



HAL
open science

**“ L’évaluation expérimentale des programmes d’emploi
et de formation aux Etats-Unis : éléments de critique
interne ”**

Coralie Perez

► **To cite this version:**

Coralie Perez. “ L’évaluation expérimentale des programmes d’emploi et de formation aux Etats-Unis : éléments de critique interne ”. *Revue française des affaires sociales*, 2000, janvier-mars (N°1), pp.145-163. halshs-00370754

HAL Id: halshs-00370754

<https://shs.hal.science/halshs-00370754>

Submitted on 25 Mar 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

« L'évaluation expérimentale des programmes d'emploi et de formation aux Etats-Unis : éléments de critique interne », *Revue Française des Affaires sociales*, Numéro spécial l'expérimentation sanitaire et sociale, janvier-mars 2000, pp.145-163.

Résumé

S'agissant des programmes d'emploi et de formation américains, les débats sont très vifs entre les tenants de l'évaluation expérimentale et ceux refusant la sélection aléatoire des participants. Tous s'accordent pourtant sur la nécessité de disposer d'un contrefactuel fiable permettant de figurer ce qu'il serait advenu des participants si le programme n'avait pas existé. En effet, il est admis qu'une évaluation doit mettre en évidence l'efficacité du programme, c'est-à-dire l'amélioration de la situation des participants consécutive et imputable au programme. En 1985 a été décidé l'évaluation du *Job Training and Partnership Act* (JTPA), programme d'emploi et de formation destiné aux « désavantagés économiques » jeunes et adultes. Cette évaluation, qui a produit ses premiers résultats en 1991, constitue une bonne illustration de l'intérêt et des limites d'une telle approche.

Introduction

Les programmes d'emploi et de formation ne représentent qu'une faible part des dépenses liées à l'emploi aux Etats-Unis¹ ; ils sont cependant systématiquement évalués et leurs résultats sont largement commentés, tant dans la littérature académique qu'entre acteurs institutionnels. Aux Etats-Unis, l'évaluation est intimement liée à l'intervention publique sur le marché du travail. Selon le paradoxe de Wilensky, "*the more evaluation, the less program development*" (1985), la réticence traditionnelle à l'interventionnisme va de pair avec la nécessité (pour qu'elle soit légitime), que la preuve soit faite de son "efficacité". Il faut ajouter que l'évaluation est un élément essentiel des jeux de pouvoir entre les branches législative et exécutive, et qu'à ce titre, son institutionnalisation fut plus rapide aux Etats-Unis qu'en France (Rouban, 1984). Le rôle que l'évaluation peut jouer dans les débats politiques n'est pas neutre sur les choix méthodologiques. En effet, les résultats des évaluations doivent être aisément appropriables par les forces politiques et "scientifiquement" valides. C'est sans doute ce qui explique en partie le succès de l'évaluation expérimentale aux Etats-Unis. Il faut ajouter à cela le rôle et les moyens dévolus aux programmes d'emploi et de formation dans le cadre américain, offrant des conditions favorables à l'acceptation sociale de cette approche.

¹ La part du PIB américain consacrée aux dépenses publiques à l'emploi est faible relativement à celle des pays européens (0.43% en 1997-98 contre 3.22% en France et 3.8% en Allemagne en 1997 – OCDE, 1999-) dont un tiers est dévolu aux politiques dites "actives". Par ailleurs, les actions entreprises par le *Department of Labor* au titre de l'emploi et de la formation ne représentent qu'un quart du budget affecté au crédit d'impôt pour les *working poor* (apRoberts, 1996).

Dans cet article, on se propose de rendre compte du cheminement de l'évaluation expérimentale aux Etats-Unis, dans le champ de l'emploi et de la formation, en exposant ses principes et son développement historique face aux démarches concurrentes d'évaluation.

Si la preuve de la causalité est consubstantielle à la définition même de l'évaluation aux Etats-Unis (1), l'approche expérimentale n'a pas imprégné d'emblée les pratiques évaluatives. S'agissant des programmes d'emploi et de formation, elle a connu des phases successives de développement (2). La décennie 80 a vu le regain d'intérêt pour l'évaluation expérimentale et le témoignage le plus significatif de ce succès est sans doute l'évaluation ambitieuse du programme JTPA (3). Toutefois, l'expérience américaine montre que l'évaluation expérimentale comporte aussi des limites. Le relatif consensus autour des attraits de l'évaluation expérimentale s'est plutôt réalisé par défaut, compte tenu du difficile usage politique des résultats des évaluations non expérimentales (4).

1. Spécificités de l'évaluation microéconomique aux Etats-Unis

1.1. La preuve de la causalité au cœur de la démarche évaluative

Etablir un lien entre l'action (la participation à un programme d'emploi et de formation) et la transformation de l'état de l'individu consécutive à ce passage, est au cœur de toute démarche évaluative. Or, comment isoler les effets produits par le programme des effets d'autres facteurs agissant de façon concomitante ? Contrôler ces effets constitue une tâche difficile que n'affrontent pas directement toutes les évaluations, soit que la recherche d'une relation causale ne soit pas leur objectif premier, soit que les conditions ne peuvent être réunies pour l'établir avec fiabilité. Ainsi, la question de la causalité a toujours été en débat parmi les théoriciens de l'évaluation. Pour D.T. Campbell et J.C. Stanley (1966), la capacité à produire des estimations non biaisées de la relation causale entre intervention et résultat (appelée validité interne) est centrale dans toute démarche évaluative. Elle traduit une vision à la fois exigeante et restrictive de l'efficacité : un programme est *efficace* si l'on peut prouver qu'il a permis l'amélioration de l'état des bénéficiaires au regard d'une situation alternative: le plus souvent, l'absence de programme.

Certes, d'autres approches concurrentes de l'évaluation existent, qui accordent un rôle moindre à la validité interne et privilégient d'autres dimensions telle que la compréhension de la mise en œuvre d'un programme (cf. Guba et Lincoln, 1989). Dans ce cas, la notion d'efficacité est plus diffuse car susceptible d'être appréciée par un éventail large de critères. Ces approches ont toutefois peu imprégné les évaluations de programmes d'emploi et de formation aux Etats-Unis (Perez, 1998c). Ainsi, et plus qu'en France, la preuve de la causalité entre intervention et résultat est au cœur de la démarche évaluative. Il est attendu d'une évaluation qu'elle démontre, par un choix méthodologique adéquat, l'amélioration de la situation des bénéficiaires consécutive au programme; autrement dit, qu'elle mesure les *impacts nets* du programme. Pour se faire, la conception expérimentale de l'évaluation, telle qu'elle est pratiquée en médecine pour évaluer l'effet d'un traitement, est retenue pour évaluer les programmes sociaux². Sont transposés dans ce champ tant le principe de la sélection aléatoire que le vocabulaire de l'expérimentation ("groupe de traitement" et "groupe de contrôle").

