

<sup>59</sup> C'est l'ambition du programme d'évaluation de la réforme de la police à Chicago. Voy. Skogan, Hartnett, 1997a, b.

Mais on attend aussi de l'évaluation qu'elle indique les formules assez prometteuses pour être transposables. La vaste revue commandée pour le Congrès des États-Unis à Lawrence Sherman (1997) poursuivait avant tout cet objectif, tout comme celle des mesures correctionnelles menée à bien par McKenzie (2006). C'est d'ailleurs pour cela que l'on recommande de concentrer l'évaluation sur certains programmes qui, par leur rigueur et leur ampleur, sont susceptibles, en cas d'évaluation positive, de resservir ensuite à une plus grande échelle<sup>60</sup>. Néanmoins, l'exercice est délicat : généraliser une expérience-pilote ne va pas de soi ; ce qui a donné de bons résultats dans un certain contexte peut s'avérer moins performant si on le transpose dans d'autres très différents. Ici, les biais de sélection peuvent jouer à plein d'où l'importance de les détecter et de les neutraliser avant de conclure à une validité externe du programme sous évaluation.

## VIII - Conclusion

En fin de compte, l'évaluation constitue un exercice cumulatif. C'est en évaluant qu'on apprend à améliorer ses façons de faire. C'est aussi en évaluant qu'on accumule progressivement un savoir pratique sur l'impact des différentes démarches de prévention... un savoir toujours révisable, comme c'est habituellement le cas en matière scientifique, mais un savoir qui permet de rendre compte de l'action publique et de l'ajuster au fur et à mesure. Il est vain de chercher à tout évaluer, surtout quand on dispose de peu d'expériences et de compétences. On ne peut déboucher que sur des pseudo-évaluations dont la valeur tient surtout de l'impression et de l'effet de mode. Mieux vaut commencer modestement en choisissant quelques programmes précis sur

---

<sup>60</sup> Ainsi Hough, Tilley (1998) conseillent de la réserver aux programmes innovateurs, élaborés et coûteux ; voy. aussi la proposition du rapport Sherman (1997).

lesquels on concentrera les moyens et les compétences disponibles. Ainsi développera-t-on peu à peu un savoir-faire et accumulera-t-on des diagnostics raisonnablement fiables. Puis, le processus fera boule de neige à travers une succession d'essais et d'erreurs.

Il ne faut pas oublier toutefois que l'évaluation est le domaine de relations particulièrement délicates entre le monde des décideurs et celui des scientifiques. Entre le refus des premiers de recourir aux seconds et, au contraire, une quasi-annexion, l'établissement d'une coopération fondée sur le respect de l'autonomie réciproque des deux sphères n'est pas facile à obtenir. Sans elle cependant l'évaluation ne serait que faux-semblant.

Pour autant, l'efficacité d'un programme ne tranche pas de la pertinence de sa localisation : les moyens que l'on dépense à un endroit risquent de manquer ailleurs où les besoins peuvent être plus urgents. Au delà de toute évaluation, on ne peut pas se dispenser d'une réflexion sur la répartition des actions, sinon les actions de sécurité peuvent bien devenir à la longue un privilège de nantis<sup>61</sup>.

Enfin, les politiques de prévention et de sécurité sont généralement incapables de venir à bout des effets dévastateurs de l'accumulation de conditions socio-économiques négatives dans certains groupes sociaux ou certaines zones urbaines à forte concentration de pauvreté<sup>62</sup>. Il ne faudrait pas qu'elles servent à masquer l'absence de politiques sociales et économiques efficaces ou, pis, la persistance d'une accumulation de décisions et de pratiques ségrégatives. Sans politique efficace de réaffiliation, elles ne seraient qu'un leurre.

---

<sup>61</sup> Crawford, 1998a, b ; également Hope (2009a).

<sup>62</sup> Sherman *et al.* (1997) y insiste à plusieurs reprises pour les États-Unis.

Ceci dit, il serait intéressant de déterminer combien d'évaluations (scientifiquement acceptables) ont eu un effet réel de modification de politiques publiques.

Au chapitre des recommandations, on pourrait suggérer à qui veut s'engager dans l'évaluation des politiques de prévention et de sécurité :

- de ne pas confondre l'évaluation – qui porte sur l'impact de ces politiques sur une cible – avec l'audit, le contrôle de programme ou le calcul coût-efficacité ;
- d'en confier la réalisation à un organe scientifique compétent et extérieur aux institutions qui sont en charge des programmes à évaluer ;
- de respecter l'extériorité mutuelle du domaine des décideurs et de celui des évaluateurs ;
- de prévoir l'évaluation avant le démarrage du programme ;
- d'y affecter des données et un savoir-faire cohérents avec la nature de l'évaluation.

## Références

BENNETT T., The effectiveness of a police-initiated fear-reducing strategy, *British Journal of Criminology*, 1991, 31, 1, 1-14.

BEAUVALLLET M., *Les stratégies absurdes. Comment faire pire en croyant faire mieux ?*, Paris, Seuil, 2009.

