

## L'évaluation « basée sur la théorie », entre rigueur scientifique et contexte politique

*Theory-based evaluation, between scientific rigor and political context*

Agathe DEVAUX-SPATARAKIS

---



**Electronic version**

URL: <http://journals.openedition.org/pmp/6962>

ISSN: 2119-4831

**Publisher**

Institut de Management Public (IDPM)

**Printed version**

Date of publication: 1 January 2014

Number of pages: 51-68

ISBN: 978-2-7430-1995-2

ISSN: 0758-1726

**Electronic reference**

Agathe DEVAUX-SPATARAKIS, « L'évaluation « basée sur la théorie »,  
entre rigueur scientifique et contexte politique », *Politiques et management public* [En ligne], Vol 31/1 |  
2014, mis en ligne le 12 décembre 2016, consulté le 10 décembre 2020. URL : [http://  
journals.openedition.org/pmp/6962](http://journals.openedition.org/pmp/6962)

---

© 2014 IDMP/Lavoisier SAS. Tous droits réservés

## L'évaluation « basée sur la théorie », entre rigueur scientifique et contexte politique

03

► **Agathe Devaux-Spatarakis**

*Doctorante en contrat Cifre*

*Centre Émile Durkheim, Sciences Po Bordeaux, 11 allée Ausone, 33607 Pessac Cedex  
Eureval, 13 bis place Jules Ferry, 69001 Lyon*

### Résumé

Cette contribution propose de revenir sur le courant de l'évaluation basée sur la théorie. Davantage une démarche évaluative qu'une méthode d'évaluation, il convient d'identifier ce qui unifie ce courant, comment il conçoit l'évaluation ainsi que sa définition de la théorie et son utilisation. S'appuyant sur les travaux de l'analyse des politiques publiques et de la sociologie de l'action publique, cette démarche propose une évaluation intégrée à son contexte politique prenant en compte son fonctionnement. La théorie peut alors devenir un outil structurant de l'évaluation contribuant à ouvrir la « boîte noire » des politiques publiques, fédérer l'apprentissage des acteurs publics, établir la causalité et même construire la transférabilité des résultats d'évaluation à d'autres contextes en s'inscrivant dans une épistémologie réaliste.

© 2014 IDMP/Lavoisier SAS. Tous droits réservés

*Mots clés* : évaluation, causalité, épistémologie réaliste, théorie du changement.

### Abstract

**Theory-based evaluation, between scientific rigor and political context.** This paper drives the attention of the reader to the stream of theory based evaluation. More an evaluative approach than an evaluation method, the author first identifies what unifies this stream, how it understands evaluation and its definition and use of theory. Grounded on academic work on public policy analysis and sociology of public action, this approach offers an evaluation integrated to its political context and taking into account its functioning. Hence, theory can become a structuring tool of evaluation contributing to open the "black box" of public policies, fostering learning from public actors, establishing causality

\*Auteur correspondant : [agathedevaux@yahoo.fr](mailto:agathedevaux@yahoo.fr)

doi :10.3166/pmp.31.51-68 © 2014 IDMP/Lavoisier SAS. Tous droits réservés

and even building transferability of evaluation results to other contexts grounded on the realist epistemology.

© 2014 IDMP/Lavoisier SAS. Tous droits réservés

**Keywords:** evaluation, causality, realist epistemology, theory of change.

---

## Introduction

L'évaluation « basée sur la théorie » connaît aujourd'hui un essor théorique notoire, notamment à travers la diffusion de l'évaluation dite « réaliste », ainsi que celle de l'analyse de « Contribution » à laquelle la revue britannique *Evaluation* a consacré l'intégralité de son numéro de juillet 2012 (Pawson et Tilley 2004 ; Mayne, 2001). D'autre part, les organisations internationales commencent à souligner l'importance d'aller au-delà de l'estimation de l'impact et d'apporter des enseignements sur les chaînes de causalité contribuant à l'efficacité d'une intervention<sup>1</sup>.

L'évaluation basée sur la théorie est une démarche structurante dans le monde de l'évaluation. Néanmoins, bien que certains de ses principes soient appliqués dans une part importante des évaluations, rares sont les incarnations fidèles de cette démarche (Rogers *et al.*, 2000 : 10). Initié à la fin des années 1970 aux États-Unis, en partie pour répondre aux échecs de la démarche d'évaluation expérimentale des grands projets américains, ce type d'évaluations propose une démarche construite sur la prise en compte des caractéristiques propres à l'intervention publique ainsi qu'au champ politique (Weiss 1997).

En France, les réflexions sur l'évaluation ont porté principalement sur son institutionnalisation (Hassenteufel 2011), sa distinction avec d'autres outils de contrôle (Toulemonde 2009 ; Gibert 2003), la participation des publics à l'acte évaluatif (Baron et Monnier 2003) et plus récemment sur une réflexion de nature méthodologique (L'Horty et Petit 2011). L'évaluation basée sur la théorie s'inscrit à la marge de ces débats tout en les traversant. Elle n'est cependant pas une méthode d'évaluation à proprement parler, mais plutôt une logique de recherche évaluative, ainsi qu'une démarche analytique (Birckmayer et Weiss, 2000 : 426). Elle regroupe un ensemble de pratiques évaluatives diverses, mobilisant des méthodes de collecte de données variées. Elle structure l'évaluation autour de l'énonciation d'une « théorie », capable de décrire comment le programme provoque les résultats escomptés ou observés et elle le représente sous la forme d'un cadre logique, d'un diagramme logique d'impact, ou d'un logigramme, ce qui est devenu désormais une pratique courante en évaluation (Rogers *et al.*, 2000 : 5).

Néanmoins, il apparaît nécessaire de clarifier les concepts mobilisés par cette démarche plurielle et de définir son ancrage épistémologique. Au-delà d'un simple outil de modélisation de l'action publique, un des intérêts particuliers de cette démarche est de s'inscrire dans l'épistémologie du *réalisme critique*, ouvrant de nouvelles pistes pour générer des connaissances, entre l'épistémologie positiviste et constructiviste (Avenier et Thomas 2011).

