

CAHIER DE RECHERCHE

JUILLET 91

■ N° 17

PRATIQUES EXEMPLAIRES OU EXEMPLES DE PRATIQUES : L'EVALUATION DANS LE SECTEUR SOCIAL AUX ETATS-UNIS

Analyse de monographies présentées

dans

"Evaluation Review" et dans "Evaluation and the Health Professions"

Patricia Croutte
Michel Legros

CRÉDOC

CREDOC

CAHIER DE RECHERCHE

**PRATIQUES EXEMPLAIRES OU EXEMPLES DE PRATIQUES :
L'ÉVALUATION DANS LE SECTEUR SOCIAL
AUX ÉTATS-UNIS**

Analyse de monographies présentées

dans

"Evaluation Review" et dans "Evaluation and the Health Professions"

Patricia CROUTTE

Michel LEGROS

Traduction : Ursula HYZY

**Département "Évaluation
des Politiques Sociales"**

Secrétariat : Christiane DUBOIS

Juillet 1991

142, rue du Chevaleret
7 5 0 1 3 - P A R I S

S O M M A I R E

	Page
INTRODUCTION : LES SEDUCTIONS AMERICAINES	5
 PREMIERE PARTIE : L'EVALUATION SUR PROGRAMME, LES PROGRAMMES D'EVALUATION	 15
I - DESCRIPTIONS DES PROGRAMMES EVALUES	17
1-1. Le nom du programme	17
1-2. La population concernée	19
1-3. Les objectifs du programme	20
1-4. Durée et localisation des programmes	22
1-5. Le contenu des programmes	23
1-6. Financement et maîtrise d'ouvrage	24
 II - LA DESCRIPTION DE L'EVALUATION	 25
2-1. Le but de l'évaluation	26
2-2. Les indicateurs retenus	27
2-3. Localisation, date et durée de l'évaluation	29
2-4. Le choix des populations	31
2-5. Le choix des méthodes et collecte de l'information	32
2-6. Le choix des méthodes de traitement de l'information	33
 III - LA MESURE DES EFFETS	 34
 SECONDE PARTIE : L'ECHIQUIER EVALUATIF	 37
 I - EVALUATION ET RECOURS A LA METHODE EXPERIMENTALE	 42
1-1. Méthode expérimentale et orthodoxie minimale	42
1-2. Méthode expérimentale et contraintes externes aux programmes	44
1-3. Effet-action : logique d'inventaire ou inventaire logique	45

II - L'EVALUATION COMME ETUDE	46
2-1. L'évaluation : une étude étroitement finalisée	46
2-2. Le temps comme opérateur principal	47
2-3. Questionnaire, grille, observation, entretien : des enquêtes par emprunts	48
III - LES APPORTS DE L'EVALUATION	50
3-1. Une pluralité de savoirs-faire	50
3-2. Evaluation et pilotage de l'action	52
3-3. Une communauté d'évaluateurs	53
CONCLUSION	55
FICHES DE PRESENTATION DES EXPERIENCES EVALUATIVES	61
ANNEXE	199

INTRODUCTION

LES SEDUCTIONS AMERICAINES

Dans l'histoire, encore bien courte, du développement de l'évaluation en France, deux rapports occupent une place particulière. Le premier, émanant d'un groupe de travail constitué par le Commissariat Général du Plan et placé sous la présidence de Michel DELEAU, définit le concept d'évaluation, propose une clarification et une information sur les différentes méthodes employées et insiste sur le rôle moteur que doivent avoir les pouvoirs publics dans l'extension des pratiques évaluatives¹. Le second rapport, rédigé par Patrick VIVERET², réintroduit le jeu des acteurs sociaux dans la légitimation des procédures évaluatives et dans la critique des produits de l'évaluation. Par cette double démarche, l'auteur de ce dernier rapport place l'évaluation au coeur du débat politique et, conséquence logique, préconise la création d'un dispositif étatique de nature à assurer le développement de l'évaluation des politiques publiques.

La complémentarité est particulièrement forte entre les deux rapports. Ainsi, lorsque le rapport de Michel DELEAU met l'accent sur les méthodes, celui de Patrick VIVERET insiste sur les procédures publiques d'accompagnement de l'évaluation. Là où le premier développe une réflexion sur la nécessaire rigueur, le second propose de mettre cette rigueur au service de l'élaboration de procédures plus démocratiques. Cette proximité des rapports ne fait qu'apporter un poids plus grand au constat fait par les deux auteurs d'un réel déficit d'évaluation dans le fonctionnement des politiques publiques. Pour apprécier la réalité de ce déficit, les deux auteurs procèdent de manière similaire en analysant, d'une part, les pratiques d'évaluation en France, et le bilan n'est guère brillant, et, d'autre part, en regardant l'état des pratiques évaluatives dans les autres pays développés. Dans ce dernier cas, les Etats-Unis constituent une référence obligée et du PPBS³ à l'évaluation des effets de la mise en oeuvre de l'impôt négatif dans le NEW JERSEY⁴, quelques exemples phares viennent démontrer l'importance des pratiques évaluatives outre-atlantique. Dans ces

¹ DELEAU M., NIOCHE J-P., PENZ Ph., POINSARD R., Evaluer les politiques publiques. Méthodes, Déontologie, Organisation. Commissariat Général du Plan, Mai 1986, 215 p., ronéoté.

² VIVERET P., L'évaluation des politiques et des actions publiques. Rapport au Premier Ministre. Juillet 1988, 104 p., ronéoté.

³ Planning Programming Budgeting System. Démarche rationalisatrice engagée par l'administration américaine dans les années 1960 et consistant à réaliser une évaluation, a priori, des effets attendus de programmes alternatifs, évaluation reposant sur des techniques de recherche opérationnelle et de calcul économique.

⁴ Sur cette évaluation, voir le rapport DELEAU, p 78

rapports et dans quelques autres travaux¹, si les expériences américaines figurent en bonne place, il ne s'agit jamais, pour leurs auteurs, de céder à une quelconque fascination pour un modèle américain de l'évaluation. Dans leurs attitudes à l'égard de la pratique évaluative aux Etats-Unis, les chercheurs français manifestent plutôt un sentiment complexe fait d'intérêt distant et de critique feutrée, sentiment alimenté par la perception de l'existence d'un fossé culturel difficilement franchissable entre les deux pays.

Cette distance, qui n'est pas que géographique, rend incertaine l'existence de filiations entre les pratiques évaluatives américaines et celles développées par les chercheurs et les praticiens français. Certes, il est toujours possible d'affirmer une continuité entre le PPBS et les techniques dites de " Rationalisation des Choix Budgétaires ", mais cette affirmation ne résiste pas à un examen attentif des conditions de mise en oeuvre de ces deux outils de management public. Il est même à craindre que la seule ressemblance réelle entre ces deux méthodes réside dans leur abandon par les deux administrations. Entre les deux pays, les échanges et les coopérations sur ce thème restent sporadiques, se limitant le plus souvent à quelques rares missions d'information. Ainsi, lorsque les parlementaires français envisagent de créer un Office Parlementaire des Choix Scientifiques et Technologiques, ils consultent leurs homologues américains. De la même manière, il arrive que des membres du General Accounting Office participent à quelque séminaire lors d'un passage à Paris, mais les échanges restent limités². Ceci explique sans doute que les programmes d'évaluation mis en oeuvre en France ne se réclament guère des approches nord-américaines. Si le dispositif législatif instaurant un Revenu Minimum d'Insertion correspond à une de nos premières "Sunset Law"³, la référence à ce type de législation et aux modes d'évaluation qui lui sont nécessairement liés n'a pas été intégrée dans le débat politique au moment du vote de la loi du 1er décembre 1988.

Dans le domaine des politiques sociales, les logiques évaluatives les plus anciennes sont à situer dans le secteur éducatif mais leurs origines sont à rechercher dans les travaux de BINET. La tentative d'acclimater dans les établissements sociaux français une méthode dite

¹ cela est particulièrement le cas de l'ouvrage d'Eric MONNIER, Evaluations de l'action des pouvoirs publics, Paris, Economica, 1987

² à titre d'exemple, une revue comme Evaluation Review ne figure pas au bulletin signalétique du CNRS.

³ Le concept de sunset law désigne aux Etats-Unis des législations adoptées par certains états et assignant un délai de vie précis à la législation. Au terme de ce délai, les effets de la loi doivent faire l'objet d'une évaluation. Un article de Luc ROUBAN dans la Revue française d'Administration Publique, (N° 29, janvier mars 1984, p 85-115) analyse la place de cette législation en rapport avec les mouvements de dérégulation.

de Programme d'Analyse de Système de Service (PASS), expérimentée dans l'Etat du Nebraska en 1969 par WOLFENBERGER puis développée au fil des années¹ constitue une exception dans le secteur social. Enfin, si certains programmes d'évaluation des politiques sociales dans le domaine de la dépendance des personnes âgées ou handicapés ont pu s'inspirer de travaux Nord-Américains, il n'en reste pas moins vrai que les fertilisations américaines au développement de l'évaluation des politiques sociales françaises se situent à un niveau mineur.

Lorsqu'un regard un peu plus attentif est porté sur la production américaine d'évaluations dans le secteur des politiques sociales, il ne peut manquer de susciter un double mouvement de séduction et de réticence chez les praticiens du social et chez les chercheurs. Antériorité, quantité et méthodes sont les attraits principaux de ces démarches évaluatives dont la "dureté" et l'absence d'enrobage théorique suscitent, en revanche, certaines craintes.

Il ne s'agit pas, dans le cadre de ce rapport destiné à présenter des travaux récents d'évaluation réalisés aux Etats-Unis, d'engager une histoire de l'évaluation -d'autres ont esquissé ce travail²- mais simplement de rappeler la réelle antériorité et surtout l'importance de l'expérience acquise par les américains depuis les années 1930 où la Works Progress Administration et la National Youth Administration financèrent des programmes de recherche destinés à servir de base à la construction de procédures évaluatives. Si l'on connaît peu de chose sur ces programmes d'évaluation mise en place au moment du New Deal, hormis le fait qu'ils aient existé, il faut relier cela à la fragilité du recueil statistique de l'époque. Le développement de l'évaluation sur une large échelle dans les politiques sociales, préparé dans le début des années 1960, se fera en réalité entre 1965 et 1975 et prendra la forme de programmes expérimentaux dont les principaux figurent dans le tableau ci-après.

¹WOLFENBERGER W., THOMAS S., Passing (Programm Analysis of Service Systems Implantation and normalization Goals) : Normalization criteria and ratings manual, Toronto, National Institute on Mental Retardation, 1983.

² voir, sur ce point, la première partie de l'ouvrage d'Eric MONNIER déjà cité.

Tableau n°1
LES GRANDS PROGRAMMES EXPERIMENTAUX

Programme	Durée	Rapport
New Jersey Income Maintenance Experiment	1968-1972	1974
Rural Negative Income Tax Experiment	1970-1972	1976
Gary Income Maintenance Experiment	1971-1974	1976
Seattle-Denver Income Maintenance	1971-1978	1983
Education Performance Contracting	1970 -1971	1972
National Health Insurance	1974 -1981	n.c.
Housing Allowances	1973 -1977	1980
National Supported Work project	1975 -1979	1980
Employment Opportunity Pilot project	1979 -1981	1982
Electricity Time-of-Use Pricing	1975 - 199?	n.c.

Source : GREENBERG D., ROBINS P.¹

En matière d'évaluation, ces programmes ont suscité de grandes espérances mais également de nombreuses critiques quant à la possibilité de s'appuyer sur leurs résultats pour pouvoir de manière rationnelle construire des politiques sociales les meilleures pour l'ensemble de la population et de la nation. Si ces critiques ont été connues en France, nous avons peut être trop eu tendance, de ce côté de l'Atlantique, à jeter l'évaluation avec l'eau des programmes expérimentaux en oubliant que ces programmes d'évaluation ont permis aux chercheurs de développer une culture de l'évaluation et surtout de construire des méthodes et des procédures. On ne peut manquer d'être surpris, à la lecture des revues consacrées en

¹ GREENBERG D., ROBINS P., The Changing Role of Social Experiments in Policy Analysis, In Evaluation studies Review Annual, Vol 10. Sage publications, 1985.

tout ou en partie à l'évaluation¹ de la diversité des évaluations engagées et surtout des débats autour des méthodes et des résultats.

Autant que l'antériorité, la quantité de travaux consacrés à l'évaluation et le nombre de personnes engagées dans ce travail ne manque pas d'étonner le praticien français. Les publications consacrées à l'évaluation sont nombreuses dans le champ du secteur social. En plus d'Evaluation Review, et d'Evaluation and the Health, d'autres revues telles Administration in Social Work, Journal of Human resources, Evaluation News et d'autres encore contribuent à la diffusion des savoirs et des méthodes et à la structuration d'un milieu professionnel. A côté de l'American Evaluation Association dont la majorité des membres appartiennent aux secteurs de l'éducation, de la psychologie et des spécialistes des méthodes statistiques, plusieurs associations rassemblent les professionnels de l'évaluation en fonction de leur champ d'intérêt. Policy Studies Organization et l'Association for Public Policy Analysis and Management témoignent de la diversité de ce mouvement associatif. Chaque année, depuis 15 ans, la publication aux éditions SAGE² d'un lourd volume sous le titre "Evaluation Studies review Annual" propose un panorama assez vaste de l'état des pratiques évaluatives aux USA.

Antériorité, quantité, la production nord-américaine, si l'on englobe le Canada et les USA dans le même mouvement, séduit également par la rigueur des méthodes employées. Le recours le plus systématiquement possible à la méthode expérimentale, le souci de quantifier les informations, y compris lorsqu'il s'agit de données qualitatives, l'appel constant aux méthodes statistiques pour le traitement des données, les appareils critiques qui entourent la publication des résultats, confèrent à ces travaux une réelle qualité.

L'attrait que l'on peut ressentir à la lecture de ces travaux ne s'en accompagne pas moins de quelques craintes qui, bien qu'elles soient souvent exprimées dans la littérature française traitant de l'évaluation aux USA, méritent notre attention. Ces critiques ont été reprises dans le rapport de Patrick VIVERET et s'articulent autour de trois idées³. Tout d'abord, l'évaluation suppose l'idée de la rationalité des objectifs, or comme chacun le sait, l'objectif affiché ne coïncide que rarement avec l'objectif réel. Ensuite, on note toujours la difficulté à construire des outils permettant de quantifier le résultat d'actions qui peuvent

¹ nous pensons ici à une revue comme Administration in Social Work du groupe The Haworth Press, Inc., revue consacrée au management des politiques sociales et dans laquelle on trouve de très nombreux articles sur le thème de l'évaluation.

² Sage Publications, 28 Banner Street, London EC1Y8QE, fax 01 253 5206

³ VIVERET P., rapport cité, p 22.

porter sur une amélioration du mieux vivre des personnes. Enfin, rien ne permet d'affirmer la force et la réalité des liaisons entre les causes et les effets, le monde du social s'accommodant mal du principe "toutes choses égales, par ailleurs".

Ces critiques sont probablement exactes, et les auteurs américains semblent être les premiers à les formuler¹ en notant régulièrement :

- la taille trop faible des échantillons,
- les biais liés aux caractéristiques personnelles des participants aux expérimentations,
- la durée trop courte des expérimentations,
- la sous estimation des effets à long terme,
- l'attrition différenciée des participants à l'évaluation,
- la présence d'effets de type Hawthorne,
- les décalages entre les résultats sur les dispositifs expérimentaux et les comportements de la population.

Au delà de ces critiques qui portent davantage sur la méthode expérimentale que sur les procédures évaluatives elles-mêmes, il est certain que les dispositifs évaluatifs américains se focalisent davantage sur les effets économiques et comportementaux et prennent moins bien en compte des effets portant, par exemple, sur la transformation des systèmes de valeurs.

Les pratiques évaluatives se sont développées dans d'autres pays que ceux cités jusqu'alors et ce développement a pu être étudié². Toutefois, la littérature spécialisée dans l'évaluation des politiques sociales reste dans sa quasi-totalité d'origine anglo-saxonne. Le retard français dans la production de littérature évaluative n'est qu'une conséquence de la très récente introduction de ces méthodes dans notre pays. Cette situation encore peu établie des techniques évaluatives en France appelle en premier lieu une explication et surtout une réflexion sur les moyens à mettre en oeuvre pour y remédier : le rapport VIVERET fournit les réponses à ces interrogations. Il nous a semblé que "le bon usage du retard français" qui

¹ voir par exemple l'article de Robert H. HAVEMAN, Social Experimentation and Social Experimentation, in Evaluation Studies, Review Annual, vol. 12, 1987, p 608 - 627.

² en particulier dans l'article de Peter WAGNER et Hellmut WOLLMANN, "les fluctuations de la recherche en évaluation : sont-elles liées aux changements de régime ?" in Revue Internationale des Sciences Sociales, N° 108, juin 1986, p 221-235.

y est évoqué trouverait un terrain d'application particulièrement riche dans cette littérature américaine de l'évaluation.

Nous avons choisi de privilégier les travaux publiés dans "Evaluation Review" en les complétant par quelques articles tirés de "Evaluation and the Health Professions". D'autres sources existent mais ne font souvent que reprendre des travaux déjà publiés, c'est le cas, par exemple, pour "Evaluation Studies, review annual". Quant à la revue de l'American Evaluation Association, Evaluation Practice, son intérêt est surtout d'ordre informatif sur la vie de l'association. Les documents disponibles regroupent des articles de réflexion ou des descriptions de pratiques évaluatives. Si les travaux de réflexion générale présentent les aspects directement politiques, méthodologiques ou relevant du domaine de l'ingénierie des enquêtes, les articles descriptifs sont essentiellement des compte-rendus d'expérience. Ce deuxième volet nous a semblé convenir plus particulièrement à l'analyse, puisqu'il peut fournir des indications précieuses sur la façon même de mener une action évaluative. De plus, nous avons éliminé l'ensemble des monographies qui ne décrivaient pas simultanément le programme sur lequel portait l'évaluation et la procédure évaluative. Ainsi 35 articles réalisés entre 1983 et 1990 ont pu être retenus.

Une fois la direction de recherche précisée, il a fallu concevoir une grille de recueil d'information qui contienne tous les éléments constitutifs de la description d'une action d'évaluation. La concision et l'exhaustivité d'une telle grille sont nécessaires à toute analyse ultérieure. Il convient de disposer de tous les paramètres susceptibles d'entrer en jeu pour décrire une évaluation. La grille s'articule en fait sur une double description : description dans un premier temps du programme évalué (programme social, de politique publique...), puis dans un deuxième temps description de l'évaluation en elle-même.

La première partie de ce rapport décrit les programmes évalués et les procédures mises en oeuvre pour leur évaluation. La seconde partie analyse, de manière thématique, les principaux axes de structuration des pratiques évaluatives décrites dans les revues. Les fiches résumant les articles analysés et les grilles de lecture des articles étudiés figurent dans la partie annexée à cette étude. Chaque article est indexé de 1 à 35 et ce numéro permet de confronter le commentaire à la fiche.

PREMIERE PARTIE

**EVALUATION SUR PROGRAMME
ET PROGRAMMES D'EVALUATION**

I - DESCRIPTION DES PROGRAMMES EVALUES

La définition précise d'un programme subissant une évaluation comprend sept dimensions. Il est cependant rare dans un article sélectionné de réunir toutes ces informations. En fait, les items les plus généralement remplis, et aussi les plus fondamentaux pour la description du programme, concernent la population visée, les objectifs de la politique, et enfin le contenu du programme.

**Les 7 points-clés de la définition
du programme évalué**

- nom du programme,
- population concernée,
- objectifs du programme,
- lieu, date et durée du programme
- contenu,
- maître d'oeuvre,
- moyens affectés.

Les compte-rendus des expériences étudiées recouvrent une importante quantité de domaines. Par conséquent, les rubriques de caractérisation des expériences associées sont très étendues.

1-1. Le nom du programme

La dénomination du programme, outre qu'elle n'existe pas toujours, n'est pas d'un grand apport à l'examen des procédures mises en place. Sur 22 dénominations, 12 intègrent la notion de plan, de programme ou de projet dans leur intitulé, cinq seulement mentionnent le nom du pays ou de l'Etat promoteur de l'expérience. La population à laquelle s'applique le programme, "handicapped children", "dislocated workers", etc..., n'est pas

systématiquement nommée. Dans la majorité des cas, la dénomination résume plutôt la nature du service proposé en insistant, à la manière d'un slogan, sur le rapprochement de deux termes, parfois redoublés, incluant souvent un nom et un qualificatif : "Better Eating for better Health" ou "Juvenile Education Program" constituent de bons exemples de cette structure sémantique qui, par ailleurs, est souvent accompagnée de son sigle.

Cette logique d'appellation des programmes se retrouve depuis quelques années également dans la langue française de manière systématique et le RMI en est un bon exemple, même si l'existence d'un complément nominal vient enrichir la structure de base : nom + qualificatif.

La liste des noms de programmes est la suivante :

- BEEP (Brookline Education Project) [1]
- Yakima Drinking and Driving Project [2]
- Fundamental Schools [6]
- Better Eating for Better Health [7]
- Delinquency Prevention Research and Development Project [8]
- JEP (Juvenile Education Program) [9]
- VHF (Voluntary Heating Fuel Budget Plan) [10]
- Retraining Delaware's Dislocated Workers [14]
- Program for Voluntary Surgical Contraception in Guatemala [16]
- CETA (Comprehensive Employment and Training Act) [17]
- Forestry Incentives Program [18]
- Program Development Evaluation [19]
- Hood River Conservation Program [20]
- OBRA (Omnibus Reconciliation Act) [21]
- Methadone Maintenance [22]
- Lifeline Electric rates [23]
- AIDS : The modern Plague [24]
- KO'ach [25]
- Primary Health Care Services Program [29]
- Adolescent Family Life Act [30]
- CFRP (Child and Family Resource Program) [31]
- Community Posthospital Follow-up Services [32]
- Program for Handicapped Children [34]

Comme il a été précisé plus avant, cette rubrique est d'un pouvoir explicatif relativement mineur dans l'étude entreprise. Si elle montre les sujets d'intérêt des évaluateurs américains, elle ne présage en rien des scénarios d'évaluation : choix d'une méthode, sélection des indicateurs en fonction de la population et du contenu du programme étudié.

1-2. La population concernée

Si nous ne disposons pas des noms de tous les programmes ayant subi une évaluation, en revanche, il a été possible d'étudier les populations concernées par les dits programmes. Des regroupements s'effectuent assez rapidement au sein de ces populations beaucoup moins disparates qu'il n'y paraît à première vue.

S'il arrive que le programme concerne sans restriction tous les individus volontaires (volontaires dans le grand public [7]), quelquefois, un seul critère simple suffit à décrire la population impliquée dans un programme.

On relève ainsi le critère de l'âge qui permet de sélectionner les participants à un programme : nourrissons [28], enfants en bas-âge [1], adolescentes [30], personnes âgées [15].

Les bénéficiaires potentiels des programmes sont définis par l'intermédiaire de leur degré de scolarisation (étroitement dépendant de l'âge des individus) : étudiants [3]-[24]-[35], élèves de primaire et secondaire [6], élèves du secondaire [8], enfants du primaire [11].

Les populations-cibles de ces programmes sont également déterminées à partir de considérations sanitaires (anciens combattants [25], patients sortis d'une unité psychiatrique [32], anciens patients psychiatriques [33], enfants handicapés [34]) ou de considérations de santé publique (volontaires en vue de stérilisation [16], toxicomanes de plus de 18 ans [22]), voire de respect de l'ordre public (conducteurs en état d'ébriété [2], jeunes délinquants [9], écoles à problèmes (délinquance...) [19]).

Par ailleurs, et dans un autre domaine d'intérêt, la littérature aborde assez régulièrement des sujets se rapportant à la production de matières premières et d'énergie, tant du point de vue du consommateur (réalisation de programmes d'économie d'énergie...)

que du producteur (contrôle de l'évolution des prix...) : fournisseurs d'eau [13], bois appartenant à des intérêts privés [18], habitations avec chauffage électrique [20], consommateurs individuels d'électricité [23].

Dans ces programmes traitant de questions sociales, les critères les plus couramment répandus sont évidemment ceux faisant référence à la pauvreté. Ceux-ci s'appuient sur les mesures des revenus des individus. On retrouve le critère de la faiblesse des revenus ou des situations économiques particulièrement défavorables dans les programmes concernant des entreprises appartenant à des personnes défavorisées [4], des jeunes défavorisés [12], des ouvriers sans emploi [14], des personnes économiquement défavorisées [17], les habitants d'un bidonville [26]).

Mais la spécificité des programmes rencontrés est telle que les populations concernées réunissent plusieurs conditions, dont, souvent, une ayant trait à une dimension de la pauvreté.

Ainsi, on trouve les clientes des services de planning familial, confrontées simultanément à des problèmes de pauvreté, de santé et de contraception (femmes avec bas revenus [5] et [27]), les personnes défavorisées chargées de famille, devant concilier des situations économiquement difficiles et l'éducation de leurs jeunes enfants (personnes défavorisées avec un enfant de 0 à 8 ans [31], et bénéficiaires plus ou moins directs de l'AFDC (Aid to Families with Dependent Children) : volontaires participant à l'AFDC [10], personnes aux revenus légèrement supérieurs à l'AFDC [29], bénéficiaires de l'AFDC [21].

1-3. Les objectifs du programme

De nombreux programmes soumis à des actions d'évaluation concernent la santé, et plus particulièrement la délivrance de soins directs (cure de désintoxication [22], traitement des troubles psychologiques [25], amélioration de l'état de santé [26], soins médicaux [29], traitement des conséquences psychologiques après la grossesse chez des adolescentes [30], amélioration de l'état de santé et baisse du taux de réhospitalisation [32], diagnostic et traitement chez des enfants handicapés [34]). Ce domaine particulier de la santé a développé une science propre, l'épidémiologie, qui étudie les rapports entre les maladies et les facteurs susceptibles d'exercer une influence sur leurs fréquences, leurs distributions ou leurs évolutions. Les actions d'évaluation sur des programmes de santé dont nous traitons ici s'inspirent très largement de ces études épidémiologiques.

Des actions d'évaluation se portent également sur des programmes relevant du planning familial [27], de la maîtrise de la contraception [5] et de la diminution du nombre de naissances [16].

Toujours dans le domaine de la santé, des programmes s'attachent à connaître et à modifier éventuellement les attitudes et comportements face à des questions de santé publique : évolution de l'attitude face à l'alcool [2], connaissance de l'hygiène alimentaire [7], campagne d'information sur le SIDA [24], sensibilisation de l'opinion à des questions de santé publique [35].

Enfin, certains programmes peuvent être rapprochés de ces programmes sanitaires par la clientèle qu'ils concernent (anciens patients psychiatriques à contacter [33]) ou par les méthodes qu'ils testent (mesure de l'exposition au tabac des nourrissons [28], offre de services et d'activités à des personnes âgées [15]).

Un autre domaine de la vie sociale qui suscite des actions d'évaluation est celui de l'enseignement, de l'éducation et de la formation. Les nouveautés et améliorations portées aux programmes scolaires ou aux types d'enseignement sont régulièrement évaluées. Les objectifs affichés par de tels programmes sont : la formation [3], le développement de systèmes d'aide aux familles pauvres avec enfants [31], une augmentation du taux de réussite scolaire [1], une amélioration de l'enseignement [6], une amélioration du niveau en mathématiques [11], un développement de l'intérêt pour l'école [8]. Ce dernier programme a un double but : outre l'inculcation du goût pour l'école, il cherche à diminuer la délinquance. En marge des programmes d'enseignement pur, il existe ainsi des programmes d'aide à la réinsertion de délinquants ou de prévention de la délinquance : diminution de la délinquance [8], réhabilitation de jeunes délinquants [9], lutte contre la délinquance [19].

Le secteur de la formation et des stages est également représenté parmi les programmes évalués. Les stages décrits ici ont fréquemment rapport à la recherche d'emploi. On note les objectifs suivants : formation [3], aide à la recherche d'emploi [12], indépendance économique par l'emploi [17].

Il reste, en fait, assez peu de programmes ayant des connotations économiques fortes. Deux programmes traitent du niveau de la production dans des secteurs économiques particuliers : augmentation de la part des commandes de l'Etat auprès des entreprises défavorisées [4], augmentation de la production de bois [18]. Ce petit nombre de programmes ayant trait à l'économie est étonnant alors qu'on avait remarqué dans les

populations étudiées une prédominance de populations ciblées par des indicateurs de pauvreté. L'objectif direct de la lutte contre la pauvreté par l'apport de prestations ou d'aides n'apparaît pratiquement pas, si ce n'est dans l'amélioration d'un programme existant déjà (l'AFDC en l'occurrence : Aids to Family with Dependant Children). Ainsi l'objectif d'un programme qui est l'augmentation des ressources des chômeurs [14] est acquis lors de stages de formation aux techniques de recherche d'emploi ou de reconversion. Plus que la distribution directe d'argent, les programmes évalués ont plutôt pour objectif de régler des situations d'endettement ou de les prévenir : traitement des arriérés de chauffage [10], économie d'énergie [20], économie d'énergie et aide aux défavorisés [23].

1-4. Durée et localisation des programmes

Evaluation Review est résolument américanophile. Sur 35 programmes évalués, 32 concernent des programmes mis en oeuvre aux Etats-Unis. 12 s'appliquent à l'ensemble des USA alors que 20 sont pris en charge par les différents Etats. Si la Californie est citée à six reprises, le Texas et le Michigan apparaissent pour 2 programmes et les autres Etats ne sont mentionnés qu'une seule fois. Le Brésil, le Guatemala et Israël sont les trois pays cités, extérieurs aux Etats-Unis.

