

L'expérimentation « telle qu'elle se fait » : leçons de trois expérimentations par assignation aléatoire
« Il s'agit d'appliquer à la politique une rigueur scientifique » (Duflo 2010)

Experiments « in the making » : lessons from three randomized controlled trials

Experimentieren in der Praxis : Erkenntnisse aus drei Experimenten mit Zufallsauswahl

La experimentación “tal como se hace” : enseñanzas de tres experimentaciones por asignación aleatoria

Agathe Devaux-Spatarakis



Édition électronique

URL : <http://journals.openedition.org/formationemploi/4180>

DOI : [10.4000/formationemploi.4180](https://doi.org/10.4000/formationemploi.4180)

ISSN : 2107-0946

Éditeur

La Documentation française

Édition imprimée

Date de publication : 10 juillet 2014

Pagination : 17-38

ISSN : 0759-6340

Référence électronique

Agathe Devaux-Spatarakis, « L'expérimentation « telle qu'elle se fait » : leçons de trois expérimentations par assignation aléatoire », *Formation emploi* [En ligne], 126 | Avril-Juin 2014, mis en ligne le 18 août 2014, consulté le 30 octobre 2020. URL : <http://journals.openedition.org/formationemploi/4180> ; DOI : <https://doi.org/10.4000/formationemploi.4180>

© Tous droits réservés

L'expérimentation « telle qu'elle se fait » : leçons de trois expérimentations par assignation aléatoire

« *Il s'agit d'appliquer à la politique une rigueur
scientifique* » (Duflo, 2010)

AGATHE DEVAUX-SPATARAKIS

Doctorante en science politique dans le cadre d'un contrat Cifre (Conventions Industrielles de Formation par la Recherche) entre le Centre Emile Durkheim (Sciences-Po Bordeaux) et le cabinet d'évaluation de politiques publiques Eureval.

Résumé

■ L'expérimentation « telle qu'elle se fait » : leçons de trois expérimentations par assignation aléatoire

Les expérimentations sociales peuvent-elles construire une logique d'apprentissage commune entre les acteurs de l'action publique et les chercheurs participant à leur évaluation ? A partir des expérimentations évaluées par la méthode expérimentale par assignation aléatoire, cette contribution étudie la capacité de cette méthode à structurer les apprentissages des acteurs. Trois expérimentations, découpées en trois séquences d'interactions, révèlent les lignes de tensions et les conditions d'apprentissage propres à la conduite de cette méthode.

Mots clés : expérimentation sociale, politique publique, mission locale, jeune, évaluation, méthodologie

Abstract

■ Experiments « in the making »: lessons from three randomized controlled trials

Can experiments lead to a joint learning process between policy actors and scientists conducting their evaluation? Grounded on experimental projects assessed by randomized controlled trials, this article studies this method ability to structure learning among its stakeholders. Three experiments, divided in three interaction sequences, reveal the tensions and the conditions of learning carried by this method of evaluation.

Keywords : social experiment, public policy, youth advisory centre, young person, evaluation, methodology

Journal of Economic Literature: H 52, I 21

Traduction : *Auteur*

Malgré leur grande diversité de mise en œuvre, toutes les expérimentations sociales poursuivent l'objectif commun de produire et d'organiser un *apprentissage* des acteurs de l'action publique¹. C'est dans cette perspective qu'à partir de 2008, Martin Hirsch, Haut-Commissaire aux solidarités actives contre la pauvreté, puis à la jeunesse, a organisé un appel à projets puis créé un Fonds d'expérimentation pour la jeunesse (FEJ), doté de 255 millions d'euros (dont 200 millions fournis par l'Etat) dédiés au financement des expérimentations. Les fondateurs du Fonds d'expérimentation ont promu une logique d'apprentissage visant à faire coïncider les intérêts, en termes d'apprentissage, des acteurs de l'action publique et des chercheurs, autour de l'estimation de l'impact d'un programme public. Cette démarche souhaitait mobiliser une méthode d'évaluation en particulier, à savoir la méthode expérimentale par assignation aléatoire². Cette méthode, empruntée aux essais cliniques dans le domaine médical, estime l'effet moyen d'une intervention publique ; à cet effet, on alloue « cette intervention, de manière aléatoire, à un groupe test et un groupe témoin, puis on compare l'évolution de chaque groupe vis-à-vis de l'effet souhaité (Banerjee et Duflo, 2009). À l'heure du bilan, il paraît alors légitime d'identifier dans quelle mesure les modalités de l'organisation de cette démonstration scientifique ont été sources d'apprentissage pour les acteurs de l'action publique.

Cette méthode d'évaluation, finalement peu utilisée puisqu'elle ne représente que 15 % des évaluations financées par ces appels à projets, a cependant concerné environ la moitié (49 %) des fonds engagés par le FEJ³. Si les utilisations de cette méthode ne sont pas représentatives de la diversité des évaluations engagées dans le cadre du FEJ, leur étude nous permet cependant d'analyser la mise en pratique de la logique d'apprentissage expérimental souhaitée par le FEJ.

Le recours à cette méthode d'évaluation a déjà suscité un ensemble de réflexions critiques. Elles pointent son instrumentalisation politique, ses difficultés à fournir des résultats probants généralisables, et son inadaptabilité à des projets sociaux complexes (Gomel et Serverin, 2009; Allègre, 2008 ; Labrousse, 2010 ; Naudet et Delarue, 2007).

Cette étude, mobilisant la sociologie de l'action publique, propose de s'intéresser à la *pratique* de l'expérimentation aléatoire. Les promoteurs de cette méthode ont souligné qu'à l'instar de toute autre entreprise scientifique, la méthode aléatoire comporte une part de

1. Nous nous inscrivons dans une acception large de la notion d'apprentissage, suivant la définition mobilisée par C. Radaelli et C. Dunlop, comme « *l'actualisation de croyances basées sur des expériences vécues ou observées, l'analyse ou l'interaction sociale* » (Dunlop et Radaelli, 2013).

2. Pour observer la préférence pour cette méthode, se référer aux différents guides ou annexes méthodologiques attenants à ces appels à projets ainsi qu'aux communiqués du FEJ jusqu'à aujourd'hui. Les motifs de cette préférence sont complexes et tiennent autant à ses caractéristiques épistémologiques qu'à la sociologie des sciences liée à cette méthode, et pourraient faire l'objet d'un article à part entière.

3. Le nombre de projets évalués par cette méthode s'élève à 33 %, puisque plusieurs projets faisaient parfois l'objet d'une évaluation commune (Fonds d'expérimentation pour la jeunesse 2012, p. 29 ; Fonds d'expérimentation pour la jeunesse, 2013, p. 13)

« *bricolages* » (Banerjee et Duflo 2009 ; L'Horty et Petit, 2010). C'est à ces ajustements que nous souhaitons nous intéresser ici. Il s'agit de comprendre quels en sont la source, les acteurs, mais aussi l'impact sur sa logique d'apprentissage.

Cette contribution présente la façon dont cet instrument d'apprentissage peut être appréhendé grâce à une conception de l'institution comme un objet malléable, dont la forme dépend des interactions entre les acteurs qui la composent. Ensuite, trois cas d'utilisation de la méthode aléatoire illustreront la portée de ce cadre d'analyse. Pour conclure, cette étude dégagera les lignes de tension autour de l'apprentissage entre les parties prenantes de ces expérimentations sociales.

1 | Un instrument d'apprentissage à la rencontre d'une diversité d'intérêts

La méthode expérimentale par assignation aléatoire peut être conceptualisée comme un *instrument* promouvant une logique d'apprentissage particulière. Néanmoins, elle s'inscrit dans une conception de l'institution dont les contours sont fortement forgés par les pratiques et les intérêts des acteurs. La rencontre autour de cet instrument entre les acteurs de l'action publique et les chercheurs statisticiens en économie appliquée, aux pratiques et intérêts différents, est donc source de tensions. L'équilibre entre le *travail institutionnel* de chacune des parties prenantes forge au final une diversité de logiques d'apprentissages qui s'éloignent du modèle promu par les défenseurs de cette méthode ; ces logiques sont illustrées par la présentation de trois cas d'expérimentations.