---

<sup>1</sup> L'Initiative Internationale pour l'Evaluation d'Impact par exemple précise que « *les études devraient clairement énoncer comment on attend qu'une intervention produise les résultats finaux et tester chaque lien de l'intervention jusqu'aux résultats. Le design de l'évaluation devrait comprendre une analyse de la chaîne causale allant de l'intervention à son impact* » 3ie impact evaluation practice: a guide for grantees. <http://www.3ieimpact.org/page.php?pg=overview>.

Il peut sembler contradictoire de vouloir structurer une évaluation autour d'une « théorie » alors que l'évaluation est souvent justement mobilisée pour savoir ce qui s'est passé « en pratique ». Dans un premier temps, nous clarifierons ce que recouvre le terme « théorie » dans cette démarche et comment celle-ci est employée pour appréhender la déclinaison d'une politique publique au plus près de son contexte. Dans un second temps, nous étudierons comment cette théorie peut constituer un outil consolidant la scientificité de cette démarche, ainsi que la synthèse de plusieurs résultats d'évaluations issus de différents contextes autour d'une épistémologie réaliste organisant d'une généralisation analytique.

## 1. Une théorie au plus près de la pratique

L'évaluation basée sur la théorie s'inscrit dans la lignée des travaux de Science politique et d'analyse des politiques publiques concevant l'intervention publique d'une part comme un processus à décomposer et d'autre part comme un processus dont la mise en œuvre peut fortement dévier de la volonté programmatique initiale (Pressman et Wildavsky, 1984). Cette conception trouve son origine dans les années 1960, dans les travaux de E. Suchman ou de D. Stufflebeam lesquels s'intéressent à observer une chaîne d'objectifs pour appréhender la politique publique (Funnell et Rogers 2011 : 15). On retrouve des traces de cette démarche dès les années 1970, dans des travaux d'évaluation aux États-Unis, mais c'est véritablement dans les décennies 1980 et 1990 que s'effectue le travail de théorisation autour des travaux de H-T Chen, L. Bickman et C. Weiss (Chen 1994 ; Weiss 1997). Au milieu des années 1990, R. Pawson et N. Tilley développent l'approche « réaliste », laquelle s'inscrit dans ce courant et l'enrichit d'une théorisation épistémologique *réaliste* autour du concept de *mécanisme*. Dans les années 2000, l'analyse de « Contribution » théorise une nouvelle incarnation de cette démarche centrée sur la démonstration de la causalité sans mobilisation de scénarios contrefactuels l'opposant ainsi à l'analyse d'attribution<sup>2</sup> (Mayne 2001).

Nous proposons de revenir sur ce qui unit ces différentes incarnations, en expliquant dans un premier temps leur programme de recherche centré sur l'ouverture de la « boîte noire » de l'action publique, puis en clarifiant ce qui est entendu par le terme de théorie et son usage. Enfin, nous étudierons ce qu'implique le fait de structurer l'évaluation autour de cette modélisation.

### 1.1. Ouvrir la « boîte noire » de l'action publique pour en comprendre les effets

Comme un enquêteur qui ouvrirait la boîte noire d'un cockpit après un accident d'avion, l'évaluation basée sur la théorie propose d'ouvrir la « boîte noire » de l'intervention publique, au sens où elle souhaite mettre à jour les composantes de cette dernière, mais aussi comprendre l'origine du problème survenu. S'appuyant sur les hypothèses du modèle de la rationalité limitée dans la prise de décision développé par H. Simon dans

---

<sup>2</sup> Une analyse d'attribution estime l'impact d'une intervention publique en comparant l'évolution d'un groupe de personnes ayant bénéficié de l'intervention avec une estimation de ce qui se serait passé sans intervention (contrefactuel). Le « contrefactuel » est estimé soit à partir de l'observation d'un groupe de non-bénéficiaires obtenu de façon expérimentale (par tirage aléatoire) ou quasi-expérimentale, soit calculé par simulation.

les années 1950, ce courant estime que les objectifs des politiques publiques sont rarement clairs puisqu'ils représentent un consensus entre des intérêts politiques divergents. En outre, les décideurs publics cherchent rarement à atteindre une vision synoptique du problème non plus qu'un optimum, mais cherchent plutôt à trouver une solution satisfaisante dans le cadre d'informations limitées (Stame 2009 : 47). Ce courant formule alors l'hypothèse que l'évaluation serait davantage utilisée si elle était guidée par une meilleure compréhension du champ politique, de la logique guidant les décideurs publics, et de la manière dont est conduite une politique publique en général (Weiss, 1990 : 171-184). L'évaluation ne se restreint donc pas à l'analyse de la politique publique (*policy*), mais prend aussi en compte la politique (*politics*) dans son acception globale. En d'autres termes, l'évaluation doit étudier à la fois l'inscription de l'action publique dans les enjeux de pouvoir du champ politique, mais aussi dans l'environnement institutionnel et contextuel de sa mise en œuvre.

L'évaluation basée sur la théorie pense que la politique publique peut être assimilée à un processus. Cette conception amène à appréhender la politique publique comme une chaîne causale allant de l'intervention initiale à l'impact final, dont le processus n'est pas linéaire, mais instable, et non maîtrisé par le décideur public. Les évaluateurs de ce courant estiment alors que la mise en œuvre participe directement à la construction des politiques publiques, puisque cette mise en œuvre peut varier de façon significative au gré des microdécisions des opérateurs de terrain (Pressman et Wildavsky, 1984).

L'évaluation basée sur la théorie a ainsi pour ambition de comprendre l'action publique dans sa globalité, sa complexité, et sa nature chaotique. La « théorie » devenant le fil rouge structurant cette compréhension, il est nécessaire d'explicitier ce que recouvre ce terme pour cette démarche.