Tous les programmes évalués ne comportent pas de mention de leur date de mise en place ou de moyens d'identifier la période de lancement. Pour les 29 programmes comportant cette indication, la répartition en fonction des années de lancement des programmes est la suivante :

1972 - 1975	4 programmes
1976 - 1980	10 programmes
1981 - 1985	11 programmes
1986 - 1988	4 programmes

La baisse du nombre de programmes sur les années 1986-1988, outre la durée plus courte retenue, est liée au décalage entre la date de réalisation de l'évaluation et la publication dans Evaluation Review et ne peut être interprétée comme un indicateur de récession des pratiques évaluatives aux USA.

Sur les 35 programmes analysés, 19 se situent dans une durée non définie par avance, 6 sont prévus pour durer entre une et cinq années et 10 concernent des actions ponctuelles. Cette indétermination dans le temps des programmes relativise la notion même de programme, conçue le plus souvent comme un ensemble d'interventions à durée déterminée, et rapproche celle-ci des formes traditionnelles prises par les politiques sociales, excluant toute idée de limitation dans le temps.

Les actions ponctuelles portent le plus souvent sur des démarches de formation sur un public restreint [3] ou sur des actions d'information s'appliquant alors à des publics plus élargis.

1-5. Le contenu des programmes

Le contenu du programme désigne l'ensemble des mesures et des actions mises en place afin de réaliser l'objectif annoncé. Il existe, a priori, une forte adéquation entre la définition des objectifs et la description qui est faite des interventions mises en oeuvre. Ainsi, les programmes à vocation sanitaire ont des contenus médicaux, de soins ou de conseils (traitement de la toxicomanie [22], soins gratuits [26], planning familial [5] et [27], services de soins et de prévention [29], services de soins, d'information et de soutien psychologique...).

Les programmes d'enseignement ou de formation développent des contenus pédagogiques ou des stages spécifiques : stage de formation [3], enseignement scolaire particulier [6], techniques pédagogiques [8], Enseignement Assisté par Ordinateur [11], méthodes d'enseignement et de maintien de la discipline [19]...

Les programmes plus économiques se distinguent par des contenus essentiellement monétaires et tarifaires, même si les conseils y ont leur place : négociation de la valeur des contrats entreprises/Etat [4], système de paiement des notes de chauffage [10], subvention au reboisement [18], conseils et équipements en matériel de chauffage [20], système d'adéquation des paiements des notes d'électricité [23].

Dans leur ensemble, les moyens mis en oeuvre ne se caractérisent pas par un très fort degré d'innovation. Il s'agit le plus souvent de proposer à une population en situation de pauvreté un ensemble de services se situant dans les secteurs de la formation ou de la santé. En matière de formation, les interventions mises en oeuvre proposent la réalisation

d'apprentissages cognitifs [35] ou mettent l'accent sur des modalités particulières d'apprentissages [11] visant, le plus souvent, à modifier des comportements [3]. Dans le domaine de la santé, il s'agit parfois de tester de nouveaux produits, mais le plus souvent de proposer des services juxtaposant un aspect social et un aspect médical [26]. Nous retrouvons là une dimension médico-sociale bien connue des travailleurs sociaux français des années 1960.

Enfin, différents programmes ne sont, en fait, que des changements de législation ou des modifications dans les règlements. Cela est particulièrement vrai pour les transformations des critères d'attribution de ressource [21] sous forme de prestations ou de contrats d'aide [4].

1-6. Financement et maîtrise d'ouvrage

En matière de financement et de maîtrise d'ouvrage, les articles cités manquent souvent de précision et les programmes décrits ne s'attachent pas nécessairement à décrire les aspects institutionnels. Lorsqu'une mention de maître d'oeuvre est indiquée, il s'agit le plus souvent d'un département de l'Etat commanditaire du programme : "Texas Department of Human Resources" [5], "Texas Department of Health" [29], etc. Les indications financières des montants des programmes sont trop parcellaires pour pouvoir être indiquées ici.

II - LA DESCRIPTION DE L'ÉVALUATION.

L'évaluateur, face aux effets d'un programme ou d'une politique à analyser, choisit les techniques d'évaluation les plus appropriées. Ces techniques sont sans aucun doute pour une grande part fonction des spécificités du programme (évoquées dans la première partie), et l'objet de la présente étude est de déterminer les paramètres de cette fonction de décision.

Une fois le but de l'évaluation fixé, l'évaluateur retient des indicateurs et c'est l'évolution de ces indicateurs, analysée par différentes méthodes statistiques, qui permet de porter le jugement sur le programme concerné par l'évaluation.

De même que le programme qu'elle concerne, l'action d'évaluation s'organise autour de sept points :

**Les 7 points-clés de la définition
de l'action d'évaluation**

- but de l'évaluation
- indicateurs retenus
- localisation, date, durée de l'évaluation
- évaluateur(s)
- choix de la population
- méthode de collecte de l'information
- outils statistiques utilisés, calculs effectués.

Pour tous les articles étudiés, les rubriques mentionnées ci-dessus n'ont pu être clairement informées. Témoignant d'une action d'évaluation, les auteurs s'attachent à en développer l'aspect qui leur paraît le plus original sans marquer toujours un réel systématisme dans la description de la démarche évaluative. Toutefois, deux rubriques apparaissent plus régulièrement, la première explicite le but de l'évaluation et la seconde précise les indicateurs retenus.

2-1. Le but de l'évaluation

Sur 35 programmes évalués, seuls 3 articles ne précisent pas les buts assignés à l'évaluation, les autres fixent à la démarche évaluative, de manière plus ou moins précise, un ou plusieurs objectifs. Sur ces 32 programmes, pour lesquels nous disposons d'information, deux fixent à l'évaluation un objectif portant sur la conduite du programme lui-même ainsi que sur les conditions de sa mise en oeuvre. Pour l'un, concernant l'évaluation d'une courte incarcération en cas de conduite en ébriété [2] cette évaluation centrée sur le programme est concomitante avec une étude d'impact. Pour l'autre, visant à renforcer la participation des élèves en classe dans une perspective de réduction de la délinquance juvénile [18], l'évaluation de la mise en oeuvre du programme constitue une démarche préalable à l'évaluation des résultats.

Trois programmes d'évaluation développent une démarche que l'on pourrait qualifier de méta-évaluative dans la mesure où ces programmes cherchent essentiellement à construire des outils d'évaluation [7, 25] ou à évaluer la qualité des programmes d'évaluation [30] Mais là encore, sur ces trois programmes, deux associent une méta-évaluation et une évaluation de premier degré portant sur la connaissance des effets de l'action engagée.

En dehors de ces cinq programmes qui, partiellement ou en totalité, échappent à une logique de mesure des effets des interventions, l'évaluation, dans les articles d'Evaluation Review, se résume toujours en une tentative de décrire et de mesurer les impacts des actions engagées. Les différences résultent dans la plupart des cas de la classification de ces effets. En effet, sur les 26 programmes restants, 7 mettent l'accent sur une recherche de la transformation des comportements des personnes qui ont subi le programme. Ce comportement peut être caractéristique d'une manière de trouver et de garder un emploi [12], de répondre à un questionnaire [33], d'organiser son travail [1] ou d'adopter un comportement délinquant [9].

6 programmes, tout en restant dans la même logique de mesure des effets, restreignent le champ de ces derniers aux seuls effets économiques, voire financiers et s'apparentent, de ce fait, plus à une démarche coûts-avantages. Il arrive enfin que l'approche économique intègre un souci d'équité sociale. Ainsi, l'Etat de Californie a introduit en 1975 une législation autorisant une tarification différentielle de la fourniture d'électricité dans l'objectif d'apporter une aide aux personnes à bas revenus [23]. Un

système identique concernant les frais de chauffage des familles bénéficiant des aides pour enfant à charge [10] a fait l'objet d'une évaluation par les autorités du Michigan.

Un seul programme évaluatif intègre les trois éléments mentionnés ci-dessus : analyse des effets, étude coût-avantage et recherche de plus grande équité. Conduit par le U. S. General Accounting Office à la demande du House Committee on Ways and Means, ce programme ne fait pas moins qu'évaluer les effets de la réforme de l'AFDC [21].

Pour améliorer les performances des élèves, est-il préférable de développer des techniques d'enseignement assisté par ordinateur ou de faire appel à des formes d'entraide à l'intérieur de la classe ou bien encore de réduire la taille des classes ?

Avec l'objectif de rendre plus autonomes des personnes en situation de pauvreté, faut-il engager des programmes de formation professionnelle ou renforcer les dispositifs monétaires d'aide sociale ? Ces questions, et d'autres encore de même nature, sont au coeur de près d'un programme évaluatif sur quatre et confortent l'image d'une méthodologie de l'évaluation voisine de l'expérimentation au sens où Claude BERNARD a pu développer ce concept.

Enfin, bien que restant dans une logique de connaissance des effets de l'intervention, deux programmes d'évaluation font explicitement référence à l'existence de référents. Pour ces deux programmes, l'intervention doit conduire à adopter des comportements ou des politiques les plus proches possibles de modèles définis a priori en fonction d'un système théorique [3].

2-2. Les indicateurs retenus

Alors que seuls quelques programmes à évaluer mettent en avant des buts économiques, la définition des objectifs de l'évaluation contribue à mieux faire apparaître les logiques économiques sous-jacentes. Mieux encore, les indicateurs servant de support à l'évaluation utilisent très fréquemment des agrégats économiques pour mesurer les effets des programmes engagés. Sur 34 évaluations pour lesquelles les indicateurs servant de base à l'évaluation sont explicitement mentionnés, on relève 9 évaluations dont les indicateurs sont strictement à caractère économique. Ainsi, dans le cas d'un programme proposant des services de planning familial à des femmes bénéficiant déjà de l'AFDC dans l'Etat du Texas [5], les évaluateurs utilisent un taux de rendement défini comme étant l'ensemble des

économies réalisées par personne grâce à la prévention de naissances non désirées (en allocations familiales, bons alimentaires...) divisées par le montant, la même années des dépenses, par personne, de planing familial. A côté de cet indicateur, les évaluateurs mentionnent une estimation du nombre de fausses-couches et d'avortements évités, sans que l'estimation de leur coût financier soit réalisée. De manière encore plus classique, d'autres évaluations utilisent des indicateurs de coût, de prix de revient, de dépenses réalisées : valeur des contrats passés par l'Etat [4], taux de rendement et économies réalisées, impact sur le coût des services sociaux [10], taux de rendement [11], variation du prix et concentration des dépenses [13], revenus et montants des assurances chômage [14], coût par opération, coût moyen et marginal, économies d'échelles [16], économies réalisées [17], taux de rendement interne et valeur actualisée [18], économies réalisées [20], étude de coût [23], taux de rendement [27], [32] et [29].

Pour 7 programmes, les indicateurs économiques se combinent à des indicateurs décrivant des faits ou des comportements de la population. Ainsi, un programme visant à évaluer l'impact de la remise d'un produit de substitution à des toxicomanes juxtapose plusieurs indicateurs parmi lesquels on note :

- le nombre des délits,
- les consommations de drogue,
- la participation du conjoint,
- l'évolution des sources de revenu [22].

De nombreux programmes sont évalués à partir d'une mesure des comportements et attitudes des individus concernés : le comportement de l'enfant avec les autres, son comportement social [1], l'attitude du public [2], la conformité à des comportements standard [3], l'opinion des élèves sur eux-mêmes et l'attitude des parents [6], le comportement des enseignants et le comportement des élèves [8], l'attitude des élèves [19], les attitudes et les comportements face au SIDA [24].

La mesure et le calcul de diverses quantités -autre que l'argent- permettent d'indiquer le degré de réalisation et d'efficacité d'un programme : nombre d'accidents dus à l'alcool [2], résultats des élèves [6], nombre de délits enregistrés [9], nombre de participants et taux de participation [15], nombre de participants et d'équipements installés [20], délits effectués et drogue consommée [22], résultat de tests [28], durée de survie [32], taux de réponse [33]. Dans leur majorité, ces indicateurs ne sont pas fortement structurés, ainsi, les indicateurs retenus pour mesurer les résultats d'un

programme scolaire seront les résultats des élèves en mathématiques et en lecture ; l'impact d'un enseignement précoce chez l'enfant sera évalué à partir d'une étude de l'organisation du travail et du comportement de l'enfant.

Les autres actions d'évaluation s'appuient sur des mesures non quantitatives mais qualitatives : organisation et efficacité du travail de l'enfant [1], connaissance par le public et application par la police et le système judiciaire [2], connaissances, convictions et habitudes alimentaires [7], caractérisation des participants [10], passé professionnel et emploi actuel [12], caractérisation des entrants/sortants de l'AFDC [21], adaptation sociale et affective [25], satisfaction [26], qualité des données [30].

Les indicateurs de satisfaction mesurée directement auprès des bénéficiaires du programme constituent un outil auxquels recourent fréquemment les différents évaluateurs.

2-3. Localisation, date et durée de l'évaluation

En matière d'évaluation, la question du temps revêt une importance suffisamment grande pour que chaque texte analysé précise les rapports entre les calendriers de l'évaluation et ceux des programmes à évaluer. 3 cas de figure se détachent nettement.

Dans le premier cas, l'évaluation prend place à un moment du déroulement d'un programme et consiste à mettre en oeuvre une démarche d'étude, le plus souvent rétrospective. Ainsi, dans le cas d'un programme visant à augmenter la proportion de réussites scolaires au moment de l'entrée à l'école primaire en fournissant différents services éducatifs aux parents [1], l'évaluation va se dérouler en utilisant une population d'enfants nés entre le printemps 1973 et l'automne 1974 au moment de leur fréquentation du jardin d'enfant -année avant l'intégration en école primaire-. Une autre évaluation, réalisée au Brésil à Paranoa, bidonville de 16 000 habitants se propose d'évaluer une année après la mise en place d'un centre médical, le degré de satisfaction des clients de ce centre médical [26].

La moitié des exemples d'évaluation rapportés dans Evaluation Review utilise ce même modèle. Dans ce cas de figure, l'évaluation et le programme d'intervention sociale sont fortement déconnectés et l'évaluation se fait le plus souvent en utilisant les données accumulées par les gestionnaires des programmes sociaux ou bien encore, en réalisant une photographie à un moment donné de l'état des personnes qui ont participé au programme.

Le second cas de figure articule davantage le programme et son évaluation. Il s'agit le plus souvent de programmes relativement courts, et concernant une population bien définie. L'exemple caractéristique de cette démarche est un programme destiné à des ouvriers ayant perdu leur emploi dans des secteurs touchés par le progrès technique et la concurrence internationale [14]. Ce programme proposait des stages de formation aux techniques de recherche d'emploi, un soutien individualisé et l'accès, le cas échéant, à des stages de reconversion. Ce programme durait 5 mois en 1983 et son impact a été mesuré 2, 5, 8 et 11 mois après le lancement.

11 programmes sur les 35 recensés utilisent cette forme d'évaluation pour laquelle l'analyse de l'impact d'une intervention sociale constitue le mécanisme dominant. Enfin, 7 programmes, proches dans leur conception des précédents recourent à un rapport différent entre le temps de programme et le temps de l'évaluation. Pour ces derniers, l'évaluation encadre le déroulement du programme. Ainsi, dans le cas d'un programme d'une durée de trois ans cherchant à atténuer les problèmes de violence et plus généralement de discipline dans des écoles secondaires situées dans des quartiers pauvres de Baltimore [19], la démarche d'évaluation se met en place par une collecte d'informations avant le lancement du programme, puis un an après et une dernière mesure est effectuée deux années après le démarrage du programme.

Cette utilisation du temps définit en fait trois approches de l'évaluation. La première est très proche d'une logique d'étude traditionnelle. Elle utilise des informations existantes ou crée ces informations par interrogation de personnes ayant bénéficié d'un programme. La mesure des degrés d'atteinte des objectifs occupe une place importante dans les résultats de cette forme d'évaluation. La seconde approche se définit mieux par son caractère d'étude d'impact, en ce sens qu'elle cherche à définir, en rapport plus ou moins direct avec les objectifs du programme, la diversité de ses effets sur les populations qui en bénéficient. Cette dimension étude d'impact nécessite un lien plus étroit, dans le temps, entre le programme et son évaluation. La troisième approche relève plus directement encore d'une définition stricte de l'évaluation en réalisant une comparaison des données recueillies avant et après le déroulement du programme.

2-4. Le choix des populations

Dans le choix des populations, un clivage dominant apparaît dans les articles analysés selon que l'évaluation fait appel ou non à l'existence de groupes-témoins. Sur 35 comptes-rendus, 18 se trouvent dans cette catégorie alors que les 17 autres utilisent le plus souvent une technique d'échantillonnage sans groupe-témoin.

Pour ces 17 programmes évaluatifs, réalisés sans groupe témoin, on distingue trois types de population. Le premier type est constitué par des organismes ou des services. Il peut prendre en compte la totalité des services relevant d'une même association [15], étudier quelques services en fonction des données disponibles [13], ou examiner un échantillon raisonné de services [29]. Le second groupe rassemble cinq évaluations réalisées à partir de l'examen de dossiers. Parmi celles-ci, se trouve, par exemple, l'évaluation d'un programme issu du Comprehensive Employment and Training Act (CETA) visant à rendre économiquement indépendantes des personnes défavorisées. Les personnes ayant bénéficié de ce programme devraient utiliser moins fréquemment les autres programmes d'aides sociales. Un échantillon de personnes économiquement défavorisées a été sélectionné dans la liste du "Continuous Longitudinal Manpower Survey" et dans la liste du "Panel Study of Income Dynamics" [17]. Un autre programme de même nature, cherchant à prévenir les naissances non désirées a été évalué à partir de l'examen des dossiers de 1 606 adolescentes et de 1 605 adultes sélectionnés aléatoirement parmi les dossiers de 78 services de planning familial au Texas [5]. Le troisième groupe relevant de la même logique rassemble des évaluations réalisées à partir de l'interrogation de personnes. Ces échantillons peuvent être très restreints et ne rechercher aucune représentativité [34], ou être constitués de manière aléatoire à partir de populations plus vastes [26].

Parmi les études évaluatives conduites en utilisant des groupes-témoins, on peut distinguer des formes plus ou moins strictes de recours à la méthode expérimentale. Ainsi, 5 programmes évaluatifs affectent, de manière aléatoire, les populations à la fois dans le groupe témoin et dans le groupe bénéficiant du programme. Dans ce programme, tous les jeunes délinquants issus de trois comtés du Kansas et destinés à participer au Juvenile Education Program sont placés, au hasard, dans le groupe d'étude ou dans le groupe témoin, ce groupe témoin étant la file d'attente pour la prochaine session du JEP [9].

D'autres études évaluatives sont réalisées à partir d'échantillons appareillés. Un programme cherchant à préparer des jeunes de milieux défavorisés à la recherche d'un emploi utilise un échantillon de 419 participants extraits de manière aléatoire d'une

population de 39 000 personnes. Le groupe-témoin est constitué par 365 sujets présentant des caractéristiques aussi proches que possible de celles des participants [12]. 5 programmes évaluatifs utilisent cette technique dans les numéros d'Evaluation Review analysés.

Enfin, 6 programmes évaluatifs présentent des populations expérimentales et des populations témoins sélectionnées en fonction de leur disponibilité ou de leur place dans les dispositifs sociaux.

2-5. Le choix des méthodes de collecte de l'information

Près du tiers des études évaluatives réalisées s'appuie sur des exploitations secondaires de données. Il peut s'agir d'information contenues dans des fichiers assurant la gestion de services, c'est le cas, par exemple, pour l'exploitation des données consultées dans le fichier de la "Delaware Division of Unemployment Insurance" -assurances chômage de l'Etat du Delaware- [14]. Il arrive parfois que ces données soient tirées non de fichiers constitués mais d'une exploitation des archives des différents services [30].

Un quart des études évaluatives est réalisé par entretiens, y compris lorsque le nombre de ces entretiens avoisine plusieurs centaines [26]. Il arrive parfois que les entretiens soient réalisés par téléphone [35]. Le recours au questionnaire s'avère moins fréquent puisqu'on ne relève que 7 études évaluatives faisant appel à ce type de recueil de l'information.

Deux études évaluatives combinent des méthodes d'analyse de dossiers et des formes autres de recueil de l'information, questionnaire ou entretien.

Enfin, quatre études évaluatives utilisent des techniques d'observation. Ainsi, dans le cadre d'un programme de formation de conseillers dans divers domaines sociaux, des séances de conseils sont filmées et les films sont visionnés par des conseillers expérimentés [3]. Un autre programme évaluatif cherchant à mesurer les effets d'un enseignement précoce sur les enfants et leurs parents conduit à observer des enfants dans les domaines suivants : organisation du travail scolaire, efficacité de l'enfant, comportement avec autrui [8].

2-6. Le choix des méthodes de traitement de l'information.

Cinq études évaluatives ne précisent pas les méthodes de traitement des données utilisées par les évaluateurs. 11 études s'essaient à construire des modèles probabilistes ou en termes de comparaisons des coûts et des avantages.

Les dix-neuf études restantes se répartissent de manière équivalente entre les méthodes suivantes :

- calcul de coefficients de corrélation,
- régressions simples ou multiples,
- analyse de variance.

Les méthodologies employées s'appuient donc principalement sur les techniques de mesures de corrélations. L'analyse des données semble peu utilisée. Il est rare qu'une étude mesure simultanément un grand nombre de critères sur une population importante.

III - LA MESURE DES EFFETS.

Loin du débat sur son caractère démocratique, ouvert et partagé, les pratiques évaluatives dont il est rendu compte dans ces revues spécialisées ont pour objectif essentiel de connaître et, si possible, de mesurer des effets. Pour ces évaluateurs, l'évaluation se définit, avant tout, comme une technique d'étude visant à connaître les changements d'états de personnes, de groupes ou de milieux sous l'influence d'interventions externes. Une place importante est, par conséquent, occupée dans chacun de ces articles pour décrire les effets générés par les actions conduites auprès des populations.

Un seul programme évaluatif [34] ne fournit pas d'information sur la nature et la quantité des effets liés à l'intervention, mais pour tous les autres, des effets sont repérés, décrits et mesurés. Ces effets sont, pour 9 cas sur 10, considérés comme satisfaisants par les auteurs des articles. Il est probable que dans le processus de sélection des articles cet aspect est valorisé au détriment d'articles plus satisfaisants sur le plan des méthodes mais moins riches quant à la production des résultats. Les effets sont présentés le plus souvent en termes de taux ou, plus simplement, de pourcentage. Ainsi, l'évaluation de la satisfaction des usagers de services médicaux montre que "95 % des personnes interrogées voulaient un centre médical dans leur commune, que celui-ci a satisfait l'attente de 34 % seulement et qu'enfin, 16 % des patients connaissaient les horaires du centre" [26].

Parmi les effets les plus fréquemment signalés par les auteurs, il faut noter l'importance accordée à l'atteinte globale des objectifs fixés au programme, que ces objectifs concernent le ciblage de la population ou le degré de transformation des caractéristiques ou des comportements de cette population sous l'effet de l'intervention. De nombreux travaux analysent également les effets différentiels des interventions en fonction des caractéristiques de la population ou, de manière plus précise, tentent de reconstruire des modèles explicatifs de l'effet de l'intervention [7]. Les comparaisons entre les différentes politiques ou actions testées dans le cadre de programmes expérimentaux se concluent, le plus souvent, par une analyse de type coûts-avantages [11].

Parmi les effets non attendus relevés au cours des procédures évaluatives, on remarque un effet de centrage autour de la norme [3], une hausse de la délinquance chez les petits délinquants confrontés à des images fortes de l'incarcération, alors que l'objectif du

programme était de créer un effet inverse [9]. Sur le plan méthodologique certains auteurs insistent sur les écarts dans la perception même des effets montrant que si un questionnaire avant/après n'est pas susceptible de décrire des évolutions, celles-ci peuvent être indiquées par le sujet dans un questionnaire de type rétrospectif. Une étude montre que la qualité des évaluations dépend fortement des financements qui leur sont consacrés [30]. Toutes les études qui abordent ce problème montrent un affaiblissement, avec le temps, de l'intensité des effets.

SECONDE PARTIE

L'ECHIQUIER EVALUATIF

Peu nombreuses, fragmentaires et partiales, les expériences évaluatives décrites dans les pages précédentes n'en dressent pas moins le portrait d'une figure possible de l'évaluation. Janus non dénuée d'ambiguïté, cette figure compose une démarche évaluative à l'aide de deux modèles. Le premier fait référence à l'expérimentation et mesure, dans la similitude des temps, la distorsion différentielle des états des personnes, des groupes ou des milieux auxquels s'appliquent, ou non, des interventions, politiques sociales ou économiques. Le second modèle utilise le temps comme opérateur principal pour rendre compte, en une démarche historiciste, des mouvements des personnes, des groupes ou des milieux. L'écart est faible entre cette seconde démarche et une pratique d'étude, plus traditionnelle, dans laquelle il s'agit, par exemple, de décrire et de comprendre l'impact d'une mesure quelconque, politique publique ou intervention privée, sur une institution ou un ensemble de personnes¹.

La distinction entre ces deux modèles ne relève pas uniquement d'une réflexion à caractère méthodologique ; elle porte en germe un débat sur la nature même de l'évaluation.

Inscrire l'évaluation dans une démarche expérimentale conduit à renforcer la dimension positiviste de celle-ci et à la détacher des contingences décisionnelles ou politiques. Ce résultat s'impose à la sphère décisionnelle car il est obtenu avec une méthode qui ne peut être remise en question. De ce point de vue, la critique que l'on peut adresser à cette approche n'est pas celle que l'on tient habituellement à l'égard du scientisme. Ce dossier valorise le résultat, alors que l'évaluation expérimentale met plutôt l'accent sur la méthode d'obtention du résultat. La filiation, pour rester cohérent avec l'origine américaine des articles, est à rechercher dans l'essor du pragmatisme comme courant philosophique aux USA tout au début du siècle² pour qui les conditions de l'élaboration de la vérité l'emportent sur la réflexion sur le sens.

¹ à titre d'exemple, mais il en existerait des centaines d'autres, une étude comme celle de Margaret GIBELMAN, National Voluntary Health Agencies in an Era of Change : Experiences and Adaptations, in Administration in Social Work, Vol 14, N° 3, 1990, pp 17-32, ne relève pas a priori du champ évaluatif et pourtant son contenu met en relation des politiques et des effets....

²Les travaux originaux de ce courant de pensée sont ceux de John DEWEY (1859-1952), Studies in logical Theory, Chicago, 1903, de William JAMES (1842-1910), Pragmatism, 1907, traduit par LE BRUN, Paris, 1968. Parmi les fondateurs, il faut citer le nom de Charles S. PIERCE (1839-1914)

En revanche, faire de l'évaluation le résultat d'un jugement contradictoire au terme d'une série d'études évaluatives conduit à réduire, voire à nier l'autonomie d'une spécificité de l'approche évaluative. Toute étude peut être alors intégrée dans la construction du jugement évaluatif. L'évaluation, dans cette hypothèse, perd toute possibilité de se construire de manière autonome. La détermination du résultat obéit à un double mécanisme intuitif. Le premier réside dans la juxtaposition des apports des différentes études dont les résultats partiels ne peuvent être emboîtés les uns dans les autres qu'au prix d'un effort d'intuition échappant, le plus souvent, à toute construction méthodologique. Le second porte sur le jugement final au terme duquel un effet d'une politique est reconnu comme satisfaisant ou non. Concernant la dimension démocratique de ce dernier jugement, il est intéressant de voir se rejoindre, par dessus les mers, les cultures et les années, le discours tenu par Patrick VIVERET¹ et un pragmatique comme DEWEY pour qui l'expérimentation ne peut trouver sa place que dans un processus démocratique. La différence entre les deux approches réside dans le moment de l'exercice de la démocratie. Pour VIVERET, il est inséparable du moment où s'exerce le jugement sur la validité des effets d'une politique alors que pour DEWEY, la démocratie est à la fois la garantie de réalisation de l'expérimentation et l'assurance du débat sur l'interprétation du résultat.

L'intérêt des travaux présentés dans Evaluation Review vient de leur caractère double où se combinent démarche expérimentale et étude classique. Dans le premier modèle, l'effet résulte directement de la dissemblance des mesures, alors que dans le second, il naît de la comparaison de plusieurs états à des moments successifs. La combinaison de ces deux modèles semble constituer, aux Etats-Unis, au travers des articles analysés, la figure dominante de l'évaluation des politiques et des interventions. Se trouverait alors validée l'hypothèse d'un maintien des méthodes d'évaluation de types expérimentaux à une époque où les grands programmes expérimentaux semblent nettement remis en question. Cette remarque se trouve corroborée par des travaux récents diffusés dans d'autres publications : "Social Work" et surtout "Administration in Social Work"². La réalité paraît plus complexe. Il n'est pas certain que cette pratique de l'évaluation occupe la totalité du champ évaluatif aux Etats-Unis ; elle doit, plutôt, être pensée comme un moment dans l'évolution des méthodes et des pratiques évaluatives. Ensuite, la démarche qui est proposée, pour être expérimentale n'en n'est pas moins atténuée par les contraintes du milieu et les contenus des programmes qui s'accommodent parfois mal de la rigueur de l'expérimentation.

¹ P. VIVERET, rapport déjà cité.

² voir, par exemple, de Howard J. DOUECK, Ann BONDANZA, Training Social Work Staff to Evaluate Practice : a pre Post. Then Comparison, Administration in Social Work, Vol 14, n°1 1990, pp 119-133.

Préciser les contenus et les limites de l'approche expérimentale telle qu'elle se trouve développée dans ces travaux, confronter les méthodes utilisées aux démarches d'études qui, tout en ne se référant pas explicitement à l'évaluation, n'en participent pas moins de la constitution de connaissances sur les effets des interventions et, enfin, situer cette démarche évaluative dans l'ensemble des pratiques évaluatives envisageables, tel est l'objet de cette seconde partie.

I - EVALUATION ET RECOURS A LA METHODE EXPERIMENTALE

Aux yeux du lecteur français, habitué à des pratiques évaluatives relevant plus d'une perspective endoformative¹, ou de la juxtaposition d'une série d'études, la découverte des nombreux dispositifs d'évaluation présentés dans les différents articles d'Evaluation Review retient d'abord l'attention en raison de leur inscription dans une perspective expérimentale dont la manifestation la plus apparente réside dans la constitution de groupes témoins. Pratique quasi-systématique, la mise en oeuvre de groupes témoins nous semble pourtant ne représenter qu'un aspect de la méthode expérimentale et fait référence à une orthodoxie minimale de l'expérimentation, prenant parfois quelques libertés à l'égard des canons de la méthode.