BRAGA A.A., WEISBURD D.L., WARING E.J., GREEN MAZEROLLE L., SPELMAN W., GAJEWSKI F., Problem-oriented policing in violent crime places: A randomized controlled experiment, *Criminology*, 1999, 37, 3, 541-580.

BREEN E. (Dir.), *Évaluer la justice*, Paris, PUF, 2002.

BRODEUR J.-P. (Ed.), *How to Recognize Good Policing: Problemes and Issues*, Thousand Oaks, Sage, 1998a.

BRODEUR J.-P., L'évaluation des programmes d'intervention sécuritaire ; quelques remarques sur les expériences canadienne et américaine, in FERRET F., OCQUETEAU F. (Eds), *Évaluer la police de proximité ? Problèmes, concepts, méthodes*, Paris, La Documentation française, 1998b, 13-22.

BRODEUR J.-P., *Les réformes de la police en Amérique du Nord : bilan des évaluations*, Eines para l'analisi de la seguretat ; Estat de la recerca, Mollet del Vallès, Escola de policia de Catalunya, 2000.

CAMPBELL D.T., Reforms as Experiments, *American Psychologist*, 1969, 24, 4, 409-429.

COOK T.D., CAMPBELL D.T., *Quasi-Experimentation*, Chicago, Rand-Mc Nelly, 1979.

CRAWFORD A., *Crime Prevention and Community Safety. Politics, Policies and Practices*, Longman, 1998a.

CRAWFORD A., Culture managériale de l'évaluation et responsabilités ; quelques leçons de l'expérience britannique des programmes locaux de sécurité in FERRET F., OCQUETEAU F. (Eds), *Évaluer la police de proximité ? Problèmes, concepts, méthodes*, Paris, La Documentation française, 1998b, 51-82.

DONZELOT J., ESTEBE PH., *L'État animateur. Essai sur la politique de la ville*, Paris, Éd. Esprit, 1993.

EKBLOM P., Safer cities and residential burglary: A summary of evaluation results, *European Journal of Criminal Policy and research*, 1996, 4, 1, 22-52.

EKBLOM P., Évaluation du programme *Safer Cities* ; impact et processus, in FERRET F., OCQUETEAU F. (Eds), *Évaluer la police de proximité ? Problèmes, concepts, méthodes*, Paris, La Documentation française, 1998, 27-50.

EKBLOM P., LAW H., SUTTON M., CRISP P., WIGGINS R., *Safer Cities and Domestic Burglary*, Home office Research Study 164, London, Home Office research and Statistics Directorate, 1996.

FARRINGTON D.P., Methodological Quality Standards for Evaluation Research, *Annals of the American Academy of Political and Social Science*, 2003, 587, 49-68.

FARRINGTON D.P., PETROSINO A., The Campbell Collaboration Crime and Justice Group. *Annals of the American Academy of Political and Social Science*, 2001, 587, 35-49.

FIELDING N.G., *Community Policing*, Oxford, Clarendon, 1995.

HOPE T., *Counting on the Counter-Factual? Evaluation and Crime Reduction in Risk Society*, Leicester, British Conference of Criminology Conference, 2000.

HOPE T., *Evaluation of Safety and Crime Prevention Policies. England and Wales*, FP6, CRIMPREV, WP7, WS3, 2009a [www.crimprev.eu].

HOPE T., *The evaluation of policies from a methodological perspective - Discussion*, FP6, CRIMPREV, WP4, WS6, Universidade de Porto, Faculty of Law, Escola de Criminologia, 2009b.

HOUGH, M., MAYHEW P., *L'évolution de la criminalité à travers deux décennies du British Crime Survey*, *Dévi-ance et Société*, 2004, 28, 3, 267-284.

HOUGH M., TILLEY N., *Auditing Crime and Disorder: Guidance for Local Partnerships*, London, Home Of- fice, 1998.

JUNGER-TAS J., *Meta-evaluation of Dutch Crime Pre- vention programs*, Den Haag, Ministry of Justice, WODC, 1989.

KESSLER M.C., LASCOUMES P., SETBON M., THOENIG J.- C. (dir.), *L'évaluation des politiques publiques*, Paris, L'Harmattan, 1998.

LAGRANGE H., SINA F., *Réflexions de méthode sur les diagnostics locaux de sécurité*, in FERRET F., OCQUE- TEAU F. (Eds), *Évaluer la police de proximité ? Prob- lèmes, concepts, méthodes*, Paris, La Documentation française, 1998, 91-114.

LASCOUMES P., SETBON M., *L'évaluation pluraliste des politiques publiques ; enjeux, pratiques, produits*, Par- is, CGP, 1995.

LEIGH A., *Policing in England and Wales: National Developments in Performance Management*, Congrès internacional sobre les eines per a l'anàlisi de la sug- eretat. L'estat de la recerca, Mollet del Vallès, Escola de policia de Catalunya, 2000.



MACKENZIE D.L., *What Works in Corrections. Reducing the Criminal Activities of Offenders and Delinquents*, Cambridge, Cambridge University Press, 2006.

MARTINSON R., What works? Questions and answers about prison reform, *Public Interest*, 1974, 35, 22-45.