## 1.2. L'intervention publique, une rencontre de théories profanes<sup>3</sup> et académiques

La décomposition analytique de l'intervention publique prend la forme d'une théorie sous-tendue par un enchaînement d'hypothèses. Tous les théoriciens de ce courant ne partagent pas les mêmes concepts, ou ne les définissent pas de la même façon. Leur acception la plus répandue, développée par C. Weiss, théoricienne historique du courant de l'évaluation basée sur la théorie, identifie deux types de théories pouvant guider un programme (Weiss 1997 : 46) :

- *La théorie de mise en œuvre* : « ce qui est requis pour traduire les objectifs en activités et opérations de programme », ce que M. Saunders appelle « petite théorie », et W. Shadish « microthéorie » (Blamey et Mackenzie, 2007 : 444 ; Shadish 1987).
- *La théorie de programme* : « les liens causaux hypothétiques entre des mécanismes délivrés par une intervention et leurs résultats escomptés », ce que M. Saunders appelle « la grande théorie », et W. Shadish « macrothéorie » (Blamey et Mackenzie, 2007 : 444 ; Shadish, 1987)

La *théorie de mise en œuvre* formule toutes les activités sur lesquelles les administrateurs de l'intervention publique exercent un contrôle, alors que la *théorie de programme* concerne les résultats et impacts moins maîtrisables et relevant de l'interaction de la théo-

<sup>3</sup> Ce terme est compris ici au sens de « non initié à la science », c'est-à-dire des théories qui n'ont pas été formulées selon les normes d'une scientificité, mais donnant néanmoins un sens à l'action pour les acteurs.

rie de mise en œuvre avec un contexte donné. Ces deux théories peuvent alors s’articuler dans une *théorie du changement* expliquant comment la *théorie de mise en œuvre* permet d’activer la *théorie de programme* en explicitant également les facteurs externes connus facilitateurs de cette activation.

Le terme « théorie » relève ici d’une double acception. La *théorie de mise en œuvre* doit être comprise au sens non scientifique du terme, c’est-à-dire d’un ensemble de suppositions implicites sous-tendant le processus prévisionnel souhaité. La *théorie de programme* est plus proche quant à elle de l’acception scientifique du terme. Elle articule souvent un ensemble de théories scientifiques rendant compte des mécanismes psychosociologiques que souhaite activer l’intervention publique. S’intéresser à la vérification d’une de ces théories est un prérequis pour s’inscrire dans la démarche de l’évaluation basée sur la théorie.

La *théorie de programme* est souvent représentée de façon simplifiée sous la forme d’un diagramme logique d’impact ou logigramme, décomposant la théorie en quatre étapes principales (Funnell et Rogers 2011 : 22):

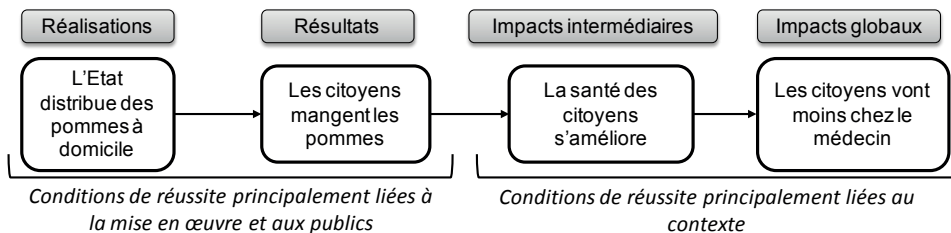
- *Les réalisations* : elles représentent les activités mises en œuvre par l’autorité publique ;
- *Les résultats* : ils sollicitent souvent l’adhésion des publics visés et dépendent de la capacité des réalisations à les mobiliser.

Ces deux premières étapes peuvent être maîtrisées par un système de suivi performant et leur atteinte est directement liée au contrôle de la mise en œuvre.

- *Les impacts intermédiaires et finaux* : ils décrivent l’impact que devrait avoir cette activité sur les bénéficiaires ainsi que la finalité escomptée de l’intervention. Leur atteinte dépend des conditions favorables environnantes ou de leur maîtrise par l’intervention.

Prenons l’exemple d’une politique fictive très simple qui prendrait au sérieux le proverbe anglais « Une pomme chaque matin, éloigne le médecin ». Nous prenons ici ce proverbe comme étant notre *théorie de programme*. Faisons l’hypothèse que les autorités publiques encouragent la consommation de pommes par la distribution à domicile d’une pomme par jour et par personne, afin d’améliorer la santé des citoyens et qu’ils aillent moins chez le médecin. On peut alors proposer une décomposition schématique et simplifiée de la *théorie de programme* de cette intervention de la façon suivante :

**Figure 1 :** Théorie de programme d’une politique d’incitation à la consommation de pommes pour diminuer les consultations médicales.



Les réalisations sont donc : une distribution de pomme qui vise à solliciter l’adhésion des publics par la consommation de celles-ci (résultats). Cette consommation a pour objectif intermédiaire l’amélioration de la santé des consommateurs, ce qui à terme devrait les mener à moins fréquenter les cabinets médicaux.

Si le champ de l'évaluation porte principalement sur la *théorie de mise en œuvre*, l'évaluateur va alors cibler ses recherches sur la partie de la théorie portant sur les réalisations et les résultats. On peut considérer les chaînes causales comme potentiellement fractales<sup>4</sup>, puisque chaque relation causale pourrait être détaillée en une nouvelle chaîne et faire l'objet d'une analyse plus minutieuse. On peut par exemple imaginer que la distribution de pommes est financée par l'État, mais que ce dernier a recours à des prestataires pour la distribution des pommes. Il est alors possible d'étudier en détail les modalités de définition de la distribution, les procédures de sélection du prestataire, le respect du cahier des charges par ce prestataire et l'accessibilité des foyers pour la livraison. Ce choix de porter l'évaluation particulièrement sur les activités de mise en œuvre peut être pertinent dans le cadre d'une évaluation précoce d'un programme dont les effets ne sont pas encore observables.

Si le champ de l'évaluation porte sur les effets d'un dispositif, alors l'évaluateur s'intéressera en outre à la *théorie de programme* s'étendant des résultats aux impacts finaux. Cette théorie s'avère être plus incertaine puisque son développement est davantage sujet aux aléas du contexte et de l'environnement de l'intervention. Elle est souvent basée sur des théories du champ académique (Chen 1994). Ici par exemple, elle mobilise des mécanismes issus de la psychologie formulant que l'accès gratuit est une incitation à la consommation, ainsi que d'autres issus par exemple de la médecine, établissant que la consommation de pommes améliore la santé.

Cette conception de l'intervention publique comme une « théorie », et la mise en lumière des hypothèses qui la sous-tendent guident l'ensemble de la démarche d'évaluation.