1.1 Une institution forgée par un travail institutionnel

Comprendre un outil de l'action publique comme un type d'institution sociale fait déjà l'objet de nombreux travaux en sociologie de l'action publique, réunis autour du concept *d'instrument d'action publique*⁴ (Le Galès et Lascoumes, 2004b). Les instruments d'action publique ont vocation à structurer les interactions sociales et certains véhiculent des logiques d'apprentissage (Coletti et Radaelli, 2013).

Suivant le modèle de l'utilisation des essais cliniques, les économistes statisticiens, qui ont promu le recours à cette méthode, l'ont présentée comme un *instrument* favorisant et organisant le développement de la prise en compte de données probantes dans la décision publique (Duflo, 2005 ; Labrousse, 2010 ; L'Horty et Petit, 2011)⁵.

4. Un instrument d'action publique est défini comme « *un dispositif technique à vocation générique, porteur d'une conception concrète du rapport politiquesociété et soutenu par une conception de la régulation* » (Le Galès et Lascoumes, 2004a, p. 14).

5. Dans son application médicale, la méthode aléatoire a été promue et étudiée comme une institution sociale dont le but était de s'assurer que les médecins utiliseraient le meilleur traitement disponible (Marks,

Dans sa conception scientifique, la méthode aléatoire est considérée comme une institution structurante pouvant être comprise comme « *un ensemble de règles et de procédures plus ou moins coordonnées qui gouvernent les interactions et comportements des acteurs et des organisations* » et qui construit un cadre stable au sein duquel les anticipations réduisent les incertitudes et structurent l'action collective » (Powell et DiMaggio, 1991 ; March et Olsen, 1989). En effet, elle constitue la seule méthode d'évaluation qui affecte les individus et organise leur accès à un traitement défini ou à des ressources, et dont la qualité des résultats dépend principalement de la conformité des individus au protocole établi. Le respect de la répartition des groupes par tirage au sort s'avère essentiel ; il en est de même du contrôle de l'interaction entre le bénéficiaire et le metteur en œuvre, afin de garantir la transparence de l'analyse des résultats.

Néanmoins, la prise en compte de la dimension interactionnelle dans la mise en œuvre concrète de cette méthode peut conduire à relativiser son caractère structurant. Elle peut amener à l'appréhender davantage comme une institution sociale dont la forme est largement déterminée par les acteurs. En effet, cet instrument structure les interactions entre acteurs pour une période de temps limitée, et les outils de contrôle pour s'assurer de la conformité des individus aux règles du protocole établis sont faibles.

Elle correspond donc davantage à une conception de l'institution proposée par O. Nay, rapprochant ce concept de la notion de *configuration*, au sens où l'application des règles de l'institution est soumise à un processus relationnel et s'inscrit dans des situations concrètes d'échange, ouvrant à l'identification de plusieurs modèles de jeux d'acteurs (Nay, 1997, p. 17). O. Nay propose alors le concept de *sites institutionnels*, selon lequel une institution n'est pas une structure composée de règles figées dans le temps et l'espace, mais s'incarne de manière différenciée selon les dispositions des parties prenantes à entrer dans ces interactions⁶. Selon cette conception, l'institution est sujette au *travail institutionnel*, c'est-à-dire à l'« *action intentionnelle des acteurs ou des organisations visant à créer, maintenir ou déstabiliser les institutions* » (Lawrence et Suddaby, 2006). Ce travail est composé d'activités stratégiques d'une multitude d'acteurs, chacun guidé par ses intérêts ou ses pratiques, s'inscrivant dans une logique de champ plus globale et s'appuyant sur un ensemble spécifique de ressources, de compétences et de perceptions du monde.

1999, p. 99).

6. Cet auteur définit les sites institutionnels comme « *des espaces sociaux identifiés par trois critères : un ensemble d'acteurs interdépendants (configuration) ; l'existence d'interactions menées en continu (dynamique de jeux) ; et enfin la présence de règles institutionnalisées exprimées dans un ensemble de manières d'agir et de penser (niveau microsociologique) et de logiques sociales (niveau macrosociologique) pesant sur l'ensemble des comportements et des usages des acteurs en situation de co-présence* » (Nay, 1997, p. 20).

1.2 Les sites institutionnels d'un instrument d'apprentissage commun

L'évaluation d'expérimentations sociales par cette méthode a été présentée comme organisant une logique d'apprentissage commune entre les scientifiques et les porteurs de projets⁷. Elle devait permettre aux premiers l'accès à de nouvelles données de terrain, et offrir aux seconds la démonstration scientifique de l'impact de leur action. Plusieurs travaux ont déjà pointé la difficile rencontre des intérêts de la science et de la politique, qui relèvent de pratiques et de règles différentes. Alors que les décideurs publics cherchent un guide pratique pour l'action, les scientifiques s'inscrivent davantage dans une perspective cumulative de la connaissance (Horowitz, 1970 ; Arendt, 1989 ; Weber, 2002). En outre, les logiques d'apprentissage dans le champ politique s'inscrivent aussi dans des enjeux de pouvoir⁸ (Robertson, 1991, p. 55).

La rencontre de ces acteurs, organisée par cette méthode, est donc sujette à des négociations visant à identifier un intérêt commun d'apprentissage. L'atteinte de ce compromis, issu de ce jeu d'acteurs, conditionne les modalités d'incarnation de cette institution ainsi que la logique d'apprentissage qu'elle véhicule. L'étude de ces négociations, au cœur de la méthode aléatoire, nous permet ainsi d'identifier l'apprentissage souhaité puis ensuite son effectivité auprès des parties prenantes. On peut identifier trois moments de négociation dans cette relation d'apprentissage à équilibre instable :

- L'étape de conception et planification de la recherche, où les deux parties doivent s'assurer que la recherche répond aux intérêts de chacun ;
- La négociation de mi-parcours, qui correspond au suivi du protocole, aux ajustements liés à la mise en œuvre, ou aux premiers résultats produits pouvant susciter une renégociation des termes de la relation ;
- L'interaction finale, le moment de la production des résultats, leur interprétation et utilisation par les parties prenantes.

Nous mobilisons cette structure en trois temps pour une analyse croisée des cas d'expérimentation (voir **encadré 1 « Méthodologie »**). Il s'agit d'étudier, pour chacun, les modalités de rencontre et les compromis entre les intérêts et les pratiques de l'économie statistique appliquée d'une part, et de l'action publique, d'autre part.

7. Les porteurs de projets sont les organisations (associations, collectivités, fédérations, universités), acteurs de l'action publique, qui conduisent des projets financés dans un cadre expérimental et s'associent aux évaluateurs.

8. Comme l'avait déjà souligné H. Heclou, les moteurs de l'apprentissage politique peuvent être, au-delà de la résolution de problèmes publics (*Puzzling*), une recherche d'augmentation du pouvoir (*Powering*). D. Robertson précise que les leçons apprises sont aussi mobilisées comme des « *armes politiques* ».

Encadré 1 : Méthodologie

Ce travail est issu d'une thèse de doctorat en sociologie de l'action publique portant sur le développement de la méthode expérimentale par assignation aléatoire, en France. Dans le cadre de cette recherche, une cinquantaine d'entretiens semi-directifs d'une heure trente en moyenne ont été réalisés entre 2012 et 2013, en face à face ou par téléphone, avec les parties prenantes des expérimentations. Ces entretiens ont été réalisés auprès des porteurs de projet, des financeurs et des évaluateurs/chercheurs autour des thèmes suivants : les explications de leur choix de conduire cette méthode, leur rencontre avec les autres parties prenantes de l'expérimentation, leurs modalités et sujets d'interaction en amont, pendant l'expérimentation et lors de la production des résultats. Ces entretiens ont été analysés dans le cadre d'une dizaine d'études de cas ayant chacune pour unité une expérimentation évaluée par la méthode expérimentale par assignation aléatoire.

Les études de cas ont été conduites par analyse croisée de la littérature grise (rapports d'évaluation, documents de restitution ou de programmation), de la littérature académique (articles scientifiques) et de ces entretiens. Grâce à leur structure commune, les entretiens ont été traités par section de questionnement, de manière croisée, afin de comparer les différentes perceptions des parties prenantes aux différents moments de l'interaction expérimentale. Le critère principal de sélection de ces cas était que l'expérimentation soit évaluée par la méthode aléatoire et menée à terme, et qu'ils permettent de rendre compte d'une variété de porteurs de projets et d'évaluateurs.