1-1. Méthode expérimentale et orthodoxie minimale²

Depuis Claude BERNARD, les principes de la démarche expérimentale ont été suffisamment explicités pour qu'il ne soit pas nécessaire d'y revenir une nouvelle fois en dehors des trois points qui nous intéressent le plus eu égard aux expériences présentées. L'expérimentation nécessite un modèle, un protocole d'expérience et un ensemble d'instruments de contrôle du déroulement de l'expérience. Par modèle, il faut entendre l'affirmation d'un lien de causalité raisonné entre deux variables. De ce point de vue, la démarche expérimentale suppose une construction théorique très élaborée. Le protocole consiste à opérationnaliser le système d'hypothèses en construisant un plan d'expérience en laboratoire ou en recréant les conditions du laboratoire sur un terrain. Les contrôles du déroulement de l'expérience s'avèrent d'autant plus nécessaires qu'un des grands principes de la démarche expérimentale réside dans le "toutes choses égales par ailleurs". Les contrôles doivent permettre de vérifier, d'une part, que le protocole est bien en accord avec le modèle mais, d'autre part, que n'interfèrent pas dans le déroulement de l'expérience des facteurs non explicitement prévus.

¹ de ce point de vue, on peut lire d'Eric MONNIER, Evaluation de l'action des pouvoirs publics, Paris, Ed. Economica, 1987.

² ce terme est emprunté à François LE POULTIER, Recherches évaluatives en travail social, Presses Universitaires de Grenoble, 1990, 247 p.

Il est relativement aisé, de ce triple point de vue, de montrer que les évaluations décrites dans Evaluation Review développent une conception minimale de l'expérimentation. On constate, à de très rares exceptions près, qu'il n'existe pas véritablement de dispositifs expérimentaux construits pour valider une hypothèse portant sur les relations entre les actions mises en oeuvre et les effets. Il s'agit plutôt de comparer les effets obtenus à partir de plusieurs logiques d'actions sans que la relation ait fait, au préalable, l'objet d'une modélisation. Lorsqu'une corrélation apparaît, c'est au terme de l'expérience et non comme un construit à vérifier. Sur le plan du choix des protocoles, on constate le plus souvent que les évaluateurs analysent les effets d'une même mesure sur des populations différentes mais qu'il leur arrive aussi de proposer un ensemble de mesures à une population définie a priori comme homogène sans que des règles explicites définissent les conditions du choix. Ces dernières semblent plutôt obéir à des règles de faisabilité. Enfin, du point de vue des contrôles, de nombreuses évaluations mentionnent des insuffisances, soit parce que des facteurs n'ont pas été pris en compte : mode de gestion, utilisation de l'espace, qualité des soins [29], soit parce que les participants aux programmes n'ont pas été sélectionnés de manière aléatoire mais en fonction de leur place dans une file d'attente, ou pire, en fonction de leur degré de motivation [7], soit enfin, parce que les effectifs trop faibles des populations ne permettent pas de se prononcer sur le caractère significatif de l'amplitude des résultats obtenus [31].

A la lecture de ces travaux, la démarche expérimentale apparaît plus comme un projet des évaluateurs et non pas comme une contrainte de méthode dont on ne pourrait s'affranchir. Il est probable que ce souci d'expérimentation est lié à des habitudes et à des cultures sectorielles. Plus un programme d'évaluation met en jeu des thèmes proches de la santé et de l'enseignement, moins les évaluateurs semblent s'écarter des rigueurs de la méthode expérimentale. Si la faisabilité du montage d'un programme évaluatif semble effectivement plus grande dans ces domaines, il faut aussi tenir compte du fait que la démarche expérimentale a longtemps trouvé un terrain de prédilection en médecine et en psychologie. La proximité aux sciences exactes rejoint probablement dans ces champs les habitudes culturelles des évaluateurs.

A cette faiblesse liée à la conduite même du processus d'expérimentation viennent s'ajouter deux autres sources de difficultés. La première est liée à la prise en compte peu importante dans ces programmes évaluatifs des déterminants exogènes de l'action et la seconde, à l'insuffisante construction analytique des relations entre les logiques d'actions et les effets observés.

1-2. Méthode expérimentale et contraintes externes aux programmes.

Un des programmes d'évaluation recensés visait à confronter les résultats d'écoles traditionnelles et d'écoles dites fondamentales, créées dans les années 1978 pour répondre aux critiques adressées au système scolaire. Ces écoles fondamentales se caractérisent par une grande importance accordée à l'enseignement de base et à la discipline. Par ailleurs, les enseignants, les parents et les élèves y sont tenus de respecter un ensemble de principes éthiques. Au terme d'une étude évaluative, il est apparu que les élèves des écoles fondamentales avaient obtenu de meilleurs résultats pour plusieurs matières, que les résultats concernant la discipline étaient favorables aux écoles fondamentales, que l'opinion des élèves sur eux-mêmes était analogue dans les deux écoles et qu'enfin les écoles fondamentales avaient une bonne réputation auprès des parents des deux groupes. A la fin de l'évaluation, les auteurs du programme s'avouent dans l'incapacité d'expliquer les facteurs qui ont permis aux écoles fondamentales d'atteindre leurs objectifs et surtout de mesurer l'impact de ces écoles sur l'ensemble du système traditionnel [6].

Cette remarque faite à propos d'un programme particulier pourrait s'appliquer à de très nombreux autres programmes, pour lesquels, dans le souci de vérifier la relation entre les interventions et les effets, se trouvent abandonnées toutes les tentatives d'expliquer les effets du programme par des variables exogènes. A deux exceptions près, aucune démarche évaluative ne s'interroge sur les causes qui ont rendu possible l'émergence de tel ou tel effet. Ce choix donne à la démarche évaluative un aspect souvent mécanique que ne viennent pas compenser la qualité et la force d'une démonstration, comme cela est souvent le cas pour les compte rendus d'expérimentation dans les sciences exactes.

Dans la majorité des cas, il n'est rien dit de l'application même du programme. Ainsi dans un programme visant à développer des systèmes d'aide aux familles pauvres avec enfants, le contenu du programme nécessite la mise en oeuvre de plans d'aide individualisés aux enfants entre la naissance et l'âge de huit ans. L'évaluation postule que les programmes ont été mis en place dans des conditions identiques de dynamisme des services, de compétence et de qualification des personnels. L'expérience française du R.M.I. ou de l'Aide Sociale à l'Enfance montre qu'il s'agit là d'une perspective optimiste et rarement mise en oeuvre sur les terrains pour lesquels toutes les choses ne sont que rarement égales par ailleurs [31].

1-3. Effet-action : logique d'inventaire ou inventaire logique

Pour décrire les effets de l'intervention publique ou privée dans le cadre du programme mis en oeuvre, les évaluateurs se situent dans une perspective d'audit ou de référentiel.

Dans une logique d'audit, les évaluateurs confrontent les effets survenus aux effets attendus et, le plus souvent, aux effets attendus et souhaités. En utilisant la technique du référentiel, les évaluateurs construisent, en dehors de toute notion d'objectif, un ensemble de changements possibles d'états des personnes ou des groupes, voire des milieux. Cette batterie de changements d'état est ensuite intégrée dans un questionnaire ou une grille d'observation. Dans la majorité des cas, ces indicateurs révèlent des comportements ou, plus souvent, reflètent des opinions sous la forme de taux de satisfaction.

Enfin, un dernier groupe d'effets se concrétise par des indicateurs financiers ou économiques.

La description des actions est d'ordinaire plus sommaire et occupe moins de place dans les comptes-rendus comme si l'effet s'appliquait essentiellement à des individus identifiables alors que les actions regroupées en un programme s'appliquaient à une population dont les seules différenciations sont celles définies par le programme d'évaluation.

On pourrait dire pour reprendre une dichotomie proposée par Raymond BOUDON¹ que les évaluateurs utilisent un paradigme de type holiste pour décrire les actions mais qu'ils lisent les effets selon un paradigme individualiste. Si la méthode expérimentale était appliquée sans faille, en particulier, dans la sélection des populations témoins et des différents échantillons, cet inconvénient pourrait s'estomper, mais comme nous l'avons indiqué plus avant, ce n'est pas toujours le cas de figure dominant.

Juxtaposant globalement des actions décrites en terme de rapport d'activité, et des effets recensés sur un mode plus individualisé, les évaluations décrites n'échappent pas à une certaine logique d'inventaire dont les critères de classement ne seraient pas formalisés et explicités pour le lecteur.

¹ in, par exemple, R. BOUDON, Effets pervers et ordre social, Paris, PUF, 1977, 286 p. et plus particulièrement dans le chapitre VII.

II - L'ÉVALUATION COMME ÉTUDE

S'il fallait situer l'évaluation dans le débat entre tenants de la recherche et partisans des études, l'évaluation telle qu'elle est décrite dans ces revues se situe résolument du côté de l'étude tant la prégnance de ses dimensions finalisées est grande. Il s'agit d'une étude particulière où le temps joue le rôle de l'opérateur final dans une démarche où les techniques sont issues des sciences sociales et ne semblent guère se renouveler.

2-1. L'évaluation : une étude étroitement finalisée

Dans un article publié dans la douzième livraison d'"Evaluation Studies - Review Annual", William R. SHADISH, Jr¹ définissait l'évaluation par rapport à d'autres démarches proches de la manière suivante :

Activité	Bénéfice attendu
Évaluation	Connaissance empirique et immédiate des avantages et des faiblesses des objectifs et du contenu des programmes
Recherche fondamentale	La constitution des savoirs
Suivi de programme	Clarification des contenus et des objectifs des programmes
Développement organisationnel	Aide à la conduite des programmes pour les rendre plus efficaces

Cette définition en recoupe bien d'autres. La démarche évaluative se situe dans l'immédiateté, dans l'empirisme et surtout dans la réponse à une demande finalisée par les décideurs ou les concepteurs des programmes.

¹ W.R. SHADISH, Jr, Sciences of Evaluation Practice - Needs, Purposes Questions and Technology, in Evaluation Studies, Review Annual, 1987, p. 520-553.

Les évaluations sont rarement utilisées pour corriger, en cours de route, le fonctionnement des programmes mais plutôt pour mesurer l'impact d'une intervention et décider de son arrêt, de sa modification ou de son éventuel prolongement. Il s'agit bien d'évaluations sommatives ou récapitulatives, ce que William R. SHADISH désigne sous le nom de "Development organisational" correspondant assez bien à ce que nous appelons évaluation dynamique ou endoformative.

Finalisées, les évaluations le sont étroitement en ce sens que leurs auteurs débordent rarement le champ du programme qui leur est imparti. Lorsqu'ils participent à l'évaluation du Revenu Minimum d'Insertion¹, les auteurs des différentes études extrapolent volontiers leurs résultats au service d'une réflexion sur l'ensemble du système de protection sociale. Ce n'est pas le cas dans les articles d'Evaluation Review dans lesquels les auteurs séparent de manière très nette les présentations des résultats des évaluations et des réflexions sur les politiques sociales qui peuvent figurer dans d'autres articles.

2-2. Le temps comme opérateur principal

Dans la majorité des comptes-rendus présentés dans Evaluation Review, le protocole évaluatif correspond au schéma suivant :

Groupe expérimental	$G_1 t_0 \rightarrow \text{intervention} \rightarrow G'_1 t_1 \dots G''_1 t_n$
Groupe témoin	$G_2 t_0 \dots \rightarrow G'_2 t_1 \dots G''_2 t_n$

Dans la mesure où le suivi du groupe G_2 qui ne participe pas à l'intervention ne peut, pour des questions de méthodes évoquées précédemment, être mesuré avec une totale fiabilité, la comparaison G_1/G_2 perd de son acuité au profit de la comparaison $G_1 t_0/G'_1 t_1$. Autant que l'intervention, la durée $T_1 T_0$ devient l'opérateur principal qui explique les changements d'états entre G_1 et G'_1 .

¹ Voir les différents rapports du programme MIRE-Plan Urbain. 1991.

Le suivi temporel peut se faire en temps réel [8] ou de manière rétrospective, cette dernière méthode est toutefois moins fréquente, le suivi longitudinal l'emportant sur l'approche rétrospective.

Changement de comportement, transformation des valeurs, augmentation ou diminution de la consommation, l'évaluation repère des inflexions, des changements, des évolutions à des moments différents. Il est toutefois paradoxal que le temps ne fasse pas l'objet ni dans Evaluation Review, ni dans d'autres publications d'une réflexion significative. Opérateur essentiel, le temps est aussi un opérateur clandestin. Les problèmes posés par le suivi dans le temps dans le cadre de panel ou par la reconstitution des itinéraires dans le cadre des études rétrospectives ne font quasiment jamais l'objet de développement dans ces comptes-rendus.

2-3. Questionnaire, grille, observation, entretien : des enquêtes par emprunts

Aucune des méthodes proposées par les évaluateurs ne relève a priori et de manière exhaustive d'une démarche évaluative. A travers ces articles, mais également dans d'autres ouvrages¹, se trouve mise en évidence la relativité des méthodes évaluatives qui s'apparentent plus à un concept de type "boîte à outils" qu'à un corpus précis d'outils utilisables dans des circonstances bien déterminées. Dans l'article évoqué précédemment William R. SHADISH Jr propose un tableau des relations entre les questions qui se trouvent posées à l'évaluateur et les technologies à sa disposition, tout en indiquant qu'il n'y a pas de liaison rationnelle entre les deux mais plutôt un ensemble de savoirs-faire validés par l'expérience.

¹ Par exemple, Edited by Marvin D. FEIT, Pb.D, Michael J. HOLOSKO, Pb.D., Evaluation of Employee Assistance Program, N.York, Haworth.Press. 1989.

LES OUTILS DE L'ÉVALUATION

<u>Question type</u>	<u>Technique utilisable</u>
Mesure d'audience	Recensement Enquête par échantillon Entretien directif ou non Observation sur site
Analyse de mise en oeuvre	Observation sur site Echantillon de clients Enquête par panel Prise d'information sur le management
Etude d'effets	Démarche expérimentale ou quasi-expérimentale Mesure avec ajustement statistique Etude de cas Investigation journalistique
Etude d'impact	Séries temporelles Données d'archives Observation participante Construction de standard
Etude de coût	Description des coûts Etude coût-bénéfice Etude coût-efficacité
Etude des causes	Création de modèle Observation participante Construction théorique

Source : W.R. SHADISH Jr. op.cit. p. 543 - Trad. M. LEGROS

III - LES APPORTS DE L'EVALUATION

L'examen des articles d'Evaluation Review et de quelques autres revues dont la vocation est de présenter des pratiques évaluatives pourrait laisser circonspect à l'égard de l'évaluation. La démarche expérimentale qui semblait, de l'extérieur, en constituer l'apport majeur se trouve fortement relativisée par les critiques méthodologiques formulées par les évaluateurs eux-mêmes. Quand à l'autre aspect de la démarche évaluative, c'est-à-dire la dimension étude, elle ne semble finalement que l'utilisation d'un ensemble de méthodes développées dans d'autres champs, dans d'autres domaines.

Cette vision pourrait sembler bien pessimiste, et, à notre avis, ce ne serait pas tout à fait juste. L'ensemble de ces travaux témoigne de l'existence d'un réel savoir-faire au service d'un pilotage des politiques sociales, maîtrisé par un corps de professionnels.

3-1. Une pluralité de savoirs-faire

Entre le problème posé à l'évaluateur et la réponse de ce dernier, l'ajustement semble toujours se faire dans de bonnes conditions. De très nombreux ouvrages sont consacrés à ce lien entre la méthode d'évaluation et le problème à traiter¹. L'expérience importante acquise par les évaluateurs américains depuis les années 1970 leur permet aujourd'hui d'évoluer d'une démarche de réponse sur mesure à une véritable logique du prêt-à-porter en utilisant des outils testés et des modalités d'interventions évaluatives adaptées à chaque situation.

L'expérience acquise porte sur l'ajustement entre les méthodes et les problèmes à traiter mais également sur l'interprétation des résultats. Les bibliographies importantes quantitativement qui accompagnent chaque sujet traité témoignent de cette correspondance singulière entre les travaux des différentes évaluations. Ceci est particulièrement renforcé

¹ On peut citer, parmi d'autres :

Thomas D. COOK, Charles S. REICHARDT, Qualitative and Quantitative Methods in Evaluation Research, Beverly Hills, 1979, Sage Publication.

Michael Q. PATTON, Practical Evaluation, Beverly Hills, 1982, Sage Publication.

dans les domaines de l'éducation, de l'action sociale et de l'évaluation des politiques sanitaires¹.

Nous avons, à plusieurs reprises, mentionné la faiblesse théorique des articles analysés. Cette vision doit être compensée par une appréciation portée sur les articles qui ne rentrent pas dans notre corpus. 35 articles mentionnaient explicitement des démarches évaluatives, soit un article sur six puisque l'ensemble du matériel dépouillé regroupait originellement quelques 212 articles différents. Parmi les autres articles dominant deux groupes différents. Le premier s'intéresse aux systèmes théoriques sous-jacents aux pratiques évaluatives. Le second propose des contributions dans le domaine des pratiques adjacentes à l'évaluation : analyse de données, étude de séries temporelles

L'ensemble de ces articles centrés sur les méthodes évaluatives et de manière plus adjacente sur l'évaluation des politiques sociales concrétise l'existence d'un savoir-faire global dans ce que nous pourrions appeler une véritable ingénierie de l'évaluation. Lorsqu'il nous est arrivé d'analyser ce terme à propos des pratiques françaises de l'évaluation² cela désignait surtout une pratique d'accompagnement du projet développé sous le nom d'évaluation dynamique par des consultants de bureaux d'études ou par des chercheurs d'équipes universitaires. En revanche, parler d'ingénierie d'évaluation, dans le contexte américain, désigne un ensemble de savoirs-faire dans le domaine de la gestion des méthodes et des outils d'évaluation qui permettent à la personne chargée d'une évaluation de répondre à des questions du type :

- l'évaluation est-elle nécessaire à ce projet ?
- à quoi doit servir l'évaluation ?
- quelles sont les questions qui doivent guider le programme d'évaluation ?
- quelles techniques sont actuellement disponibles pour répondre à ces questions ?

Le réponse à ces questions constitue le véritable sujet des articles d'Evaluation Review et forme le savoir spécifique des chercheurs et techniciens responsables de la conduite des projets d'évaluation.

¹ Un bon exemple de ces discussions des résultats en fonction des méthodes employées dans le domaine sanitaire est donné par un article de John W. FINNEY et Rudolph H. MOOS, Matching Patients with Treatments. Conceptual and Methodological Issues. Evaluation Studies, 12, 1987, p. 555-567.

² Michel LEGROS, Du management à l'évaluation : les outils de gestion des systèmes hospitaliers ; Cahiers de Promofaf, Paris, décembre 1990, 64 p.

3-2. Evaluation et pilotage de l'action

L'évaluation de ces programmes n'est pas une activité de recherche désintéressée conduite par des équipes dont le seul souci serait de mieux comprendre le fonctionnement social. Il s'agit, bien au contraire, d'une activité parfaitement insérée dans le programme lui-même et à laquelle participent les gestionnaires du programme.

Les évolutions ne mentionnent pas systématiquement dans les articles la place des résultats de leur évaluation dans le processus de décision. Toutefois, sans que cette liaison soit complètement explicitée, de très nombreux indices montrent qu'une liaison forte existe entre évaluation et pérennité des programmes. Cette liaison peut être affirmée dès la description du protocole d'évaluation. C'est le cas, par exemple, pour une évaluation conduite au Guatemala et portant sur la rentabilité de différents types de services appliquant des programmes de stérilisations chirurgicales volontaires [16], elle peut être également vérifiée lors de la présentation des résultats de l'évaluation. C'est le cas de l'évaluation de l'introduction d'une courte peine de prison pour les conducteurs en état d'ébriété qui conclut à l'inefficacité de cette sanction sur les comportements des conducteurs [2].

En aucune manière, la liaison entre évaluation et pilotage de l'action relève d'une liaison mécanique voire automatique. Dans le cas du programme évoqué plus avant et portant sur les campagnes de stérilisation volontaire, les auteurs montrent que le coût le plus faible est généré par des cliniques, centralisant la demande d'intervention. L'évaluation montre par ailleurs qu'à ce coût plus faible est associé un fort inconvénient lié à la perte de proximité des services, les gestionnaires du programme semblent en avoir conclu à l'intérêt, en dépit d'un coût plus élevé, du maintien des services de proximité [16].

Notons que cette vision un peu idyllique des relations entre évaluation et pilotage de l'action, vision qui peut être dégagée des articles analysés, est parfois nuancée dans des articles de réflexions et non par des comptes-rendus d'expérience. Léonard SAXE montrait, par exemple, que la majeure partie des travaux évaluatifs avaient un double effet : permettre des prises de décision mais également donner aux décideurs une information sur la complexité des problèmes auxquels ils se trouvent confrontés¹. De ce point de vue, la liaison entre évaluation et décision semble moins linéaire que la lecture des articles d'Evaluation Review semble le laisser croire.

¹ Un article de Leonard SAXE, Policymakers'Use of Social Science Research, Technology Assessment in the U.S. Congress, Evaluation Studies, Vol 12, op. cit., p.214-233.

3-3. Une communauté d'évaluateurs

Si leur conduite, leur finalité et leur construction mêmes les rapprochent davantage des programmes d'études que des activités de recherches fondamentales, les activités d'évaluation ne semblent pas être reléguées aux mains d'évaluateurs, obscurs tâcherons de travaux d'études. Les articles sont systématiquement signés par des auteurs membres d'équipes universitaires, professeurs ou assistants professeurs, et plus rarement d'étudiants en cours de Ph.D.

Ce sont ces mêmes personnes qui, à côté des articles, décrivent des expériences d'évaluation, rédigent des papiers plus fondamentaux sur les politiques publiques ou les méthodologies évaluatives.

A titre d'exemple, sur les 72 membres du comité éditorial d'Evaluation Review et dont les noms reviennent assez fréquemment parmi les articles signés, on retrouve 46 représentants d'universités, 12 représentants d'administrations fédérales y compris un membre du Federal Bureau of Investigations.

La composition du comité directeur de l'American Evaluation Association est structurée de manière voisine puisque sur 10 membres, on retrouve 7 professeurs d'université, 1 membre du General Accounting Office et 2 membres de grands cabinets privés.

Démarche utilitariste relevant d'une activité d'étude, l'évaluation, aux Etat-Unis, n'en est pas moins profondément investie dans ces pratiques et dans ces organismes par la communauté universitaire de façon massive.

CONCLUSION

Les travaux dont nous avons rendu compte dans cette note ont été dans la majorité extraits d'*Evaluation Review* et plus minoritairement, d'*Evaluation and the Health Professions*. Ils ne peuvent en aucune manière être considérés comme représentatifs de l'ensemble des travaux d'évaluation réalisés par les évaluateurs américains ces dernières années. D'abord parce que ces travaux ne font pas l'objet de publications systématiques dans les revues et de ce point de vue les travaux publiés rendent plutôt compte d'une pratique institutionnelle de l'évaluation, et, d'autre part, nous n'avons retenu qu'une publication à caractère fédéral alors que les différents Etats et surtout les universités disposent de bulletins propres dans lesquels ils réalisent des présentations d'études évaluatives.

A cette absence de représentativité de l'ensemble du champ, il faut ajouter une absence de représentativité des dynamiques temporelles à l'oeuvre dans l'évolution des pratiques évaluatives. Ainsi, au moment où l'ouvrage de base, presque le manuel, de Peter H. ROSSI et Howard E. FREEMAN¹ connaît sa quatrième édition et que celui de Michael Quinn PATTON² approche de sa troisième édition, on voit arriver des ouvrages plus iconoclastes rompant avec une vision néo-positiviste de l'évaluation, tel le livre de Egon G. GUBA et de Yvonne S. LINCOLN³. Pour ces auteurs, en matière d'évaluation, le paradigme positiviste est insuffisant pour supporter l'ensemble des perspectives de développement de l'évaluation en ce sens qu'il ne peut accepter de prendre en compte les importantes marges d'incertitudes de chaque programme. Incertitudes qui tiennent autant aux stratégies des acteurs qu'aux impossibilités de trouver des méthodes quantifiant la totalité des objets à évaluer. S'appuyant sur des approches tentant de formaliser l'incertitude et la complexité⁴ ou réintroduisant des méthodes qualitatives dans les processus évaluatifs, ces auteurs, qui ont reçu en 1987 le Paul LAZARFELD Award de l'Association Américaine d'Evaluation pour leur contribution à la recherche évaluative, se trouvent singulièrement proches des travaux français sur les approches dynamiques de l'évaluation⁵.

¹ Peter H. ROSSI, Howard E. FREEMAN, Evaluation, a Systematic Approach, SAGE Publication, 1989, 496 p. 4ème édition.

² Michael Quinn PATTON, Creative Evaluation, SAGE Publication, 1987, 264 p. 2nd édition.

³ Egon G. GUBA, Yvonne S. LINCOLN, Fourth Generation Evaluation, SAGE Publication, 1990, 294 p. 2nd édition.

⁴ Douglas R. HOFSTADER. N.Y. Basic Books, 1979.

⁵ M. BONETTI, V. de GAULEJAC, J. FRAISSE, L'évaluation dynamique des organisations publiques. Paris, Ed. d'organisation, 1988.

Si la démarche à dominante expérimentaliste occupe donc un poids très grand dans la recherche évaluative, son caractère homogène attesté par les comptes-rendus d'Evaluation Review n'en tolère pas moins d'autres approches. En caricaturant, on pourrait affirmer que les rapports de force entre approche expérimentale et approche dynamique ou qualitative sont exactement inversés aux Etats-Unis, et en France où à l'exception de chercheurs comme Jean-Marc DUTRENIT¹ et François LE POULTIER² l'approche expérimentale n'est guère développée.

Cela s'explique probablement par un décalage de culture et un réel malentendu. L'analyse du décalage des cultures échappe par son ambition à cette note mais force est de reconnaître que les philosophies analytiques et pragmatiques n'ont guère pu se développer en Europe alors qu'elles ont servi de support intellectuel à la mise en oeuvre de l'évaluation dans les pays anglo-saxons.

Le malentendu, en revanche, nous semble plus analysable. Les évaluations américaines mettent en avant une méthode expérimentale à laquelle ils ne recourent que de manière imparfaite et souvent secondaire et privilégient une démarche d'étude étroitement finalisée. En Europe, et particulièrement en France, nous avons écarté délibérément cette référence à l'expérimentation contrôlée pour privilégier une démarche d'évaluation dont les rapports avec la recherche, pour affirmés qu'ils soient par les évaluateurs -chercheurs- restent fortement tenus.

Si les positions des praticiens de l'évaluation semblent bien éloignés on pourrait affirmer, de manière paradoxale, que leur pratique de l'évaluation et la demande du marché tendent à les rapprocher.

La pratique de l'évaluation en France comme aux U.S.A. se caractérise dans la réalisation d'études mobilisant des services, des méthodes et outils émanant des sciences sociales. Au delà des habillages américains et français, cette réalité de l'étude évaluative semble s'imposer de manière très forte. Les travaux anglo-saxons mettent plus nettement l'accent dans leur pratique sur une véritable ingénierie de l'évaluation, là où faute d'expérience et de marché nous en restons souvent dans les ajustements exploratoires et les tâtonnements de la pré-série. Vue sous cet angle, l'expérience américaine nous semble particulièrement utile.

¹ J.M. DUTRENIT. Gestion et évaluation des services sociaux. Paris, éd. Economica, 1989.

² F. LE POULTIER, Recherches évaluatives en travail social. P.U.G., Grenoble, 1990.

Le demande d'évaluation produisant des résultats opératoires, intégrables dans les décisions, semble au fil des années gagner le marché français de l'évaluation qui n'est plus nécessairement conçue par les demandeurs comme un lieu de promotion de recherche, mais bien comme une demande de services. De ce point de vue, l'intervention des collectivités territoriales et des différentes missions d'évaluation devrait contribuer à renforcer cette évaluation.

Quel que soient les rapprochements possibles entre les deux pratiques, un décalage de taille et de structuration subsistera entre les deux marchés. Si l'expérience américaine ne nous semble guère utile pour augmenter le marché de l'évaluation en France, en revanche, la structuration des offreurs d'évaluation en association scientifique dotée d'organes de presse, réunissant régulièrement ses membres en des congrès dont les actes font l'objet d'une large diffusion nous semble devoir constituer une perspective point trop hors de portée dans la situation française.

**FICHES DE PRESENTATION
DES EXPERIENCES EVALUATIVES**

REVUE : Evaluation Review

TITRE : The Impact of Early Education

AUTEURS : D. E. Pierson, M. B. Bronson, J. P. Swartz, D. K. Walker, E. Dromey,
T. Tivnan

Volume 7 ; N° 2 ; 1983 ; p.191-216

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : Brookline Education Project (BEEP).

POPULATION : Enfants, de la naissance au jardin d'enfants.

OBJECTIFS DU PROGRAMME :

Augmenter la proportion de réussites scolaires au moment de l'entrée à l'école primaire.

LOCALISATION-DATE-DUREE :

Brookline, Massachusetts, USA. Lancé en 1972.

CONTENU DU PROGRAMME :

Il accorde une grande importance au milieu familial.

Services fournis : - examens médicaux réguliers,
- formation des parents,
- programmes éducatifs pour les parents.

MAITRE D'OEUVRE : N.C.

MOYENS DU PROGRAMME : N.C.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE :

Ecoles publiques de Brookline, automne et printemps de l'année de jardin d'enfants.

EVALUATEURS :

Médecins, psychologues, parents, observateurs formés, enseignants.