### **1.3. La théorie comme échafaudage sous-tendant la démarche évaluative**

La théorie structure à la fois le cadrage du champ de l'évaluation, l'organisation de la collecte de données, l'analyse des résultats et impacts, et la façon dont est porté le jugement.

La première étape d'une évaluation basée sur la théorie consiste à rendre explicites la ou les théories implicites guidant le programme ou la politique publique évaluée. L'absence d'une théorie préétablie explicitant l'intervention publique n'est pas un obstacle à l'application de cette démarche (Birckmayer, Weiss, 2000 ; VanDerKnaap, 2004). En son absence, le travail préliminaire de l'évaluateur consiste à faire émerger la théorie sous-jacente ayant jusqu'à présent guidé implicitement l'intervention publique. Pour ce faire, il peut réunir l'ensemble des textes officiels relatifs à cette intervention afin de reconstruire les éléments clés de la *théorie de programme*. M. Q. Patton identifie trois types d'approches pour élaborer une théorie. Une approche déductive qui consiste à partir d'une théorie issue des disciplines académiques et à construire des modèles de mise en relation entre les interventions et leurs effets ; une approche inductive qui implique un travail de terrain préalable pour que l'évaluateur génère une théorie à partir de ses observations et enfin une approche orientée vers les utilisateurs qui exige l'implication des parties prenantes de la politique dans ce processus d'élaboration de la théorie (Patton, 1997 : 344-346). Le travail de l'évaluateur est souvent un compromis entre ces différentes approches. Si parfois une démarche déductive, issue de théories profanes ou académiques, peut être un point de départ, elle est souvent enrichie

---

<sup>4</sup> Une forme fractale a pour caractéristique d'être composée de la même structure géométrique à différentes échelles, comme on peut l'observer sur la structure du chou romanesco par exemple.

par une autre approche inductive pour comprendre comment la théorie a été effectivement déployée en pratique, et cette approche inductive requiert aussi régulièrement la recherche d'un éclairage par les parties prenantes de l'intervention évaluée.

Ce travail d'élaboration est l'occasion d'un moment d'échange avec les décideurs publics et les parties prenantes, aboutissant à une théorie faisant mutuellement sens pour les parties prenantes de cette intervention (Birckmayer and Weiss 2000). Cette vision commune de l'intégralité de la logique d'intervention est un élément précieux pour des politiques publiques au caractère compliqué ou complexe. Dans le cadre de la mise en place d'un nouveau dispositif ou d'une nouvelle procédure, la co-élaboration de la théorie avec les parties prenantes de cette intervention contribue à rendre intelligible, voire à légitimer cette nouvelle logique d'intervention. Par exemple, une politique visant à favoriser la formation en apprentissage chez les jeunes déscolarisés, mobilise à la fois les Missions Locales, l'Education Nationale, les chambres consulaires, les entreprises et l'équipe éducative de l'établissement. Chacune de ces parties prenantes a une vision parcellaire de l'intervention, mais leur coordination est néanmoins indispensable. La théorie du changement permet d'explicitier le rôle successif de chacun dans cette intervention et d'établir des outils de suivi afin d'identifier à quels moments les difficultés dans la mise en œuvre du programme sont apparues. Cette co-construction est particulièrement présente dans le cadre d'évaluations dont les instances de suivi sont pluralistes et donnent la parole à une diversité de parties prenantes (Baron et Monnier 2003).

Une fois la théorie de programme explicitée, celle-ci devient « l'échafaudage » sous-tendant les questions évaluatives, la collecte de données et l'analyse. Organiser l'évaluation autour de la théorie d'action est un moyen pour l'évaluateur de se concentrer autour d'une logique faisant sens pour les acteurs concernés et favorise l'appropriation des résultats ainsi que l'apprentissage qui en découlera (Weiss, 2001).

Les questions d'évaluation sont choisies pour vérifier non seulement dans quelle mesure l'intervention publique a eu les effets escomptés, mais aussi pour repérer comment les différents maillons de la chaîne causale reconstruite se sont articulés et pourquoi. On hiérarchise ensuite ces questions en favorisant celles portant sur les maillons les plus problématiques ou les plus importants pour les parties prenantes. La théorie guide aussi le choix d'indicateurs que l'on pense ici comme des implications, observables empiriquement, des hypothèses de fonctionnement de la *théorie du changement*. Ensuite, la collecte de données est organisée afin de réunir un maximum d'éléments de preuve pour invalider ou valider chaque étape de la théorie et renseigner les indicateurs. Aucune méthode particulière de collecte de données n'est recommandée par les praticiens de cette démarche. La règle en vigueur est de choisir des méthodes selon leur adaptabilité au contexte et aux questions posées. La tendance est aussi à privilégier l'utilisation de plusieurs outils de collecte afin de confronter des données qualitatives et/ou quantitatives. Une place importante est accordée aux données pouvant éclairer l'évaluateur sur les particularités du contexte dans lequel s'inscrit l'intervention publique et comment il interfère avec la théorie identifiée (Stame, 2004 : 59 ; Blamey et Mackenzie, 2007 : 447).

À la lumière des données collectées, l'équipe d'évaluation révisé la théorie de l'intervention en expliquant les étapes ou les transitions qui n'ont pas eu lieu, séparant celles liées à la conception de l'intervention de celles liées à son environnement. Elle peut alors formuler des recommandations pour une intervention future.



Ce courant dit « basé sur la théorie » apparaît alors comme une démarche analytique qui, paradoxalement, a pour préoccupation d'être au plus près du contexte politique et social dans lequel une intervention publique s'inscrit et de rendre compte le plus justement possible des aléas de la mise en œuvre de cette intervention. Mais qu'en est-il de la rigueur scientifique de cette démarche ? Comment établit-elle qu'une politique fonctionne dans un contexte donné ? Et surtout dans quelle mesure parvient-elle à s'extraire du cas localisé et à produire des enseignements transposables à d'autres contextes ?

## **2. Une construction de la scientificité autour d'une déconstruction analytique**

Cette seconde partie propose de faire état de la nature des connaissances produites par l'évaluation basée sur la théorie, dans le cadre d'une évaluation ou de synthèse de plusieurs résultats d'évaluations distinctes. Au-delà de son adaptabilité au champ politique, elle peut aussi contribuer à fournir des connaissances rigoureuses sur l'action publique.