Ces études de cas ont fait l'objet d'une analyse croisée entre elles, par chacune des trois étapes de l'expérimentation mentionnées précédemment. Il s'agissait de dégager des tendances communes en procédant par induction analytique ; la collecte a été interrompue lorsque les nouveaux cas n'apportaient qu'une information redondante. L'étude de cas est la méthode la plus pertinente pour rendre compte d'un phénomène contemporain peu répandu et dont les éléments contextuels et le phénomène étudié sont difficiles à distinguer. C'est sur la redondance des informations entre les différentes sources et entre les cas, en d'autres termes la triangulation, que repose la rigueur des résultats.

1.3 Trois cas d'expérimentations sociales contre le « décrochage scolaire »

Nous avons éprouvé ce cadre d'analyse sur plus d'une dizaine de cas ayant eu recours à la méthode expérimentale par assignation aléatoire. Nous en avons ici choisi trois témoignant d'une variété de configurations et donc de logiques d'apprentissage différentielles.

Le cas de *la Mallette des parents* donne à voir un projet orienté fortement, dès sa conception, par des problématiques de recherches. Il permet de comprendre les facteurs de succès d'un apprentissage commun entre acteurs politiques et chercheurs.

L'expérimentation de *la Mission locale de Tulle* illustre les difficultés du maintien d'un apprentissage commun lorsque le projet expérimental préexiste au protocole d'évaluation et que les évaluateurs ont peu d'influence sur le projet. Elle nous aide à comprendre les dissensions entre les apprentissages des parties prenantes, à l'issue de l'expérimentation,

qui ont mené à des avis différenciés sur l'efficacité du projet expérimental, témoignant de la confrontation de deux temporalités.

Enfin, *le projet Apipaq (Mission locale de Marseille)* illustre un projet expérimental évalué par un évaluateur ayant des motivations de recherches différentes des deux cas précédents. Cette expérimentation fut peu formatée par les chercheurs, et l'apprentissage des parties prenantes fut aussi différent et perturbé par les pratiques des acteurs de l'action publique.

Ces trois cas ne constituent pas un échantillon représentatif de l'ensemble des cas d'expérimentation évalués par cette méthode d'évaluation ; cependant, ils présentent des sites institutionnels illustrant leur diversité d'incarnation selon les intérêts en jeu.

Ces trois expérimentations s'étendent sur une période allant de 2008 à 2012. Elles sont toutes destinées à favoriser la réussite scolaire et à lutter contre le décrochage.

Le rectorat de Créteil, porteur du projet *la Mallette*, se caractérise, à l'époque de cette expérimentation, par un dynamisme expérimental particulier ; il a en effet amorcé trois expérimentations destinées à être évaluées par assignation aléatoire. Cette dynamique est portée par le recteur, qui souhaite développer de nouvelles modalités d'intervention, ainsi que par l'alliance avec une démarche scientifique. Ceci se traduit par le recrutement d'une chargée de mission détentrice d'un doctorat qui a joué un rôle d'intermédiaire avec l'équipe de recherche chargée de l'évaluation.

Selon ses acteurs, *la Mission locale de Tulle* a fait du développement de projets sa « *marque de fabrique* » et initiait des « *expérimentations* », au sens non scientifique du terme, avant les appels à projets du Haut-Commissariat. L'expérimentation pour la sécurisation des apprentis a été portée par son directeur et une salariée « *chargée du développement* » ainsi que par le directeur du CFA concerné par l'expérimentation.

La Mission locale de Marseille est la plus importante de France. Dès 2005, elle avait initié des actions expérimentales de prévention du décrochage. *Le projet Apipaq* était porté particulièrement par le chargé de développement des partenariats de cette Mission locale et mis en œuvre par un organisme privé de formation.

- Dans le premier cas, *la Mallette des parents*, l'expérimentation, issue du premier appel à projets du Haut-Commissariat aux solidarités actives, a été organisée sur deux années scolaires à partir de septembre 2008. Ce projet était porté par le Rectorat de Créteil et fut évalué par les chercheurs de l'École d'Économie de Paris. Cette expérimentation est présentée par le FEJ comme un succès puisqu'elle a fait l'objet d'une généralisation, dont nous étudierons les conditions, et d'un essaimage à partir de 2010. Elle proposait aux parents d'élèves de 6^{ème} de plusieurs collèges de l'académie des séances d'information sur l'organisation de la scolarité de leurs enfants. L'objectif était, entre autres, d'améliorer la présence en classe de ces derniers, ainsi que leur comportement et l'investissement des parents au collège.

- Le deuxième cas est issu du même appel à projets et s'intitule « Sécurisation du parcours des jeunes s'engageant dans l'apprentissage ». Cette expérimentation fut portée par la Mission Locale de Tulle et évaluée conjointement par le CREST (Centre de Recherche en Economie et Statistique) et la DARES (Direction de l'animation de la recherche, des études et des statistiques). L'objectif de cette expérimentation était d'augmenter le taux d'entrée de jeunes en apprentissage et de prévenir les ruptures de contrats d'apprentissage et de parcours de formation. Engagée en avril 2008, elle s'est terminée fin 2011. Selon les évaluateurs, les résultats de cette expérimentation n'ont pas été assez probants statistiquement pour conclure à l'efficacité du dispositif. Ils ont débouché sur la conduite de nouvelles expérimentations de cette intervention par le Conseil National des Missions Locales dans les régions Centre, Pays de la Loire et Rhône-Alpes.

- Le troisième cas est *Apipaq* (Accompagnement Partenarial d'Itinéraires Personnalisés d'Accès à la Qualification). C'est une expérimentation issue du premier appel à projets du FEJ en 2009. Ce projet était porté par la Mission Locale de Marseille et a été évalué par le Céreq (Centre d'Etude et de Recherche sur les Qualifications). L'intervention, mise en place en décembre 2009, a duré deux ans et demi. Son objectif était d'apporter une nouvelle réponse au décrochage scolaire, en coordonnant sa détection, et en proposant une réponse personnalisée aux jeunes *via* l'organisation d'un Espace d'Accueil Personnalisé (EAP), pendant un an, dans un organisme de formation. Au vu des résultats, les évaluateurs n'ont pas préconisé la généralisation du dispositif. Néanmoins, un dispositif de plateforme de suivi et d'appui aux décrocheurs a été mis en place par les instances nationales avant la fin de l'expérimentation⁹.

2 | Une diversité de logiques d'apprentissages issues de jeux d'acteurs

Le découpage analytique de l'expérimentation en trois séquences permet une comparaison croisée de chaque séquence entre les différents cas. Nous abordons, dans un premier temps, la phase préliminaire de rencontre autour de l'expérimentation, séquence permettant de comprendre les motivations des parties prenantes et leurs intérêts, puis la phase de mise en œuvre, lieu du travail institutionnel des acteurs le plus intense, et enfin la phase de production des résultats et des modalités de leur prise en compte.

2.1 L'expérimentation aléatoire à la croisée des intérêts

La première séquence de l'expérimentation aléatoire est la rencontre d'une variété d'intérêts guidant les acteurs dans cette entreprise commune. Certains sont liés à une recherche

9. Dès décembre 2010, a été instauré le Système interministériel d'échange d'information (SIEI), suivi de la mise en place des plateformes et de la Circulaire n° 2011-028 du 9 février 2011.

d'apprentissage, mais d'autres préoccupations sont aussi présentes, et ces intérêts varient selon les parties prenantes des expérimentations.

Pour les porteurs de projets, la décision de conduire une expérimentation évaluée par cette méthode représentait une opportunité d'apport de connaissances, mais aussi une source de financements dans un contexte de restriction budgétaire.