BUT DE L'EVALUATION :

Déterminer l'effet d'un enseignement précoce sur les enfants et leurs parents. L'article ne décrit que la première partie d'une évaluation de plus grande ampleur.

INDICATEURS RETENUS :

L'enfant est jugé dans les domaines suivants :

- l'organisation de son travail, son efficacité, son comportement avec les autres (I),
- ses facultés motrices et intellectuelles, son comportement social (II).

CHOIX DE LA POPULATION :

Echantillon : 132 enfants nés entre le printemps 1973 et l'automne 1974 ayant participé depuis leur naissance au BEEP.

Groupe témoin : 366 enfants nés entre 1970 et 1974. (Le choix, ici, d'enfants nés avant le lancement du BEEP est destiné à réduire l'effet de biais dû à la différence de motivation entre les familles ayant choisi de participer et celles qui refusent un programme proposé.)

METHODE DE COLLECTE DE L'INFORMATION :

- Observations en classe par des observateurs formés (I)
- Grille d'évaluation par l'enseignant (II)

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Pour chaque variable une valeur de référence a été établie afin de déterminer si les résultats révèlent ou non un problème chez l'enfant.

On compare pour chaque variable le pourcentage d'enfants ayant des difficultés dans le groupe des participants et dans le groupe témoin.

Régressions.

LES EFFETS DU PROGRAMME**LES TYPES D'EFFETS :**

Les résultats indiquent que le BEEP a atteint son but . Dans le cas de l'évaluation par les enseignants, les résultats ne sont favorables aux participants que dans certains sous groupes d'enfants : ceux qui ont le plus de risques d'avoir des difficultés scolaires, du fait de leur milieu familial par exemple.

LES BIAIS :

- Voir "groupe témoin".
- Des enseignants étaient au courant de la participation de certains enfants au BEEP.

LES CRITIQUES DU PROGRAMME : N.C.

LES CRITIQUES DE L'EVALUATION : N.C.

OBSERVATIONS : N.C.

REVUE : Evaluation Review

TITRE : A "Mandatory" Jail Sentence for Drinking and Driving.

AUTEURS : J. W. Grube, K.A. Kearney.

Volume 7 ; N° 2 ; 1983 ; p.235-246

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : Yakima Drinking and Driving Project.

POPULATION : Conducteurs en état d'ébriété.

OBJECTIFS DU PROGRAMME :

Changer l'attitude du public envers l'alcool au volant et réduire le nombre de cas de conduite en état d'ébriété.

LOCALISATION-DATE-DUREE : Etats Unis, Etat de Washington, 1978.

CONTENU DU PROGRAMME :

Introduction d'une peine de deux jours de prison pour les coupables.

MAITRE D'OEUVRE : N.C.

MOYENS DU PROGRAMME : N.C.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE : Etats-Unis, Comté de Yakima (Etat de Washington), 1979.

EVALUATEURS : N.C.

BUT DE L'EVALUATION :

Evaluer la mise en oeuvre du programme, d'une part, et son impact, d'autre part.2

INDICATEURS RETENUS :

Indicateurs de mise en oeuvre : connaissance du programme par le public (1.1), son application par la police (1.2), son application par le système judiciaire (1.3).

Indicateurs d'impact : attitude du public (2.1), nombre d'accidents de la route dûs à l'alcool (2.2) entre 1976 et 1978 et pour l'année 1979.

CHOIX DE LA POPULATION :

(1.1) 315 numéros de téléphone ont été choisis. 238 ménages ont pu être interrogés.

(1.2) 32 officiers de police.

(1.3) Des échantillons de personnes ayant conduit en état d'ébriété ont été choisis au hasard parmi les données du test d'alcoolémie "Breathalyzer".

(2.1) Même échantillon que pour (1.1).

METHODE DE COLLECTE DE L'INFORMATION :

- Entretiens téléphoniques

- Questionnaires

- Les données de l'alcootest "Breathalyzer" pour déterminer le pourcentage de personnes déclarées coupables et de personnes condamnées.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Test du Chi-deux

LES EFFETS DU PROGRAMME

LES TYPES D'EFFETS :

- Le programme était mal connu du public : 40 % des personnes interrogées ne le connaissaient pas.

- Il a été mal appliqué par le système judiciaire : 37,3 % des personnes arrêtées ont passé moins de deux jours en prison et 28,9 % n'y sont pas allées.

- Rien n'indique qu'il ait réduit la fréquence des cas de conduite en état d'ébriété.

LES BIAIS : N.C.

LES CRITIQUES DU PROGRAMME : N.C.

LES CRITIQUES DE L'EVALUATION : N.C.

OBSERVATIONS : N.C.

REVUE : Evaluation Review

TITRE : Evaluating Microcounseling Training.

AUTEURS : S.J. Gill, C.F. Berger, G.L. Logar.

Volume 7 ; N° 2 ; 1983 ; p.247-256

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : N.C.

POPULATION : Etudiants.

OBJECTIFS DU PROGRAMME : N.C.

LOCALISATION-DATE-DUREE : Etats Unis.

CONTENU DU PROGRAMME :

Stage de formation de conseillers (dans divers domaines : éducation, santé ...). Des comportements précis à adopter face au client sont enseignés aux étudiants. Cette méthode comporte la critique de séances de conseil filmées en vidéo.

MAITRE D'OEUVRE : N.C.

MOYENS DU PROGRAMME : N.C.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE :

Un cours de formation de conseillers dans une université américaine.

EVALUATEURS : N.C.

BUT DE L'EVALUATION :

Vérifier l'hypothèse selon laquelle le comportement des étudiants lors d'une séance de conseil sera très différent après le stage et se rapprochera d'un modèle.

INDICATEURS RETENUS :

7 comportements standards à adopter face au client sont définis. Ils sont notés en pourcentage des réponses et réactions du conseiller face au client.

CHOIX DE LA POPULATION :

10 étudiants sont suivis. Il n'y a pas de groupe témoin afin d'éviter plusieurs difficultés : le nombre insuffisant de sujets, le fait de priver certains du bénéfice d'un stage, le risque d'interférence entre les deux groupes par des contacts extérieurs.

METHODE DE COLLECTE DE L'INFORMATION :

Trois séances de conseil sont filmées, le stage ayant lieu entre la deuxième et la troisième séance (étude avant/après). Les films sont visionnés par des conseillers expérimentés.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Les scores pour chaque comportement et chaque élève sont calculés en pourcentage du total des réponses de tous les étudiants. Puis est calculée la différence entre ces scores et les scores de référence.

Techniques statistiques : analyse de variance, tests.

LES EFFETS DU PROGRAMME**LES TYPES D'EFFETS :**

L'hypothèse de départ a été confirmée en partie par les résultats.

Les étudiants les plus proches du modèle au départ s'en éloignent un peu, après le stage. Les autres s'en rapprochent.

LES BIAIS : N.C.

LES CRITIQUES DU PROGRAMME : N.C.

LES CRITIQUES DE L'EVALUATION : N.C.

OBSERVATIONS : N.C.

REVUE : Evaluation Review

TITRE : Effectiveness of the Mandatory Minority Business Set-aside Contracting Goals

AUTEURS : D.E. Black

Volume 7 ; N° 3 ; 1983 ; p.321-336

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : N.C.

POPULATION :

Entreprises possédées ou dirigées par des personnes socialement et économiquement défavorisées.

OBJECTIFS DU PROGRAMME :

Augmenter la part de ce type d'entreprises dans les commandes d'Etat.

LOCALISATION-DATE-DUREE : Etats-Unis, lancé en 1978.

CONTENU DU PROGRAMME :

Les dirigeants d'organismes publics doivent négocier chaque année avec le "Small Business Administration" un objectif quant à la valeur totale des contrats qu'ils accorderont à ce type d'entreprises.

MAITRE D'OEUVRE :

Une loi votée par le Congrès le 24 octobre 1978 est à l'origine de ce programme (Public Law 95-507). L'article 221 de cette loi concerne le programme évalué.

MOYENS DU PROGRAMME : N.C.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE :

Etats-Unis, 1981.

Etude sur cinq ans : 1977 et étude des quatre premières années de fonctionnement.

EVALUATEURS : N.C.

BUT DE L'EVALUATION :

Déterminer si la loi a contribué à accroître de façon significative la valeur des contrats accordés par des organismes publics à ce type d'entreprises et si d'autres variables interviennent. (Voir outils statistiques, calculs).

INDICATEURS RETENUS :

Y_t - Valeur totale des contrats accordés à des entreprises défavorisées par un organisme d'Etat

CHOIX DE LA POPULATION :

22 organismes publics civils et militaires pour lesquels des données suffisantes ont pu être recueillies.

METHODE DE COLLECTE DE L'INFORMATION :

Deux sources principales : Minority Business Development Agency (MBDA), Small Business Administration.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Régression multiple : $Y_t = b_1 + b_2X_{2t} + b_3X_{3t} + b_4X_{4t} + b_5X_{5t} + c_t$

X_{1t} - Valeur totale des contrats.

X_{3t} $X_{3t} = 1$ la loi est déjà appliquée par l'organisme au moment de l'étude.

$X_{3t} = 0$ les observations sont faites avant que la loi ne soit appliquée.

X_{4t} - Nombre total d'entreprises défavorisées accessibles par l'organisme pendant l'année budgétaire.

X_{5t} - L'objectif (en Dollars) négocié avec le Small Business Administration quant aux contrats à accorder à ces entreprises

LES EFFETS DU PROGRAMME**LES TYPES D'EFFETS :**

X_{2t} , X_{4t} , X_{5t} ont un effet positif sur Y_t alors que X_{3t} a un effet négatif assez important.

X_{2t} : pour chaque Dollar de plus en commandes publiques 0,01\$ revient aux entreprises défavorisées.

Le programme n'a pas eu d'effet significatif sur la somme totale des contrats accordés aux entreprises concernées.

LES BIAIS :**LES CRITIQUES DU PROGRAMME :**

- Son application entraîne une augmentation des frais administratifs de l'organisme public et une augmentation de ses coûts du fait que les entreprises concernées ne subissent pas la concurrence.

- L'absence de mesures incitatives auprès des organismes publics pour contrebalancer ces désavantages aggrave leur attitude vis-à-vis des entreprises défavorisées.

LES CRITIQUES DE L'EVALUATION : N.C.

OBSERVATIONS : N.C.

REVUE : Evaluation Review

TITRE : The Costs and Benefits of Title XX and Title XIX Family Planning Services in Texas.

AUTEURS : D. Malitz

Volume 8 ; N° 4 ; 1984 ; p.519-536

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : N.C.

POPULATION : Femmes à bas revenu. 250 000 patientes.

OBJECTIFS DU PROGRAMME : N.C.

LOCALISATION-DATE-DUREE : Etats-Unis, Texas.

CONTENU DU PROGRAMME :

Services de planning familial fournis dans le cadre de deux programmes distincts "Title XX" et "Title XIX", ce dernier s'adressant aux femmes bénéficiant déjà du programme AFDC.

MAITRE D'OEUVRE :

Texas département of Human Resources (TDMR), Texas Department of Wealth (Departements des ressources humaines et de la santé du Texas)

MOYENS DU PROGRAMME :

22 millions de Dollars pour l'année budgétaire 1981.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE : Texas. L'étude concerne l'année budgétaire 1981.

EVALUATEURS :

L'évaluation a été demandée par le "Texas Département of Human Resources" (TDHR)

BUT DE L'EVALUATION :

Evaluer la rentabilité financière de ces services de planning familial.

INDICATEURS RETENUS :

Taux de rendement : Economies réalisées par personne grâce à la prévention de naissances non désirées (en allocations familiales, bons alimentaires ...) divisées par les dépenses de planning familial par personne en 1981.

De plus, des estimations du nombre de fausses couches et avortements évités ont été effectuées sans estimation de leur coût financier.

CHOIX DE LA POPULATION :

Pour le programme "Title XX", l'échantillon a été choisi au hasard parmi les dossiers de 78 services de planning familial au Texas (1 606 adolescentes, 1 605 adultes).

Pour "Title XIX" aucun échantillon n'a pu être formé. L'évaluation a été faite d'après les données de "Title XX".

METHODE DE COLLECTE DE L'INFORMATION :

Des questionnaires sur chaque patiente ont été remplis par le personnel des services de planning familial.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Le nombre de grossesses évitées a été estimé à l'aide de données sur l'utilisation des différentes méthodes de contraception par les patientes avant leur première visite et après leur dernière visite au service de planning familial.

Puis les économies réalisées ont été évaluées en calculant le nombre de naissances évitées d'après des données sur le pourcentage de naissances, avortements et fausses couches parmi les grossesses indésirées chez des femmes à faible revenu (Dry Foos, 1982).

LES EFFETS DU PROGRAMME**LES TYPES D'EFFETS :**

Les taux de rendement calculés ont été de : 2,44 pour les adolescentes et 1,73 pour les adultes dans le cadre du programme "Title XX" ; de 3,29 pour les adolescentes et 2,93 pour les adultes de "Title XIX".

Pour un total d'environ 272 000 femmes, près de 22 000 avortements et 7 000 fausses couches ont été évités.

LES BIAIS : N.C.

LES CRITIQUES DU PROGRAMME : N.C.

LES CRITIQUES DE L'EVALUATION :

De nombreux avantages humains ou avantages à long terme n'ont pas été pris en compte.

Les dépenses encourues par les villes ou les comtés n'ont pas été prises en compte. Pour les calculs, de nombreuses hypothèses et estimations ont été utilisées.

OBSERVATIONS : -

REVUE : Evaluation Review

TITRE : An Evaluation of Fundamental Schools

AUTEURS : L J. Weber, J K. Mc Bee, J H. Lyles

Volume 8 ; N° 5 ; 1984 ; p.595-614

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME :

Ecoles dites "fondamentales" (Fundamentale Schools). Elles font partie de l'enseignement public.

POPULATION : Elèves de l'enseignement primaire et secondaire.

OBJECTIFS DU PROGRAMME :

Ces écoles ont été créées pour répondre aux critiques adressées au système scolaire considéré comme médiocre et inapte à préparer les élèves aux besoins de la société.

LOCALISATION-DATE-DUREE :

Les premières écoles de ce type sont apparues en 1978. Etats-Unis

CONTENU DU PROGRAMME :

Une grande importance est accordée à l'enseignement de base (mathématiques, lecture, sciences ...), à la discipline.

Les enseignants, les parents et les élèves sont tenus de respecter un ensemble de principes.

MAITRE D'OEUVRE : N.C.

MOYENS DU PROGRAMME : N.C.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE : Etats-Unis

BUT DE L'EVALUATION :

Comparer les écoles fondamentales et les écoles traditionnelles.

INDICATEURS RETENUS :

- (1) - Résultats des élèves.
- (2) - Ambiance et conditions de travail des élèves.
- (3) - Opinion des élèves sur eux-mêmes.
- (4) - Attitude des parents.

CHOIX DE LA POPULATION :

Quatre écoles traditionnelles ont été choisies pour être comparées aux quatre écoles fondamentales existant au moment de l'évaluation. Les écoles étaient équivalentes.

METHODE DE COLLECTE DE L'INFORMATION :

- (1) Deux épreuves écrites.
- (2) - Questionnaire adressé aux enseignants,
 - Consultation des listes d'exclusions temporaires d'élèves.
 - Observations en classe.
- (3) Questionnaire adressé aux élèves.
- (4) Questionnaire adressé aux parents.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

- (1) Analyses de covariance, tests pour comparer les résultats des deux groupes.
- (2) Tests t
- (3) Tests t

LES EFFETS DU PROGRAMME**LES TYPES D'EFFETS :**

- (1) Les élèves des écoles fondamentales ont obtenu de meilleurs résultats pour plusieurs matières ou des résultats analogues à ceux du groupe témoin. Les différences sont significatives.
- (2) Les résultats concernant la discipline sont favorables aux écoles fondamentales.
- (3) L'opinion des élèves sur eux-mêmes est analogue dans les deux groupes.
- (4) Les écoles fondamentales ont bonne réputation auprès des parents des deux groupes.

LES BIAIS : N.C.**LES CRITIQUES DU PROGRAMME :**

On observe peu de contacts entre les élèves dans les écoles fondamentales.

LES CRITIQUES DE L'EVALUATION :

Cette étude ne s'interroge pas sur les facteurs qui ont permis aux écoles fondamentales d'atteindre leurs objectifs, ni sur l'influence de ces écoles sur les établissements traditionnels.

OBSERVATIONS : N.C.

REVUE : Evaluation Review

TITRE : Nutrition Behavior Change : Outcomes of an Educational Approach

AUTEURS : P K. Edwards, A C. Acock, R L. Johnston

Volume 9 ; N° 4 ; 1985 ; p.441-460

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : "Better Eating for Better Health".

POPULATION :

Le programme s'adresse au grand public. Les participants sont des volontaires.

OBJECTIFS DU PROGRAMME :

Améliorer les connaissances en matière d'hygiène alimentaire et les habitudes alimentaires du public.

LOCALISATION-DATE-DUREE : 3 000 centres de la Croix Rouge Américaine.

CONTENU DU PROGRAMME : Cours sur l'hygiène alimentaire.

MAITRE D'OEUVRE :

La Croix Rouge Américaine et le département de l'agriculture des Etats-Unis (USDA).

MOYENS DU PROGRAMME : N.C.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE : 51 centres de la Croix Rouge

BUT DE L'EVALUATION :

Déterminer s'il est possible d'effectuer des mesures fiables des trois indicateurs retenus, si un tel programme peut avoir un impact sur le grand public, si les effets sont durables. Déterminer la relation existant entre les changements des trois indicateurs.

INDICATEURS RETENUS :

- les connaissances des sujets sur l'hygiène alimentaire.
- leurs convictions dans ce domaine.
- leurs habitudes alimentaires.

CHOIX DE LA POPULATION :

51 centres de la Croix Rouge ont été choisis.

Le groupe témoin est formé de volontaires qui suivaient d'autres cours de la Croix Rouge pendant la même période. Il n'a pas été choisi au hasard.

METHODE DE COLLECTE DE L'INFORMATION :

Deux questionnaires, un avant le cours (N = 1 461), un juste après (N = 1 301). (212 et 133 réponses pour le groupe témoin).

Un entretien par téléphone (200 réponses sur 248 appels).

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Calcul de coefficients de corrélation entre les changements des trois indicateurs.

Coefficients alpha de Cronbach pour déterminer la fiabilité des mesures.

LES EFFETS DU PROGRAMME

LES TYPES D'EFFETS :

Le cours a eu un effet positif sur l'ensemble des participants. Il n'y a pas de différence significative dans les progrès réalisés en fonction du sexe, de l'origine ethnique et du revenu.

Les connaissances disparaissent avec le temps. Cependant, on remarque une nette amélioration des habitudes alimentaires dix semaines après le cours.

Les méthodes de mesure des connaissances et du comportement sont fiables.

LES BIAIS :

Les participants sont des volontaires motivés. Les conclusions de l'expérience ne peuvent donc être généralisées.

LES CRITIQUES DU PROGRAMME : N.C.

LES CRITIQUES DE L'EVALUATION : N.C.

OBSERVATIONS : N.C.

REVUE : Evaluation Review

TITRE : Measuring Program Implementation with a Classroom Observation Instrument

AUTEURS : D.M. Kerr, L. Kent , T.C.M. Lam

Volume 9 ; N° 4 ; 1985 ; p. 461-482.

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : Delinquency Prevention Research and Development Project.

POPULATION : Elèves de l'enseignement secondaire.

OBJECTIFS DU PROGRAMME :

Augmenter la participation, l'attention des élèves en classe et leur intérêt pour l'école.
L'objectif final est de prévenir la délinquance.

LOCALISATION-DATE-DUREE : Etats-Unis (6 sites).

CONTENU DU PROGRAMME :

Ce programme est fondé sur le "Social Development Model" (Hawkins et Weis, 1980).
Des enseignants sont formés à une série de techniques pédagogiques destinées à améliorer la participation des élèves.

MAITRE D'OEUVRE : N.C.

MOYENS DU PROGRAMME : N.C.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE : Etats-Unis (6 sites)

EVALUATEURS :

Kerr et Cummings ont mis au point en 1983 une grille d'observations en classe l'"Interactive Teaching Map".

BUT DE L'EVALUATION :

Evaluer la mise en oeuvre du programme avant d'en évaluer les résultats.

INDICATEURS RETENUS :

5 comportements-type ont été définis pour les élèves (l'élève travaille, rêve, bavarde...) : indicateurs de résultat du programme ; 11 comportements ou actes des enseignants (travail écrit, explications...) : indicateurs de mise en oeuvre du programme.

CHOIX DE LA POPULATION :

Des enseignants ont été choisis au hasard afin d'être formés. Les autres ont servi de groupe témoin.

METHODE DE COLLECTE DE L'INFORMATION :

L'"Interactive Teaching Map" permet de noter les comportements-type pendant chaque minute du cours.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Pour vérifier la validité des observations en classe, les résultats de trois observateurs ont été comparés (coefficient de Kappa).

Calcul de coefficients de corrélation entre le degré de mise en oeuvre du programme et le nombre moyen de minutes pendant lesquelles :

- l'élève travaille (coefficient > 0)
- l'élève ne travaille pas (coefficient < 0)

LES EFFETS DU PROGRAMME

LES TYPES D'EFFETS :

Les enseignants du groupe expérimental appliquent mieux le programme. Cependant l'effet s'estompe avec le temps. Leurs élèves sont plus souvent actifs.

LES BIAIS :

Une même personne évalue le comportement de l'enseignant (mise en oeuvre du programme) et celui des élèves (résultat du programme).

LES CRITIQUES DU PROGRAMME : N.C.

LES CRITIQUES DE L'EVALUATION : N.C.

OBSERVATIONS :

L'article souligne en citant Sorg (1980) et Cook et Poole (1982) que les évaluateurs omettent souvent de vérifier la bonne mise en oeuvre d'une politique avant d'en évaluer les résultats. Des programmes sont jugés sans effet alors qu'ils n'ont jamais été réellement appliqués. L'"Interactive Teaching Map" est ici destinée à résoudre ce problème.

REVUE : Evaluation Review

TITRE : An Evaluation of a Juvenile Education Program in a State Penitentiary.

AUTEURS : T P. Locke, G H. Johnson, K. Kirigin-Ramp, J D. Atwater, M. Gerrard
(University of Kansas).

Volume 10 ; N° 3 ; 1986 ; p.281-298

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : Juvenile Education Program (J E P).

POPULATION : Jeunes délinquants

OBJECTIFS DU PROGRAMME :

Réhabilitation de jeunes délinquants. Il s'agit de leur faire connaître la réalité de la vie en prison dans un but dissuasif.

LOCALISATION-DATE-DUREE :

Prison de Lansing (Michigan) ; depuis octobre 1978.

CONTENU DU PROGRAMME :

De jeunes délinquants rencontrent des prisonniers sélectionnés pour le programme. La particularité de J E P est de tenter d'associer à chaque sujet un prisonnier qui lui soit proche dans son style de vie.

MAITRE D'OEUVRE : N.C.

MOYENS DU PROGRAMME : N.C.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE : 1980.

BUT DE L'EVALUATION :

Déterminer l'impact du programme sur le taux de criminalité des jeunes.

INDICATEURS RETENUS :

Nombre de délits déclarés par le sujet avant et après l'expérience.

CHOIX DE LA POPULATION :

Tous les jeunes délinquants (de juin à octobre 1980) en liberté surveillée issus de trois comtés du Kansas et destinés à suivre le programme JEP sont inclus dans l'expérience. Sur 53 au départ, 32 ont réellement participé à l'évaluation. Les sujets sont placés au hasard dans un groupe d'étude ou un groupe témoin (liste d'attente du JEP).

METHODE DE COLLECTE DE L'INFORMATION :

Les sujets sont interrogés avant et après l'expérience.

- questions générales (famille, école),
- déclaration du nombre de délits commis,
- deux questionnaires sur les valeurs morales admises par le sujet.

Les casiers judiciaires des sujets ont été consultés.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Analyses de variance et de covariance.

LES EFFETS DU PROGRAMME

LES TYPES D'EFFETS :

Augmentation ou diminution du nombre de délits commis.

Groupe d'étude/groupe témoin.

Petits délinquants/grands délinquants.

LES BIAIS : N.C.

LES CRITIQUES DU PROGRAMME :

On enregistre une hausse du nombre de délits chez les petits délinquants. Ce type de programme peut avoir une influence néfaste sur eux.

LES CRITIQUES DE L'EVALUATION :

Les échantillons étant petits, les résultats doivent être analysés avec prudence.

OBSERVATIONS :

Des évaluations de projets de réhabilitation existent depuis 1920.

La première évaluation de programmes du même type que JEP a été réalisée en 1970 (Brodsky).

REVUE : Evaluation Review

TITRE : Financing Alternatives in Low-Income Energy Assistance

AUTEURS : Karol Kerns

Volume 10 ; N° 4 ; 1986 ; p.434-454

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : Voluntary Heating Fuel Budget Plan (VHF)

POPULATION :

Les participants sont des volontaires bénéficiant des aides pour familles avec enfants à charge (AFDC).

OBJECTIFS DU PROGRAMME :

Eviter les arriérés dans le paiement des frais de chauffage dûs à la mauvaise gestion de leurs allocations par certaines familles. L'objectif est de lutter contre le risque de coupures de chauffage et les dépenses supplémentaires en aides d'urgence qu'il entraîne.

LOCALISATION-DATE-DUREE : Etats-Unis, Michigan 1978-1981 (3 ans)

CONTENU DU PROGRAMME :

Il s'agit d'un nouveau système de paiement des frais de chauffage. La particularité de VHF est que le paiement est déduit directement des allocations reçues par la famille. Dans le cadre des autres systèmes, la famille recevait une allocation mensuelle pour les frais de chauffage.

MAITRE D'OEUVRE : N.C.

MOYENS DU PROGRAMME : Le programme est financé par la sécurité sociale.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE : Michigan.

BUT DE L'EVALUATION :

Déterminer l'efficacité du programme (voir indicateurs).

INDICATEURS RETENUS :

- Caractéristiques des participants.
- Impact du programme sur la consommation d'énergie.

CHOIX DE LA POPULATION :

Les échantillons ont été choisis parmi les sujets ayant le droit de bénéficier de ce type de programme, ayant bénéficié de façon constante des AFDC et n'ayant pas changé de domicile (afin d'éviter qu'un changement dans la consommation ne soit dû à un changement de type de logement).

METHODE DE COLLECTE DE L'INFORMATION :

- Les dossiers de l'AFDC.
- La comptabilité des fournisseurs de gaz.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Tests, régression multiple.

Trois estimations différentes de coûts en l'absence de VHF ont été effectuées pour déterminer l'impact de ce programme sur les coûts des services sociaux.

LES EFFETS DU PROGRAMME

LES TYPES D'EFFETS :

Les plus grandes différences entre participants et non participants avant le lancement du projet concernent la consommation d'énergie et l'utilisation faite des allocations. Les participants consomment plus que les non participants. Le programme a attiré les familles ayant le plus de problèmes de paiement et faisant donc le plus appel aux aides d'urgence.

Des économies considérables ont été réalisées par les services sociaux grâce à la combinaison de l'ancien système et du système VHF.

LES BIAIS : N.C.

LES CRITIQUES DU PROGRAMME : N.C.

LES CRITIQUES DE L'EVALUATION : N.C.

OBSERVATIONS : N.C.

REVUE : Evaluation Review

TITRE : Cost-Effectiveness of Computer-Assisted Instruction.

AUTEURS : H M. Levin, G V. Glass, G R. Meister.

Volume 11 ; N° 1 ; 1987 ; p.50-72

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : N.C.

POPULATION : Enfants à l'école primaire.

OBJECTIFS DU PROGRAMME :

Améliorer le niveau en mathématiques et en lecture des élèves.

LOCALISATION-DATE-DUREE : Los Angelès, 1976-1980.

CONTENU DU PROGRAMME :

Enseignement assisté par ordinateur sous forme d'exercices d'entraînement et d'enseignement de matières.

Séances quotidiennes de dix minutes en mathématiques, lecture, expression.

MAITRE D'OEUVRE :

MOYENS DU PROGRAMME :

Le programme a été financé par l'"Educational Testing Service" et le "Los Angelès Unified School District".

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE : Los Angelès, 1976-1980.

EVALUATEURS : N.C.

BUT DE L'EVALUATION :

Comparer l'enseignement assisté par ordinateur à d'autres méthodes:

- 1) - Entraide : les plus âgés aident les plus jeunes ou aide apportée par un adulte.
- 2) - Réduction de la taille des classes
- 3) - Allongement de l'année et de la journée scolaires.

INDICATEURS RETENUS :

- Résultats en mathématiques et en lecture.
- Pour l'évaluation du coût et du rendement :
 - * coût d'application par élève de chacune de ces politiques sur un nouveau site.
 - * taux de rendement : impact sur les résultats en mathématiques et lecture de 100\$ dépensés par élève et par matière.

CHOIX DE LA POPULATION : N.C.

METHODE DE COLLECTE DE L'INFORMATION :

Observation des résultats des élèves pendant les séances de dix minutes.

Pour la réduction de la taille des classes et l'augmentation du nombre d'heures, des travaux déjà parus sur le sujet ont été consultés.

Pour évaluer la méthode d'entraide un programme de ce type a été réalisé dans l'Idaho.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES : N.C.

LES EFFETS DU PROGRAMME**LES TYPES D'EFFETS :**

L'enseignement assisté par ordinateur est en première place pour l'amélioration des résultats en mathématiques suivi de la méthode d'entraide entre élèves. Cette dernière est la plus efficace pour l'enseignement de la lecture.

L'entraide entre élèves a une efficacité par rapport aux coûts bien supérieure à celle de l'enseignement assisté par ordinateur.

L'allongement de l'année scolaire ne semble pas être une solution rentable.

LES BIAIS : N.C.

LES CRITIQUES DU PROGRAMME : N.C.

LES CRITIQUES DE L'EVALUATION : N.C.

OBSERVATIONS : N.C.

REVUE : Evaluation Review

TITRE : The Effects of Youth Employment Program Participation on Later Employment.