Cette démarche n'a pas vocation à élaborer des lois générales qui établiraient quelle intervention marche ou ne marche pas. Elle cherche à établir, autant dans le cadre d'une évaluation isolée que lors d'une synthèse de résultats de plusieurs évaluations, ce qui marche, pour qui, comment et dans quelles conditions (Pawson et Tilley, 2004 : 1). Cette ambition modeste d'apport de connaissances s'inscrit dans les travaux de C. Lindblom sur la décision publique, lequel considère que le changement en politique s'effectue au pas-à-pas, à la marge (Lindblom, 1959). Cela implique donc de produire des résultats donnant des clés pour ajuster des interventions publiques, les moduler et non pas les invalider ou les valider dans leur intégralité (Birckmayer et Weiss 2000 : 424). Cette conception de la capacité à générer des connaissances s'inscrit aussi dans des réflexions du champ scientifique. Nous présentons dans un premier temps la construction de la validité interne des résultats et de la démonstration de la causalité produites par ce type d'évaluation, autour d'une déconstruction analytique pouvant s'émanciper de l'élaboration d'un scénario contrefactuel. Dans un deuxième temps nous proposerons des pistes de réflexion quant à l'organisation de la validité externe des résultats, autour d'une « épistémologie réaliste » mobilisant la logique de comparaison d'études de cas.

### **2.1. Une validité interne basée sur la décomposition d'une chaîne causale et l'exclusion d'explications rivales**

Si l'on considère l'intervention publique comme une « boîte noire » en s'intéressant uniquement aux activités (*input*) et à l'impact final (*output*), on prend le risque d'interpréter des résultats d'évaluation de manière incorrecte. Les évaluateurs n'auront pas les outils pour déterminer si l'absence de l'impact escompté est imputable aux hypothèses sous-tendant cette intervention, à une des étapes de sa mise en œuvre, ou à un contexte particulier. Si l'on reprend notre proverbe britannique, W. Churchill en proposait la version alternative suivante : « Une pomme chaque matin, éloigne le médecin... surtout si vous visez bien ». Ainsi on arrive au même impact final – la diminution des consultations médicales – sans passer par la consommation des pommes, mais par leur utilisation en projectile. Une décomposition théorique de cette action permet de déterminer avec précision le ou les chemins empruntés par les publics ciblés les ayant menés à la situation finale.

La décomposition analytique permet d'établir la différence entre une invalidation des hypothèses sous-tendant l'intervention, et un échec lié à une mise en œuvre non adéquate ainsi que d'identifier les facteurs de succès ou d'échec (Weiss, 1997 : 43 ; Suchman, 1967).

Il existe deux temps d'analyse d'une théorie guidant un programme. L'analyse « normative » que nous appellerons ici *analyse du changement*, puisque le terme normatif en français fait référence à une opération de jugement qui n'est pas le sens voulu de ce terme dans cette analyse. Cette *analyse du changement* consiste à effectuer une comparaison entre la théorie développée et les réalisations effectives, les résultats du programme et les changements constatés par rapport à la situation préalable. Ensuite intervient l'analyse « causale » qui consiste à vérifier que ce sont bien des mécanismes activés par l'intervention qui ont produit ces changements (Chen, 1994 :13).

L'analyse du changement consiste à vérifier que chaque étape de la chaîne causale a bien eu lieu, mais aussi à mesurer l'atteinte des objectifs fixés. La rigueur de cette démarche ne tient pas à l'utilisation d'une méthode de collecte en particulier, mais au croisement, à la triangulation de plusieurs méthodes. Quatre types de triangulation peuvent être utilisés : on peut croiser plusieurs méthodes de collectes de données ; plusieurs sources de natures différentes issues de la même méthode, par exemple plusieurs entretiens ; plusieurs analystes pour corriger les biais d'interprétation ; et même plusieurs théories explicatives (Denzin, 1978 ; M. Patton, 2001). On estime que la robustesse d'un élément de preuve augmente lorsqu'il est corroboré par plusieurs outils ou plusieurs sources.

Lors de l'analyse de la correspondance des observations de terrain avec la théorie escomptée, on peut aboutir à des résultats variés donnant lieu à des pistes d'interprétation différentes :

**Figure 2 :** Différentes interprétations possibles selon les résultats de l'analyse du changement

Réalisations	Résultats	Impacts intermédiaires	Impacts finaux	Pistes d'interprétations
L'intervention a-t-elle été mise en œuvre de façon adéquate ?	Les publics ont-ils adhéré à ce programme ?	Les objectifs intermédiaires ont-ils été atteints ?	Les objectifs finaux ont-ils été atteints ?	
x	x	x	x	a. Échec de la théorie de mise en œuvre (réalisation non conforme)
✓	x	x	x	b. Échec dans l'adhésion des publics (1 <sup>er</sup> lien causal)
✓	✓	x	x	c. Échec de la théorie de programme
✓	✓	✓	x	
✓	✓	✓	✓	d. Vérification de la théorie du changement (des 2 théories)
✓	✓	x/✓	x/✓	e. Échec partiel de théorie de programme (contextes favorables ou défavorables )
✓	✓	x	✓	f. Échec de la théorie du programme, succès dû à un autre chemin causal

Lecture : x absence de vérification  
 ✓ vérification

Adaptation de la Source : Funnell, S., et P. Rogers. 2011. *Purposeful program theory, effective use of theories of change and logic models.* Jossey-bass. p 478.

On observe dans le tableau ci-dessus que trois cas (b, c, e) présentent une configuration similaire, en cumulant un succès de réalisations et un échec final, mais elles donnent cependant lieu à des interprétations diverses selon le maillon de la chaîne où s'est produit l'échec. Néanmoins, toutes les pistes d'interprétation dues à la présence ou à l'absence des effets voulus à chaque maillon de la chaîne restent uniquement basées sur des relations de corrélation, et n'établissent pas encore de relations de causalité. Aujourd'hui, encore de nombreuses déclinaisons de la démarche basée sur la théorie arrêtent l'analyse évaluative à ce stade (Rogers *et al.*, 2000 : 11).