On retrouve deux logiques, l'une consistait à profiter du financement pour organiser un nouveau dispositif, comme ce fut le cas pour *la Mallette*, l'autre à donner une nouvelle ampleur et structurer des projets existants, comme pour les deux autres cas. Cet apport était conséquent puisque cette subvention du FEJ ou du Haut-commissariat a atteint 620 669 euros pour *la Mallette* et 241 585 euros pour *Apipaq*¹⁰. Un porteur de projet résume : « *Il y avait moyen d'avoir des interlocuteurs de plus haut niveau et d'aller plus loin dans la démarche, (...) et il ne faut pas se leurrer, on était aussi en recherche de financements. À cette époque, il y avait un vrai coup d'arrêt des financements* ».

La démarche expérimentale s'inscrivait aussi dans une logique d'apprentissage qui consistait à produire une estimation de l'impact d'une intervention. L'objectif était de conférer une légitimité scientifique au projet pour garantir la pérennité de son financement, voire sa généralisation par des autorités décisionnaires. Cette légitimité était d'autant plus forte qu'elle était apportée par la méthode aléatoire ; en effet, celle-ci était présentée, dans le cadre de ces appels à projets, comme étant la plus rigoureuse pour estimer l'impact d'une intervention, par rapport à d'autres méthodes quasi-expérimentales ou qualitatives (Le Haut-Commissaire aux solidarités actives contre la pauvreté, 2007 ; Conseil scientifique du FEJ, 2009). De fait, choisir un partenariat avec une équipe de recherche souhaitant pratiquer cette méthode s'inscrivait, d'une part, dans une stratégie d'amélioration des chances de voir son projet sélectionné et financé et d'autre part, permettait de renforcer la légitimité de ce dernier si son impact était estimé positif par cette méthode. Les porteurs de projet de l'expérimentation de Tulle illustrent cette stratégie par leurs propos : « *J'ai compris ce qu'ils entendaient comme type d'évaluation. (...) J'ai compris qu'il y avait des évaluateurs certifiés, entre guillemets, par le Haut-Commissariat et que notre projet ne pourrait être retenu que si on avait ce type d'évaluateur*¹¹ » Ils poursuivent : « *Quitte à trouver un évaluateur, autant trouver le plus sérieux qui soit, qui ne soit pas contestable (...) pour établir l'efficacité du projet.* »

Les entretiens conduits auprès des évaluateurs révèlent plusieurs raisons à leur engagement dans ce processus coûteux en temps et en énergie. On note deux situations types : d'une part, les chercheurs qui souhaitaient avant tout expérimenter cette méthode

10. Source : entretien avec le porteur du projet – nous ne disposons pas des chiffres exacts pour le financement de la *Mission Locale de Tulle* – et note de restitution finale du projet de la *Mallette des parents* (Rectorat de Créteil, 2011).

11. Il s'est avéré au final que les critères de sélection ont été beaucoup plus souples, mais les porteurs ne le savaient pas lors de leur candidature.

d'évaluation ; d'autre part, les chercheurs qui, tout en s'intéressant à cette méthode, l'envisageaient principalement comme un moyen de maximiser leurs chances de sélection par les organisateurs de l'appel à projets.

Le premier groupe de chercheurs se composait d'économistes appliqués, statisticiens, pour la plupart administrateurs de l'INSEE, du Laboratoire du CREST ou de l'École d'Économie de Paris, proches du Cabinet du Haut-Commissaire. Certains d'entre eux, qui faisaient partie du conseil scientifique du FEJ, ont participé au cadrage méthodologique des appels à projets favorables à la méthode aléatoire. Le deuxième groupe était constitué des chercheurs n'ayant pas participé en amont à la définition des appels à projets et ayant eu recours à cette méthode aussi pour maximiser leur chance de sélection.

Cette méthode présentait un intérêt académique double. Les chercheurs ont choisi les expérimentations auxquelles ils participaient afin qu'elles nourrissent leur travail de recherche ou qu'elles leur permettent de s'inscrire dans des débats académiques contemporains. Ainsi, ils ne se perçoivent pas comme des prestataires d'évaluation, mais avant tout comme des scientifiques pouvant aider à résoudre des questions économiques liées à l'intervention publique. Ils peuvent, dans certains cas, orienter une partie du projet pour tester des théories débattues en économie, voire participer à leur conception, comme ce fut le cas pour *la Mallette*. Une autre motivation était de prouver que cette méthode pouvait être appliquée au contexte français de l'action publique. Depuis la modification constitutionnelle du 1^{er} août 2003 inscrivant dans la loi la possibilité d'expérimenter et autorisant pour la première fois, sous conditions, le traitement différencié des citoyens, un groupement de statisticiens a considéré le développement de cette méthode comme une partie intégrante de leur projet scientifique. Des économistes du CREST expliquent : « *Je suis rentré en France avec une seule envie, c'était de faire ça et donc c'est ce que j'ai essayé de faire de multiples façons.* », « *Le Graal des économètres, c'était de résoudre les problèmes de sélection, et l'expérimentation était l'outil phare.* »¹² Plusieurs articles témoignent de l'engouement de ces chercheurs pour cette méthode, portée à l'époque par la médiatisation des travaux d'Esther Duflo (Duflo, 2005 ; L'Horty et Petit, 2011 ; Zamora, 2011 ; Parienté, 2008).

Ensuite, cette méthode présentait l'avantage de pouvoir être utilisée pour l'évaluation, tout en pouvant être convertie en production scientifique correspondant aux standards des revues scientifiques internationales. Ainsi, la pratique de la méthode aléatoire appliquée à des problématiques d'action publique permettait l'apport de financements externes pour les équipes de chercheurs, afin de financer les frais d'une recherche particulièrement

12. Les problèmes ou plutôt « biais de sélection » sont ici des erreurs induites par des méthodes quasi-expérimentales (donc sans tirage aléatoire) employées pour composer des groupes de comparaison. Ils peuvent amener à constituer deux échantillons de personnes ne possédant pas des caractéristiques similaires entre elles ou non représentatives de la population des bénéficiaires potentiels, et fragiliser ainsi l'estimation de l'impact.

couteuse en enquête de suivi des bénéficiaires. Néanmoins, cette alliance nécessaire avec l'évaluation des projets était aussi vécue comme une contrainte par ces chercheurs¹³.

Enfin, cette méthode représentait un moyen de réunir des fonds pour les équipes et centres de recherche, dans une période de réduction des dotations d'État. Cette méthode conférait aux scientifiques un accès aux marchés de l'évaluation. Elle leur offrait même un avantage concurrentiel, puisqu'elle requérait une maîtrise technique dont ils n'étaient qu'un nombre limité à disposer. L'ampleur des appels à projets a ouvert ce marché à d'autres centres de recherche pour lesquels le choix de pratiquer cette méthode s'inscrivait dans une stratégie de maximisation de leur chance d'accéder aux financements des appels à projets. Un évaluateur du Céreq se souvient : « *C'était la méthode préconisée par le Fonds d'expérimentation (...) Dès le départ, on l'a posée comme une contrainte forte, on leur a dit (aux porteurs de projet) "on va faire comme ça", il n'y avait pas vraiment de discussion possible puisqu'on pensait que pour que ce soit accepté par les financeurs, il fallait en passer par là.* » Ce témoignage peut paraître surprenant au regard du faible nombre de projets ayant eu recours à cette méthode d'évaluation parmi les projets financés par le FEJ. Ce paradoxe s'explique par la révision à la baisse des critères méthodologiques des jurys du FEJ, après réceptions des candidatures, devant le manque d'offres d'évaluations de qualité ayant recours à la méthode expérimentale aléatoire.

Les partenariats porteurs de projets/chercheurs se sont formés soit par la sollicitation de la structure par le chercheur, comme dans le cas de *la Mallette*¹⁴, soit par la sollicitation du centre de recherche par le porteur de projets, comme dans le cas de *d'Apipaq*, soit lors de rencontres organisées par les initiateurs des appels à projets, comme dans le cas de l'expérimentation de Tulle¹⁵.