² Cette dénomination est fréquemment employée aux Etats-Unis; elle souligne la proximité entre les programmes d'assistance sociale (*Welfare*) et les programmes d'emploi et de formation.

1.2. L'évaluation expérimentale et ses alternatives : l'enjeu du principe de sélection aléatoire

Depuis les années 60, la pratique évaluative américaine a vu le développement de trois démarches concurrentes pour évaluer les programmes d'emploi et de formation. Leur manière d'envisager l'application du principe de sélection aléatoire les distingue.

L'évaluation expérimentale³ se distingue de toute autre démarche évaluative par l'affectation aléatoire des individus éligibles à deux groupes, un *groupe de traitement* (les participants au programme) et un *groupe de contrôle* composé de non bénéficiaires. Le recueil d'informations sur ces deux groupes peut se faire sur une base longitudinale et permet de comparer en tout point du temps la situation des participants avec celle de leurs homologues.

La comparaison directe des situations postérieures à la participation s'effectue directement, les deux groupes étant statistiquement équivalents. Par exemple, la différence entre les revenus moyens de chacun des groupes fournit une estimation valide de l'impact du programme. Cette différence est comparée à ce qu'il serait attendu sur la base de la chance grâce à des tests statistiques traditionnels de signification. Toute différence significative peut être attribuée au programme, l'affectation aléatoire assurant l'expérimentateur d'observer, *toutes choses égales par ailleurs*, l'effet non biaisé du programme sur les participants.

Certaines évaluations, dites non expérimentales parce qu'elles n'ont pas recours à la sélection aléatoire, peuvent tenter de comparer la situation subséquente au programme de participants et de non participants. Elles se heurtent dans ce cas au problème du biais de sélection. En effet, si les participants à un programme sont repérés par un certain nombre de caractéristiques observables, des facteurs non observables (tels que la motivation ou l'attachement à l'emploi) déterminent pour une part la participation au programme, et sont probablement corrélés avec des facteurs non observables influant sur la variable de résultat (les gains par exemple) : il en résulte un biais de sélection. Autrement dit, même sans participer au programme, il est permis de penser que les bénéficiaires n'auraient pas connu les mêmes revenus moyens que les non-participants.

Pour contourner ce problème, il peut être tentant de reproduire la logique de l'évaluation expérimentale en constituant non aléatoirement un groupe de comparaison, aussi proche que possible du groupe de participants. Pour se faire, deux grandes options (non exclusives) sont envisageables : l'appariement (*matching*) des groupes de participants et de non-participants sur la base de caractéristiques supposées liées à la variable de résultat, et le contrôle statistique des variables influant a priori sur la variable de résultat. Ces évaluations sont qualifiées de quasi-expérimentales.

Encart 1

³ L'expérimentation est généralement associée à l'évaluation d'un programme-pilote, susceptible d'être généralisé à l'ensemble du territoire. Toutefois, les problèmes "techniques" posés étant en grande partie de nature identique, on parlera indifféremment d'expérimentation ou d'évaluations sur données expérimentales.

Les programmes d'emploi et de formation destinés aux "désavantagés économiques" dans le cadre du CETA et du JTPA

Les programmes d'emploi et de formation américains gérés par le *Department of Labor* s'inscrivent dans des lois-cadres exprimant les grandes orientations fixées par le Congrès.

Deux lois-cadres se sont succédées depuis 1973: le *Comprehensive Employment and Training Act* et le *Job Training and Partnership Act*. Dans ce cadre sont versées des subventions globales aux états afin qu'ils conçoivent et mettent en œuvre leurs propres programmes, et sont instaurés des programmes catégoriels dirigés vers les populations les plus vulnérables.

En tendance, les actions en faveur de l'emploi et de la formation se concentrent sur les caractéristiques productives des individus vus comme structurellement inadaptés (les "laissés pour compte économique") et prennent la forme de programmes catégoriels dits de "seconde chance" (*remedial system*). Précisément ciblés (amérindiens, jeunes issus de familles pauvres, travailleurs âgés et disposant de faibles ressources...) et ayant des objectifs opérationnels définis (réduire la dépendance à l'aide sociale, accroître le revenu et augmenter les opportunités d'emploi), ces programmes se prêtent bien aux évaluations expérimentales.

Les évaluations mentionnées dans ce texte se réfèrent aux programmes destinés aux "désavantagés économiques" jeunes et adultes dans le cadre du CETA et du JTPA. Dans cette dernière loi-cadre, ces programmes ont été regroupés sous le titre IIA. Le principe commun est qu'une fois vérifiées les conditions d'éligibilité (essentiellement liées aux ressources) et estimés leurs besoins, les participants sont orientés dans principalement quatre catégories de "services": l'aide à la recherche d'emploi (*job search assistance*), la formation générale (*classroom training*), la formation en entreprise (*on-the-job training*) et l'expérience de travail (*work experience*) (Perez, 1998a et b).

Depuis les années 70, les programmes d'emploi et de formation ont constitué un champ privilégié de développement d'évaluations microéconomiques sophistiquées. Ces évaluations sont facilitées, dans ce domaine, par la conception même de ces programmes alliant une définition des objectifs opérationnels, une durée d'activité, un public-cible et un budget précis⁴. Dans ce contexte, l'évaluation expérimentale a rencontré un intérêt croissant, accentué par les vertus qu'on lui a prêté au regard des autres démarches d'évaluation.

2. Diffusion de l'évaluation expérimentale dans le champ de l'emploi et de la formation

Les travaux de Campbell et Stanley ont largement influencé l'évaluation des programmes sociaux à tel point que la décennie 70 a été qualifiée de "*Decade of Social Experiment*" (Greenberg, Robins, 1986). Les expérimentations de première génération, conduites à l'échelle fédérale, visaient l'estimation de paramètres tels que l'élasticité de l'offre de travail en vue de l'évaluation ex ante ou ex post de programmes fédéraux⁵. L'ambition théorique de ces travaux s'est heurtée à la complexité du réel et au coût de réalisation; leur nombre a décliné jusqu'au début des années 80.

⁴ La *Sunset Legislation* de 1976 suspend la durée de vie d'un programme aux résultats de son évaluation. L'évaluation est "*la clé de voûte dans la recherche d'une meilleure accountability*" (Rouban, 1984).