AUTEURS : R.B. Ekstrom, N.E. Freeberg, D.A. Rock.

Volume 11 ; N° 1 ; 1987 ; p. 84-101

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : N.C.

POPULATION : Jeunes défavorisés.

OBJECTIFS DU PROGRAMME :

Préparer des jeunes défavorisés à la recherche d'un emploi.

LOCALISATION-DATE-DUREE : Etats-Unis, 370 sites. Lancé en 1979.

CONTENU DU PROGRAMME :

10 types différents de stages selon le site. Information, entraînement à la recherche d'un emploi, stages, très peu d'enseignement fondamental.

MAITRE D'OEUVRE : US Department of Labor (Ministère du travail).

MOYENS DU PROGRAMME : Financé par le Ministère du travail.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE :

Faute de moyens financiers, seule la première phase de l'évaluation a pu être effectuée : 1982, Etats-Unis, 370 sites.

EVALUATEURS : N.C.

BUT DE L'EVALUATION :

Déterminer si ce type de programme permet à de jeunes adultes de trouver et de garder un emploi. Aider à l'élaboration de programmes de ce type.

INDICATEURS RETENUS :

- Nombre de mois de travail effectués pendant les trois dernières années
- Salaire horaire actuel
- Degré de complexité du travail (échelle de 0 à 10)
- Eventuels échecs dans la recherche d'un emploi pendant la période de trois ans et raison de ces échecs
 - Manière de chercher un emploi
 - Satisfaction quant à leur emploi actuel ou leur dernier emploi
 - Statut professionnel

CHOIX DE LA POPULATION :

L'échantillon total était de 39 000 sujets. L'échantillon étudié est de 419 participants ayant terminé le stage en été 1979 (étude trois ans après).

Groupe témoin : 356 sujets présentant des caractéristiques aussi proches que possible de celles des participants

METHODE DE COLLECTE DE L'INFORMATION : Les sujets ont été interrogés.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

- Calcul de coefficients de regression
- Variables explicatives : âge, origine ethnique, situation économique de la famille, sexe, niveau de lecture avant le stage, taux de chômage des jeunes dans la région, appartenance au groupe des participants ou au groupe témoin
- Variables dépendantes : 5 indicateurs (statut professionnel actuel, nombre de mois de travail, degré de complexité du travail, salaire horaire, satisfaction).

LES EFFETS DU PROGRAMME

LES TYPES D'EFFETS :

Les participants ont été employés plus longtemps pendant les trois années étudiées que les sujets du groupe témoin.

De tels stages peuvent avoir des effets sensibles sur l'emploi. Cependant, ces effets sont faibles comparés à ceux du manque de niveau de formation, des origines, de la situation de l'emploi dans la région.

LES BIAIS :

Les trois années de suivi ont correspondu à une période de récession.

LES CRITIQUES DU PROGRAMME :

Il donne trop peu d'importance à l'enseignement de base.

LES CRITIQUES DE L'EVALUATION :

Les différents types de programmes étudiés étaient très inégaux.

OBSERVATIONS : N.C.

REVUE : Evaluation Review

TITRE : How Fair are Water Tariffs ? Some Evidence from California.

AUTEURS : W. Douglas Morgan.

Volume 11 ; N° 1 ; 1987 ; p.143-150

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : N.C.

POPULATION : N.C.

OBJECTIFS DU PROGRAMME : N.C.

LOCALISATION-DATE-DUREE : Californie.

CONTENU DU PROGRAMME : N.C.

MAITRE D'OEUVRE : N.C.

MOYENS DU PROGRAMME : N.C.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE :

Californie. Comparaison entre le début des années 70 et le début des années 80.

EVALUATEURS : N.C.

BUT DE L'EVALUATION :

Analyser l'échelle de tarification de fournisseurs d'eau courante sur dix années afin de déterminer si celle-ci a évolué vers plus de justice.

INDICATEURS RETENUS :

Ecarts maxima de prix.

Coefficient de Gini.

Variation du prix moyen sur la période de dix ans.

CHOIX DE LA POPULATION :

13 fournisseurs d'eau en Californie ayant pu fournir des données utilisables par les évaluateurs. L'échantillon dont rien ne prouve qu'il est représentatif des fournisseurs d'eau Californiens est cependant varié (entreprises privées, publiques, ...).

METHODE DE COLLECTE DE L'INFORMATION :

Les archives de ces 13 entreprises ont été consultées : consommation d'eau, chiffre d'affaires par type de consommateur du début des années 70 au début des années 80.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Courbe de Lorenz.

Ordonnées : pourcentages cumulés de la consommation d'eau.

Abcisses : pourcentages cumulés du chiffre d'affaires.

Calcul du coefficient de Gini pour chaque fournisseur au début des années 70 et au début des années 80.

LES EFFETS DU PROGRAMME

RESULTATS :

On constate une diminution du coefficient de Gini chez tous les fournisseurs (cette diminution varie de 15 % à 16 %).

Tous ont augmenté leurs prix moyens de 50 % pendant la période considérée (la hausse des prix à la consommation sur ces dix années a été de 106 %).

Les écarts maxima entre les différents prix de chaque fournisseur ont diminué.

Les échelles de tarification sont donc devenues plus "justes".

LES BIAIS : N.C.

LES CRITIQUES DU PROGRAMME : N.C.

LES CRITIQUES DE L'EVALUATION :

Les différences de coût marginal (du fait de la distance, de l'altitude) n'ont pas été prises en compte.

Dans chaque catégorie les consommateurs sont considérés comme équivalents (propriétaires, locataires...)

OBSERVATIONS : N.C.

REVUE : Evaluation Review

TITRE : Lessons from the Delaware Dislocated Worker Pilot Program

AUTEURS : Howard S. Bloom (New-York University)

Volume 11 ; N° 2 ; 1987 ; p.157-177

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : "Retraining Delaware's Dislocated Workers"

POPULATION :

Ouvriers ayant perdu leur emploi dans des secteurs victimes du progrès technique et de la concurrence internationale (voir page suivante "choix de la population").

OBJECTIFS DU PROGRAMME :

Programme expérimental, destiné à fournir des informations pour l'organisation des programmes issus du "Job Training Partnership Act", voté par le Congrès, et à augmenter les revenus des participants.

LOCALISATION-DATE-DUREE :

Etats-Unis, Etat du Delaware, janvier-juillet 1983, 5 mois.

CONTENU DU PROGRAMME :

Stage de formation aux techniques de recherche d'un emploi.

Conseils individualisés.

Recherche d'un emploi.

Stage de reconversion si nécessaire.

MAITRE D'OEUVRE : Delaware Department of Labor.

MOYENS DU PROGRAMME : N.C.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE : Etat du Delaware janvier-juillet 1983, 5 mois.

EVALUATEURS : L'auteur de l'article.

BUT DE L'EVALUATION :

INDICATEURS RETENUS :

Assurances chômage perçues.

Revenus.

CHOIX DE LA POPULATION :

Personnes ayant bénéficié d'assurances chômage pendant 7 à 12 semaines de suite, ayant accompli au moins dix ans de scolarité, ayant accès à un moyen de transport, n'ayant pas de chance d'être réemployées par leur ancien employeur. 175 personnes avaient le droit de bénéficier du programme ; 65 d'entre elles ont été choisies au hasard pour y participer. Les autres ont servi de groupe témoin.

METHODE DE COLLECTE DE L'INFORMATION :

Consultation des fichiers du "Delaware Division of Unemployment Insurance" (Assurances chômage du Delaware).

L'impact du programme a été mesuré 2, 5, 8 et 11 mois après son lancement.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES : N.C.

LES EFFETS DU PROGRAMME

LES TYPES D'EFFETS :

- Les assurances chômage perçues ne varient pas entre le groupe témoin et les participants sauf pendant le premier trimestre où elles sont supérieures pour ces derniers (probablement du fait du temps passé en séances de conseil et de formation à la recherche d'un emploi).

- Il semble que tout au long de la période d'étude les revenus des participants ont été inférieurs à ceux des sujets du groupe témoin. Cependant les différences ne sont pas significatives.

- 80 % des participants ont trouvé un emploi , 16 (25 %) des personnes choisies n'ont pas ou très peu participé au programme.

LES BIAIS : N.C.

LES CRITIQUES DU PROGRAMME :

Les conseillers étaient trop nombreux et mal préparés. Une plus grande part de responsabilité devrait être exigée des participants dans la recherche d'un emploi.

LES CRITIQUES DE L'EVALUATION :

Les raisons de désistement ou de faible participation devraient être examinées de plus près.

Cette évaluation concerne un programme récent. Les faibles résultats sont en partie dus aux difficultés de mise en route.

OBSERVATIONS : N.C.

REVUE : Evaluation Review

TITRE : Applying Theories of Aging to Evaluation at Social Programs for the Elderly

AUTEURS : Howard Litwin

Volume 11 ; N° 3 ; 1987 ; p.267-280

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : N.C.

POPULATION : Personnes âgées

OBJECTIFS DU PROGRAMME :

Fournir divers services et proposer des activités aux personnes âgées.

LOCALISATION-DATE-DUREE :

Israël, centres sociaux créés à la fin des années 60.

CONTENU DU PROGRAMME :

Ces centres sociaux proposent des activités culturelles pour tous les âges. Cette évaluation concerne les services destinés aux personnes âgées :

I - programmes sociaux et culturels.

II - bénévolat de personnes âgées et pour les personnes âgées

III - services spécialisés

MAITRE D'OEUVRE : N.C.

MOYENS DU PROGRAMME :

Chaque centre est autonome bien qu'il existe un organisme central de coordination.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE : Israël

EVALUATEURS : H. Litwin (auteur de l'article)

L'évaluation lui a été demandée par l'Association israélienne des centres sociaux.

BUT DE L'EVALUATION :

Cette évaluation est fondée sur deux théories du vieillissement permettant d'élaborer une classification des centres :

1) Théorie du désengagement : les personnes âgées ont moins de relations sociales, moins de responsabilités. Elles gagnent en liberté.

2) Théorie de la vieillesse active : une personne âgée doit jouer un rôle dans la société.

INDICATEURS RETENUS :

Pour chaque centre, l'ampleur de chaque type de services ou d'activités (I, II, et III) proposés.

NB - I, II et III relèvent respectivement de la théorie 1); de la théorie 2), des deux à la fois.

3 critères sont retenus :

- nombre de participants au type de programme étudié,
- taux de participation,
- diversité des activités

et services proposés par cette catégorie de programme.

CHOIX DE LA POPULATION :

Les 60 centres qui proposaient des programmes pour personnes âgées.

METHODE DE COLLECTE DE L'INFORMATION :

Il a été demandé au personnel de ces centres de faire un inventaire des activités proposées. Données sur les services offerts aux personnes âgées en 1983-84.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Calcul de coefficients de régression pour les facteurs déterminant l'ampleur d'un type de programme :

- âge du programme,
- nombre de personnes âgées dans la région,
- existence d'un autre club du troisième âge dans la région (coefficient négatif).

LES EFFETS DU PROGRAMME

LES TYPES D'EFFETS :

Les programmes de type I sont les plus développés. Les centres sociaux d'Israël relèvent donc plutôt de la théorie de la vieillesse active.

LES BIAIS : N.C.

LES CRITIQUES DU PROGRAMME : N.C.

LES CRITIQUES DE L'EVALUATION : N.C.

OBSERVATIONS :

L'intérêt de cette évaluation est de montrer l'avantage qu'il peut y avoir à partir d'une base théorique afin de pouvoir dégager des critères d'évaluation.

REVUE : Evaluation Review

TITRE : Cost Effectiveness of the APROFAM.
Program for Voluntary Surgical Contraception in Guatemala

AUTEURS : M.E. Mc Bride, J.T. Bertrand, R. Santiso, Victor Hugo Fernandez

Volume 11 ; N° 3 ; 1987 ; p.300-326

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : Programa consolidado

POPULATION : N.C.

OBJECTIFS DU PROGRAMME : (Voir contenu)

NB - Le taux de croissance naturelle de la population au Guatemala était de 3,5 % en 1985.

LOCALISATION-DATE-DUREE : Guatemala, depuis 1973.

CONTENU DU PROGRAMME :

Stérilisations chirurgicales volontaires effectuées :

* au Guatemala au centre chirurgical de la capitale, dans des cliniques privées.

* A l'intérieur du pays :

- dans une clinique APROFAM, par une équipe APROFAM ambulante,

- dans une clinique privée par une équipe privée ambulante.

MAITRE D'OEUVRE : Asociacion Pro Bienestar de la Familia (APROFAM).

MOYENS DU PROGRAMME :

Financé en majeure partie par l'association pour la stérilisation chirurgicale volontaire.

Le nombre de services appliquant le programme pendant la période d'étude variait de 23 à 32.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE : Guatemala. Etude de la période 1979-1984.

BUT DE L'EVALUATION :

Comparer la rentabilité des différents types de services appliquant le programme afin d'orienter le développement futur de celui-ci.

INDICATEURS RETENUS :

- Coût par opération.
- Coût moyen, coût marginal, économie d'échelle.

CHOIX DE LA POPULATION : N.C.

METHODE DE COLLECTE DE L'INFORMATION : N.C.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Régression multiple : relation entre le coût total mensuel et le nombre d'opérations effectuées dans la capitale et à l'intérieur du pays.

LES EFFETS DU PROGRAMME

LES TYPES D'EFFETS :

Les équipes ambulantes et issues de cliniques privées comptent pour la majeure partie le nombre d'opérations effectuées. Elles ont un coût beaucoup plus élevé que les autres services.

Le coût moyen est décroissant, le coût marginal croissant.

Le programme dans son ensemble réalise des économies d'échelle.

LES BIAIS : N.C.

LES CRITIQUES DU PROGRAMME : N.C.

LES CRITIQUES DE L'EVALUATION : N.C.

OBSERVATIONS :

L'article ne rend compte que d'une partie d'un plus vaste projet d'évaluation.

Une des questions que devait résoudre cette évaluation était de savoir s'il fallait remplacer les services ambulants par des cliniques. Bien que leur coût soit plus élevé il a été décidé de garder les premiers du fait de l'avantage de la proximité.

REVUE : Evaluation Review

TITRE : Training the Disadvantaged

AUTEURS : Laurie J. Bassi

Volume 11 ; N° 4 ; 1987 ; p.493-509

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME :

Programmes issus du CETA ("Comprehensive Employment and Training Act").

POPULATION : Personnes économiquement défavorisées.

OBJECTIFS DU PROGRAMME :

Rendre économiquement indépendantes des personnes défavorisées.

LOCALISATION-DATE-DUREE : Etats-Unis.

CONTENU DU PROGRAMME : Formation professionnelle.

MAITRE D'OEUVRE : N.C.

MOYENS DU PROGRAMME : N.C.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE :

Etats-Unis, l'étude concerne les années 1976-1978.

EVALUATEURS :

Etude demandée par l'"U.S. Department of Labor" à l'Urban Institute.

BUT DE L'EVALUATION :

Déterminer si ce type de formation est rentable, si elle réduit la nécessité pour les participants de faire appel à des aides sociales. Préparer l'élaboration de la base de données d'un autre programme de ce type : le JTPA, Job Training Partnership Act.

INDICATEURS RETENUS :

Les économies en aides sociales réalisées

CHOIX DE LA POPULATION :

Un échantillon d'individus économiquement défavorisés a été choisi dans la liste du "CLMS", d'une part, et dans celle du "PSID" d'autre part. (La définition de sujet économiquement défavorisé n'est pas exactement la même dans les deux sources).

METHODE DE COLLECTE DE L'INFORMATION :

Données du "Continuous Longitudinal Manpower Survey" (CLMS) (participants au CETA).

Les données du "Panel Study of income Dynamics" PSID ont servi de comparaison.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Deux relations ont été établies :

- l'une pour déterminer à quel point le CETA a réduit pour un sujet la probabilité de devoir dépendre d'aides sociales :

$$D_{it} = \beta_0 + X_{it} \beta_1 + P_i \beta_3 + \mu_{it}$$

D : variable dummy 1 : le sujet perçoit des aides sociales
 0 : le sujet ne perçoit pas d'aides sociales

X : vecteur de caractéristiques déterminant les revenus potentiels de l'individu (niveau d'études, problèmes de santé : 1/0, âge ...),

Y : vecteur de caractéristiques de l'individu déterminant les aides sociales potentielles.

P : variable dummy 1 : participant
 0 : le sujet appartient au groupe témoin
 (sujet i, temps t) : terme d'erreur.

- l'autre pour estimer les économies en aides sociales réalisées

$$B_{it} = \beta_0 + X_{it} \beta_1 + Y_{it} \beta_2 + P_i \beta_3 + \mu_{it}$$

LES EFFETS DU PROGRAMME**LES TYPES D'EFFETS :**

Il semble que la participation au CETA aide les femmes à s'approcher d'une situation d'indépendance économique sans toutefois l'atteindre.

Les économies en aides sociales pour les hommes sont insignifiantes.

Les sujets du groupe témoin hommes et femmes confondus ont perçu en moyenne plus d'aides sociales.

LES BIAIS :

Les listes du PSID comportent certainement des sujets ayant, dans le passé, participé au CETA.

LES CRITIQUES DU PROGRAMME : N.C.

LES CRITIQUES DE L'EVALUATION :

Les données du CLMS n'indiquent pas l'Etat de résidence du sujet, ce qui rend impossible un calcul du montant précis des aides sociales perçues par chaque individu.

OBSERVATIONS : N.C.

REVUE : Evaluation Review

TITRE : Economic Evaluation of a Federal Natural Resources Program : The Case of the Foresty Incentives program

AUTEURS : P.V. Elletson, C.D. Risbrudt

Volume 11 ; N° 5 ; 1987 ; p.660-669

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : Foresty incentives Program.

POPULATION :

Surfaces boisées destinées à la vente de bois possédées par des intérêts privés à caractère non industriel (58 % des surfaces boisées aux Etats-Unis). Certaines conditions sont requises pour pouvoir bénéficier du programme.

OBJECTIFS DU PROGRAMME :

Remédier au manque d'investissement sur ce type d'exploitations afin d'assurer un bon approvisionnement du pays en bois à long terme.

LOCALISATION-DATE-DUREE : Etats-Unis.

CONTENU DU PROGRAMME :

Participation de l'Etat pouvant atteindre 65 % des coûts de plantation d'arbres et d'amélioration des surfaces boisées.

MAITRE D'OEUVRE : L'Etat

MOYENS DU PROGRAMME :

Investissements annuels de l'Etat de 10 à 15 millions de Dollars.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE : Etats-Unis, 1987.

BUT DE L'EVALUATION :

Evaluer la rentabilité financière des investissements effectués en 1979.

INDICATEURS RETENUS :

- Taux de rendement interne.
 - Valeur actualisée des impôts sur les ventes à venir.
- Le rendement avec et sans le programme a été estimé.

CHOIX DE LA POPULATION :

Un échantillon de 676 cas a été choisi parmi 8 622 exploitations ayant eu accès au programme en 1979.

METHODE DE COLLECTE DE L'INFORMATION : N.C.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

- Taux de rendement interne de l'investissement public et privé (8,6 %),
- Valeur actualisée nette des ventes calculée avec un taux d'intérêt de 4 %.
- Valeur actuelle des impôts (au taux de 10 %).

LES EFFETS DU PROGRAMME

LES TYPES D'EFFETS :

Le taux de rendement interne de la somme de l'investissement public et privé est de 8,6 %.

Le revenu pour l'Etat est évalué à 2,5 fois l'investissement annuel de l'Etat pour ce programme.

Ce programme a permis la survie d'exploitations en difficulté.

LES BIAIS : N.C.

LES CRITIQUES DU PROGRAMME : N.C.

LES CRITIQUES DE L'EVALUATION :

Ce type d'évaluations présente certaines difficultés : prise en compte des risques à long terme (incendies ...), difficulté dans la collecte des données du fait de l'éparpillement des exploitations, participation de nombreux organismes à la gestion du programme.

OBSERVATIONS : N.C.

REVUE : Evaluation Review

TITRE : An Evaluation of an Organization Development approach to Reducing School disorder.

AUTEURS : Denise C. Gottfredson

Volume 11 ; N° 6 ; 1987 ; p.739-760

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME :

Les mesures prises pour atteindre les objectifs du programme s'inspirent d'une méthode précise : "Program Development Evaluation".

POPULATION :

Deux écoles secondaires de Baltimore ayant de graves problèmes de discipline, de violence ... Les deux écoles sont situées dans des quartiers pauvres.

OBJECTIFS DU PROGRAMME : Atténuer les problèmes de discipline.

LOCALISATION-DATE-DUREE :

Etats-Unis, Baltimore, 3 ans à partir de l'automne 1982.

La première année était une année de préparation du programme.

CONTENU DU PROGRAMME :

- Nouvelles méthodes de maintien de la discipline en classe selon des techniques précises comportant des réunions avec les élèves, des récompenses matérielles et morales.

- Nouvelles méthodes d'enseignement : compétitions entre des équipes d'élèves.

- Autres aspects : information régulière des parents, interventions de personnes extérieures venant parler de leur profession.

MAITRE D'OEUVRE : N.C.

MOYENS DU PROGRAMME : N.C.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE : Baltimore 1982-1985.

BUT DE L'EVALUATION : N.C.

INDICATEURS RETENUS :

- Organisation de l'école : moral des enseignants, administration ...
- Indiscipline, délinquance, convocations au bureau, renvois ...
- Attitude des élèves : intégration, attachement à l'école et aux enseignants, récompense, leurs attentes quant au déroulement de leurs études.

CHOIX DE LA POPULATION :

Seule une des deux écoles a réellement appliqué le programme. L'autre a servi de témoin.

A l'intérieur de la première école certaines classes ont servi de groupe expérimental et ont été suivies pendant les trois années.

METHODE DE COLLECTE DE L'INFORMATION :

Essentiellement par des enquêtes auprès de tous les enseignants et élèves de l'école.

Les mesures ont été effectuées avant le lancement du programme, puis un an et deux ans après.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES : N.C.

LES EFFETS DU PROGRAMME

LES TYPES D'EFFETS :

On enregistre une baisse de la délinquance dans toutes les classes de la première école et une forte amélioration de son organisation.

Les attentes des élèves quant au déroulement de leurs études sont devenues moins ambitieuses (résultat inattendu).

Dans la première école les différences entre classes expérimentales et classes témoin sont faibles.

LES BIAIS : N.C.

LES CRITIQUES DU PROGRAMME : N.C.

LES CRITIQUES DE L'EVALUATION : N.C.

OBSERVATIONS : N.C.

REVUE : Evaluation Review

TITRE : The Mood River Conservation Project : An Evaluator's Dream

AUTEURS : Eric Hirst

Volume 12 ; N° 3 ; 1988 ; p.310-325

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : Mood River Conservation Project (MRCP).

POPULATION :

Les habitations de Mood River, équipées d'un système de chauffage électrique pouvant bénéficier de ce programme (3 500).

OBJECTIFS DU PROGRAMME :

Programme expérimental de recherche sur les économies d'énergie.

LOCALISATION-DATE-DUREE : Mood River, Oregon - 1982-1986, 5 ans.

CONTENU DU PROGRAMME :

Il comporte deux volets :

- Conseils et installation d'équipements dans des foyers de Mood River.
- Collecte de données.

MAITRE D'OEUVRE : Natural Resources Defense Council

MOYENS DU PROGRAMME :

Il a été financé par le "Bonneville Power Administration" (20 millions de Dollars).

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE : Mood River, 1982-1986.

BUT DE L'EVALUATION :

Apporter des informations sur l'efficacité des programmes d'économies d'énergie de ce type.

INDICATEURS RETENUS :

- Nombre de foyers ayant participé.
- Nombre d'équipements conseillés réellement installés par les participants.
- Economies d'électricité réalisées.

CHOIX DE LA POPULATION :

Groupe d'étude : les 3 500 habitations de Mood River bénéficiant du programme.

Groupe témoin : deux communes comparables à Mood River.

METHODE DE COLLECTE DE L'INFORMATION :

- Entretiens avec des habitants de Mood River et des communes témoin et envoi de questionnaires sur des échantillons choisis au hasard.
- Entretiens avec les fournisseurs d'électricité et des contractants du programme.
- Des données météorologiques ainsi que des données sur la consommation de bois ont été recueillies.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES : N.C.

LES EFFETS DU PROGRAMME

LES TYPES D'EFFETS :

On note une participation très élevée (85 % des foyers pouvant bénéficier de ce programme ont installé des équipements).

80 % des équipements conseillés ont été installés.

Les économies d'électricité ont été très inférieures à ce qui était prévu. Plusieurs raisons sont invoquées : baisse de la consommation d'électricité avant le projet du fait d'une hausse du prix de l'électricité, changements de comportement après le programme (moins de consommation de bois et hausse de la température pour plus de confort).

LES BIAIS : N.C

LES CRITIQUES DU PROGRAMME : N.C.

LES CRITIQUES DE L'EVALUATION : N.C.

OBSERVATIONS :

Certaines particularités de la méthode d'évaluation ont contribué à la qualité de cette étude :

- la collecte des données a commencé une année avant le lancement du projet et s'est terminée plus d'une année après l'installation des équipements,
- la variété et la quantité des données,
- la vérification constante de la qualité des données,
- la coopération entre divers organismes

REVUE : Evaluation Review

TITRE : Multiple Questions require Multiple Designs. An Evaluation of the 1981 Changes to the AFDC Program.

AUTEURS : T.E. Hedrick, S.L. Snipman

Volume 12 ; N° 4 ; 1988 ; p.427-448

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME :

Modifications apportées à l'AFDC en 1981 par l'"Omnibus Reconciliation Act" (OBRA)

POPULATION : Les bénéficiaires des AFDC.

OBJECTIFS DU PROGRAMME :

- Réduire le coût de l'AFDC.
- Redistribuer ses ressources au profit des plus pauvres.
- Eliminer une injustice : les familles de l'AFDC qui commençaient à travailler pouvaient continuer à bénéficier des aides, tandis que d'autres familles dans la même situation financière ne pouvaient y avoir accès.
- Décourager la "dépendance" vis-à-vis des aides sociales.

LOCALISATION-DATE-DUREE : Etats-Unis, 1981.

CONTENU DU PROGRAMME : 22 modifications à l'AFDC.

MAITRE D'OEUVRE : N.C.

MOYENS DU PROGRAMME : N.C.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE :

Etats-Unis. Le rapport final a été publié en 1985.

EVALUATEUR :

U.S. General Accounting Office (GAO) à la demande du "House Committee on Ways and Means"

BUT DE L'EVALUATION :

Apporter une réponse aux questions suscitées par ce changement :

1 : au niveau national les modifications vont-elles diminuer le nombre de bénéficiaires et réduire les coûts ?

2 : Les modifications vont-elles avoir un effet pervers en encourageant les bénéficiaires à ne pas travailler pour conserver leur droit aux allocations AFDC et surtout à l'assistance "Medicaid" !

3 : Les modifications ont-elles un impact sur le bien-être des familles ayant perdu le bénéfice de l'AFDC.

INDICATEURS RETENUS :

1 : Nombre de familles bénéficiaires de 1973 à 1984.

- total national des paiements d'allocations AFDC de 1973 à 1984.

2 : Nombre de personnes travaillant/ne travaillant pas admises ou ayant quitté l'AFDC avant/après les changements institués par l'OBRA.

3 : Revenus des personnes ayant perdu le bénéfice des aides de l'AFDC (avant/après).

CHOIX DE LA POPULATION :

- 1 : Etude au niveau national.
- 2 : Etude faite dans les Etats ayant recueilli des données suffisantes.
- 3 : Echantillon de familles n'ayant plus droit aux allocations AFDC du fait des modifications de 1981 (sur 5 sites).

METHODE DE COLLECTE DE L'INFORMATION :

- 1 : Données nationales.
- 2 : Données des Etats étudiés.
- 3 : Entretiens avec les familles.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

- 2 : ARIMA - (Moyennes mobiles)

LES EFFETS DU PROGRAMME

LES TYPES D'EFFETS :

- Les modifications apportées à l'AFDC ont entraîné les diminutions du nombre de bénéficiaires et des coûts attendus.
- Rien n'indique qu'elles aient incité les personnes bénéficiant d'aides sociales à ne pas travailler.
- Les familles ayant perdu le bénéfice de l'AFDC ont constaté une forte diminution de leur revenu et des difficultés dues au fait qu'elles perdent automatiquement tout droit à l'assistance "Medicaid".!

LES BIAIS : N.C.

LES CRITIQUES DU PROGRAMME :

Certains Etats ont retardé l'application des modifications ou augmenté leur seuil de pauvreté.

LES CRITIQUES DE L'EVALUATION : N.C.

OBSERVATIONS :

Dès la publication des premiers résultats de l'évaluation, en 1984, le Congrès a voté une hausse du seuil de revenu donnant droit aux allocations de l'AFDC et a décidé d'instaurer une période de transition pendant laquelle les personnes ne bénéficiant plus de l'AFDC sont couvertes par l'assistance "Medicaid".

REVUE : Evaluation Review

TITRE : Evaluation of a Drug Abuse Treatment

AUTEURS : Yih-Ing Hser, M.D. Anglin, Chih-Ping Chou

Volume 12 ; N° 5 ; 1988 ; p. 547-570

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : Methadone Maintenance

POPULATION :

Toxicomanes de plus de 18 ans, drogués depuis au moins deux ans et s'étant déjà inscrits au moins deux fois à d'autres types de traitements.

OBJECTIFS DU PROGRAMME : N.C.

LOCALISATION-DATE-DUREE : Etats-Unis, introduit dans les années 60.

CONTENU DU PROGRAMME :

Traitement pour toxicomanes. Methadone : produit de remplacement de l'héroïne.

MAITRE D'OEUVRE : N.D.

MOYENS DU PROGRAMME : N.D.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE : Des cliniques du sud de la Californie, 1978-1981.

EVALUATEURS : N.C.