Pour conduire l'analyse causale, on s'intéresse aux mécanismes entre chaque maillon de la chaîne afin d'établir des relations de causalité. Le but de cette démarche est de vérifier que le changement observé est bien imputable à l'intervention et non à des explications rivales. Pour estimer l'apport d'une intervention, certaines méthodes, par exemple la méthode expérimentale et les méthodes quasi-expérimentales, reconstruisent ou observent des scénarii contrefactuels pour comparer la situation avec et sans l'intervention publique (Duflo, 2005 : 188). Cette méthode a une certaine robustesse, mais nécessite de porter l'étude sur un maillon de la chaîne bien précis afin de contrôler que l'effet estimé émerge bien du mécanisme étudié. Par exemple, dans l'étude d'un dispositif d'offre de services pour accompagner le retour à l'emploi, on peut l'utiliser pour s'intéresser au maillon de mise en œuvre « modes d'incitation pour participer au dispositif » pour déterminer avec précision le mode de sollicitation le plus efficace.

La démarche favorisée par l'évaluation basée sur la théorie s'inscrit dans une conception de la causalité différente, laquelle implique des méthodes de démonstration alternatives. L'évaluation basée sur la théorie soulève deux questionnements principaux pour aborder la causalité. Comment établir une causalité complexe prenant en compte à la fois le contexte d'intervention et l'intervention dans sa décomposition analytique ? Comment appréhender leurs effets séparément ?

L'épistémologie du « réalisme critique » apporte des réponses à ce questionnement en proposant des modalités d'apport de connaissances dans des systèmes sociaux ouverts. Il comprend le « réel » comme un ensemble de *structures* et de *mécanismes* reliés de manière contingente à des événements empiriques (Tsang et Kwan 1999). Les *structures* sont composées d'objets inter-reliés qui définissent leur identité les uns par rapport aux autres. Les *mécanismes* de structure génèrent des actions entre ces objets selon des conditions intrinsèques propres à chaque *mécanisme*, et des conditions extrinsèques variant selon les contextes d'applications (Avenier et Thomas, 2011 : 8). Ils peuvent intervenir à un niveau micro (individu), meso (gouvernance ou réseau) ou macro (institution plus large) (George et Bennett 2005 : 147). Alors que cette démarche a normalement pour ambition d'identifier les *mécanismes* explicatifs à l'origine des phénomènes observés lorsqu'elle s'applique dans un contexte d'évaluation, son but est alors de vérifier que les *mécanismes* mis en place par l'action publique ont provoqué les phénomènes souhaités.

Cette démarche s'intéressant à l'interaction d'une intervention publique et d'un contexte pour la production d'un effet, ne s'inscrit pas dans une conception unicausale et linéaire de la causalité. Elle s'inscrit dans la notion de *champ causal* et théorise le lien non pas entre un effet et une cause unique, mais entre un effet et un *faisceau causal*<sup>5</sup> réunissant un ensemble

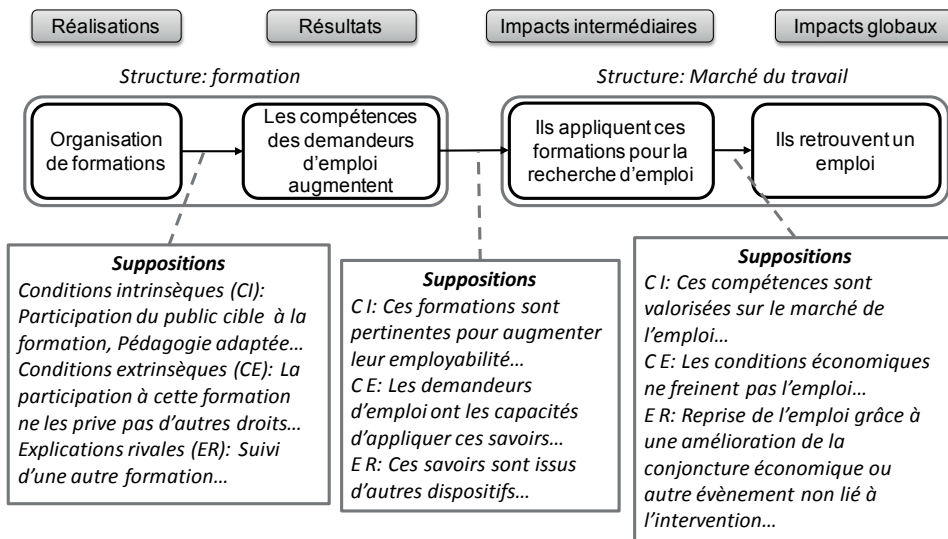
<sup>5</sup> Merci à Jacques Toulemonde pour cette proposition de traduction du terme anglais original « causal package »

de causes, dont chaque cause ne peut à elle seule provoquer l'effet final, mais dont la synergie avec les autres causes est susceptible de créer l'effet en question (Befani, 2012 : 12).

L'évaluateur doit donc déterminer si l'intervention publique a agi comme un *mécanisme* activé et activant les autres éléments du faisceau causal pour atteindre l'effet souhaité. Pour vérifier l'imputabilité du changement constaté à l'intervention publique, l'évaluateur doit identifier les facteurs de contexte, ou causes extérieures, facilitant ou compliquant la production de l'effet, et leur interaction avec l'intervention publique. Il doit ensuite vérifier que d'autres facteurs non liés à l'intervention n'ont pas été la cause principale du passage d'un maillon à l'autre de la chaîne. Cette opération consistant à évincer les explications rivales permet aussi d'éviter l'écueil souvent reproché à l'évaluation basée sur la théorie, de s'enfermer dans une vision « tunnel » de la politique, qui pourrait amener à occulter d'autres facteurs importants non pris en compte dans la théorie initiale (VanDerKnaap, 2004 : 21). L'efficacité des *mécanismes* d'une intervention reste dépendante du contexte qui l'environne et peut favoriser ou freiner son enclenchement.

À titre d'exemple, étudions un dispositif de formation professionnelle pour favoriser le retour à l'activité des demandeurs d'emploi.