⁵ Le cas emblématique étant l'évaluation du *New Jersey-Negative Income Tax program* en 1968.

En matière d'évaluations des programmes d'emploi et de formation (cf. encart 1), les travaux menés dans les années 60 et 70 se limitaient généralement à repérer la situation d'emploi et à mesurer les revenus subséquents à la participation au programme. Le dispositif évaluatif du *Comprehensive Employment and Training Act* (CETA), mis en œuvre en 1973, constitue la première réalisation d'un système coordonné et régulier de mesure d'impacts nets. Il s'agit d'une évaluation quasi-expérimentale, c'est-à-dire élaborant statistiquement un contrefactuel au groupe des participants. L'originalité du dispositif est de s'appuyer sur une base de données spécifique sur les caractéristiques des participants, le *Continuous Longitudinal Manpower Survey* (CLMS). Les évaluateurs sont convaincus que, pour obtenir une estimation fiable des impacts du programme (et contourner le problème du biais de sélection), ils doivent disposer des revenus antérieurs à l'évaluation sur une période suffisamment longue, et ce pour les membres des groupes de participants et de non participants. Les données disponibles rendent alors possible l'application d'une variété de procédures d'appariement des membres des deux groupes (Barnow, 1987). Parallèlement, des évaluations expérimentales ponctuelles des programmes d'emploi et de formation ont lieu dans les années 70. Ce sont des expérimentations que l'on peut qualifier de seconde génération, c'est-à-dire focalisées sur la mesure ex post d'impacts nets, le plus souvent pour des programmes pilotes ayant vocation à être étendu à l'ensemble du territoire.

Lorsque la seconde loi-cadre sur l'emploi et la formation est votée en 1982, le *Job Training and Partnership Act* (JTPA), un projet d'évaluation est bâti dans la continuité de celui du CETA. Une évaluation quasi-expérimentale est alors envisagée, basée sur la constitution de nouvelles bases de données⁶. Les polémiques autour des résultats de l'évaluation du CETA vont pourtant mettre fin à ce projet. En effet, comme l'ont montré L. Bassi (1983) et B. Barnow (1987), la variabilité des résultats obtenus par l'évaluation non expérimentale compromettrait toute conclusion sur l'efficacité du programme (cf. encart 2). En réaction, l'évaluation expérimentale du JTPA titre IIA est alors décidée et, dans son sillage, le *design* expérimental étendu à d'autres projets d'évaluation de programmes d'emploi et de formation (comme le JOBSTART ou le Job Corps...).

Encart 2

La non fiabilité des évaluations quasi-expérimentales : exemple du CETA

Les évaluations successives du programme CETA fournissent un exemple de l'inconsistance des résultats fournis par les évaluations non expérimentales tentant de comparer participants et non-participants sans avoir recours à la sélection aléatoire.

Les trois évaluations successives réalisées par la firme Westat Inc. présentent des impacts globaux (en termes de différentiels de revenu) s'échelonnant entre 129\$ et 600\$, selon la cohorte retenue. Face à la variabilité de la première vague de résultats, le *Department of Labor* (DOL) et le *Congressional Budget Office* décident le financement de nouvelles évaluations reposant sur la même base de données (CLMS). Elles ne réussirent pas à lever le doute sur les résultats du programme. Un groupe d'évaluateurs de cette seconde vague conclut : "*Bien que ces évaluations furent toutes basées sur le même ensemble de données,*

⁶ Une base de données détaillées sur les caractéristiques des participants fut envisagée (*Job Training Longitudinal Survey* : JTLS) ainsi qu'une enquête spécifique nationale pour constituer le groupe de comparaison (*Survey of History of Work* : SHOW).

elles ont produit un éventail très large d'impacts estimés du programme. En fait, selon l'étude retenue, il était possible de conclure à un accroissement des revenus des participants consécutif au programme, ou bien au contraire, à une réduction de ces revenus au regard de ceux des non-participants" (Dickinson et al., 1986, pp. 452-453).

Effectuant une "méta-évaluation" du CETA, B. Barnow (1987) discerne trois sources de variabilité des résultats entre les évaluateurs :

- les différentes techniques d'appariement choisies par les évaluateurs contribuent, dans une faible mesure, aux écarts dans les estimations des effets;
- la constitution des groupes de comparaison produit de la variabilité car, bien que s'appuyant sur la même base de données, les auteurs n'ont pas retenu les mêmes critères (âge, durée de présence dans le programme, type de mesures et leur éventuelle combinaison...);
- les spécifications économétriques différentes du modèle des gains et de la décision de participer au programme; elles constituent, selon B. Barnow, la source de variabilité la plus importante. Les différences renvoient essentiellement à des hypothèses différentes sur les variables à introduire et sur la manière de maîtriser le biais de sélection⁷.

L'évaluation du CETA a donc ravivé les débats entre partisans et opposants de l'évaluation expérimentale en donnant l'avantage aux premiers. Le projet d'évaluation du JTPA en 1986 est pour eux, l'occasion de réaliser une évaluation expérimentale ambitieuse puisqu'elle s'applique à un programme fédéral en fonctionnement (*on going program*) et mis en œuvre dans l'ensemble des états⁸. Cette évaluation constitue ainsi une bonne illustration du regain d'intérêt pour la méthode expérimentale et des débats qui l'entourent.

3. Le regain d'intérêt pour l'évaluation expérimentale: l'évaluation du JTPA titre IIA

3.1. Arguments pour l'expérimentation

Avant de lancer, sur la demande du Congrès, l'évaluation de ce programme, le DOL sollicite en 1985 l'avis d'un panel d'experts composé de chercheurs reconnus pour leurs travaux sur l'emploi, la formation et l'évaluation⁹. *Ce Job Training Longitudinal Survey Advisory Panel* tire la conclusion, sur la base de l'évaluation du CETA, qu'aucun résultat fiable n'est produit par les évaluations non expérimentales. En conséquence, ils recommandent la méthode de la sélection aléatoire pour l'évaluation du JTPA, alliant fiabilité et clarté des résultats. Leur rapport final est un plébiscite pour l'expérimentation dont nombre d'arguments sont encore mobilisés aujourd'hui¹⁰.

Les critiques traditionnellement portées à l'encontre d'une telle démarche telles que la faisabilité ou le coût, sans être niées, sont mises en balance avec les échecs jugés coûteux des

⁷ Les estimations pour une même cohorte sont extrêmement sensibles à la constitution des groupes de comparaison. Ni un éventuel changement de conjoncture économique, ni des modifications notables du programme ne peuvent expliquer totalement les grands écarts observés d'une cohorte à l'autre.