PATTON M.Q., *Qualitative Evaluation and Research Methods*, Newbury Park, Ca, 1980.

PAWSON R., TILLEY N., What works in evaluation research?, *British Journal of Criminology*, 1994, 34, 3, 291-306.

PAWSON R., TILLEY N., *Realistic Evaluation*, London, Sage, 1997.

POLDER W., VAN VLAARDINGEN F.J.C., Preventiestrategieën in de praktijk, een meta-evaluatie van criminaliteitpreventieprojecten, *Justitiele Verkenningen*, 1992, 2, 5-104.

ROBERT PH., Évaluer la prévention, *Archives de politique criminelle*, 1994, 16, 53-70.

ROBERT PH., The evaluation of prevention policies, *European Journal of Crime, Criminal Law and Criminal Justice*, 2003, 1, 114-130.

ROBERT PH., *L'évaluation des politiques de prévention et de sécurité examinée sous une perspective méthodologique*, FP6, CRIMPREV, WP4, WS6, Universidade de Porto, Faculty of Law, Escola de Criminologia, 2009.

ROBERT PH., LASCOUMES P., *Les bandes d'adolescents ; une théorie de la ségrégation*, Paris, Éditions ouvrières, 1974.

ROCHÉ S., *Police de proximité. Nos politiques de sécurité*, Paris, Seuil, 2005.

ROSENBAUM D.P. (Ed.), *Community Crime Prevention; Does It Work?*, Beverley Hills, Sage, 1986.

ROSENBAUM D.P., *The Challenge of Community Policing: Testing the Promises*, Thousand Oaks, Sage, 1994.

SHERMAN L.W., FARRINGTON D.P., WELSH B.C., MACKENZIE D.L. (Eds), *Evidence-based crime prevention*. London/New York, Routledge, 2002.

SHERMAN L.W., FARRINGTON D.P., WELSH B.C., MACKENZIE D.L. (Eds), *Evidence-based crime prevention (revised edition)*, London, Routledge, 2006.

SHERMAN L.W., GOTTFREDSON D., MAC KENZIE D., ECK J., REUTER P., BUSHWAY S., *Preventing Crime: What Works, What Doesn't, What's Promising*, Washington DC, NIJ, 1997.

SKOGAN W.G., *Disorder and Decline: Crime and the Spiral of Decay in American Cities*, NY, Free Press, 1990.

SKOGAN W.G., HARTNETT S.M., *Community Policing, Chicago Style*, Oxford, Oxford University Press, 1997a.

SKOGAN W.G., HARTNETT S.M. AND THE CHICAGO COMMUNITY POLICING EVALUATION CONSORTIUM, *Community Policing in Chicago, Year Four: An Interim Report*, Illinois Criminal justice Information Authority, 1997b.

SMEETS S., TANGE C., *L'évaluation des politiques de sécurité et de prévention en Belgique*, FP6, CRIMPREV, WP7, WS3, 2009 [w [www.crimprev.eu](http://www.crimprev.eu)].

SOETENHORST DE SAVORNIN LOHMAN J., *New Forms of crime Prevention: the Dutch Experience*, in HEBBERECHT P., SACK F. (dir.), *La prevention de la délinquance en Europe ; nouvelles stratégies*, Paris, L'Harmattan, 1997, 83-100.

TILLEY N., *Understanding Car Parks, Crime and CCTV: Evaluations and Lessons from Safer Cities*, London, HMSO, 1993a.

TILLEY N., *After Kirkholt – Theory, Method and Results of Replication Evaluations*, Crime prevention Unit Series, Paper 47, London, Home Office, 1993b.

TILLEY N., *Thinking about Crime Prevention Performance Indicators*, London, Home Office, 1995.

VAN DIJK J.J.M., The State of Knowledge in the Netherlands, in ROBERT PH. (Ed.), *Crime and Prevention Policy*, Eigenverlag Max-Planck-Institut für ausländisches und internationales Strafrecht, 1993, 243-263.

VAN NOIJE L., WITTEBROOD K., *Sociale veiligheid ontseuteld: veronderstelde en werkelijke effecten van veiligheidsbeleid*, Den Haag, Sociaal en Cultureel Planbureau, 2008.

WITTEBROOD K., *Evaluation of safety and crime prevention policies in the Netherlands*, FP6, CRIMPREV, WP7, WS3, 2009 [[www.crimprev.eu](http://www.crimprev.eu)].

WYVEKENS A., Quand l'urbain rencontre la sécurité : la prévention situationnelle en France aujourd'hui, *Revue de droit pénal et de criminologie*, 2008, 9-10, 897-900.

WYVEKENS A., *L'évaluation des politiques de prévention et de sécurité en France : une activité problématique*, FP6, CRIMPREV, WP7, WS3, 2009 [[www.crimprev.eu](http://www.crimprev.eu)].

ZAUBERMAN R., ROBERT PH., POTTIER M.-L., Risque de proximité ou risque lié au style de vie ; Enquêtes et évaluation de la sécurité urbaine, *Cahiers de la sécurité intérieure*, 2000, 42, 193-220.