**Figure 3:** Théorie du changement d'un dispositif de formation pour augmenter l'employabilité des demandeurs d'emploi



On cherche à intervenir sur deux *mécanismes* : celui de l'apprentissage entre le formateur et l'apprenant (qui composent une structure) ; et celui de la rencontre d'une offre et d'une demande entre un demandeur d'emploi et un employeur (qui composent une autre structure). Si l'on s'intéresse au *mécanisme* « apprentissage », l'évaluateur va vérifier les suppositions sous-jacentes à ce *mécanisme*, c'est-à-dire les conditions intrinsèques et extrinsèques ont été réunies pour qu'il soit activé, et ainsi comprendre les succès ou les échecs de l'intervention. En conditions intrinsèques, on peut identifier par exemple l'organisation effective de la formation, la sélection de formateurs de qualité, la participation effective des apprenants

aux séances, et leur compréhension de cette formation. En conditions extrinsèques, on peut identifier, entre autres, les motivations des apprenants, les modalités techniques d'organisation des formations, la cohérence de ce dispositif avec les autres dispositifs auxquels peuvent avoir recours ces bénéficiaires. Il faut aussi vérifier que ces résultats sont bien imputables à l'intervention et non pas à une explication rivale, comme le suivi d'une autre formation, ou l'apprentissage de connaissances par ailleurs. De fait, une attention spéciale doit être donnée aux indicateurs qui permettent de séparer ce qui tient de la *théorie du changement* et non des théories rivales, afin de déterminer l'opérationnalité du *mécanisme* (Hall 2006 : 27). Plus l'intervention est décomposée analytiquement en petites entités, plus la relation de causalité pourra être déterminée avec précision. Cette opération est à répéter pour chaque *mécanisme* de l'intervention.

Le résultat d'une évaluation basée sur la théorie devrait fournir, d'une part, une *théorie du changement* confirmée ou révisée comprenant des indications sur les éléments de contexte et de mise en œuvre favorables ou hostiles à la concrétisation de cette « théorie » sur le terrain. Ainsi, l'évaluateur ne donne pas « des bonnes recettes » ou des politiques « clés en main » à répliquer dans d'autres contextes, mais propose plutôt « un code de la route » indiquant à quoi il faut être attentif pour que l'intervention provoque l'effet souhaité (Pawson et Tilley, 2004 : 21)

## **2.2. Une validité externe structurée par une généralisation analytique de comparaison d'études de cas**

Établir la validité externe d'un résultat consiste à délimiter dans quelle mesure les enseignements qui en sont issus peuvent faire l'objet d'une montée en généralisation et être transposables à d'autres contextes. Aujourd'hui, en évaluation, la validité externe est typiquement établie de manière statistique par des revues systématiques (*systematic review*), ou méta-analyses comparant des résultats de plusieurs évaluations portant sur des types d'intervention proches, voire identiques. Cette démarche s'inscrit dans une épistémologie positiviste de réplication d'une expérimentation contrôlée afin d'établir une loi générale.

L'épistémologie « réaliste » propose quant à elle une généralisation *analytique* et structure la comparaison des résultats d'évaluation autour des *mécanismes* (Pawson 2002). Le *mécanisme* permet une montée conceptuelle de l'observation empirique vers une théorie intermédiaire qui devient la structure de comparaison des différents cas. Ainsi, il est alors possible de comparer les effets de plusieurs politiques publiques, si ces dernières s'inscrivent dans une théorie de moyenne portée commune ou mobilisent un *mécanisme* semblable. Ensuite, les variations de contexte et de mise en œuvre propre à chaque intervention permettent d'établir des typologies de configurations efficaces ou non, au vu des régularités d'échec ou de succès rencontrés par la mise en œuvre de cette intervention (Ridde, 2012 : 4). À la différence d'une épistémologie « positiviste » qui invalide un *mécanisme* s'il ne produit pas l'effet souhaité dans un certain contexte, l'épistémologie réaliste peut expliquer cette absence d'effet par des conditions extrinsèques ou intrinsèques non réunies, ou par le voisinage avec un *mécanisme* compensateur, comme par exemple un manque de cohérence avec une autre intervention (Tsang et Kwan 1999 : 169).

Cette logique de comparaison centre l'étude de cas sur le questionnement de la rencontre d'une théorie et d'un contexte particulier, en s'inscrivant dans une interrogation, en

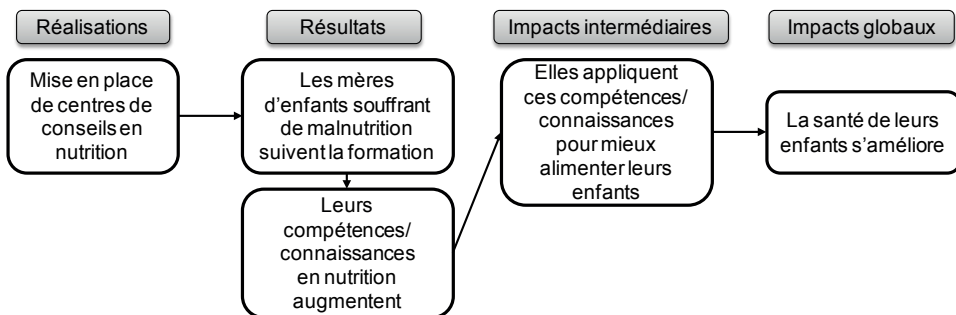
retour, de cette théorie par les données empiriques du cas (David 2004 : 3). Cet échange entre la théorie et le terrain, contribue à enrichir la connaissance sur chaque *mécanisme*, et notamment à mettre à jour les suppositions l’accompagnant, c’est-à-dire les conditions intrinsèques et extrinsèques favorisant son bon fonctionnement. Ici le contexte n’est pas une variable à contrôler, mais une source d’informations précieuses pour délimiter le périmètre de validité du *mécanisme* ou de la théorie (Eisenhardt et Graebner 2007 : 25).

Il est alors possible de générer de la connaissance à partir d’une théorie intermédiaire, composée du même ensemble de *mécanismes*, dans des contextes différents. Par exemple pour des politiques de formation en faisant varier le lieu de formation, le profil des publics ou le type de pédagogie.

L’approche par *mécanisme* permet aussi de comparer ce type d’intervention avec des programmes sensiblement différents, mais correspondant à la même logique. Par exemple, un programme développé au Bangladesh de formation à la nutrition auprès de mères d’enfants souffrant de malnutrition, peut apporter une connaissance sur des suppositions intrinsèques ou extrinsèques favorisant l’activation d’un *mécanisme* présent aussi dans un dispositif de formation pour augmenter l’employabilité des demandeurs d’emploi.