⁸ Le programme JTPA titre IIA concerne en 1992 près de 960.000 personnes et dispose d'un budget de 988 millions de dollars (US DOL, 1995). Bien que le budget affecté à l'évaluation du JTPA n'ait jamais été formellement rendu public, il est estimé à environ 5% de la dotation annuelle du programme.

⁹ Dont R. Lalonde, L. Orr, G. Burtless, T. Fraker, R. Maynard...

¹⁰ Stromsdorfer et al., 1985, *Recommandations of the Job Training Longitudinal Survey Research Advisory Panel*, DOL.

évaluations non-expérimentales précédentes. En substance et selon ses partisans, l'évaluation expérimentale autorise un nombre limité de questions mais auxquelles elle peut répondre de manière précise.

Par ailleurs, l'expérimentation pose un problème éthique car son protocole l'amène à refuser l'accès au programme à des personnes éligibles. Dans l'argumentaire des tenants de l'expérimentation, ce problème est contourné par le *volontariat* de la participation à l'expérimentation (c'est-à-dire l'accord préalable des personnes éligibles à participer au processus de sélection aléatoire). Les experts expliquent, non sans cynisme, que les gagnants et les perdants de l'expérimentation ne sont pas connus a priori¹¹. Les résultats produits par l'évaluation expérimentale sont perçus comme lisibles et scientifiquement valides. Ils apparaissent adaptés à éclairer la prise de décision en fournissant un avis scientifique sur l'opportunité de maintenir un programme¹², même si cet usage comporte le risque de substituer un calcul économique au débat politique.

Sur les conseils de ce panel d'experts, le DOL opte pour une conception expérimentale de l'évaluation du JTPA titre IIA (cf. encart 1) et confie sa réalisation à une firme privée d'évaluation (Abt Associates Inc.).

3.2. Le recueil des données expérimentales

Le recueil des données expérimentales nécessaires à l'évaluation du JTPA s'est effectué selon une procédure spécifique (cf. annexe). Sur la base des candidatures à l'entrée dans le programme, les opérateurs vérifient dans un premier temps que les conditions d'éligibilité soient bien remplies. Une fois l'éligibilité avérée, les candidats sont évalués par les opérateurs qui vont déterminer les "services" utiles à fournir (*program activities*). Rappelons que ces services sont initialement: l'aide à la recherche d'emploi, formation en entreprise, la formation générale et socialisation au travail.

Cependant, dans le fonctionnement réel du programme, les opérateurs conseillent aux individus une combinaison de ces services. Il n'était donc pas possible d'isoler les impacts relatifs à chacun des services sans perturber le fonctionnement du programme sur les sites. C'est la raison pour laquelle trois groupes de services (*service strategy*) ont été constitués :

- la formation générale qui peut s'accompagner des autres services à l'exception de la formation en entreprise;
- la formation en entreprise et l'aide à la recherche d'emploi pouvant s'accompagner d'autres services à l'exception de la formation générale;
- les autres services regroupant les individus qui n'auraient été orientés ni vers la formation générale ni vers la formation en entreprise.

¹¹ "Se pose, après tout, un problème éthique quand le programme ou traitement ne fonctionne pas. [Dans ce cas] le groupe de non-participants voit sa situation inchangée, tandis que celui des participants a perdu son temps, supporté des coûts psychologiques, vu ses espoirs et ses attentes déçus" (dans Stromsdorfer et al.).

¹² "Parce que les décideurs peuvent aisément se saisir des résultats d'une expérimentation simple et de leur signification, ils sont plus à même de se concentrer sur les implications de ces résultats sur la politique publique" (Burtless, 1995, p. 69). Des auteurs comme Aaron (1978) et Gueron (1991) ont développé l'idée selon laquelle la complexité des phénomènes observés mise en évidence dans certaines évaluations peut entraver l'action publique en produisant des débats scientifiques obscurs pour les profanes.

Pendant ce processus d'orientation, les opérateurs expliquent aux "candidats" que tous n'auront pas accès au programme du fait de l'évaluation, et qu'un tirage au sort déterminera les participants. Les individus consentant à participer à l'évaluation doivent signifier leur compréhension des conditions et leur accord (notamment sur le recueil de données administratives les concernant) par la signature d'un document¹³ : ils sont ainsi déclarés "volontaires" pour participer à l'expérimentation.

Dans un second temps, les dossiers sont transmis à l'équipe d'évaluation qui procède au tirage au sort pour chaque groupe de services. Dans chacun d'eux, deux tiers des candidats sont affectés à un groupe de traitement (les futurs participants), un tiers au groupe de contrôle (ceux dont l'accès au programme est interdit pendant la durée de l'évaluation). Afin de ne pas biaiser l'observation ultérieure des situations des individus, les membres du groupe de contrôle doivent, en théorie, ne pas bénéficier de services d'emploi et de formation analogues. Ce processus assure de comparer *ceteris paribus* ce qu'il advient des participants au programme au regard de leurs homologues, c'est-à-dire ceux qui s'étaient vus recommander (puis refuser) une combinaison identique de services.

Le processus de recueil des données expérimentales envisagé par les évaluateurs suppose la collaboration des opérateurs du programme. Cependant, ces derniers manifestent rapidement leurs réticences et des compromis doivent être trouvés pour maintenir la réalisation de l'évaluation expérimentale et garantir sa validité scientifique.

3.3. Réticences et compromis dans la mise en œuvre de l'évaluation expérimentale

La décision d'autoriser une expérimentation est délicate sous deux aspects : il faut dénier l'accès au programme à des "ayants droit" et autoriser l'accès à des individus auxquels on ne peut garantir le sens des effets du traitement. Ce dilemme, bien connu en médecine, resurgit dans le champ des expérimentations sociales (où pourtant les enjeux sont moins importants, d'après G. Burtless 1995). Deux arguments sont utilisés à l'appui de la méthode : premièrement, les individus doivent être informés des risques encourus et un dédommagement financier peut être versé à l'issue du programme en cas de détérioration de leur situation initiale. Deuxièmement, la nécessité d'évaluer de manière expérimentale le programme est renforcée si l'on est incapable de dire avec certitude le sens et l'ampleur des effets attendus –les résultats fournis permettant de répondre précisément à cette question –.

Un troisième argument, plus cynique, s'appuie sur l'insuffisante capacité d'accueil des programmes d'emploi et de formation. Tous les candidats au programme évalué n'auraient de toute façon pas pu participer au programme, les moyens étant insuffisants et l'offre restreinte. L'expérimentation est alors présentée aux administrateurs comme "*un moyen équitable et rationnel de décider de la répartition des ressources qui leur sont comptées*" (Gueron, 1991, p. 197).