La méthode aléatoire a organisé une logique d'apprentissage nouvelle en évaluation. En effet, l'évaluation n'était pas uniquement orientée vers l'apprentissage des acteurs de l'action publique, mais devait aussi explicitement satisfaire les intérêts de recherche des scientifiques¹⁶. Cette méthode était ainsi sous l'influence d'une diversité d'intérêts d'apprentissages et stratégiques. Dans l'ensemble des cas étudiés, l'évaluation aléatoire a été présentée aux porteurs de projets comme une condition augmentant considérablement

13. Comme l'explique un des chercheurs du CREST : « *le gros intérêt de la recherche non randomisée, c'est que tu n'as pas vraiment à demander l'autorisation ou à avoir l'accord des gens qui mettent en place le projet pour le faire.* »

14. Étaient sollicitées par les appels à projets des « *équipes de recherches, les collectivités, les administrations, les entreprises, les associations* » (Haut-Commissaire aux solidarités actives contre la pauvreté, 2007). Les chercheurs pouvaient donc aussi être à l'initiative de projets.

15. Pour que des partenariats porteurs de projets/évaluateurs se forment, la conférence « Les rencontres de l'expérimentation », en novembre 2007, fut organisée par le Haut-Commissariat (Dares, 2008, p. 8)

16. Cette recherche d'apprentissage est présente à différents degrés lorsque des chercheurs choisissent de conduire des évaluations, quelle que soit la méthode employée ; néanmoins, les prérequis de l'apprentissage des chercheurs par la méthode expérimentale sont les plus contraignants pour la conduite du projet.

leurs chances d'accéder aux financements, amenant ces derniers à l'accepter sans que les implications ne soient clairement définies. Face à eux, les chercheurs qui, pour la plupart, avaient peu l'expérience de cette méthode, avaient sous-estimé l'ampleur du travail institutionnel nécessaire à sa mise en œuvre pour assurer aux résultats produits un intérêt académique. Le développement suivant présente d'une part, la nature des compromis entre ces deux parties ; d'autre part, leur recherche d'un terrain d'entente entre les exigences scientifiques du protocole expérimental et les pratiques de mise en œuvre.

2.2 La mise en œuvre de la sélection aléatoire

La méthode expérimentale par assignation aléatoire, en tant qu'institution scientifique, requiert une conformité forte de la mise en œuvre au protocole préalablement défini, incluant un contrôle de l'administration du « traitement » ainsi que de la composition des groupes de comparaison. Ces exigences ont été sujettes au plus fort travail institutionnel des parties prenantes de l'expérimentation. Elles ont abouti à des sites institutionnels présentant des incarnations variées de cette méthode. Nous présentons, dans un premier temps, la confrontation des pratiques des acteurs scientifiques et de l'action publique autour du tirage aléatoire et, dans un second temps, autour du contrôle des modalités d'intervention.

2.2.1 Le tirage aléatoire, une pratique scientifique structurante

Le respect de l'allocation aléatoire des individus au sein du groupe test et du groupe témoin constitue la clé de voûte garantissant la rigueur des résultats scientifiques produits ; il est donc essentiel pour l'évaluateur/chercheur.

Cependant, cette condition est justement celle qui interfère le plus avec la pratique des acteurs de la mise en œuvre. Au-delà des préoccupations éthiques relatives à la privation d'accès à un dispositif d'un public défavorisé, certains acteurs de la mise en œuvre ont vécu cette condition comme une remise en cause de leurs compétences et de leurs pratiques professionnelles d'analyse des besoins et d'orientation des bénéficiaires. Ce pouvoir d'interprétation de la norme et de décision d'accès à la prestation des « *street level-bureaucrats* » – des acteurs de la mise en œuvre au contact des bénéficiaires – est souligné depuis les années 80 par les travaux français et anglo-saxons de science politique (Lipsky, 2010; Dubois, 1999 ; Dubois, 2013). Il s'avère particulièrement structurant pour les opérateurs de l'action sociale ou éducative. Le porteur du projet Apipaq explique : « *Il y a eu une levée de boucliers des conseillers. (...) En tant que prescripteurs, ils ne pouvaient pas orienter et faire jouer la chance ou le hasard, c'était incompatible, dans la prise de conscience d'un professionnel de considérer que si j'oriente, je ne suis pas sûr que le jeune en bénéficiera. Ça, ce n'est pas dans*

ma pratique. Si j'oriente, c'est que j'ai des compétences pour apprécier, évaluer l'attente du public et si j'oriente vers tel dispositif, je joue mon rôle de professionnel et j'attends que l'offre de service se réalise (...) Il y avait un décalage entre ce que disaient les chercheurs sur l'intérêt de la randomisation et le professionnel au quotidien avec le public (...) Il y avait quelque part un sentiment d'injustice, de grand n'importe quoi, on était opposés entre l'approche scientifique et l'approche du terrain des professionnels. » Cette réaction était plus vive si le dispositif proposait un traitement important, par exemple les résistances furent fortes pour le dispositif *Apipaq* parce que « *les acteurs misaient beaucoup sur ce projet* », alors qu'elles furent moindres pour *la Mallette* parce que, pour reprendre les termes d'un évaluateur, « *on n'offrait pas aux classes bénéficiaires quelque chose de fou* », puisque cette intervention prenait la forme de trois réunions avec les parents d'élèves.

Cette disposition a été l'objet d'un travail institutionnel conséquent de la part des évaluateurs qui eurent recours à plusieurs moyens pour contrôler ce processus. D'abord, les risques pouvaient être limités en modifiant le niveau de la répartition aléatoire. En comparant au niveau des classes au lieu des individus, comme ce fut le cas pour *la Mallette*, ou en faisant intervenir le tirage au sort en amont de l'intervention, c'est-à-dire avant l'intervention des acteurs de la mise en œuvre, comme dans le cas de Tulle. Le tirage aléatoire devenait alors moins visible, voire invisible, et ne venait plus perturber les pratiques d'interactions entre acteurs de la mise en œuvre de l'action publique et bénéficiaires. Le contrôle s'est aussi opéré en aval par une vérification systématique, par l'évaluateur, que les bénéficiaires sélectionnés étaient bien ceux qui avaient reçu la prestation. Un porteur de projet explique ce contrôle : « *Ils avaient une armée d'assistants de recherche pour aller assister à tous les débats et vérifier que les parents dans les débats étaient ceux tirés au sort.* »

En outre, les chercheurs doivent alors s'investir dans un travail institutionnel de nature cognitive. Ce travail de « pédagogie » s'effectue au niveau du porteur de projets, de ses chargés de mission ou des opérateurs au contact des bénéficiaires, selon le niveau du tirage au sort (au niveau des localités, des établissements, ou des individus par exemple). Il s'effectue au cours de réunions parfois conflictuelles où l'équipe d'évaluation rappelle la nécessité de l'emploi de cette méthode pour produire des résultats crédibles, et légitime le recours au tirage au sort.

Les évaluateurs doivent aussi s'assurer que l'expérimentation comprend un nombre suffisant d'individus pour garantir la rigueur du résultat scientifique produit¹⁷. Une moitié de l'échantillon devant être témoin, il est nécessaire d'identifier le double de bénéficiaires potentiels par rapport au nombre d'individus pouvant être traités. Cette

17. En effet, dans la méthode expérimentale, la puissance statistique de la démonstration, c'est-à-dire son aptitude à obtenir un résultat statistiquement significatif, dépend du nombre d'individus participant à l'expérimentation.

condition requiert d'avoir une très bonne connaissance du public ciblé, de savoir comment le mobiliser et d'avoir une capacité à « délivrer le traitement » à un nombre important de bénéficiaires, conditions rarement réunies dans le lancement de nouvelles interventions. Cette difficulté était particulièrement présente dans les projets visant les « décrocheurs du système scolaire », population échappant à un recensement systématique. Cette condition est encore plus difficile à remplir dans les cas de non-recours au dispositif, même s'il a été proposé au bénéficiaire potentiel. Les évaluateurs ont dû, dans plusieurs cas, intervenir pour sécuriser la taille de l'échantillon, soit en prolongeant la durée de l'évaluation, soit en demandant aux porteurs de projets d'étendre le programme à d'autres territoires, comme ce fut le cas pour l'expérimentation de Tulle.

2.2.2 Le contrôle scientifique du traitement

La particularité de la méthode expérimentale est de requérir un contrôle complet sur l'intervention qu'elle évalue. Il s'agit d'une part, de s'assurer que l'administration du « traitement » est bien la même, indépendamment du partenariat d'acteurs, du conseiller et du bénéficiaire, et d'autre part, de pouvoir prétendre à une généralisation des résultats par la suite. Cette participation des chercheurs dans la définition du projet dépend de l'avancement de la formalisation du projet lors de l'association avec le porteur de projets. Cette pratique scientifique de maîtrise de l'administration du « traitement » entre en conflit avec la pratique d'ajustement des acteurs de la mise en œuvre.