BUT DE L'EVALUATION :

Déterminer l'impact du traitement en fonction de certaines caractéristiques des patients : sexe, origine ethnique, degré de participation au traitement (variable indépendante des deux premières). Apporter des améliorations aux évaluations de ce type.

INDICATEURS RETENUS :

- délits, trafic de drogue...
- consommation de drogues diverses
- relations sociales, mariage, consommation de drogue par le conjoint, participation du conjoint au traitement.
- sources de revenu : emploi, perception d'allocations.

CHOIX DE LA POPULATION :

720 toxicomanes ayant commencé le traitement de 1971 à 1978 (coupe transversale sur deux listes de clients). 251 hommes "blancs", 283 femmes "blanches", 141 hommes d'origine mexicaine, 45 femmes d'origine mexicaine.

METHODE DE COLLECTE DE L'INFORMATION :

Un interrogatoire rétrospectif (adaptation d'un schéma de Nurco (1975), décrit par Mc Glothlin (1977)).

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Tous les indicateurs sont mesurés par le pourcentage moyen de temps passé dans chacune des situations qu'ils représentent, pour chacune des trois périodes prises en compte :

- du début de la consommation quotidienne d'héroïne au premier traitement
- l'ensemble des périodes de traitement du premier traitement à l'interrogatoire
- l'ensemble des périodes hors traitement du premier traitement à l'interrogatoire.

Analyses de variance pour chaque indicateur.

LES EFFETS DU PROGRAMME

LES TYPES D'EFFETS :

Globalement, on enregistre des progrès significatifs pendant les périodes de traitement indépendamment du sexe, de l'origine ethnique ou du degré de participation au programme. Cette amélioration se maintient en partie pendant les périodes hors traitement.

Pour certains indicateurs les résultats varient en fonction du sexe et de l'origine ethnique (par exemple, les patients d'origine mexicaine ont plus de chance de revenir à une consommation quotidienne de drogue).

LES BIAIS : N.C.

LES CRITIQUES DU PROGRAMME : Une plus grande participation devrait être encouragée.

LES CRITIQUES DE L'EVALUATION :

Cette étude se limite à des cas graves. Les résultats ne doivent pas être généralisés .

OBSERVATIONS :

Cette expérience apporte des éléments nouveaux par rapport aux méthodes des évaluations précédentes dans ce domaine :

- La période de référence avant traitement est plus longue.
- L'effet cumulatif des différentes périodes de traitement est évalué plutôt que l'effet de périodes isolées.
- Les comparaisons avant/pendant le traitement ne tenaient pas compte d'autres sources d'évolution (mûrissement de l'individu avec l'âge...) liées au temps.

REVUE : Evaluation Review

TITRE : Lifeline Rates in California Pricing Electricity to Attain Social Goals

AUTEURS : M. Hennessy, D.M. Keane

Volume 13 ; N° 2 ; 1989 ; p.123-140

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME :

Nom de la méthode de tarification de l'électricité : "Lifeline Electric Rates" appelés ultérieurement "Baseline electric Rates".

POPULATION : Les consommateurs individuels d'électricité en Californie.

OBJECTIFS DU PROGRAMME :

Eviter le gaspillage auquel conduisaient les tarifs dégressifs en vigueur à l'époque.

Fournir une quantité d'électricité, définie comme le minimum nécessaire aux besoins essentiels, à un tarif avantageux.

Aider les ménages à bas revenus.

LOCALISATION-DATE-DUREE : Californie, depuis 1975.

CONTENU DU PROGRAMME :

La première tranche, à laquelle est appliquée le tarif réduit, est calculée en pourcentage de la consommation moyenne d'électricité des particuliers (de 50 à 70 %).

Le tarif appliqué à la deuxième tranche est supérieur à la moyenne des tarifs pour compenser la perte.

Ce système a pour effet de faire subventionner les particuliers petits consommateurs d'électricité par d'autres particuliers qui en consomment plus.

MAITRE D'OEUVRE : N.C.

MOYENS DU PROGRAMME :

Ce système de tarification a été introduit par la législature de l'Etat de Californie.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE :

Californie. Les données utilisées concernent la période novembre 1985-octobre 1986

BUT DE L'EVALUATION :

Déterminer l'effet de ce système de tarification sur les bas revenus.

INDICATEURS RETENUS :

Différences entre les montants de factures d'électricité avec le système existant d'une part, et avec l'un ou l'autre des systèmes simulés, d'autre part.

CHOIX DE LA POPULATION : N.C.

METHODE DE COLLECTE DE L'INFORMATION :

- Consultation des factures d'électricité des particuliers (11/85-10/86).
- Enquête effectuée en 1985 auprès des clients (nombre d'appareils ménagers, informations démographiques ...).

Echantillon représentatif des clients de Pacific Gas and Electricity Company.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Deux autres systèmes de tarification, tenant compte plus directement des ménages à bas revenus, ont été simulés :

Systeme A : Pour la première tranche le tarif réduit n'est appliqué qu'aux clients à bas revenus. Un tarif constant est appliqué au delà quel que soit le niveau de vie du client. Ce tarif est inférieur au tarif appliqué à la deuxième tranche en Californie.

Systeme B : Un tarif constant est appliqué quelle que soit la consommation, plus un forfait mensuel non payable par les ménages à bas revenus. Le tarif de B est situé entre celui de A et le tarif existant.

LES EFFETS DU PROGRAMME**LES TYPES D'EFFETS :**

Les effets sont analysés en fonction du revenu et de la consommation d'électricité des ménages : bas revenus ou non, consommation comprise dans la première tranche/consommation supérieure à la première tranche.

LES BIAIS : N.C.

LES CRITIQUES DU PROGRAMME :

Cette aide sociale est financée par des particuliers, les entreprises n'interviennent pas.

Les exigences d'économies d'énergie ne sont plus les mêmes que dans les années 70 : le tarif appliqué au delà de la première tranche est très supérieur aux coûts marginaux actuels.

LES CRITIQUES DE L'EVALUATION : N.C.

OBSERVATIONS : N.C.

REVUE : Evaluation Review

TITRE : An Evaluation of an Undergraduate Course on AIDS

AUTEURS : P.R. Abramson, J.C. Sekler, R. Berk, M.Y. Cloud

Volume 13 ; N° 5 ; 1989 ; p.516-532

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : "AIDS : The Modern Plague"

POPULATION : Etudiants

OBJECTIFS DU PROGRAMME :

- Diminuer les risques de contamination.
- Informer les étudiants sur les politiques publiques concernant le SIDA.
- Encourager la tolérance envers les personnes séropositives ou atteintes du SIDA.

LOCALISATION-DATE-DUREE :

University of California, Los Angelès (UCLA), 2ème trimestre 1987-88 - 10 semaines.

CONTENU DU PROGRAMME : Une série de cours sur le SIDA.

MAITRE D'OEUVRE : N.C.

MOYENS DU PROGRAMME : N.C.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE :

University of California, Los Angelès (UCLA). 2ème trimestre 1987-88.

BUT DE L'EVALUATION :

Déterminer l'impact du cours sur l'attitude et le comportement des étudiants face au SIDA.

INDICATEURS RETENUS :

- Connaissance sur le SIDA.
- Attitudes envers le SIDA.
- Comportement face au risque de contamination.

CHOIX DE LA POPULATION :

Il n'a pas été possible de former un groupe d'étude et un groupe témoin au hasard.

Groupe d'étude : 525 étudiants.

Groupe témoin : 522 étudiants suivant un cours d'astronomie.

METHODE DE COLLECTE DE L'INFORMATION :

Un examen en début de session et un autre immédiatement après la fin des cours.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Régressions pour déterminer la relation entre certaines réponses et les caractéristiques du sujet.

LES EFFETS DU PROGRAMME

LES TYPES D'EFFETS :

Pour tous les indicateurs le cours semble avoir eu un effet positif.

LES BIAIS :

Les étudiants s'inscrivant au cours sont généralement plus motivés. Pendant les dix semaines que dure la session, ils vont s'informer en dehors des cours et leurs progrès ne seront pas attribuables uniquement au programme.

LES CRITIQUES DU PROGRAMME : N.C.

LES CRITIQUES DE L'EVALUATION :

Les échantillons n'ont pas été choisis au hasard. Une seule évaluation a été effectuée après la mise en oeuvre du programme. Il y a eu un taux élevé de désistements parmi les sujets des échantillons.

OBSERVATIONS : N.C.

REVUE : Evaluation Review

TITRE : Self Reported Change Versus Change in Self-Report

AUTEURS : S.E. Spiro, A. Shalev, Z. Solomon, M. Kotler.

Volume 13 ; N° 5 ; 1989 ; p.533-549.

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : "Ko'ach"

POPULATION : Anciens combattants

OBJECTIFS DU PROGRAMME :

Soigner les soldats souffrant toujours de troubles psychiques quatre ans après les combats.

LOCALISATION-DATE-DUREE :

Israël, un camp d'entraînement militaire, juillet 1986.

CONTENU DU PROGRAMME :

41 anciens combattants ont passé quatre semaines dans un camp d'entraînement militaire.

Ils ont été exposés à des situations, des bruits, rappelant l'événement qui les a traumatisés.

Ils étaient soutenus par leur femme, des amis, des supérieurs.

Ils ont reçu un enseignement : humour, maîtrise de soi ...

MAITRE D'OEUVRE :

Service psychiatrique des forces armées israéliennes

MOYENS DU PROGRAMME : N.C.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE :

Israël, camp d'entraînement militaire. De 12 mois avant à 9 mois après le programme.

BUT DE L'EVALUATION :

Déterminer l'impact d'un tel programme sur les participants. L'auteur de l'article utilise cette expérience comme base de réflexion sur l'utilisation, comme instrument d'évaluation d'un programme, des déclarations des participants à ce programme.

INDICATEURS RETENUS :

- Présence des différents symptômes.
- Adaptation sociale et affective de l'individu.

CHOIX DE LA POPULATION :

Les participants ont été choisis selon leur besoin d'un traitement, leur motivation (41).

Les sujets du groupe témoin, équivalent sur le plan socio-économique, ont été sélectionnés parmi d'autres victimes de traumatismes.

METHODE DE COLLECTE DE L'INFORMATION :

- Une série de questionnaires avant/après.
- Un questionnaire rétrospectif 9 mois après uniquement pour les participants sur les effets du traitement et l'aide qui leur a été apportée par leur entourage, leur satisfaction quant au traitement.
- L'avis des médecins.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Etude de corrélations entre différentes mesures de leur état par les participants.

LES EFFETS DU PROGRAMME

LES TYPES D'EFFETS :

D'après la série de questionnaires avant/après la majorité des participants ne manifestent aucune amélioration. L'état de certains se serait aggravé.

Par contre, dans le questionnaire rétrospectif la plupart des sujets déclarent se sentir mieux.

Les médecins constatent une amélioration.

LES BIAIS :

Le même organisme offrait ce traitement et demandait aux participants de l'évaluer 9 mois après. Ceci a pu fausser les réponses au questionnaire rétrospectif.

Les sujets sont plus sensibles à leurs problèmes pendant le traitement ce qui a pu fausser les réponses aux questionnaires avant/après sur leur état.

LES CRITIQUES DU PROGRAMME :

LES CRITIQUES DE L'EVALUATION :

OBSERVATIONS : N.C.

REVUE : Evaluation and the Health Profession

TITRE : Satisfaction with Medical Services in a Low-Income Brazilian Community

AUTEURS : P. Paine, M. da Gloria Wright

Volume 11 ; N° 3 ; 1988 ; p 388-398

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : N.C.

POPULATION : Les 16 000 habitants d'un bidonville.

OBJECTIFS DU PROGRAMME : Voir contenu

LOCALISATION-DATE-DUREE :

Brésil, Paranoa (bidonville de 16 000 habitants). Le centre médical a été inauguré en 1985.

CONTENU DU PROGRAMME :

Centre médical offrant gratuitement les services suivants :

- vaccinations,
- soins prénataux et pédiatrie,
- médecine générale,
- soins d'urgence sans gravité.

MAITRE D'OEUVRE :

Le problème de sécurité sociale du Brésil.

MOYENS DU PROGRAMME :

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE : Brésil, Paranoa. Avril-mai 1986 (8 semaines).

BUT DE L'EVALUATION :

L'inefficacité d'un service médical peut s'expliquer en partie par un fort taux d'insatisfaction parmi les clients. Des clients satisfaits auraient plus tendance à prendre des mesures préventives, à suivre les conseils du médecin. Cette étude a pour but d'évaluer la satisfaction des clients d'un service médical au Brésil.

INDICATEURS RETENUS :

Satisfaction quant à l'accessibilité (horaires, attentes ...) du service.

Satisfaction quant aux médecins.

Satisfaction quant au centre médical, et connaissance des horaires.

CHOIX DE LA POPULATION :

260 des 2 600 baraques en bois ont été choisies au hasard. 245 étaient des habitations et les chefs de famille ou leur épouse ont été interrogés. Echantillon final : 172 femmes (26 % d'illettrées) - 73 hommes (36 % d'illettrés).

METHODE DE COLLECTE DE L'INFORMATION :

Les sujets ont été interrogés par quatre membres du personnel médical habitant dans cette commune.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES : N.C.

LES EFFETS DU PROGRAMME

LES TYPES D'EFFETS :

- 1) Accessibilité : 43 % des patients considèrent que l'attente est trop longue.
- 2) Médecins : La majorité des patients sont satisfaits de leur dernière consultation. Cependant, 65 % des personnes interrogées doutent de ce que leur dit le médecin.
- 3) Centre médical : 95 % des personnes interrogées voulaient un centre médical dans leur commune. Celui-ci a satisfait l'attente de 34 % seulement. 16 % des patients, seulement, connaissaient les horaires du centre.

LES BIAIS : N.C.

LES CRITIQUES DU PROGRAMME :

L'information circule mal.

Les longues attentes représentent un coût d'opportunité considérable pour des ouvriers payés à l'heure de travail. Le manque de confiance résulte d'une mauvaise communication entre médecins et patients de milieux socio-culturels différents.

LES CRITIQUES DE L'EVALUATION : N.C.

OBSERVATIONS : N.C.

REVUE : Evaluation and the Health Profession

TITRE : A. Benefit-Cost Analysis of Family Planning Services in Iowa

AUTEURS : L.M. Levey, J.A. Nyman, J. Maugoard

Volume 11 ; N° 4 ; 1988 ; p.403-424

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : N.C.

POPULATION : Femmes à bas revenu.

OBJECTIFS DU PROGRAMME : N.C.

LOCALISATION-DATE-DUREE : Etats-Unis, Iowa.

CONTENU DU PROGRAMME : Service de planning familial

MAITRE D'OEUVRE : Etat du Texas.

MOYENS DU PROGRAMME : N.C.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE :

Etats-Unis, Iowa 1984.

L'étude est faite sur les cinq années 1978-1983 et par année sur cette période.

EVALUATEURS : Family Planning Council at Iowa.

BUT DE L'EVALUATION :

Déterminer la rentabilité des services de planning familial pour l'Etat afin de permettre à l'Etat d'Iowa de décider s'il doit ou non augmenter ses dépenses en services de planning familial dans une situation de diminution des aides de l'Etat fédéral.

INDICATEURS RETENUS :

Taux de rendement : Economies réalisées par l'Etat grâce à la prévention d'une naissance divisées par les dépenses en planning familial pour une femme.

NB : dépenses évitées : dépenses en aides sociales considérées comme une conséquence de la naissance d'un enfant.

Coûts : Coût moyen par femme en 1983 des services de planning familial.

CHOIX DE LA POPULATION :

Les femmes bénéficiant déjà d'aides sociales ou celles qui en bénéficiaient en cas de maternité. Quatre classes d'âge ont été définies : 14-19 ans, 20-29 ans, 30-34 ans; 35-49 ans.

METHODE DE COLLECTE DE L'INFORMATION :

1) Données de l'Aide aux familles avec enfants à charge (AFDC) et données sur les bons alimentaires ; les fichiers de Medicaid.

2) Données d'administrations de la santé de l'Etat d'Iowa.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Pour le calcul des dépenses évitées, des probabilités par classe d'âge d'éviter une naissance non désirée ont été utilisées. (Alan Guttmacher Institute et Forest (1981)).

LES EFFETS DU PROGRAMME**LES TYPES D'EFFETS :**

Pour les femmes bénéficiant déjà d'aides sociales, les économies réalisées par l'Etat sont supérieures aux coûts pour les femmes de 14-19 ans et 20-29 ans et sont une fonction décroissante de l'âge.

Les dépenses évitées sur 5 ans sont 6,6 fois supérieures aux économies calculées sur un an pour les adolescentes (3,3 fois pour les adultes).

LES BIAIS :

Les calculs effectués sur 5 ans n'ont pas tenu compte du fait que certaines femmes n'auraient plus bénéficié d'aides sociales s'il n'y avait pas eu de naissance. Pour celles-ci les dépenses non liées directement à la maternité en sont quand même la conséquence.

LES CRITIQUES DU PROGRAMME : N.C.

LES CRITIQUES DE L'EVALUATION : N.C.

OBSERVATIONS : N.C

REVUE : Evaluation and the Health Profession

TITRE : A Comparison of Biochemical and Interview Measures of The Exposure of Infants to environmental Tobacco smoke.

AUTEURS : K.E. Bauman, V.J. Strecher, R.A. Greenberg, N.J. Haley.

Volume 12 ; N° 2 ; 1989 ; p. 179-191.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE :

1985, Memorial Hospital, Chapel Hill, Caroline du Nord, Etats-Unis.

EVALUATEURS : -

BUT DE L'EVALUATION :

Comparer deux techniques de mesures de la quantité de fumée de cigarettes inhalée par des nourrissons :

- une analyse d'urine de l'enfant
- une enquête auprès de la mère.

INDICATEURS RETENUS : N.C.

CHOIX DE LA POPULATION :

Echantillon total : 94 enfants nés entre le 6 juin et le 30 octobre 1985.

Groupe d'étude : les 37 enfants pour lesquels une analyse d'urine a pu être effectuée.

METHODE DE COLLECTE DE L'INFORMATION :

- Un entretien avec les mères à l'hôpital sur des questions générales et sur les habitudes des fumeurs vivant sous le même toit que l'enfant.
- Analyse d'urine de l'enfant.
- Deuxième entretien avec la mère : questions précises sur le temps pendant lequel l'enfant est exposé à la fumée, l'aération de la pièce...

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Les résultats biochimiques et ceux des entretiens avec les mères sont étudiés de deux manières :

- comme des variables binaires (résultat positif : l'enfant est un fumeur passif / résultat négatif) :

ENTRETIEN	ANALYSE		TOTAL
	positif	négatif	
Positif	9	2	11
Négatif	3	23	26
TOTAL	12	25	37

On constate que les deux techniques de mesure sont en accord.

- comme des variables continues : on obtient un coefficient de corrélation entre les deux méthodes de mesure de 0,67.

OBSERVATIONS :

L'évaluation sous forme d'entretien apporte plus d'informations sur les sources de fumée dans l'environnement de l'enfant (causes). L'évaluation biochimique peut être considérée comme une mesure de l'absorption de fumée (effets).

REVUE : Evaluation and the Health Profession

TITRE : A Cost-Effectiveness Evaluation of Primary Health Care Project for the Poor

AUTEURS : C.E. Begley, C. Mc Kinnon Dowd, R. Mc Candless

Volume 12 ; N° 4 ; 1989 ; p.437-452

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME :

Primary Health Care Services Program (PHCSP).

POPULATION :

Personnes dont les revenus sont inférieurs à 150 % du niveau de pauvreté et supérieurs aux revenus servant de référence à l'Assistance aux familles avec enfants à charge (AFDC). Les personnes ayant souscrit une assurance privée ou bénéficiant de Médecine ou Medicaid sont exclues.

OBJECTIFS DU PROGRAMME :

Fournir des services médicaux aux personnes à bas revenu dans les régions mal desservies en contractant avec des services privés et publics.

LOCALISATION-DATE-DUREE :

Etat du Texas, régions urbaines et rurales. Lancé en 1985.

CONTENU DU PROGRAMME :

26 services répartis dans tout le Texas effectuent au minimum les six types d'actes médicaux suivants :

- diagnostics et traitement - examens préventifs,
- services d'urgence - information médicale,
- planning familial - analyse, radiographies ...

MAITRE D'OEUVRE :

L'Etat du Texas par l'intermédiaire de son Département de la Santé (Texas Department of Health, TDH).

MOYENS DU PROGRAMME :

8 millions de Dollars pour les deux années 1985-86,

15,4 millions de Dollars pour les deux années 1987-88, accordés par l'Etat du Texas au TDH.

Les 26 services sont aussi sponsorisés par divers organismes.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE : Texas, 9 sites, trimestre mars-mai 1987.

EVALUATEURS :

Centre d'Etude des politiques de santé à l'Université du Texas (Center for Health Policy Studies). L'étude est financée par le Bureau of Dentaleans Chronic Discase Prevention.

BUT DE L'EVALUATION :

Evaluer l'efficacité-coût des différents services appliquant ce programme les uns par rapport aux autres.

INDICATEURS RETENUS :

Indicateur de "produit" des services : nombre de consultations pour diagnostic/traitements, urgences, planning familial, prévention, information médicale (les autres actes, ne sont pas inclus dans les calculs).

Indicateur d'efficacité/coût : taux de rendement (= coût moyen d'une consultation).

CHOIX DE LA POPULATION :

9 des 26 services appliquant le programme ont été choisis comme étant représentatifs des différents types d'organisations.

METHODE DE COLLECTE DE L'INFORMATION :

Documentation fournie au TDH par les 9 sites.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Etudes de corrélation entre le type de service (public/privé et selon l'organisme qui le sponsorise), la structure du personnel (équivalent nombre de médecins à plein temps : équivalent nombre de non médecins à plein temps) d'une part et les coûts ajustés, d'autre part.

Le taux de rendement a été calculé de quatre manières :

- en prenant en compte les dépenses réellement effectuées,
- en égalisant pour tous services les salaires et les prix,
- en évaluant le nombre de consultations que peut assurer le service avec les ressources dont il dispose (pleine capacité),
- avec prix et salaires égalisés et à pleine capacité.

LES EFFETS DU PROGRAMME**LES TYPES D'EFFETS :**

Etude des effets de trois facteurs : nombre de patients, structure du personnel, le type de service.

- * En moyenne, si les services fonctionnaient à leur capacité maximale les coûts moyens seraient très inférieurs,

- * Les services ayant une proportion de médecins élevée ont des coûts moyens élevés.

- * Les cliniques publiques présentent les coûts moyens les plus bas.

LES BIAIS :

L'existence d'une corrélation éventuelle entre structure du personnel et type d'organisme n'a pas été vérifiée.

L'échantillon est petit.

LES CRITIQUES DU PROGRAMME :

La plupart des services fonctionnent bien au dessous de leur capacité. Il faut noter cependant que le programme était récent.

LES CRITIQUES DE L'EVALUATION :

De nombreux facteurs n'ont pas été pris en compte : gestion, utilisation de l'espace, qualité des soins.

OBSERVATIONS : N.C.

REVUE : Evaluation and the Health Profession

TITRE : Assessing the quality of Evaluations of adolescent Pregnancy care Programs

AUTEURS : J.P. Ducette, Denis Mc Bride, Gerald J. Stahler

Volume 12 ; N° 4 ; 1989 ; p.453-466

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME :

Adolescent Family Life Act Programs (plusieurs programmes)

POPULATION : Adolescentes

OBJECTIFS DU PROGRAMME :

Réduire ou éliminer les conséquences négatives d'une grossesse chez les adolescentes.

LOCALISATION-DATE-DUREE :

Etat-Unis

Hôpitaux, écoles, écoles spécialisées, autres services médicaux ou sociaux.

CONTENU DU PROGRAMME :

Série de programmes offrant divers services : soins, enseignement, services de psychologie ...

Ces programmes comportent des évaluations (c'est sur ces évaluations que se concentre l'article). De 1 % à 5 % du budget de ces programmes doivent être consacrés à des évaluations.

MAITRE D'OEUVRE : N.C.

MOYENS DU PROGRAMME : N.C.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE : N.C.

EVALUATEURS :

Evaluation demandée par l'U.S. Office of Adolescent Pregnancy Programs (OAPP).
Réalisée par les auteurs de l'article.

BUT DE L'EVALUATION :

Evaluer la qualité des évaluations de ces programmes.

INDICATEURS RETENUS :

* Qualité des données, des instruments de collecte de l'information, des techniques statistiques, des conclusions : les hypothèses de départ sont-elles bien définies ? (indicateurs notés de 1 à 5). Note globale attribuée aux évaluations du fonctionnement des programmes et aux évaluations de leur résultat.

* Classification des évaluations : complète/incomplète ; expérimentation, quasi-expérimentation, avant/après

CHOIX DE LA POPULATION :

45 dossiers de demandes de prolongation de programme soumises à l'OAPP pour l'année budgétaire 1987-88 ont été analysés.

METHODE DE COLLECTE DE L'INFORMATION :

Voir "Choix de la population".

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Analyses de variance.

RESULTATS :

- Les évaluations de programmes appliquées dans des hôpitaux sont de bien meilleure qualité. Le budget de ces programmes est plus élevé. A budget égal, la différence de qualité serait minime.

Les notes moyennes pour chaque indicateur sont faibles.

- 27 % seulement des évaluations d'impact des programmes étaient achevées et utilisables.

- 80 % des évaluations de fonctionnement des programmes étaient complets.

_ Les différentes évaluations étaient très variées dans leurs méthodes et de qualité très inégale.

LES EFFETS DU PROGRAMME

LES TYPES D'EFFETS : N.C.

LES BIAIS : N.C.

LES CRITIQUES DU PROGRAMME : N.C.

LES CRITIQUES DE L'EVALUATION : N.C.

OBSERVATIONS :

Aucune des évaluations ne comportait d'analyse de coûts.

De nombreuses évaluations n'ont pas vraiment tenu compte des objectifs du programme.

La limitation à 5 % de la part du budget consacré à l'évaluation est contraignante pour les évaluations de programme à faible budget.

REVUE : Evaluation Review

TITRE : Evaluation of the Child and Family Resource Program. Early Evidence of Parent-Child interaction Effects

AUTEURS : D.P. Affholter, D. Connell, M.J. Nauta

Volume 10 ; N° 3 ; 1986 ; p.281-298

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : N.C.

POPULATION :

Familles défavorisées avec un enfant dont le suivi est assuré de 0 à 8 ans.

OBJECTIFS DU PROGRAMME :

Développer des systèmes d'aide aux familles pauvres avec enfants, adaptables à d'autres projets.

LOCALISATION-DATE-DUREE :

Etats-Unis (11 sites), début 1973, évalué en 1977

CONTENU DU PROGRAMME :

Aide au développement de l'enfant :

- . en mettant l'accent sur la structure familiale,
- . plans d'aide individualisés,
- . continuité du programme : 0-8 ans.

MATRE D'OEUVRE :

Administration for Children, Youth and Families

MOYENS DU PROGRAMME : N.C.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE : N.C.

EVALUATEURS : Abt Associates Inc.

BUT DE L'EVALUATION : N.C.

INDICATEURS RETENUS :

Comportement de l'enfant, notamment vis-à-vis des parents (cf. film l'enfant).

CHOIX DE LA POPULATION :

- 1) 6 sites : dans chaque 40 CRFP + 35-60 familles témoins.
- 2) 2 sites : dans chaque 15 CRFP + 15 témoins.
- 3) in fine : 25 paires (CRFP - témoins) étudiées.

METHODE DE COLLECTE DE L'INFORMATION :

Les familles ont été filmées à deux reprises : la caméra suit l'enfant concerné par l'étude. Des systèmes de codage permettent de mesurer tous les types d'activité rencontrés.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

- 1) Pour l'apariement des familles : calcul de distance multivariée.
- 2) Tris croisés, pourcentages, tests.

LES EFFETS DU PROGRAMME

LES TYPES D'EFFETS :

Augmentation du temps passé aux relations parent-enfant ; provoque un développement plus précoce de l'enfant.

LES VARIABLES EXPLICATIVES :

Dans l'apariement : âge enfant ; âge mère ; nombre d'enfants ; revenu annuel par tête ; études supérieures pour la mère (oui/non) ; mère travaille (oui/non) ; deux parents (oui/non).

LES BIAIS :

Que des familles noires.

LES CRITIQUES DU PROGRAMME : N.C.

LES CRITIQUES DE L'EVALUATION :

Significativité de l'amplitude des résultats observés.

OBSERVATIONS : N.C.

REVUE : Evaluation and the Health Profession

TITRE : Community Posthospital Follow-up Services

AUTEURS : Ann Solberg

Volume 7 ; N° 1 ; 1983 ; p.96-109

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME : N.C.

POPULATION :

Patients des unités psychiatriques qui viennent d'être hospitalisés.

OBJECTIFS DU PROGRAMME :

Améliorer l'état de santé et éviter une nouvelle hospitalisation.

LOCALISATION-DATE-DUREE :

Etats-Unis, Comté de Fresno, Californie ; unité de psychiatrie aiguë

CONTENU DU PROGRAMME :

. Suivi des patients, de leur famille dans le but d'augmenter leur implication dans les programmes post-hospitaliers.

. Première visite à domicile 24 heures après la sortie d'hôpital.

MAITRE D'OEUVRE :

Département Santé de l'Etat de Californie.

MOYENS DU PROGRAMME :

DESCRIPTION DE L'EVALUATION

EVALUATEURS :

Sur le terrain : 4 travailleurs sociaux en psychiatrie.

LOCALISATION-DATE-DUREE :

Unité psychiatrique du Comté de Fresno, Californie.
3 mois, fin en juillet 1980.

BUT DE L'EVALUATION :

Evaluer l'efficacité du suivi à la fois dans la réduction de la réhospitalisation et dans le coût total de la prise en charge du patient.

INDICATEURS RETENUS :

. Durée de suivi : temps écoulé entre la sortie de l'hôpital et une éventuelle réhospitalisation.

. Analyse des coûts.