En dépit de ces arguments, l'évaluation expérimentale du JTPA a heurté les opérateurs sollicités pour mettre en œuvre le recueil des données expérimentales.

¹³ Les rapports d'évaluation ne précisent pas le taux d'acceptation de ces conditions.

Tout d'abord, un modèle probabiliste avait initialement fourni un échantillon de 76 premiers sites¹⁴, représentatifs de l'ensemble, sur lesquels devaient être prélevées les informations. Cependant, aucun des administrateurs des sites sélectionnés n'accepta de participer à l'expérimentation, et ce en dépit d'un dédommagement financier prévu par le DOL. La préoccupation éthique et d'image publique des opérateurs face à la nécessité de refuser l'accès au programme à des candidats éligibles a constitué la raison essentielle de ce refus. Un autre motif était la crainte de ne pouvoir atteindre les critères de performance fixés par le gouvernement étatique, ces critères servant à apprécier la bonne gestion locale du programme¹⁵. Finalement, seulement 16 sites ont accepté de participer à l'évaluation¹⁶. Ce compromis est susceptible de rendre délicate la généralisation des résultats de l'évaluation.

Par ailleurs, et pour atténuer les réticences des opérateurs, il a été décidé qu'un service minimal de formation et d'aide à la recherche d'emploi serait fourni aux membres du groupe de contrôle, sanctionnés par le tirage au sort. Bien que dispensée en dehors du programme JTPA, la prestation fournie était de nature similaire alors qu'en théorie, les membres du groupe de contrôle ne devaient pas avoir accès à de tels services. Ce compromis implique de commenter les résultats observés lors de l'évaluation en termes de "valeur ajoutée" (*incremental effect*) du programme JTPA, au regard des autres programmes dispensés sur le territoire.

Enfin, et toujours dans l'objectif d'une meilleure coopération avec les opérateurs, le ratio de la sélection aléatoire a été porté à deux pour un (deux participants pour un membre du groupe de contrôle). Cette décision résulte d'un compromis entre la nécessaire précision statistique (le ratio optimum étant de un participant pour un non participant) et le souhait des opérateurs de minimiser le nombre de dénis d'admission au programme.

Au total, l'évaluation a porté sur un échantillon de 20601 personnes (jeunes et adultes) éligibles au programme JTPA titre IIA, échantillon recueilli sur 16 sites entre novembre 1987 et septembre 1989, dont les deux tiers ont eu accès au programme (groupe de traitement).

3.4. Exemples des résultats de l'évaluation expérimentale du JTPA

Impact positif sur les gains des adultes, non significatif pour les jeunes

Le tableau 1 expose les résultats de l'évaluation de ce programme en termes d'impacts sur les revenus, agrégés sur la période d'observation (soit 30 mois à compter de l'affectation aléatoire aux groupes de traitement et de contrôle).

¹⁴ Le programme JTPA est dispensé sur la base d'un découpage territorial infra-étatique : 649 *Service Delivery Areas*. Nous les désignerons par le terme de sites.

¹⁵ Deux autres motifs ont pu être évoqués : d'une part, les résultats des évaluations seraient apparus jusqu'alors, pour les opérateurs, comme un "bien public" auquel ils pouvaient avoir accès sans supporter le coût d'une implication directe. D'autre part, le dédommagement financier prévu n'auraient pas été jugé suffisant pour compenser les coûts explicites (administratifs) et implicites (image de marque...) de l'évaluation expérimentale (NCEP, 1993).

¹⁶ "Nous avons sélectionné les sites selon les critères suivants: leur diversité, leur volonté de participer, leur capacité à mettre en œuvre l'expérimentation, la taille de l'échantillon qu'ils pouvaient fournir et la composition de cet échantillon" (US DOL, 1993).

Tableau 1**Evaluation de l'impact sur les revenus du JTPA Titre IIA par groupe démographique.**

	<i>Revenus moyens</i>		<i>Impact par inscrit</i>	
	Groupe de traitement [1]	Groupe de contrôle [2]	En dollars [3]	En % de [2]
Femmes adultes	13 417\$	12 241\$	1 176 \$***	9.6%
Hommes adultes	19 474\$	18496\$	978*	5.3%
Jeunes femmes	10 241\$	10106\$	135	1.3%
Jeunes hommes	15786\$	16375\$	-589	-3.6%

* statistiquement significatif au seuil de 10%, **au seuil de 5%, ***au seuil de 1% (test bilatéral)

Source : USDOL, 1993.

On repère que les participantes adultes ont gagné, en moyenne et pendant deux ans et demi, l'équivalent de 5367\$ annuels (soit 13.417\$ sur l'ensemble de la période), tandis que les membres du groupe de contrôle gagnaient en moyenne 4896\$ par an. Les participantes ont ainsi gagné 9.6% de plus que ce qu'elles auraient perçu sans participer au programme¹⁷. Ce gain est supérieur à celui obtenu par les participants adultes (+5.3% seulement)¹⁸.

Les jeunes participantes bénéficient d'un revenu à peine supérieur aux jeunes femmes du groupe de contrôle (+135\$ sur 2.5 ans). Cette différence, non significative, permet de conclure que le programme n'a pas eu d'impact sur leur revenu et, sur cette base, a provoqué des débats sur l'opportunité de maintenir ce programme pour les jeunes.

Conformément à l'esprit de l'évaluation expérimentale, l'accent est mis sur la *différence* entre les gains moyens des deux groupes et non sur le *montant absolu* des revenus. L'efficacité du programme s'apprécie ainsi au regard d'une alternative (l'absence de programme), et pas sur le fait de "sortir" les participants d'une situation de pauvreté.

Par ailleurs, des résultats plus détaillés font apparaître que les différentiels de revenus (entre groupe de traitement et de contrôle) varient substantiellement selon le temps et les services proposés. Ainsi, les gains issus de la formation générale de base sont négatifs en début de période puis lentement croissants, tandis que la formation en entreprise produit des gains substantiels et rapides se maintenant sur la période d'observation. Pour autant, l'évaluation expérimentale ne permet pas de comparer l'efficacité relative des services proposés puisque l'orientation dans les mesures ne résulte pas d'un processus aléatoire.