La littérature de l'analyse des politiques publiques souligne, depuis les années 70, que la mise en œuvre fait partie intégrante du processus de construction des politiques publiques et qu'elle est le fruit d'une succession de micro-décisions, de négociations, d'apprentissages qui en font un processus évolutif, chaotique et non linéaire dans la durée (Pressman et Wildavsky, 1984)¹⁸. Un porteur de l'expérimentation de Tulle explique : « *Ils ont évalué deux groupes, mais toutes choses n'étaient pas égales par ailleurs, chaque partenaire continuait à avancer, on a des conventions pluriannuelles d'objectifs qui nous obligent à avancer. Donc, comment continuer à faire une évaluation dans ces conditions-là ?* » Cet acteur fait référence à l'environnement changeant de l'expérimentation, aux ajustements des accompagnements au sein du CFA ainsi qu'aux orientations nationales interférant avec le projet, comme par exemple la création des « développeurs de l'apprentissage », en 2009, dans les chambres consulaires, qu'il a fallu intégrer au projet, ou encore le changement d'orientation d'objectifs pour les Chambres de commerce, de la création de contrats d'apprentissage à celle de contrats professionnels.

La phase de mise en œuvre a été le théâtre d'un travail institutionnel émanant de chacune de ces pratiques en conflit. Cette recherche identifie deux situations que l'on retrouve dans les cas présentés.

18. On peut d'ailleurs souligner que ces travaux ont émergé en réaction aux conduites des premières expérimentations aux Etats-Unis dans les années 70.

Première situation, le projet a été initié *ex nihilo* à l'occasion de l'association avec les chercheurs, et ces derniers ont pu influencer la conception du projet. Dans ce cas, la mise en œuvre a été pensée selon la méthode aléatoire dès le départ et le maximum de *process* sont établis et sécurisés. Pour *la Mallette*, les scientifiques ont influencé la nature de l'intervention dès la conception du projet, un évaluateur explique : « *J'avais une revue de la littérature pluri-disciplinaire au moment où l'on interagissait avec les chefs d'établissement sur le thème de l'investissement des parents. (...) On a transféré cette information au responsable du rectorat qui s'occupait de concevoir le dispositif.* » Ici, dès l'origine, la communauté d'intérêts d'apprentissage entre scientifiques et acteurs de l'action publique est organisée.

À l'opposé, on retrouve une autre situation dans l'expérimentation portée par la Mission Locale de Tulle, où l'accompagnement des conseillers des Missions Locales, qui existait avant l'expérimentation, est resté plus opaque et moins maîtrisable par l'évaluateur qui admet : « *Réellement, qu'est-ce qu'ils faisaient ? Je n'en sais rien.* » Sur ce site institutionnel s'observe un décalage entre le travail des évaluateurs et la recherche d'apprentissage des porteurs de projets qui souhaitaient estimer l'impact de leur action.

Ainsi, la séquence d'interaction de la mise en œuvre de l'expérimentation révèle que la recherche d'apprentissage des chercheurs nécessite une conformité de l'action publique à la méthode scientifique expérimentale de production de connaissance. Lorsque l'intervention des porteurs de projets est davantage inscrite dans leurs normes et leurs pratiques usuelles, l'apprentissage académique devient alors compromis et les conditions de l'expérimentation aléatoire sont moins réunies.

2.3 La difficile convergence des logiques d'apprentissage scientifiques et de l'action publique

Suite à une faible puissance statistique, le volet de l'évaluation quantitative de l'expérimentation *Apipaq* s'est appuyé sur des redressements statistiques pour conclure à un écart non significatif entre le groupe test et le groupe témoin quant au maintien en formation et l'obtention de diplôme, ainsi que sur l'insertion professionnelle deux ans après le décrochage¹⁹ (Maillard et *al.*, 2012). Les porteurs de projets ont accepté ces résultats ; cependant, ils ont été surpris et déçus par ces derniers qui allaient à l'encontre de leur ressenti sur le terrain ; certains indicateurs ont d'ailleurs été questionnés par le prestataire, notamment sur ce que l'on entendait par « *un retour positif en formation* ». L'évaluateur de ce projet explique en partie ce décalage par le fait que l'évaluation aléatoire portait uniquement sur l'impact de ce projet sur le retour en formation et pas

19. Ces résultats étaient fragiles puisque sujets à un taux de réponse à l'enquête limité. Un peu plus de 40 % des publics participants ont répondu, autant dans le groupe témoin que dans le groupe test, mais parmi les répondants du groupe test, seuls 74 répondants sur 199 furent effectivement traités. La comparaison des groupes a donc été renforcée par un redressement statistique par pondération normée (voir rapport d'évaluation).

sur d'autres dimensions du projet. Il précise : « *Je ne pourrais pas vous dire si c'est bien ou pas Apipaq, la seule chose que je peux vous dire, c'est que ça a été fait dans l'objectif de faire retourner les jeunes vers une formation qualifiante et que par rapport à cet objectif-là, le contrat n'est pas rempli.* » L'évaluateur déplore aussi sa propre méconnaissance du contenu de l'intervention en tant que telle à l'issue de l'évaluation, en supposant : « *S'il y avait eu un impact positif, la valise de généralisation aurait été peu précise.* » En d'autres termes, l'évaluateur n'aurait pas été en capacité de proposer un projet standardisé transposable à d'autres contextes en l'absence d'informations sur le dispositif en tant que tel et sa diversité. Entre temps, l'État a installé des plateformes de suivi des décrocheurs, relayant la mise en commun de bases de données de repérage de décrocheurs initiée entre la Mission Locale et l'Éducation nationale, et renforcée par l'expérimentation *Apipaq* comme le souligne le porteur de projet, « *Apipaq était une étape, pas une parenthèse* ». Ce site institutionnel révèle donc un décalage d'apprentissage entre, d'une part, l'évaluateur, qui estime un impact non positif selon un objectif défini sur un indicateur unique et, d'autre part, un porteur de projets qui, par cette action, estime avoir renforcé positivement ses partenariats. Les résultats n'ont pas servi à éclairer une nouvelle action puisque les pratiques de repérage des décrocheurs ont été entérinées par une décision nationale prise avant la fin de l'expérimentation. On observe ici des décalages possibles entre le temps de l'expérimentation et la celui de la programmation politique, qui conduisent à prendre des décisions avant la fin de l'expérimentation et la production d'apprentissages.

Les résultats de l'expérimentation sur la sécurisation des parcours des apprentis ont engendré des désaccords entre les porteurs de projets et les évaluateurs du dispositif. Les évaluateurs estiment que l'accompagnement en amont de l'apprentissage, ainsi que pendant, n'a mené ni à un accès accru à l'apprentissage (seulement au pré-apprentissage), ni à une réduction du taux d'abandon (Crest-Dares, 2012). C'est dans la distinction des groupes témoin et test que sont apparues des difficultés. La puissance statistique étant déjà faible, elle n'a pas permis une quantification précise de l'accompagnement. L'expérimentation ne s'est pas déroulée de manière binaire (avec/sans), mais plutôt suivant un *continuum* de rencontres plus ou moins fréquentes avec les missions locales, puisque 40 % des jeunes du groupe témoin de la phase 1 et 50 % du groupe témoin de la phase 2 ont eu au moins un contact avec les missions locales, ce qui a contribué à diluer l'effet du dispositif. L'évaluateur déplore : « *Ils ne se rendent absolument pas compte des enjeux scientifiques, pas de l'enjeu, mais des exigences scientifiques, il faut vraiment que l'évaluation soit complètement circonscrite, qu'elle soit complètement nouvelle et que seuls les jeunes assignés en bénéficient. Donc il faut vraiment qu'elle soit différente de ce qu'on fait habituellement.* » Les évaluateurs préconisent, en conclusion, de mener davantage d'expérimentations de ce dispositif avant de pouvoir se prononcer. Ce qu'ils ont pu mettre d'ailleurs en application, puisque le Conseil national des missions locales a décidé d'essayer ce dispositif en proposant des expérimentations financées par le FEJ. La logique de ces chercheurs n'est pas d'apporter un éclairage sur le fonctionnement d'un dispositif en particulier, mais de commencer à accumuler des preuves pour nourrir ensuite des compilations de résultats produits par

des dispositifs similaires – une perspective scientifique cumulative assumée par le service Evaluation du Fonds d'expérimentation pour la jeunesse (Valdenaire et Bérard, 2013).