CHOIX DE LA POPULATION :

71 patients ont subi un suivi post-hospitalier particulier (30 jours) +

72 patients ont reçu les suivis post-hospitaliers qu'ils réclamaient

Au total 143 patients assignés au hasard dans l'un des deux groupes.

METHODE DE COLLECTE DE L'INFORMATION : N.C.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Pourcentages, médianes.

LES EFFETS DU PROGRAMME

LES TYPES D'EFFETS :

. On mesure des délais de réhospitalisation plus importants pour les personnes suivies.

. L'économie en niveau des coûts pour les personnes suivies (les coûts du suivi sont largement financés par les économies en réhospitalisation).

LES BIAIS : N.C.

LES CRITIQUES DU PROGRAMME :

En fait, améliorations qui ont été apportées : période de suivi dès qu'un patient sort de l'hôpital ; période de suivi qui passe à 40 jours.

LES CRITIQUES DE L'EVALUATION : N.C.

OBSERVATIONS : N.C.

REVUE : Evaluation Review

TITRE : Preliminary Contact with a Mailed Follow up Survey

AUTEUR : Michael H. Miner

Volume 7 ; N° 3 ; 1983 ; p. 385-396

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE :

Un service psychiatrique à Saint-Louis, dans le Missouri, 1980-1981.

EVALUATEURS : N.C.

BUT DE L'EVALUATION :

Déterminer l'impact d'une prise de contact préliminaire (par téléphone/par courrier) sur le taux de réponse à un questionnaire envoyé à d'anciens patients d'un service psychiatrique.

INDICATEURS RETENUS : Taux de réponse au questionnaire.

CHOIX DE LA POPULATION :

Les parents ou tuteurs de 145 enfants destinés à être contrôlés, entre le 1/12/80 et le 30/05/81, six ou dix-huit mois après leur dernière visite au service psychiatrique.

Une condition d'expérimentation parmi trois est appliquée au hasard à chaque sujet :

- une lettre est envoyée au sujet une semaine avant l'envoi du questionnaire
- le sujet est contacté par téléphone trois jours avant l'envoi du questionnaire.

Trois essais au total sont effectués, trois jours différents de la semaine et à trois heures différentes de la journée.

- aucune prise de contact préliminaire n'est effectuée.
- Les trois groupes ainsi formés étaient respectivement de 49, 47 et 50 individus.

N.B. La lettre ou l'appel téléphonique précise au sujet le but de l'enquête, l'importance de sa participation et l'informe qu'une indemnisation de 10 \$ lui est versée s'il répond.

RESULTATS :

Taux de réponse en fonction de la méthode de prise de contact :

Par courrier : 63%

- Par téléphone :
- ayant pu être contactés : 81%
 - n'ayant pas pu être contactés : 47 %

Groupe témoin : 55 %

N.B. taux de réponse = nombre de réponses / (nombre de quest. envoyés - nombre de "retours à l'expéditeur").

OUTILS STATISTIQUES UTILISES :

Un test du chi-deux effectué en éliminant le groupe des sujets n'ayant pu être joints au téléphone indique qu'une prise de contact préliminaire a un fort impact sur le taux de réponse.

Un test du chi-deux effectué sur les groupes des sujets contactés au téléphone et par courrier indique une faible influence de la méthode employée sur le taux de réponse.

CRITIQUES DE L'EVALUATION :

Certains facteurs pouvant influencer les résultats n'ont pas été pris en compte :

- La distinction entre patients contrôlés après six ou après dix-huit mois.
- La distinction entre parents naturels, parents adoptifs et institutions.

OBSERVATIONS :

Le résultat selon lequel le taux de réponse à l'enquête augmente lorsque le sujet est contacté auparavant est en accord avec les résultats d'autres expériences (Duncan, 1979 ; Kerin et Peterson, 1977 ; Linsky, 1975).

Par contre, le fait qu'une prise de contact par téléphone ne soit pas beaucoup plus efficace ne correspond pas aux résultats d'études antérieures (le nombre important d'appels n'ayant pas abouti peut en être la raison).

REVUE : Evaluation Review

TITRE : Evaluation of Programs for Handicapped Children. A Strategic Constituent Approach

AUTEURS : Ronald W. Thompson, Kenneth D. Orton

Volume 9 ; N° 6 ; 1985 ; p.701-715

DESCRIPTION DU PROGRAMME

NOM DU PROGRAMME :

Programme diagnostique pour enfants avec problème de communication.

POPULATION :

Enfants ayant un problème pour communiquer (écoute et langage)

OBJECTIFS DU PROGRAMME :

Des spécialistes diagnostiquent l'état de l'enfant (pédiatres, sociologues, ophtalmos, ORL ...) + psychiatre.

LOCALISATION-DATE-DUREE : U.S.A.

CONTENU DU PROGRAMME : N.C.

MAITRE D'OEUVRE : N.C.

MOYENS DU PROGRAMME : N.C.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE : N.C.

EVALUATEURS : N.C.

BUT DE L'EVALUATION :

Evaluer l'efficacité des cliniques diagnostic qui s'occupent d'enfants handicapés.

INDICATEURS RETENUS :

D'après phase 1 : 100 critères d'efficacité :
 qualité des services
 qualité du personnel
 satisfaits du Cr
 satisfaits des employés
 support

Phase 2 : voir méthode de collecte

CHOIX DE LA POPULATION :

Première phase	8 superviseurs + 6 membres du personnel + 8 couples de parents + 8 responsables éducation ou scolaires. Au hasard.
Deuxième phase	35 responsables et cliniciens - 31 réponses puis 28 (80 % de participation)

METHODE DE COLLECTE DE L'INFORMATION :

Première phase entretien structuré (entre entretien libre et questionnaire)
 > experts + et - de la clinique.

Deuxième phase questionnaire : "Ce point est-il important pour qu'un établissement soit efficace ?" (oui/non).

"Ce point est-il rempli dans cette clinique ?" (de 1 à 5).

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Calcul des corrélations.

ANOVA : moyenne, écart-type.

LES EFFETS DU PROGRAMME

LES TYPES D'EFFETS : N.C.

LES BIAIS : N.C.

LES CRITIQUES DU PROGRAMME : N.C.

LES CRITIQUES DE L'EVALUATION : N.C.

OBSERVATIONS : N.C.

REVUE : Evaluation Review

TITRE : Evaluation of an STD Education Media Campaign

AUTEURS : J. Mc Killip; K. Baldwin

Volume 14 ; N° 4 ; 1990 ; p. 331-346.

DESCRIPTION DU PROGRAMME EVALUE

NOM DU PROGRAMME : N.C.

POPULATION :

Etudiants de l'université de Carbondale, Illinois du Sud.

OBJECTIFS DU PROGRAMME :

Sensibiliser l'opinion à des questions de santé publique.

LOCALISATION-DATE-DUREE :

USA (Illinois), une semaine précise entre fin janvier à mi-mars 1987.

CONTENU DU PROGRAMME :

Campagne d'information sur les MST, l'usage du préservatif (articles dans la presse, affichage...). Deux autres mesures qui auraient pu être le sujet d'une campagne de presse sont également étudiées, il s'agit de la consommation d'alcool et des bienfaits issus de la pratique régulière d'une activité physique.

MAITRE D'OEUVRE : N.C.

MOYENS DU PROGRAMME : N.C.

DESCRIPTION DE L'EVALUATION

LOCALISATION-DATE-DUREE :

USA (Illinois), 8 semaines dont 4 semaines de pré-campagne, une semaine de campagne et 3 semaines post-campagne.

EVALUATEURS : N.C.

BUT DE L'EVALUATION :

Déterminer l'impact d'une campagne de presse sur un thème de santé publique, en comparant les réactions du public sur ce sujet traité dans les médias aux réactions de ce même public sur des sujets de même ordre mais non concernés par une campagne d'information.

INDICATEURS RETENUS :

Cette enquête étudie simultanément l'impact de la campagne médiatique sur les étudiants et les processus causaux impliqués. Ces processus sont au nombre de trois :

- (a) évaluation du changement dans le contenu des médias;
- (b) évaluation de l'audience du public;
- (c) évaluation des effets de la campagne sur le sujet.

L'impact des médias a été mesuré ici par les attitudes des sujets, leurs croyances et les intentions qu'ils affichaient concernant le programme étudié; de même on a mesuré la fréquence d'apparition du sujet dans les conversations eues la semaine passée avec des amis.

CHOIX DE LA POPULATION :

Sur 1617 étudiants tirés au hasard, 560 ont été effectivement interrogés. Huit échantillons distincts ont été formés pour chacune des semaines d'étude. Au départ, on avait pris le même nombre d'hommes et de femmes, et le même nombre d'étudiants de chaque année.

METHODE DE COLLECTE DE L'INFORMATION :

Les étudiants ont été contactés par téléphone. Les enquêteurs (5 hommes et 5 femmes) avaient reçu neuf heures de formation à partir du manuel de Guenzel, Berckmans et Cannell. 5% des interviewés pris au hasard étaient appelés pour vérifier que l'entretien avait été bien fait.

OUTILS STATISTIQUES UTILISES, CALCULS EFFECTUES :

Méthode de l'analyse de la variance, avec des poids différents accordés au trois niveaux intra-sujet de santé (usage du préservatif, consommation d'alcool et pratique d'un exercice régulier) et des poids différents accordés aux 8 niveaux inter-sujet (les 8 semaines d'interrogation).

Poids intra : - +2 pour l'usage du préservatif
- -1 pour la consommation d'alcool et l'usage d'un exercice physique

Poids inter : - -3 pour les quatre semaines pré-campagne
- +6 pour la semaine de la campagne
- +3 à +1 pour les trois semaines post-campagne.

LES EFFETS DU PROGRAMME**LES TYPES D'EFFETS :**

En ce qui concerne le contenu des médias, on note une augmentation importante des articles concernant les MST et l'usage du préservatif pendant la semaine de la campagne. La couverture des deux autres sujets reste constante.

L'audience a suivi les mêmes évolutions.

L'impact des médias sur les attitudes, les croyances, les intentions et l'occurrence dans les conversations du sujet a été mesuré.

LES BIAIS : N.C.

LES CRITIQUES DU PROGRAMME : N.C.

LES CRITIQUES DE L'EVALUATION : N.C.

OBSERVATIONS :

Ce type d'évaluation est surtout remarquable par le choix fait de la comparaison entre trois sujets proches (usage du préservatif, consommation d'alcool et pratique d'exercices physiques réguliers) dont l'un seulement a fait l'objet d'une campagne de sensibilisation. On choisit des sujets proches du point de vue de l'intérêt historique, des avantages sociaux, de la fiabilité des mesures, mais qui ne seront pas concernés par l'intervention.

On peut d'ailleurs comparer cette technique aux études cas-témoins.

A N N E X E

Articles correspondant à la fiche n°9

Revue : Evaluation Review

Titre : An évaluation of a Juvenile Education Program in a State Penitentiary.

Auteurs : T.P. Locke, G.H. Johnson, K. Kirgin-Ramp, J.D. Atwater, M. Gerrard (University of Kansas).

Volume 10 ; N°3 ; p. 281-298

Comités éditoriaux

- Evaluation Review
- Evaluation Practice
- Evaluation and the Health Professions

A controlled study of the impact of a juvenile education program on the recidivism rates of juveniles was performed. The program involved introducing the juveniles to the realities of prison life. No significant differences between experimental and control groups in the mean number of self-reported status or criminal offenses committed during the premeasure or follow-up period was found. A finding that youths categorized as more delinquent were affected differently by program attendance compared to youths categorized as less delinquent is offered as a tentative explanation of the conflicting results of prior studies of this type of intervention.

AN EVALUATION OF A JUVENILE EDUCATION PROGRAM IN A STATE PENITENTIARY

THOMAS P. LOCKE
GLENN M. JOHNSON
KATHRYN KIRIGIN-RAMP
JAY D. ATWATER
MEG GERRARD

University of Kansas

Thomas P. Locke is Staff Clinical Psychologist on the Children's Service of the Bert Nash Mental Health Center in Lawrence, Kansas.

Glenn M. Johnson, has a master's degree and is currently completing his pre-doctoral internship in clinical psychology at the Norfolk Regional Center in Norfolk, Nebraska.

Kathryn Kirigin-Ramp is Associate Professor in the Department of Human Development and Family Life and Co-Director of the Achievement Place Research Project at the University of Kansas, Lawrence.

Jay D. Atwater is a master's degree candidate in the Department of Human Development and Family Life at the University of Kansas, Lawrence.

Meg Gerrard is Assistant Professor in the psychology Department of the University of Kansas, Lawrence.

Although the adult criminal justice system has been charged with protecting society, implementing the processes of punishment and retribution, and the rehabilitation of convicted offenders, the task of the juvenile justice system in this country is more circumscribed. According to Romig (1978: xxii), the juvenile system "is clear in its authorized mission—to rehabilitate," that is, to divert the young offender from further delinquent activity and subsequent adult criminal behavior.

The ability to evaluate accurately the effects of rehabilitative efforts is essential both because of budgetary considerations and because of the potential problems associated with intervening in young lives. Unfortunately, of the 982 studies published since 1920 that were reviewed by Romig, only 170 met the essential requirements of either random or

matched assignment of subjects to appropriate control group(s) and behavioral outcome measures.

One type of intervention that has received considerable attention is the diversion program in which youthful offenders are taken into a penitentiary setting to meet with selected convicts from the inmate population. Programs of this type are designed to inform the juveniles about the reality of prison life and, it is hoped, to dissuade them, whether by information or intimidation, from further illegal activity. As is often the case the effects of these programs are not easily empirically investigated. Early evaluative information consisted of informal remarks and letters collected from program participants. Subsequent empirical evaluations have gleaned conflicting results but have pointed to a variety of design considerations.

The earliest evaluation of this type of program was a study by Brodsky (1970) involving 85 subjects, most of whom were not delinquent. The participants were taken through the Illinois program. The dependent variable was the extent of attitude change toward punishment of criminals and toward institutions. Though the attitude change for delinquent youth was not statistically significant, it was in the direction of more favorable attitudes toward prisons. The program had more of an impact on the college students than on the delinquent and pre-delinquent subjects. Unfortunately, no control groups were used and no measure of pre- or postdelinquent activity was provided.

A more recent attempt to empirically evaluate an exposure-to-prison program was reported by Serpas et al. (1979). From January to June, 1979, a Juvenile Awareness Program operated in Orleans Parish, Louisiana. A total of 90 juveniles who had been in contact with the court participated; 69 of these youth were involved in the assessment of the program impact. Serpas et al. found a 52% decrease in the absolute number of arrests for these youth from the one-year pre-test period to a one-year posttest period. Unfortunately, there was no equivalent control group in this study. In light of the well-documented decrease in the frequency of delinquent behavior with maturation (Elliott, 1977; Maltz et al., 1980; Wolfgang et al., 1972), these results must be seen as merely suggestive.

The Juvenile Offenders Learn Truth program at the state prison in Lansing, Michigan, has been evaluated by the Michigan Department of Corrections (National Council on Crime and Delinquency, 1976). No statistically significant differences in the number or type of subsequent offenses were found between the experimental ($n = 78$) and control ($n = 85$) groups at either a three- or a six-month follow-up.

Certainly no program has received as much national attention as the Juvenile Awareness Program at the Rahway State Prison in New Jersey. It was this program, presented to a national television audience as "Scared Straight," for which unqualified—and unsubstantiated—claims of 80-90% success rates were made. Three empirical studies of the effectiveness of this program were published. Two studies (Finckennauer, 1979; Finckennauer and Storti, 1978) raise questions about the appropriateness of such programs for certain juvenile subgroups.

The 1978 report by Finckennauer and Storti concerned an evaluation of the attitude changes brought about by this prison program. Subjects were drawn from counseling agencies, police departments, educational institutions, and employment and recreation organization. The youths were assigned to the experimental treatment or a no-treatment comparison group condition. Although the initial intent was for random assignment, practical limitations precluded its implementation. Many of the youth had no official contact with the juvenile justice system. As expected, the experimental group members showed a significant negative shift in their attitudes toward crime. This shift was maintained at the follow-up evaluation, which ranged from one day to almost nine months. No significant differences were found between the experimental and comparison groups on attitudes toward law, justice, self-perception, policemen, prisons, punishment, or obeying the law. Finckennauer and Storti concluded that it is "probably simplistic and unrealistic to expect that a two or three hour visit to Rahway can counteract the longterm effects" of a multitude of other complex factors (1978: 27). Moreover, this mixture of delinquent and nondelinquent youth in Finckennauer and Storti's subject pool and the lack of additional analyses within the delinquent subgroup distorted the applicability of their findings to a more homogeneous delinquent population.

Finckennauer (1979) conducted a second evaluation based on a behavioral outcome measure of subsequent offenses of the sample used in the Finckennauer and Storti study. Juvenile court records were examined for a minimum of six months after the experimental group visited Rahway and after the control group was pretested. Measures of frequency and seriousness of delinquent activity were collected for each youth. Prior to treatment, 41% of the experimental group and 60% of the control group had had *no* contact with the juvenile court system. Although this difference was not statistically significant, it did suggest that Finckennauer's results might be more generalizable to a nondelinquent population than to a delinquent population. Results indicated that a significantly higher proportion of the juveniles who did *not* attend

the program did better in terms of recidivism than did the youth in the experimental group. This was true for both frequency and seriousness of subsequent offenses, and indicated that this type of program may have harmful effects in some instances.

In a third evaluation of the Rahway program, Langer (1980) obtained results that contradicted Finckennauer's findings. Each of his experimental group subjects had committed one or more recorded offenses, and each had to have gone through the program at least 10 months before the evaluation began. An attempt was made to select randomly a matched group of offenders from official police and juvenile court files based on age, race, sex, and aggregate number and seriousness of offenses. Both groups contained male and female offenders. An analysis of short-term effects based on a 10-month follow-up showed no statistically significant differences between the groups. The mean seriousness scores increased for both groups, and the increase was more substantial for the control group. However, the mean seriousness score of the experimental group was significantly lower than the same score for the control group at a 22-month follow-up period. Given the results after 22 months, Langer concluded that the Lifers' Program had indeed made a significant contribution.

EVALUATING A MIDWESTERN JUVENILE EDUCATION PROGRAM

In October, 1978, an inmate program was begun at a midwestern state penitentiary. This program, known as the Juvenile Education Program (JEP), was similar in some respects to the Rahway program but was designed to be more educational and less confrontative. One aspect of the program that seemed unique to the JEP was an attempt to match the juvenile offender with an inmate in terms of lifestyle similarities in the hope that the juvenile would be able to identify with and thus respond better to that inmate's admonitions against criminal activity. Additionally, more individualized attention was available through this program than was apparent in other similar programs.

Although the program had been operational since October 1978 and had served approximately 300 male juveniles, no systematic evaluation had been carried out to assess the program's effect on juvenile crime. The purpose of this study was to examine the impact of this JEP on the recidivism rates of juvenile and young adult offenders.

METHOD

SUBJECTS

From June to October, 1980, all juveniles from three Kansas counties who were on probation and scheduled to attend the JEP at Lansing State Penitentiary were referred into the program evaluation. The population of the three counties (Shawnee, Wyandotte, and Johnson) as of 1980 was 154,800, 171,121, and 268,157, respectively. All subjects from Shawnee County (15) were from Topeka, a metropolitan area of 115,000. The majority of subjects in the Wyandotte County sample (12) resided within Kansas City, Kansas. Subjects from Johnson County (5) lived in one of five suburban Kansas City areas, all with populations in excess of 25,000.

A total of 53 subjects participated in the initial data collection. Of those, 36 completed the follow-up evaluation conducted from January to April, 1981. Of the 17 subjects not in the follow-up evaluation, 6 had moved from their county of origin, 6 could not be contacted by their court service worker, and 2 refused to participate; 2 individuals who had been assigned to the experimental group failed to attend the JEP and were not available at the follow-up; and 1 control group subject was taken out of the evaluation after attending the JEP without his court service worker's knowledge. Additionally, 4 youth who were incarcerated throughout the follow-up period for offenses committed *before* the evaluation began were eliminated. Consequently, 32 subjects were involved in the study. Individuals dropped from the evaluation were not significantly different from those remaining in regard to age, race, or number of status, miscreant and delinquent offenses committed during the six-month period (the pre-measure period) prior to the collection of the preliminary data. However, the mean Jesness value orientation score (Jesness, 1966) of the subjects dropped from the evaluation was significantly more deviant ($p < .05$) than was the mean value orientation score of the remaining subjects.

All subjects were males between 14 and 19 years of age who (1) were on probation following the commission of a criminal offense (one that would be considered a misdemeanor or a felony if the youth were an adult) and (2) had agreed to participate in the evaluation with parental permission. Specific statewide guidelines for referral to the JEP were not available; thus, the court service workers were asked to utilize the same idiosyncratic criteria for referral that they had used prior to the

evaluation. Such subjective criteria as attitude, degree of emotional stability, and acceptance of responsibility for one's behavior were reportedly used. Although some court officials felt the best candidate for the program was the youth who saw his offense as serious and accepted responsibility for his actions, other workers saw the program as best suited for "the juvenile who considers the court process, police affiliation, school, and probation as a joke" (Lucas, 1979; personal communication). In spite of the discretionary nature of the referrals, there was no evidence to indicate that the sample on which these analyses were based was unrepresentative of the population of juveniles on probation in each of the three counties.

The sample consisted of 65% white, and 33% black juveniles, with the remainder either American Indian or Mexican American. Though the difference was not significant, 13 (81%) of the control group subjects were white, whereas only 8 (50%) of the experimental group subjects were white ($z = 1.84, p = .06$). The age range for the sample was from 14.1 years to 19.8 years; mean age for the control group subjects was 16.7 years and 16.4 years for the experimental group. The difference was not statistically significant.

During the six-month pre-measure period, the experimental group subjects had an average of 1.25 charges filed against them in juvenile court. For the control group the mean was 1.66. The differences between groups were not statistically significant.

DEPENDENT MEASURES

For each of the two data collection sessions, individual subjects were queried about demographic, family, school-related information, self-reported delinquency, and attitudes. Self-reported delinquency rates were obtained for specific periods of time using a questionnaire designed by Ageton and Elliott (1978) for the National Youth Survey. After all self-reported data were collected, the juvenile court records, including police offense reports, were reviewed by two trained evaluators using the standardized recording procedures described by Kirigin et al. (1974). Of the 36 interviews conducted to collect follow-up data, 12 were audiotaped and scored to establish the interrater reliability of the self-reported delinquency measure. The correlation between raters was .96. Interrater reliability for the official court data collected during the premeasure period was .99, and for the follow-up period it was .94. The correlation between self-reported delinquency and court charges filed

during the pre-measure period was .145; for the follow-up period it was $-.064$. Neither relationship was statistically significant. Similarly, the rank-order correlation coefficient for court records and self-reported delinquency was not significant. Erickson and Empey (1976) suggested that one measure of the validity of self-reported delinquency is the extent to which the offenses found in police records and court files are self-reported. Of the 96 delinquent and adult charges brought against subjects in this evaluation, 78 (81%) were self-reported.

For many delinquency theorists, value orientation is fundamental in the development and maintenance of delinquent behavior. If that assumption were true, it would seem reasonable to expect values to play a central role in the success or failure of any type of intervention. From Cochrane's (1974) perspective, a positive change in value orientation is a necessary condition for the success of any diversion or corrections program. Two instruments were used to assess individual value structure. Because several studies have shown a relationship between delinquent activity and lower-class value orientation (Landis and Scarpitti, 1965), we wanted to evaluate that relationship. The value orientation scale of the Jesness Inventory was used to establish the degree to which each subject maintained a value system similar to that held by lower-class juveniles. This paper-and-pencil measure consists of 39 true-false questions. The second value orientation measure was developed by Cernkovich (1978) to assess the extent to which individuals held both subterranean or deviant and conventional values simultaneously. The instrument consists of 42 statements to which the subject is asked to strongly agree, agree, be undecided about, disagree or strongly disagree. They were asked to use the undecided category only if unable to select another alternative. If a subject stated he had difficulty reading, the statements and alternatives were read aloud by the experimenter.

PROCEDURE

County program coordinators were responsible for obtaining signed parental consent forms and submitting a list of the individuals eligible to attend the JEP during each of the five months in which pre-measure data were collected. These individuals were then randomly assigned to one of two experimenters who explained the purpose of the program evaluation, obtained individual consent, and administered the three questionnaires. The self-reported delinquency questionnaire was always administered first. The order of the value orientation measures was

alternated for each succeeding subject so that the Jesness scale was given before the Cernkovich scale 50% of the time.

Subjects from each county for each month were randomly assigned to attend the JEP or to be placed in a waiting list control group. As all subjects were on probation, they continued to meet with their court service worker periodically. Assignment to groups occurred prior to the scoring of any of the questionnaires and was conducted by an individual who had not been involved in the data collection. Three exceptions to this random assignment procedure occurred. One of two brothers referred to the JEP was randomly assigned to the experimental group and the other was placed by the experimenter in the control group. In addition, two control group subjects were sent to the JEP by a county program coordinator. One youth was sent as a replacement for an injured experimental group subject who was unable to attend the program, and the other was erroneously sent to the JEP after the initial data collection but prior to the follow-up evaluation. All comparisons between groups on premeasure scores were run with and without the latter two subjects with similar results. The subjects were left in the experimental group in order to increase the sample size.

Collection of the follow-up data took place six to eight months after the preliminary data were collected. In no case was the time between the pre-measure and follow-up data collections less than six months. Because 5 of the 32 subjects were incarcerated at some time during the six-month pre-measure period and/or the six-to-eight-month follow-up period, the number of offenses per day at risk in the community was calculated for each subject.

For the follow-up interview, subjects were again randomly assigned to one of two experimenters. The procedure for administering the three questionnaires was identical to that used during the pre-measure assessment. Self-reported status, miscreant, and delinquent offenses from the time of the pre-measure evaluation to the date of the follow-up were collected and served as the dependent measure by which treatment effects were assessed along with official court records.

RESULTS

There were no statistically significant differences between experimental and control groups in the mean number of self-reported status,

miscreant, or delinquent offenses committed during the pre-measure period, nor were the mean number of police or court contacts different. The mean value orientation scores on both the Jesness and Cernkovich measures were also comparable. Across all subjects neither value measure, when dichotomized into high/low and scored for degree of value deviancy, was significantly correlated with frequency of self-reported status or criminal offenses committed during the pre-measure period. The correlation between self-reported delinquency and the degree of value conventionality as measured by the Cernkovich measure was negative but not significantly so. Jesness and Cernkovich deviancy scores were positively correlated [$r(32) = .61, p < .05$] and the relationship between self-reported criminal activity during the pre-measure and follow-up periods was significant [$r(32) = .47, p < .05$].

A repeated measures analysis of variance (ANOVA) of the self-reported status offenses found no effect for treatment but a decreasing trend in the mean number of status offenses from pre-measure to follow-up [$F(1, 30) = 3.21, P = .08$]. Miscreant and delinquent offenses were summed and divided by days at risk to obtain an overall measure of criminal activity for each subject. An ANOVA found no effect for treatment [$F(1, 30) = .75, p = .39$], but a highly significant decrease in criminal behavior for both groups over time [$F(1, 30) = 9.53, p = .004$]. The Jesness scores [$F(1, 30) = 14.71, p = .0006$] and the Cernkovich scores [$F(1, 30) = 11.51, p = .002$] became significantly less deviant from pre-measure to follow-up but were unaffected by treatment. The repeated measures ANOVA was rerun with miscreant offenses weighted (1) and delinquent offenses weighted (2) with comparable results. Finally, a repeated measures ANOVA on the unweighted, combined, miscreant, and delinquent juvenile court charges resulted in a significance decrease with time [$F(1, 30) = 6.85, p < .05$] but no other significance. All subsequent analyses were based on self-reported data.

LEVEL OF DELINQUENCY EFFECTS

Because value orientation was not significantly related to the number of offenses committed during the pre-measure period, an analysis of the relationship between treatment and level of criminal activity was performed. Experimental and control group subjects were divided at the median into high-delinquent and low-delinquent groups based on the self-reported delinquency level during the pre-measure period. Figure 1

is a plot of the means and Table 1 is the table of significance for a treatment by high/low delinquency by time ANOVA.

As illustrated in Figure 1 and Table 1, within the high-delinquency group, the mean number of offenses for both experimental and control subjects decreased significantly over time [F(1, 14) = 29.88, p = .001]. The treatment by time interaction was also significant [F(1, 14) = 6.22, p = .03]. Although the mean number of pre-measure offenses per day at risk for the experimental group (2.6) was significantly higher than the control mean (1.9), [t(28) = 2.74, p < .05], at the follow-up the experimental group mean (.58) was lower than the control group mean (1.12) [t(28) = 2.02, p = .06]. Both experimental [t(28) = 9.62, p < .001] and control [t(28) = 4.63, p < .001] group means decreased significantly over time. The problem of unequal group means on the pre-measure for the high delinquency experimental and control groups was addressed in a subsequent analysis.

Within the less delinquent group, neither the treatment effect, [F(1, 14) = .12, p = .73] nor the main effect for time [F(1, 14) = .08, p = .78] was significant. The treatment by time interaction was marginally significant [F(1, 14) = 3.82, p = .07]. The pre-measure means for the experimental and control groups were .15 and .26, respectively. This ordinal pattern is reversed at follow-up with the experimental mean (.33) exceeding the control mean (.13).

TABLE 1
Repeated Measures Analyses of Variance

Source	Sum of Squares	Degrees of Freedom	Mean Square	F	Probability
Mean	46.56261	1	46.56261	78.70	0.0000
Treatment	0.07701	1	0.07701	0.13	0.7210
LowHi	26.36817	1	26.36817	44.57	0.0000
TL	0.00935	1	0.00935	0.02	0.9009
Error	16.56644	28	0.59166		
Crim	7.01066	1	7.01066	26.39	0.0000
CT	0.86160	1	0.86160	3.24	0.0825
CL	7.46948	1	7.46948	28.11	0.0000
CTL	2.33268	1	2.33268	8.78	0.0062
Error	7.43913	28	0.26568		

NOTE: Treatment is experimental or control group; LowHi is low-delinquency or high-delinquency group; and Crim is the repeated measure of criminal activity with two levels.