La prédominance du critère de revenu

La variable d'intérêt des évaluations de programmes d'emploi et de formation (expérimentales ou non) destinés aux "désavantagés économiques" est le revenu : "*l'évolution des probabilités d'emploi, de même que celle des statuts d'activité est, le plus souvent, une variable secondaire*" (Ermakoff, 1991, p.22). Il faut rappeler que l'accroissement des gains monétaires (comme la réduction de la dépendance à l'aide sociale qui est son corollaire), est un objectif explicite de ces programmes. De plus, aux Etats-Unis, l'emploi a peu de visibilité, une notion sans contenu, une pure révélation du travail révélée par son "prix", le salaire. Ainsi, les

¹⁷ Les moyennes présentées incluent 0\$ pour les personnes n'ayant pas travaillé pendant la période considérée.

¹⁸ Comme le montrait déjà l'évaluation du programme CETA, les femmes tirent relativement plus de bénéfices du programme que les hommes.

critères de taux de retour à l'emploi (aidé ou non), au chômage, à l'inactivité, n'ont pas de réalité dans le cadre de l'évaluation des programmes d'emploi et de formation américains.

Le revenu, et plus précisément le différentiel de revenu entre participants et non participants, fut donc naturellement le principal critère d'évaluation du JTPA. Cette évaluation présentant, dès son origine, un caractère ambitieux et exemplaire, un de ses résultats fut de décomposer le différentiel de revenu en quatre dimensions : l'effet "pur" d'emploi, un indicateur de stabilité de l'emploi, un indicateur d'intensité de l'emploi et un changement du taux horaire de salaire. L'évaluation a ainsi mis en évidence que la moitié du gain de revenu enregistré par les participantes était dû à une augmentation du taux horaire de salaire, mais que le programme ne permettait ni une plus grande stabilité de l'emploi (nombre de semaines travaillées par emploi déclaré), ni une plus grande intensité de l'emploi (nombre d'heures travaillées par semaine). Il faut cependant noter que la plupart des évaluations expérimentales sont moins sophistiquées, ne retiennent que le critère de revenu et, éventuellement, d'autres variables secondaires telles que l'obtention d'un certificat d'études (*General Education Diploma*), un taux de délinquance (pour les jeunes "désavantagés économiques"), ou encore le fait de recevoir une aide sociale.

Si l'évaluation expérimentale repose en théorie sur un principe simple, l'affectation aléatoire, et produit des résultats aisément appropriables bien que frustes (au regard du nombre limités de variables de résultat), elle n'est pas exempte de limites voire de biais, comme nous le soulignons à présent.

4. L'évaluation expérimentale : une panacée ?

Comme on l'a suggéré, c'est l'inconsistance des résultats des évaluations quasi-expérimentales qui a constitué un argument essentiel pour prôner l'affectation aléatoire. Pour autant, des difficultés d'origines diverses rendent les estimations des évaluations expérimentales au mieux imprécises, au pire biaisées.

Au moins trois principes fondamentaux sont à la base de l'expérimentation : la validité interne (exigeant l'affectation aléatoire), la validité externe (problème de représentativité) et la reproductibilité. Or, les conditions concrètes de réalisation de l'évaluation peuvent constituer une entrave à leur atteinte. Par ailleurs, l'évaluation expérimentale est confrontée à des problèmes de faisabilité et de nature des questions posées. Ces difficultés et limites sont illustrées dans le cadre de l'évaluation du programme JTPA Titre IIA.

4.1. Les entraves au principe de validité interne

La validité interne est, nous l'avons dit, la capacité à produire des estimations non biaisées de la relation causale entre intervention et résultat. L'affectation aléatoire des candidats éligibles à un groupe de traitement ou de contrôle assure de disposer d'un contrefactuel fiable. Plusieurs difficultés peuvent compromettre le respect de ce premier principe.

Tout d'abord, les évaluations expérimentales (comme les évaluations non expérimentales) sont confrontées à des problèmes classiques de représentativité des échantillons. Ainsi lors de la collecte des données, certains individus ne peuvent être interrogés,

pour des raisons de localisation par exemple. L'importance de ces non-réponses peut compromettre la validité de l'analyse sur les données restantes. Dans l'évaluation du JTPA, 30% de l'échantillon n'ont pu être recontactés entre les deux interrogations (espacées de 18 mois). L'utilisation des fichiers administratifs (notamment celui de l'assurance-chômage) a été un moyen efficace de tester le biais de non réponse et de le corriger. L'abandon différentiel de membres des groupes de traitement et de contrôle pendant l'expérimentation constitue une autre limite. Ce biais d'attrition a pour conséquence que les membres du groupe de traitement risquent de ne pas être représentatifs de l'ensemble des bénéficiaires potentiels. Ce problème, bien connu des analystes, a été réglé par des redressements d'échantillon adéquats dans le cadre de l'évaluation du JTPA.

D'autres problèmes, plus spécifiques à l'expérimentation, se posent. Il en va ainsi du refus de participation, soit que les candidats éligibles ne souhaitent pas subir l'affectation aléatoire, soit qu'ils refusent de participer à un programme expérimental dont la durée est temporaire et les résultats encore incertains. La parade pour les évaluateurs peut résider dans des formules de compensation financière, à condition toutefois qu'elles n'affectent pas l'attitude des individus face au programme.

En lien avec ce problème, une hypothèse implicite de l'expérimentation est qu'elle ne modifie pas les décisions de participation ou les comportements des agents impliqués. Or, un biais engendré par le processus d'affectation aléatoire provient du fait que les administrateurs du programme impliqués dans l'expérimentation sont contraints à recruter plus largement, afin de pouvoir constituer des groupes (de traitement et de contrôle) de taille suffisante (Heckman, Smith, 1995). Les participants à l'expérimentation n'ont pas, dans ce cas, les mêmes caractéristiques que ceux qui auraient participé au programme si l'expérimentation n'avait pas eu lieu. Ainsi, la crainte des évaluateurs du JTPA était l'appel à des candidats supplémentaires pour constituer le groupe de contrôle (en cas de refus massif face à l'expérimentation). Cet effort (inhabituel) de recrutement, à une période où le taux de chômage moyen des zones observées diminuait, aurait pu biaiser les résultats. Si les évaluateurs du JTPA ont reconnu la difficulté, ils n'ont pas pu en mesurer les conséquences.

Une autre hypothèse nécessaire à l'expérimentation est l'absence de "substitut" proche au programme évalué pour les membres du groupe de contrôle. La non réalisation de cette hypothèse entraîne un biais de substitution (Heckman, Smith, 1995). Autrement dit, la différence entre les résultats des deux groupes ne fournit pas une estimation valide de l'impact net du programme sur les participants. Nous avons dit qu'un compromis avait dû être trouvé entre les évaluateurs et les opérateurs locaux du programme JTPA, compromis résidant dans la fourniture de services similaires. Les impacts estimés du programme JTPA ont donc été biaisés "à la baisse" puisque la situation de référence n'était pas l'absence de programme.