Les porteurs de projets ont estimé que les résultats de cette évaluation ne reflétaient pas la réalité du terrain. Comme il n'y a pas eu de véritable groupe témoin, ils ont pris en compte l'évolution globale de la situation des apprentis entre décembre 2007 et décembre 2012. Selon les indicateurs du CFA, le nombre d'apprentis a augmenté de 21 %, les offres d'apprentissage de 8 % et surtout les ruptures de contrat sont passées de 23 % à 18,7 % et les ruptures sans solution non liées à des événements personnels de 8,2 % à 3,12 %. Néanmoins, dans une démarche comparative entre le groupe témoin et le groupe test, les évaluateurs n'ont pu se prononcer sur la significativité des résultats. Les porteurs de projets ont décidé de continuer le projet, malgré l'absence de résultats. L'un de ces porteurs explique : « *Pour nous, l'action ne s'est jamais terminée, personne n'a voulu l'arrêter, (...) on a eu une réunion à la fin de l'action avec les financeurs locaux, ils ont dit "nous on veut bien voir, mais on veut savoir les résultats". Or on n'en avait pas, il y avait des résultats partiels, mais peu concluants.* » Ce cas témoigne du même décalage d'apprentissage que le précédent, entre des chercheurs s'intéressant à un impact particulier et un porteur de projets avec une vision globale du dispositif, qui consolide des processus de mise en œuvre au cours de l'expérimentation et estime que le processus est stabilisé à la fin de la période expérimentale²⁰. En outre, cette expérimentation s'est trouvée en tension entre deux pratiques d'apprentissage aux temporalités différentes. D'une part, les porteurs de projets sont dans l'obligation d'action permanente, laquelle requiert des prises de décision rapides, et donc des résultats probants sur leur action, suffisamment précoces pour défendre le financement de leur projet ; d'autre part, la logique scientifique est elle faite de prudence et de cumulativité des éléments probants avant de recommander la généralisation.

La Mallette des parents est présentée par le FEJ comme le modèle vertueux d'expérimentation ayant donné lieu à une généralisation à 1 300 collèges, en 2010, au vu de ses résultats positifs sur la participation des parents et les comportements et résultats des élèves²¹ (Valdenaire et Bérard, 2013). Néanmoins, la généralisation de cette expérimentation s'inscrit dans le contexte particulier de l'accession du porteur de projets, le recteur de Créteil, au poste de DGESCO (directeur général de l'Enseignement scolaire) au ministère de l'Éducation nationale, en décembre 2009. Celui-ci précise : « *Comme je suis devenu DGESCO dans l'intervalle, on s'est dit qu'il fallait le développer dans toute la France, d'autant plus qu'on voyait toutes les vertus, l'une d'entre elles étant que ce n'était pas très coûteux par rapport à l'effet choisi. Donc on attendait l'évaluation pour envisager cette généralisation que*

20. Un des porteurs de ce projet explique : « *Le travail s'est stabilisé maintenant, il rentrerait dans l'esprit de personne de modifier les pratiques* ».

21. Par exemple, les parents des classes bénéficiaires rencontrent davantage les enseignants (30 % au lieu de 24 % dans le groupe témoin) et les enfants des classes bénéficiaires ont plus de distinctions au conseil de classe (38,5 % contre 34,5 %), mais l'effet sur les résultats scolaires est limité. Nous renvoyons le lecteur intéressé à la synthèse du rapport d'évaluation de cette expérimentation.

j'ai proposée au ministre, qui a accepté. (...) Le fait que moi-même je sois devenu une sorte d'abeille en passant de l'échelle Créteil à l'échelle nationale est un cas particulier. » En outre, cette généralisation s'inscrit dans le contexte de la loi Ciotti, de septembre 2010, qui mettait la participation des parents à la scolarité de leurs enfants en haut de l'agenda politique²². Un observateur interne explique : « *Martin Hirsch était clairement contre la loi Ciotti et pour la Mallette des parents, et Nicolas Sarkozy l'a repris à son compte pour dire : "On fait la loi Ciotti, mais on fait aussi la Mallette des parents, ce n'est pas comme si on était uniquement sur la punition".* »

Il semble ainsi que la généralisation de cette expérimentation relève davantage d'une logique politique que d'une pratique scientifique où une généralisation s'effectue au regard de la validité externe des résultats produits. En effet, en examinant de plus près cette expérimentation, il apparaît que la durée d'expérimentation a été réduite de trois ans à deux ans et que *la Mallette*, en tant que telle, n'a été évaluée rigoureusement que lors de la première année d'intervention sur un échantillon de 200 classes dont 100 tests, parmi lesquelles 543 parents déjà volontaires ont été sollicités et dont au final 58 % ont participé à au moins une réunion sur trois. En effet, le tirage aléatoire en deuxième année permettait d'estimer rigoureusement uniquement l'impact de modes différents de sollicitations des parents sur leur participation aux réunions *Mallette* et ne disposait plus de la puissance statistique suffisante pour confirmer les effets de *la Mallette* sur une deuxième année²³.

Conclusion

Ainsi l'application de ces outils théoriques à différents cas d'utilisation de la méthode expérimentale par assignation aléatoire confirme que cette méthode n'est pas une institution rigide favorisant une logique d'apprentissage univoque.

L'étude des pratiques de l'expérimentation aléatoire démontre que divers intérêts poussent porteurs de projets et chercheurs à s'engager dans une expérimentation ; ces intérêts transparaissent dans la mise en œuvre de cette dernière et resurgissent lors de la production et de l'utilisation des résultats de l'évaluation du projet.

À la lumière des trois cas présentés, il s'avère que la méthode expérimentale par assignation aléatoire ne peut s'inscrire dans un modèle d'apprentissage commun entre l'action publique et la recherche que lorsque les intérêts de l'action publique et les intérêts scientifiques en jeu convergent. Elle nécessite, dès sa conception, de réduire la marge de

22. Pour mémoire, la loi Ciotti proposait de lutter contre l'absentéisme scolaire en supprimant les allocations familiales aux parents des enfants ayant cumulé au moins quatre demi-journées d'absence irrégulière à l'école dans le mois. Elle fut abrogée en janvier 2013.

23. Le tirage au sort porte sur le mode de sollicitation, et on ne peut plus comparer l'impact du programme entre des classes tests ou témoins, mais uniquement entre des classes ayant un différentiel de participation des parents de cinq points ; ce qui incite les évaluateurs à être beaucoup plus réservés sur ces résultats, qu'ils estiment non exempts d'aléas potentiels (Ecole d'Economie de Paris, 2011).

manœuvre des acteurs de l'action publique afin de permettre l'estimation rigoureuse de l'impact et l'apprentissage scientifique. Néanmoins, pour les porteurs de projets, l'apprentissage intervient au cours de leur pratique de la mise en œuvre et se manifeste par des ajustements de l'intervention. Celle-ci apparaît alors souvent, pour les porteurs de projets, comme stabilisée et efficace à la fin de l'expérimentation ; la non prise en compte de ce résultat par l'estimation de l'impact engendre un décalage de perception entre les deux parties.