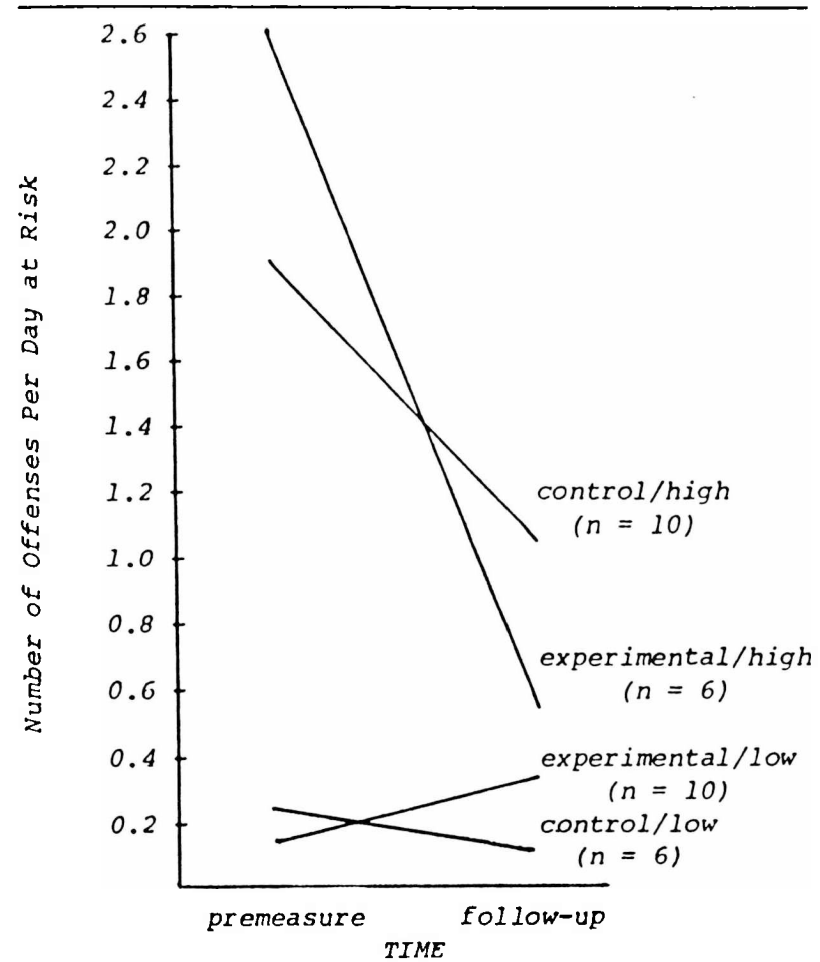


Figure 1: Mean Number of Self-Reported Offenses per Day at Risk for Treatment by Level of Delinquency Groups

Given the small and unequal cell frequencies, a test of the assumption of homogeneity of the variance-covariance matrices was performed. The assumption underlying the use of ANOVA was not met. Glass et al. (1972) have investigated the consequences of violating the assumptions underlying the fixed-effects model and concluded that (1) when ns are unequal and variances are heterogeneous as they are in these data, and (2) when samples with smaller ns come from populations with smaller

variances, the actual significance level may be greatly *exceeded* by the nominal significance level. In these data, the cell with the larger variance has the larger *n*, whereas the cell with the smaller variance has the smaller *n*. Consequently, although the homogeneity assumption has not been met, the test of the effects is probably conservative.

As mentioned earlier, the experimental and control means of the high-delinquency group were not equivalent on the pre-measure. A MANOVA was run to assess more efficiently the treatment by level of delinquency interaction. Once the variance attributable to pre-measure delinquency scores, membership in the low-/high-delinquency groups, and membership in the treatment/control groups was removed, the level of delinquency by treatment interaction accounted for a significant proportion of the residual variance [$F(1, 24) = 4.45, p = .05$]. The test for parallelism was not significant, making the following ANCOVA statistically appropriate.

To validate further the treatment by level of delinquency interaction, a 2-way ANCOVA was performed, equating groups on the pre-measure level of criminal activity. Although the parallelism test indicates the procedure is statistically appropriate, the reader may well question the logical absurdity of equalizing the high- and low-delinquency groups. However, the ANCOVA was *not* used to assess the main effects but rather to test the treatment by high-/low-delinquency group membership interaction once the effect for pre-measure criminal activity had been removed. The ANCOVA significance table is shown in Table 2, and the plot of adjusted means is depicted in Figure 2. Consistent with the two preceding analyses, the interaction of treatment by degree of deviancy was significant [$F(1, 27) = 4.71, p = .04$]. In the low-delinquency group, the adjusted experimental group mean (.81) was greater than the adjusted control group mean (.56), [$t(27) = .726, p < .05$]. The inverse relationship occurred in the high-delinquency group. The adjusted experimental group mean (-.11) was less than the adjusted control group mean (.78) [$t(27) = 2.54, p < .05$].

DISCUSSION

This study was designed to assess the extent to which the JEP would effect the recidivism rates of juvenile and young adult offenders.

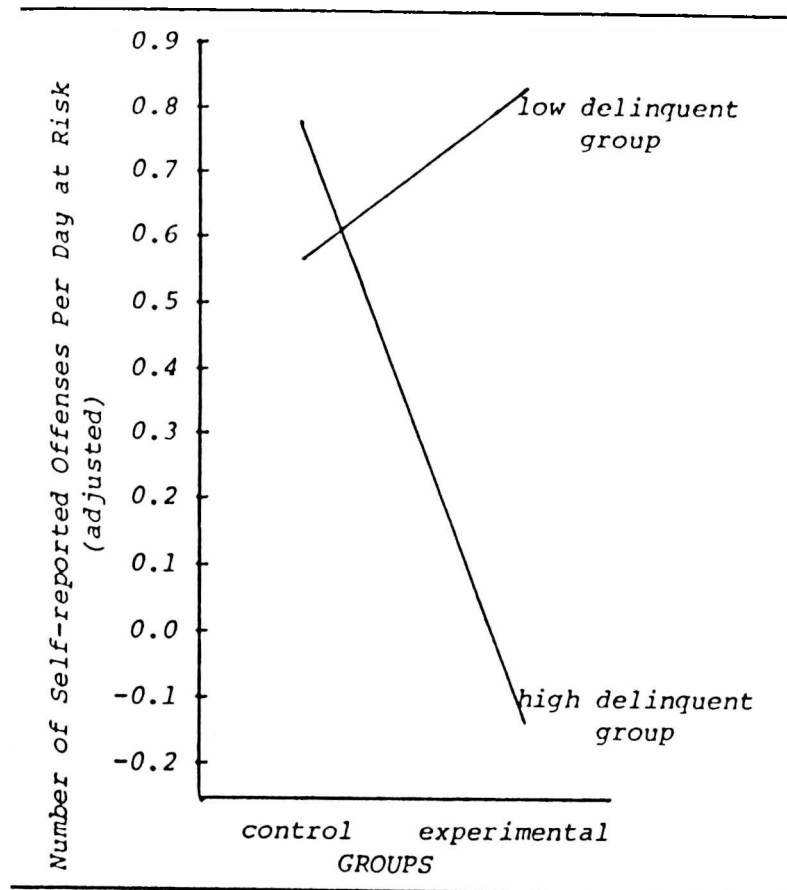


Figure 2: Plot of Adjusted Means

Subjects were assigned randomly to experimental or control groups. In spite of both court system administrative concerns and attrition during the follow-up, which resulted in a smaller than desirable sample size, the primary hypothesis was supported. However, given the small sample size, results should be interpreted with some caution. There was no difference between experimental and control subjects in the mean number of status or criminal offenses committed during the pre-measure or follow-up period. However, unlike the Langer (1980) results, the mean number of offenses per day at risk for all subjects decreased significantly over time. The decrease was apparent in both self-reported

TABLE 2
Analysis of Covariance

Source	Sum of Squares	Degrees of Freedom	Mean Square	F	Probability
Mean	0.01085	1	0.01085	0.02	0.8784
Treatment	0.69960	1	0.69960	1.54	0.2255
LowHi	0.21771	1	0.21771	0.48	0.4949
TL	2.13970	1	2.13970	4.71	0.0390
Covar	2.18151	1	2.18151	4.80	0.0373
Error	12.27702	27	0.45470		

NOTE: Treatment is experimental or control group; LowHi is low-delinquency or high-delinquency group; and Covar is the pre-measure of criminal activity.

delinquency and official police and court records. The results are less puzzling when one remembers that the mean age of subjects in the present study was 16.5 years, nearly 3 years older than that for Langer's sample. The concept of maturational reform suggests that although illegal activity may accelerate during the early teenage years, it typically begins to decrease by age 17, independent of court system efforts to intervene. It is interesting to note the maturation effect in self-reported activity as well as official court contacts.

The finding that the more delinquent youth were affected differently by JEP attendance than were the less delinquent youth may offer an explanation of the conflicting results presented by the two Rahway evaluations. If one keeps in mind that the Rahway studies utilized police and court records whereas the current study was concerned with self-reported delinquency as well as court records, these results offer at least tentative support for the seemingly conflicting conclusions of prior studies.

Consonant with Finckennauer's findings (1979) that juveniles tended to increase their illegal behavior following program attendance, individuals in the present study with initially low levels of self-reported delinquency tended to increase the frequency of criminal activity, relative to a low-delinquency control group, after JEP attendance. The comparison is made more tenable when one recalls that 41%-61% of Finckennauer's sample had had no official contact with the juvenile justice system prior to the program evaluation. Labeling theory suggests that once a youth has been formally labeled "delinquent," society—

particularly the legal system—focuses on the individual as deviant. As a consequence, the probability of detecting further criminal behavior, if not the actual frequency of the behavior, increases. For the youth exhibiting relatively infrequent criminal behavior, personal acceptance or rejection of the delinquent label may be as instrumental in bringing about further criminal behavior as is society's reaction to the label. For these youth, participation in programs in which they are told by hardened adult criminals who ought to know that their present behavior is a likely precursor of adult criminal activity may reinforce what has been only a tenuous delinquent self-concept. If a nebulous delinquent self-concept becomes more reified after attending these sorts of programs, the juveniles have been shown excellent role models on which to pattern their criminal behavior. Although the inmates seem sincere in their admonitions against their deviant lifestyle, they often do so in a very aggressive, defiant manner. An adolescent male unsure of his masculinity may admire the very macho image projected by the inmates. When an inmate says one must be tough and ruthless to survive in a prison environment, some juveniles may want to be seen by their peers in that manner and aspire, in part, to the attainment of such roles. For these individuals, continued involvement in illegal activities may be status enhancing. At the same time, this low-delinquency group may have difficulty realizing that life in an adult prison is a likely consequence of their behavior. They may attend more to the very masculine role models offered by the inmates and less to the potential consequences of their own actions.

In contrast to the results within the low-delinquency group, high-delinquency subjects attending the JEP self-reported fewer offenses during the follow-up than did the high-delinquency control group. These results are consistent with Langer's conclusions (1980) in that all subjects in both studies had to have committed one or more miscreant or delinquent offenses before being included in the studies. Perhaps these individuals are more sure of their delinquent identity and see more clearly the link between juvenile and adult crime of which the inmates spoke. For this group, life in an adult prison may seem a more likely consequence of their present behavior, and that realization could increase the deterrent effect of such programs. A more subtle explanation of the program effect for this group might be the perceived change in status that would follow incarceration. If one can assume that the highly delinquent youth is afforded a level of status—perhaps based on

fear—within the adolescent population, then life as an adult inmate involves more than a loss of personal freedom. As the inmates speak of intimidation, violence, and homosexual coercion directed toward all new prisoners, it must be obvious that their status as delinquents would be meaningless in the new environment. Although some delinquents could view prison life as a challenge, others may be unwilling to move from their high position in the juvenile status order to the demeaning position that would be theirs as new inmates. The perceived loss of status may dissuade these teenagers from further illegal activities.

IMPLICATIONS

The overall results of this study appear to be consistent with the myriad studies evaluating delinquency treatment programs in general and those fewer evaluations of juvenile awareness programs. On the outcomes measured, there were no statistically significant differences between the experimental and control groups. Given the consistency of the finding of no treatment effect, a number of authorities have suggested alternative analyses based on offender typology (Empey, 1980) to determine if indeed programs may be producing differential impacts.

In this study, an analysis of outcomes by "high" and "low" offender type was carried out and the results seem suggestive. The high-delinquency youths seemed to show greater pre-to-post-study changes in self-reported delinquencies relative to their controls than did the low-delinquency youths. These results, however, can only be viewed as suggestive given the relatively small sample size available for these analyses.

The overall findings clearly underscore the need for ongoing program evaluations of juvenile interventions. To date, no single treatment modality has produced an enduring impact on criminal behavior (Elliott, 1980). At the same time, the findings of this research seem to point to the importance of looking beyond the repeated reaffirmation of the null hypothesis for possible differential impacts. As Empey (1980: 169) has stated, "It strains the imagination to believe that all programs affect all offenders in exactly the same way."

Clearly, more research attention is needed to develop reliable and standardized typologies that in turn can be used in analyses of treatment impacts.

REFERENCES

- AGETON, S. S. and D. S. ELLIOTT (1978) *The Incidence of Delinquent Behavior in a National Sample of Adolescents*. Boulder, CO: Behavior Research Institute.
- BRODSKY, S. L. (1970) "The prisoner as agent of attitude change: a study of prison profile effects." *British J. of Criminology* 10: 280-285.
- CERNKOVICH, S. A. (1978) "Value orientations and delinquency involvement." *Criminology* 15: 443-458.
- COCHRANE, R. (1974) "The impact of a training school experience on the value of systems of young offenders." *British J. of Criminology* 14: 336-344.
- ELLIOTT, D. S. (1977) "Self delinquency estimate—1977." Unpublished manuscript, Behavioral Research Institute, Boulder, CO.
- (1980) "A repertoire of impact measures," in M. Klein and K. Teilmann (eds.) *Handbook of Criminal Justice Evaluation*. Beverly Hills, CA: Sage.
- EMPEY, L. T. (1980) "Field experimentation in criminal justice: Rationale and design," in M. Klein and K. Teilmann (eds.) *Handbook of Criminal Justice Evaluation*. Beverly Hills, CA: Sage.
- ERICKSON, M. L. and L. T. EMPEY (1976) "Court records, undetected delinquency and decision making," in R. Giallombardo (ed.) *Juvenile Delinquency: A Book of Readings*. New York: John Wiley.
- FINCKENAUER, J. D. (1979) *Juvenile Awareness Project Evaluation. Report 2*. Newark, NJ: Newark School of Criminal Justice.
- and J. R. STORTI (1978) *Juvenile Awareness Project Evaluation. Report 1*. Newark, NJ: Newark School of Criminal Justice.
- GLASS, G. V., P. D. PECKHAM, and J. R. SAUNDERS (1972) "Consequences of failure to meet assumptions underlying the fixed effects of variance and covariance." *Rev. of Educ. Research* 42: 237-285.
- JESNESS, C. F. (1966) *The Jesness Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- KIRIGIN, K. A., D. L. FIXSEN, E. L. PHILLIPS, and M. M. WOLF (1974) "An evaluation manual for collecting follow-up information on youths in trouble." Unpublished manuscript, University of Kansas.
- LANDIS, J. R. and F. R. SCARPITTI (1965) "Perceptions regarding value orientation and legitimate opportunity: delinquents and non-delinquents." *Social Forces* 44: 83-91.
- LANGER, S. (1980) "The Rahway State Prison lifer group: a critical analysis." Unpublished manuscript.
- MALTZ, M. D., A. C. GORDON, D. McDOWALL and R. McCLEARY (1980). "An artifact in pretest-posttest designs: how it can mistakenly make delinquency programs look effective?" *Evaluation Review* 4: 225-240.
- National Council on Crime and Delinquency (1976) *Criminal Justice Newsletter*, Volume 10.
- ROMIG, D. A. (1978) *Justice for our Children: An Examination of Juvenile Delinquent Rehabilitation Programs*. Lexington, MA: D. C. Heath.
- SERPAS, F. R., G. D. LITTON, Jr., and J. L. ASHCROFT (1979) "A closeout report of the Orleans Parish juvenile awareness program." Unpublished manuscript, available from the Mayor's Criminal Justice Coordinating Council, Orleans Parish, Louisiana.
- WOLFGANG, M. E., R. M. FIGLIO, and T. SELLIN (1972) *Delinquency in a Birth Cohort*. Chicago: University of Chicago Press.

EVALUATION & THE HEALTH PROFESSIONS

EDITOR: R. BARKER BAUSELL
University of Maryland

EDITORIAL BOARD

STEPHEN ABRAHAMSON, *Medical Education, University of Southern California*
GERALD L. BARKDOLL, *Food and Drug Administration*
CAROLE R. BAUSELL, *Maryland Alliance for Early Intervention*
DAVID W. CHAMBERS, *Dentistry, University of the Pacific*
LAWRENCE H. COHEN, *Psychological Services, University of Delaware*
RAYMOND M. COSTELLO, *Health Science, University of Texas at San Antonio*
MICHAEL E. DEBAKEY, *Baylor College of Medicine*
JAMES J. DIAMOND, *American College of Radiology*
RAYNARD DOOLEY, *Dentistry, University of Illinois Medical Center*
JOHN D. ENGEL, *National Board of Medical Examiners*
DAVID G. GIL, *Social Welfare, Brandeis University*
GENE V. GLASS, *Education, Arizona State University*
LEON J. GROSS, *National Board of Examiners in Optometry*
RONALD K. HAMBLETON, *Education, University of Massachusetts*
PRAFULLA JOGLEKAR, *Business Administration, LaSalle College*
JEAN JOHNSON, *Nursing, Wayne State University*
HAROLD G. LEVINE, *Medical Education, University of Texas at Galveston*
JEROME P. LYSAUGHT, *University of Rochester*
G. DEAN MAC EWEN, *Alfred I. du Pont Institute, Wilmington, Delaware*

HENRI R. MANASSE, *Pharmacy, University of Illinois Medical Center*
WILLIAM C. MCGAGHIE, *Medicine, University of North Carolina*
M. CLINTON MILLER, *Medical University of South Carolina*
EVAN G. PATTISHALL, *Medicine, The Pennsylvania State University*
W. JAMES POPHAM, *Education, University of California at Los Angeles*
PATRICIA PRESCOTT, *Nursing, University of Maryland*
MARK RAYMOND, *American College Testing Program*
JAY PRICE, *Education, University of Wisconsin, Stevens Point*
ARTHUR ROTHMAN, *Medical Education, University of Toronto*
T. JOSEPH SHEEHAN, *Medicine, University of Connecticut*
CARL H. SLATER, *Public Health, University of Texas at Houston*
KAREN SOEKEN, *Nursing, University of Maryland*
JACK TURNER, *Huntsville-Montgomery County Mental Health Center*
ROBERT A. WALKINGTON, *Health and Human Services*
GEORGE D. WEBSTER, *American Board of Internal Medicine*
MARY L. WOLFE, *Nursing, University of Maryland*
PIERRE WOOG, *Nursing, Adelphi University*
PAUL M. WORTMAN, *Psychology, University of Michigan*

For SAGE Periodicals Press: Catherine M. Chilton

FOUNDING EDITORS: R. BARKER BAUSELL
CAROLYN F. WALTZ

EVALUATION & THE HEALTH PROFESSIONS

Vol. 12 No. 3, September 1989

CONTENTS

- An Evaluation of the Quality Assurance System for Skilled Nursing Facilities in New York State**
EDWARD L. HANNAN, JOSEPH F. O'DONNELL,
and WENDY K. LEFKOWICH 235
- Measuring Patient Intensity: A Reliability Study**
PATRICIA A. PRESCOTT, KAREN L. SOEKEN,
and JUDITH W. RYAN 255
- Evaluation Issues in the Development of Expert Systems in Medicine**
ETA S. BERNER, C. MICHAEL BROOKS,
RANDOLPH A. MILLER, FRED E. MASARIE, Jr.,
and JAMES R. JACKSON 270
- Interviewer Preferences for Respondent Groups**
ROSALIND J. DWORKIN 282
- Medical Education in Prepaid Settings:
Synthesis of the Literature Using
the Case Survey Method**
BARBARA BARZANSKY and JANET PERLOFF 300
- Methodologist's Corner*
- The Effects of Nonresponse and Late Response on a Survey of Physician Attitudes**
EDWARD GUADAGNOLI and SUSAN CUNNINGHAM 318
- Comparing Physicians' Responses to the First and Second Mailings of a Questionnaire**
JEFFERY SOBAL and KEVIN SCOTT FERENTZ 329

EVALUATION REVIEW

A JOURNAL OF APPLIED SOCIAL RESEARCH

Editors: Richard A. Berk, *University of California, Los Angeles*, and
Howard E. Freeman, *University of California, Los Angeles*

Advisory Editors:

Alfred Blumstein, *Carnegie-Mellon University*
Robert Boruch, *Northwestern University*
Wyatt C. Jones, *Brandeis University*
Katherine Lyall, *University of Wisconsin*

Peter H. Rossi, *University of Massachusetts*
Lee Sechrest, *University of Arizona*
Ernst W. Stromsdorfer, *Washington State University*
Ann D. Witte, *Wellesley College*

Associate Editors:

Marvin C. Alkin, *University of California, Los Angeles*
Gerald L. Barkdoll, *Food and Drug Administration*
Kurt J. Beron, *University of Texas at Dallas*
Linda Berry, *Oak Ridge National Laboratory*
Charles L. Betsey, *U.S. Sentencing Commission*
Kenneth L. Bickel, *Aid Association for Lutherans*
Barri A. Braddy, *Research Triangle Institute*
John A. Burghardt, *Mathematica Policy Research*
Anthony G. Cahill, *Pennsylvania State University*
Joseph P. Carbonari, *University of Houston*
Lawrence H. Cohen, *University of Delaware*
Phoebe H. Cottingham, *The Rockefeller Foundation*
Barbara J. Craig, *Virginia Polytechnic Institute & State University*
Robert T. Deacon, *University of California, Santa Barbara*
Dan Durning, *Duke University*
Catherine C. Eckel, *Virginia Polytechnic Institute*
Paul V. Ellefson, *University of Minnesota*
R. Scott Frey, *Kansas State University*
Joel H. Garner, *National Institute of Justice*
Howard H. Goldman, *University of Maryland*
Gary D. Gottfredson, *Johns Hopkins University*
Denis O. Gray, *North Carolina State University*
Risa P. Gressard, *University of Virginia*
George F. Grob, *Department of Health and Human Services*
Harry P. Hatry, *The Urban Institute*
Gary T. Henry, *Education Secretariat of Virginia*
Laura Kann, *Centers for Disease Control*
Kenneth M. Keating, *Bonneville Power Administration*
Barbara H. Kehler, *The Henry J. Kaiser Family Foundation*
Janet E. Kohlhase, *University of Houston*
Helen F. Ladd, *Duke University*
Pamela K. Lattimore, *U.S. Department of Justice*
Louis A. Le Blanc, *Indiana University*
Duane E. Leigh, *Washington State University*
Frances Marcus Lewis, *University of Washington*
Joseph K. Lowery, *U.S. Sentencing Commission*
Michael Mangano, *U.S. Department of Health and Human Services*

Karen R. Mangelsdorf, *Florida State University*
Peter J. May, *University of Washington*
Rebecca A. Maynard, *Mathematica Policy Research, Inc.*
Charles McClintock, *Cornell University*
Joan E. McLaughlin, *Food and Nutrition Service, USDA*
Eric G. Moore, *Queen's University, Canada*
Gary D. Nelson, *Center for Disease Control*
Saundra Murray Nettles, *The Johns Hopkins University*
Jeri Nowakowski, *Northern Illinois University*
Randall J. Olsen, *Ohio State University*
Georgine M. Pion, *American Psychological Association*
Robert D. Plotnick, *University of Washington*
William R. Prosser, *U.S. Department of Health and Human Services*
John M. Quigley, *University of California, Berkeley*
R. James Rasmussen, *California Department of Justice*
Stephen W. Raudenbush, *Michigan State University*
Leslie J. C. Riggins, *U.S. General Accounting Office*
Ray C. Rist, *U.S. General Accounting Office*
John S. Robey, *University of Texas*
Eugene A. Rosa, *Washington State University*
Lawrence F. Salmen, *The World Bank*
William Scarbrough, *U.S. General Accounting Office*
Seymour I. Schwartz, *University of California, Davis*
Lawrence W. Sherman, *University of Maryland*
Andrejs Skaburskis, *Queen's University, Canada*
Frank J. Smith, *North Carolina State University*
C. Matthew Snipp, *University of Wisconsin*
Richard C. Sonnichsen, *Federal Bureau of Investigation*
Michael A. Stoto, *Institute of Medicine, National Academy of Sciences*
Maria Teresa Tatto, *Michigan State University*
James V. Terry, *Consultant*
Robert Tillman, *Wheaton College*
Edward L. Vine, *Lawrence Berkeley Laboratory*
Brian L. Wilcox, *American Psychological Association*
Richard A. Windsor, *University of Alabama*

Editorial Assistant:
Marilyn Garvin

For Sage Periodicals Press:
Lois Lyons and Liann Lech

EVALUATION REVIEW

A JOURNAL OF APPLIED SOCIAL RESEARCH

Volume 13, Number 6

December 1989

CONTENTS

The Politics of Evaluation with Privatized Programs: <i>Who Is the Audience?</i>	JODY L. FITZPATRICK	563
Revitalizing Program Evaluation: <i>The U.S. Department of Education Experience</i>	ALAN L. GINSBURG	579
An Assessment of Victim Service Needs	THOMAS G. BLOMBERG, GORDON P. WALDO, and CAROLA A. BULLOCK	598
Analysis of No-Difference Findings in Evaluation Research	GEORGE JULNES and LAWRENCE B. MOHR	628
Index		657

Sage Periodicals Press  A Division of SAGE Publications, Inc.
Newbury Park • London • New Delhi

EVALUATION PRACTICE

Sponsored by the American Evaluation Association

M. F. "Midge" Smith, Editor
University of Maryland
College Park, MD

Carl Wisler, Associate Editor
General Accounting Office
Washington, D.C.

Advisory Board Members

Stanley Bolin
Department of Defense
Washington, D.C.

Barbara Searle
World Bank
Washington, D.C.

Clifford W. Graves
GRIGSBY/GRAVES Environmental
Direction
San Francisco, CA

Richard C. Sonnichsen
Federal Bureau of Investigation
Washington, D.C.

John Heinberg
Department of Labor
Washington, D.C.

Douglas L. Sporn
U.S. Food and Drug Administration
Washington, D.C.

Ernest R. House
University of Colorado
Boulder, CO

Chris Wye
Housing and Urban Development
Washington, D.C.

Conrad Katzenmeyer
Department of Education
Washington, D.C.

AMERICAN EVALUATION ASSOCIATION BOARD MEMBERS

Ross Conner, *President*
University of California
Irvine, California

Oliver W. Cummings
Arthur Andersen & Co.
St. Charles, Illinois

Michael Quinn Patton, *Past President*
University of Minnesota
Minneapolis, Minnesota

Barbara Gross Davis
University of California
Berkeley, California

Yvonna S. Lincoln, *President-Elect*
Vanderbilt University
Nashville, Tennessee

Peter J. Gray
Syracuse University
Syracuse, New York

Nancy Kingsbury, *Secretary-Treasurer*
U.S. General Accounting Office
Washington, D.C.

Anna Madison
Texas Southern University
Houston, Texas

Larry A. Braskamp
University of Illinois
Champaign, Illinois

Debra J. Rog
COSMOS
Washington, D.C.

EVALUATION PRACTICE

EP • Evaluation Practice • Volume 10, Number 4 • November 1989

IN THIS ISSUE . . .

From the Editor 5

Features

Mediation Evaluation: Status Report
and Challenges for the Future
Hugh O'Doherty and the Faculty Consortium
on Dispute Resolution Research 8

Articles

A Review of Comprehensive Examinations in
Selected Evaluation Training Programs
Diane M. Hauer and Elisa J. Slee 20

Content Analysis of the *Evaluation Studies*
Review Annual, 1975-1986
Nick L. Smith and Prachee Mukherjee 26

Forum

The Politics of Verifying Trustworthiness
in Evaluation Auditing
Thomas A. Schwandt 33

Good Organizational Reasons for Bad
Evaluation Research
Michael Hennessy and Michael J. Sullivan 41

Interviews and Panel Discussions

Discussion on Assessment in Higher Education
L. A. Braskamp 51

Book Review

Group Techniques for Idea Building,
by Carl M. Moore
Reviewed by Julia A. Gamon 57

CAHIER DE RECHERCHE

Récemment parus :

Les opinions des Français fin 1989 et une comparaison 1981-1989, Enquête "Conditions de vie et Aspirations des Français", par Laurent Clerc, Olivier Martin, sous la direction de Georges Hatchuel, N° 11, Février 1991.

Enquête sur mémoires - Etudes et Recherches sans le secteur social : Lexicographie de la littérature grise en action sociale - Volume 1, par Séverine Binard, Michel Legros, N° 12, Mars 1991.

Construction d'un corpus et perte d'information en analyse lexicale : méthodes et pratiques, par Sébastien Lion, sous la direction de Saadi Lahlou, N° 13, Avril 1991.

Penser l'insertion - Méthodes et critères : Contribution à une analyse des critères de l'insertion dans les réseaux de prise en charge des jeunes en difficulté, par Michel Legros, N° 14, Avril 1991.

L'analyse propositionnelle du discours, par Michel Messu, N° 15, Mai 1991.

Classification dichotomique descendante, par Sébastien Lion, N° 16, Mai 1991.

Président : Bernard SCHAEFER Directeur : Robert ROCHEFORT
142, rue du Chevaleret, 75013 PARIS - Tél. : (1) 40.77.85.00

CRÉDOC

Centre de recherche pour l'Étude et l'Observation des Conditions de Vie