Enfin, une autre condition est l'invariance du traitement pendant la phase expérimentale. Or, la durée d'une évaluation expérimentale peut être très longue, comme le prouve l'évaluation du JTPA décidée en 1985 et achevée en 1993. Ainsi, le JTPA, comme la loi-cadre précédente, a subi des modifications substantielles en pleine phase d'évaluation (par exemple, la loi d'amendement du programme portant sur le ciblage du programme en 1992). Ceci limite la portée des résultats de l'évaluation mais interpelle aussi du point de vue du rapport entre évaluation et prise de décision.

4.2. Les limites à la validité externe de l'évaluation

La validité externe est la capacité à généraliser les résultats de l'évaluation à l'ensemble de la population concernée par le programme. Celle-ci peut être limitée dans plusieurs cas.

Tout d'abord, le mode de sélection des personnes rendu nécessaire par le dispositif expérimental est toujours susceptible d'introduire un biais qui en compromet la validité externe. Ainsi, dans le cas du JTPA, la participation à l'expérimentation s'est effectuée sur la base du "volontariat". On peut alors soupçonner un biais d'auto-sélection.

De plus, l'évaluation expérimentale confère aux membres du groupe de traitement un avantage sur les non-participants, avantage qui serait moins important si toute la population éligible pouvait bénéficier du traitement. Les effets mesurés du programme peuvent être biaisés "à la hausse" au regard des effets attendus de la généralisation du programme à l'ensemble de la population éligible. Cette critique est cependant commune aux évaluations microéconomiques qui raisonnent en équilibre partiel.

Enfin, les contraintes qui ont pesé sur la sélection des sites expérimentés ont pu, dans l'évaluation du JTPA, compromettre la généralisation des résultats à l'ensemble du territoire.

4.3. La reproductibilité

L'expérimentation est une situation "artificielle" qui peut conduire à sa non-reproductibilité. Ainsi, il n'est pas sûr que le même programme dispensé à une population volontaire puis non volontaire produise les mêmes effets. Se pose aussi le problème bien connu de l'effet Hawthorne, selon lequel les participants à un programme expérimental se comportent différemment puisqu'ils se sentent "observés".

De plus, les individus soumis à l'expérience, sachant que celle-ci est de durée limitée (deux à cinq ans), peuvent adapter leur comportement. Toutefois, ce biais est plus probable dans le cas de programmes reposant sur des transferts financiers que dans celui de programmes d'emploi et de formation (excepté peut-être le cas de subventions pures à l'emploi) où le temps passé en formation est court, et le "traitement" donné en une fois.

Enfin, la diversité des pratiques mises en œuvre dans les sites observés (ou dans le site unique s'il s'agit d'un programme-pilote) est un obstacle à la reproduction de l'expérimentation. Il est ainsi important que le programme expérimenté soit pris en charge par les mêmes structures et agences que celles appelées à fonctionner couramment.

4.4. La faisabilité

La faisabilité est la capacité à réaliser l'évaluation dans les délais et avec les moyens fixés. Les délais font l'objet de compromis entre le commanditaire, soumis aux contingences de l'agenda politique (élections, vote de budget...) et l'évaluateur soucieux d'exigences scientifiques. De ce point de vue, l'évaluation expérimentale est une entreprise longue et coûteuse (les premiers résultats n'ont été publiés qu'en 1991). L'affectation aléatoire d'environ 20.000 candidats éligibles au programme a duré, à elle seule, deux ans. La mise en œuvre du processus de sélection aléatoire est onéreuse car elle nécessite de former du personnel et une

assistance technique dans chaque site pour contrôler les affectations et vérifier, pendant l'expérimentation, que le protocole est bien suivi. C'est le prix, plaident les évaluateurs, de la précision statistique. Mais il faut toutefois ajouter qu'il est extrêmement difficile de connaître le coût d'une évaluation, quel soit la méthodologie adoptée, et donc d'établir des comparaisons.

Enfin, la faisabilité d'une expérimentation a un coût politique, lié aux préoccupations éthiques des commanditaires et des opérateurs, dont l'évaluation du JTPA constitue un bon exemple.

4.5. L'évaluation expérimentale : un questionnement limité ?

L'évaluation expérimentale laisse dans l'ombre un certain nombre de questions, soit que des conditions pratiques l'empêchent d'y répondre, soit que cela dépasse son ambition.

Ainsi, outre les limites liées au critère prédominant d'évaluation (le différentiel de revenu), si la question de l'impact de l'accès au programme sur les revenus ou l'emploi des participants est bien traitée par l'évaluation expérimentale, on ne connaît pas suffisamment les processus qui conduisent un individu à participer au programme et déterminent la durée de sa participation. De plus, la focalisation sur les *impacts moyens* d'un programme ne dit rien sur la *distribution* de ces gains ; autrement dit, qui sont les individus bénéficiant le plus du programme ? (Heckman et Smith, 1995).

Une critique plus générale vise le peu de rapports établis entre les résultats de telles évaluations et le fonctionnement du programme évalué¹⁹. Autrement dit, ces évaluations permettent de conclure si le programme fonctionne ou pas, mais ne fournissent pas d'indications aux administrateurs sur les raisons. Cette critique peut avoir des implications sur l'efficacité de l'action publique : en effet, il peut être coûteux, sur la base de résultats médiocres, de supprimer totalement un programme alors qu'une correction de son ciblage ou de son fonctionnement, aurait suffi.

Conclusion

Les évaluations expérimentales de programmes d'emploi et de formation sont fréquentes aux États-Unis, plus souvent réalisées pour des programmes-pilotes que pour des programmes en cours²⁰. L'évaluation du JTPA, tout en étant une bonne illustration d'une évaluation expérimentale de grande ampleur, reste de ce point de vue exceptionnelle. Les débats qui opposent les économistes américains sur les meilleures méthodes d'évaluation sont très vifs et ne sont pas sans impacts puisqu'ils influencent les orientations du DOL dans le financement des évaluations²¹. Ces débats sont-ils pour autant spécifiques aux États-Unis ?

¹⁹ L'insuffisante relation entre évaluation des effets sur les bénéficiaires, évaluation de type organisationnel et évaluation par les bénéficiaires de leur passage par le programme plaide pour le développement d'évaluation "*comprehensive*". Ce terme, employé notamment par A. Blalock, recouvre l'articulation de différents regards portés sur l'efficacité d'un programme public.