La particularité de cette méthode est d'organiser la confrontation de deux temporalités et logiques d'apprentissage. L'une, scientifique, basée sur le respect d'un protocole strict défini au préalable, inscrite dans le long terme, s'appuie sur l'accumulation de résultats de plusieurs expérimentations qui, s'ils sont probants, ouvrent sur une généralisation. L'autre, issue de l'action publique, sans cesse réactualisée par l'ajustement de pratiques, à court terme, inscrite dans un contexte particulier, et dont la généralisation de résultats probants dépend principalement de la possibilité de financement et du positionnement dans l'agenda politique. Il arrive parfois que, comme dans le cas de *la Mallette*, ces deux logiques convergent autour d'un intérêt d'apprentissage commun. Néanmoins, à ce jour, la faible prise en compte des résultats produits par ces expérimentations, soulignée par d'autres travaux, appelle à délimiter le périmètre de l'utilisation pertinente de cette méthode (Gomel et Serverin, 2013). Nos résultats confirment l'importance de la prise en compte des intérêts et pratiques des parties prenantes de l'apprentissage. Ils suggèrent que la méthode expérimentale par assignation aléatoire ne peut, à elle seule, constituer l'instrument d'un apprentissage de l'action publique guidée par des données probantes (*evidence-based policy*) ; c'est uniquement lorsque l'intérêt scientifique et l'intérêt politique convergent, dès la conception même de l'action, et que les résultats scientifiques soutiennent l'intérêt politique, qu'un apprentissage commun est possible.

■ Bibliographie

Allègre G. (2008), « L'expérimentation sociale des incitations financières à l'emploi, questions méthodologiques et leçons des expériences nord-américaines », *Document de travail de l'OFCE* n° 2008-22.

Arendt H. (1989), « Vérité et politique », in *La Crise de la culture : huit exercices de pensée politique*, Paris, Gallimard.

Banerjee A. et. Duflo. E (2009), « L'approche expérimentale en économie du développement », *Revue d'économie politique* 119 (5), pp. 691-726.

Boussaguet L., Jacquot S. et Ravinet P. (2010), *Dictionnaire des politiques publiques*, Paris, Presses de Sciences Po.

Code de l'éducation - Article L313-8. 2010.

- Coletti P. et Radaelli C.-M. (2013), « Economic Rationales, Learning, and Regulatory Policy Instruments », *Public Administration* 91 (4) : 1056-70. doi:10.1111/padm.12006.
- Conseil scientifique du FEJ (2009), « Guide méthodologique pour l'évaluation des expérimentations sociales à l'intention des porteurs de projets ».
- Crest-Dares (2012), *Sécurisation du parcours des jeunes s'engageant dans l'apprentissage : bilan d'une expérimentation aléatoire contrôlée conduite en Corrèze*, Rapport d'évaluation, ministère des Sports, de la Jeunesse, de l'Education populaire et de la Vie associative, Fonds d'expérimentation pour la jeunesse.
- Dares (2008), « Colloque sur les expérimentations pour les politiques publiques de l'emploi et de la formation », n° 168, Paris.
- Dubois V. (1999), *La vie au guichet*, Paris, Économica.
- Dubois V (2013), « Introduction, The sociology of administrative work, a study of "street-level bureaucracy à la française », in *La France et ses administrations: un état des savoirs*, by Eymeri-Douzan J.-M et Bouckaert G., pp. 169-177, Bruxelles : Bruylant.
- Duflo E. (2005), « Evaluer l'impact des programmes d'aide au développement : le rôle des évaluations par assignation aléatoire », *Revue d'économie du développement* 19 (2) : pp. 185-226.
- Duflo E. (2010), « Introduction », in *Conférence nationale de l'expérimentation sociale, réformes contre placebo ?*, Paris.
- Dunlop C.A. et Radaelli C.M. (2013), « Systematising Policy Learning: From Monolith to Dimensions », *Political Studies* 61 (3), pp. 599-619. doi:10.1111/j.1467-9248.2012.00982.x.
- Ecole d'Economie de Paris (2011), *Quels effets attendre d'une politique d'implication des parents dans les collèges, évaluation de l'impact de la Mallette des parents*, Fonds d'expérimentations pour la jeunesse.
- Fonds d'expérimentation pour la jeunesse (2012), *Rapport d'activité 2009-2011*.
- Fonds d'expérimentation pour la jeunesse (2013), *Rapport d'activité 2012*, ministère des Sports, de la Jeunesse, de l'Education populaire et de la Vie associative.
- Gomel B. et Serverin E. (2009), « Expérimenter pour décider ? Le RSA en débat ». Document de travail, Centre d'étude de l'emploi, n° 119.
- Gomel B. et Serverin E. (2013), « L'expérimentation sociale aléatoire en France en trois questions », *Travail et Emploi*, n° 135, septembre, pp. 57-69.
- Horowitz I.L. (1970) « Social Science Mandarins: Policymaking as a Political Formula », *Policy Sciences* 1 (3): pp. 339-360. doi :10.1007/BF00145217.

- L'Horty Y. et Petit P. (2010), « Evaluation aléatoire et expérimentations sociales », 135. *Document de travail*, Centre d'études de l'emploi.
- L'Horty Y. et Petit P. (2011), *L'évaluation aléatoire : un succès qui ne doit rien au hasard*, La vie des idées.
- Labrousse A. (2010), « Nouvelle économie du développement et essais cliniques randomisés: une mise en perspective d'un outil de preuve et de gouvernement », *Revue de la régulation* 7.
- Lawrence T.B. et Roy S. (2006), « Institutions and Institutional Work », in *The Sage handbook of organization studies*, by Clegg S., Hardy C. et Lawrence T.B., 2nd ed, pp. 215-254, London, Thousand Oaks, Calif : Sage Publications.
- Le Galès P. et Lascoumes P. (2004a), « L'action publique saisie par ses instruments », in *Gouverner par les instruments*, pp. 1-44. Presse de Science po.
- Le Galès P. et Lascoumes P. (2004b.), *Gouverner par les instruments*, Presse de Science po.
- Le Haut-Commissaire aux solidarités actives contre la pauvreté (2007), *Appel à projet d'expérimentation sociale 2008*.
- Lipsky M. (2010), *Street-Level Bureaucracy : Dilemmas of the Individual in Public Services*. New York : Russell Sage Foundation.
- Maillard D., Mora V., Rouaud P., Théry M., Veneau P., et Lehberger T. (2012), « Evaluation d'un dispositif d'accompagnement de jeunes en rupture scolaire », *Net.Doc*, n° 97, août, Céreq.
- March J.G et Olsen J.P (1989), *Rediscovering Institutions : The Organizational Basis of Politics*. New York : Free Press.
- Marks H. (1999), *La médecine des preuves : histoire et anthropologie des essais cliniques : 1900-1990*, Le Plessis-Robinson: Institut Synthélabo pour le progrès de la connaissance.
- Naudet J.-D. et Delarue J. (2007), « Promouvoir les évaluations d'impact à l'Agence française de Développement : renforcer l'appropriation et l'apprentissage institutionnels », 02. Série Notes méthodologiques Ex Post. Département de la Recherche, Division Evaluation et Capitalisation, Agence française de développement.
- Nay O. (1997), *La région, une institution : la représentation, le pouvoir et la règle dans l'espace régional*, Collection « Logiques politiques », Paris, L'Harmattan.
- Parienté W. (2008), « Analyse d'impact : l'apport des évaluations aléatoires », *STATECO*, n° 103.
- Powell W. et DiMaggio P. (1991) *The New Institutionalism in Organizational Analysis*. Chicago:University of Chicago Press.

- Pressman J.L et Wildavsky A.B (1984), *Implementation: How Great Expectations in Washington Are Dashed in Oakland : Or, Why It's Amazing That Federal Programs Work at All, This Being a Saga of the Economic Development Administration as Told by Two Sympathetic Observers Who Seek to Build Morals on a Foundation Of*. Berkeley, Calif.: University of California.
- Rectorat de Créteil (2011), « La Mallette des parents, APDIESES-11 », Note de restitution finale remise par le Rectorat de Créteil, porteur du projet, au Fonds d'expérimentations pour la jeunesse, dans le cadre de l'appel à projets lancé en 2008 par le ministère en charge de la Jeunesse, Rapport d'évaluation.
- Robertson D. (1991), « Political Conflict and Lesson Drawing », *Journal of Public Policy* 11 (1): pp. 55-91.
- Valdenaire M. et Bérard J. (2013), « L'expérimentation pour renouveler les politiques publiques ? » *La Vie des idées*, juin 25. <http://www.laviedesidees.fr/L-experimentation-pour-renouveler.html>.
- Weber M. (2002), *Le savant et le politique*, Paris, Plon.
- Zamora P. (2011), « La méthode d'évaluation aléatoire : apports et limites », *Tracés*, Hors-Série, pp. 175-186.