²⁰ Les programmes pilotes (*demonstration program*) posent aux évaluateurs moins de problèmes éthiques et techniques (application localisée).

²¹ Le DOL finance la recherche sur les méthodes d'évaluation en octroyant des bourses doctorales et en subventionnant des programmes de recherche universitaires.

En France, toutes les évaluations des dispositifs d'emploi et de formation sont non expérimentales. Les évaluations quasi-expérimentales existent mais sont généralement peu sophistiquées au regard de leurs homologues américaines. Le recours à des méthodes comparatives (reposant sur l'usage d'un contrefactuel) est moins systématique en France qu'aux Etats-Unis. Cela peut s'expliquer par le moindre développement des bases de données adaptées à l'évaluation de ces programmes. Ainsi, quelle que soit la méthode utilisée, tous les évaluateurs soulignent l'importance de la disponibilité et de la qualité des données pour réaliser des estimations fiables. De ce point de vue, la constitution de panels de bénéficiaires et de panels "témoins"²² ouvre des perspectives en France. Mais d'autres raisons, liées au contexte sociétal dans lequel s'inscrivent les évaluations, peuvent être évoquées. Ainsi, les dispositifs d'emploi et de formation sont, en France, de nature plus complexe, tant dans leur conception (cibles moins étroites, articulation de financements, pas de durée de vie fixée à priori...) que dans leurs objectifs ; en conséquence, ils nécessitent d'être évalués au regard d'une combinaison de critères. Cette différence dans le recours aux données expérimentales s'explique sans doute aussi par le fait qu'en France, la preuve de la causalité n'est pas considérée comme une question centrale pour l'évaluation. Faut-il le regretter ? La poursuite de la maîtrise scientifique des effets des programmes n'est-elle pas une quête vaine qui rappelle l'expérience de la Rationalisation des Choix Budgétaires et les prétentions d'une rationalisation scientifique du politique ?

Bibliographie

AARON H., (1978), *Politics and the Professors. The Great Society in Perspective*. The Brookings Institutions, Washington D.C.

ApROBERTS L., (1996), "Quelle politique pour les travailleurs pauvres ?", *IRES Chronique Internationale*, n°40, mai, pp.28-32.

BARNOW B. (1987), "The Impact of CETA Programs on Earnings. A Review of Literature", *The Journal of Human Resources*, vol. XXII, n°2, pp.157-193.

BASSI L. (1983), "The Effect of CETA on the Postprogram Earnings of Participants", *The Journal of Human Resources*, vol. XVIII, n°4, pp.539-606.

BURTLESS G., (1995), "The Case for Randomized Field Trials in Economic and Policy Research", *Industrial and Labor Relations Review*, vol.39, n°1, pp.105-114.

CAMPBELL D.T., STANLEY J.C., (1966), *Experiments and Quasi-experimental Design for Research*, Rand Mc Nally.

DICKINSON K., JONHSON T., WEST R., (1986), "An Analysis of the Impact of CETA on Participants' Earnings", *The Journal of Human Resources*, vol.XXI, pp.64-91.

²² Des panels "loupes" (des panels de jeunes et de chômeurs en mesures) et des panels "témoins" (le panel "jeunes sortants" du Céreq et le panel "chômeurs de longue durée" de l'ANPE) ont été mis en œuvre par la DARES pour mieux apprécier, en dynamique et en comparatif, les effets de la politique d'emploi.

ERMAKOFF I., (1991), Les travaux d'évaluation des politiques de l'emploi aux Etats-Unis : éléments d'une approche critique, Séminaire d'Economie du Travail, Université de Paris I, septembre.

GREENBERG D.H., ROBINS P.K. (1986), "The Changing Role of Social Experiments in Policy Analysis", *Journal of Policy Analysis and Management*, vol.5, n°2, pp.340-363.

GUBA E.G., LINCOLN Y.S., (1989), *Fourth Generation Evaluation*, Sage Publications.

GUERON J. (1991), "Evaluation et action : les programmes pour l'emploi et la formation aux Etats-Unis" dans OCDE, *L'évaluation des programmes pour l'emploi et des mesures sociales*, Paris, pp.191-199.

HAM J., LaLONDE R. (1994), "Looking into the Black Box : Using Experimental Data to Find Out How Training Works", *International Journal of Manpower*, vol.15, n°5, pp.32-37.

HECKMAN J.J., SMITH J.A., (1995), "Assessing the Case of Social Experiments", *Journal of Economic Perspectives*, vol.9, n°2, pp.85-110.

NCEP (National Commission For Employment Policy) (1993), *Evaluating JTPA Programs for Economically Disadvantaged Adults : a Case Study of Utah and General Findings*, Research Report n°92-02, Washington D.C.

OCDE (1999), Les Perspectives de l'emploi de l'OCDE, juin 1999.

PEREZ C., (1998a), "La politique publique d'emploi américaine" dans BARBIER J.C., GAUTIE J. (dir.), *Les politiques publiques d'emploi en Europe et aux Etats-Unis*, PUF, Collection Les Cahiers du Centre d'Etudes de l'Emploi, pp. 197-217.

PEREZ C. (1998b), "Les politiques catégorielles aux Etats-Unis. Fonctions et limites du ciblage des programmes d'emploi et de formation américains", dans Méhaut P., Mossé P., (coord.), *Fondements, portées et limites des politiques sociales catégorielles*, actes des XVIIIème journées de l'AES, Marseille, 10-11 septembre, L'Harmattan.

PEREZ C. (1998c), L'évaluation des programmes d'emploi et de formation aux Etats-Unis, dans Conseil Scientifique de l'Evaluation, *L'évaluation en développement 1997*, Rapport annuel sur l'évolution des pratiques d'évaluation des politiques publiques, La Documentation Française, pp.109-151.

ROUBAN L. (1984), "Evaluation des politiques publiques et mouvement de dérégulation aux Etats-Unis", *Revue Française d'Administration Publique*, n°29, janvier-mars, pp.85-115.

US DOL (1993), *The National JTPA Study Title IIA. Impacts on Earnings and Employment at 18 months*. Research and Report Series 93-C, Washington D.C.

US DOL (1995), Training and Employment Report of the Secretary of Labor, US DOL, ETA, Washington D.C.

WILENSKY H. (1985), "Nothing Fails Like Success : The Evaluation-Research Industry and Labor Market Policy", *Industrial Relations*, vol. 24, n°1, pp.1-19.

Annexe

Le processus de sélection aléatoire des participants au programme JTPA (source: USDOL, 1993)