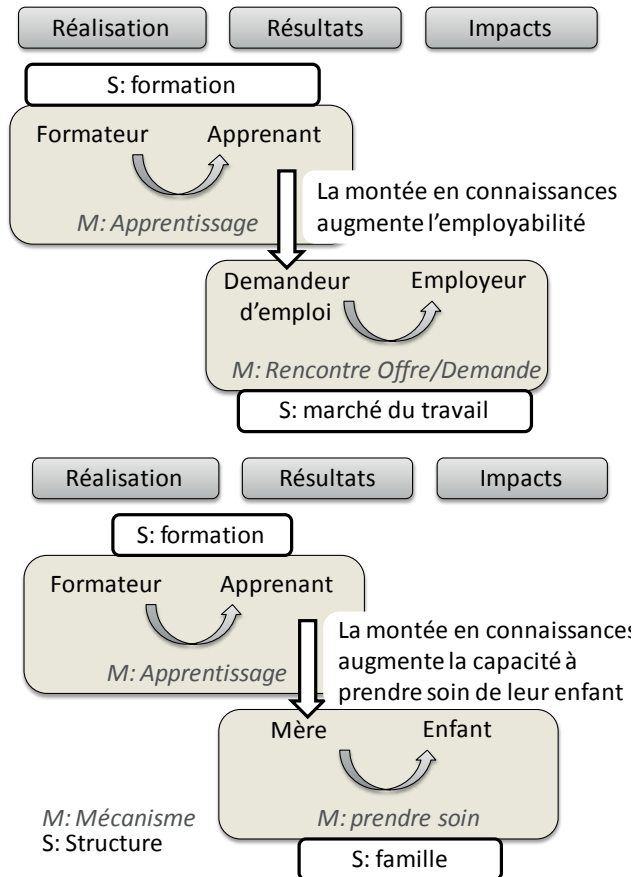
Prenons cette intervention financée par la Banque mondiale. D’abord expérimentée en Inde, elle fut étendue au Bangladesh sur la base des premiers résultats d’étude d’impact positifs bien qu’ils fussent ensuite mis en cause par d’autres résultats plus mitigés (White, 2009 : 4). Suite à cette ambiguïté des résultats, les commanditaires ont décidé d’avoir recours à une évaluation basée sur la théorie afin d’identifier quelle dimension de la *théorie du programme* connaissait des dysfonctionnements (Funnell et Rogers, 2011 : 483). Ce programme proposait un apport de connaissance aux publics ciblés afin de modifier leurs comportements alimentaires, et proposait aussi un apport de nourriture. La dimension la plus importante du programme était le changement de comportement engendré par l’apport de connaissance, et donc notre centre d’intérêt ici. La *théorie du changement* du programme d’éducation à la nutrition peut être présentée de la manière suivante :

**Figure 4 :** Théorie du changement du programme de formation à la nutrition



Si l’on monte au niveau conceptuel du *mécanisme* on peut convertir cette logique d’action en une théorie intermédiaire composée d’un ensemble de *mécanismes* permettant de la comparer à d’autres interventions. Quels enseignements issus de cette évaluation peuvent éclairer d’autres interventions publiques ?

**Figure 5 :** Théories du changement et mécanismes d'un programme de formation des demandeurs d'emploi (en haut) et de formation à la nutrition (en bas).



Les évaluateurs ont identifié plusieurs anomalies dans le déroulement de la théorie. Ils ont remarqué dans la formation à la nutrition que le passage entre les réalisations et les résultats n'était pas stable, puisque certains des conseillers n'avaient pas les outils adéquats pour diagnostiquer les enfants les plus malnutris et former les mères les plus en difficultés. Ils ont aussi remarqué que le suivi de formation et l'acquisition de connaissance, dans de nombreux cas, n'avaient pas abouti à un changement de comportement ni à une amélioration de la santé des enfants ciblés. Une étude ethnologique révéla que le public de départ avait été mal ciblé, les mères du Bangladesh n'étant ni responsables des courses alimentaires – tâche qui incombe à leurs maris – ni responsables des menus de leurs enfants – tâche qui incombe aux grands-mères dans un contexte de foyer intergénérationnel (White, 2009 : 6).

Ces enseignements sur l'intervention, d'apparence très liés au contexte local, semblent difficilement être sujets à une généralisation et apporter des connaissances mobilisables à une autre échelle et dans d'autres contextes. Néanmoins, une montée en conceptualisation permet de rapprocher ce programme d'autres types d'intervention.

La mise en place de centres de conseils en nutrition est une autre modalité d'organisation d'une formation qui requiert le suivi de la formation par les publics ciblés et l'acquisition de connaissances et de compétences. L'objectif intermédiaire est bien l'utilisation de cet acquis dans le cadre souhaité. Il est possible alors de convertir les enseignements locaux en enseignements communs, aux différents programmes correspondants à ces mécanismes similaires, et pouvant éclairer de manière opérationnelle leur conception et leur mise en œuvre.

Premièrement, le constat qu'une partie importante des nourrissons participants n'étaient pas les plus malnutris peut être converti en enseignement général qui préconiserait la conception d'outils communs et fiables mis à disposition des opérateurs de terrain pour déterminer les publics cibles les mieux à même de bénéficier de ce type d'intervention.

Deuxièmement, le constat du manque d'influence des mères bangladaises sur le choix et la préparation des repas de leur enfant peut être converti en enseignement général précisant que le bon fonctionnement de cette théorie requiert que les publics ciblés et suivant la formation aient les capacités d'appliquer dans la structure souhaitée (marché de l'emploi ou famille) leurs nouvelles connaissances et compétences acquises pour atteindre l'objectif souhaité. Par exemple si l'on forme des demandeurs d'emploi à la rédaction de CV pour augmenter leur employabilité, il faut au préalable avoir analysé la *structure* du marché de l'emploi dans lequel ils doivent appliquer ces compétences. Si l'on postule une relation directe employeurs/employés, on formera les demandeurs d'emploi à rédiger des CV très ciblés ce qui peut s'avérer être contre-productif si cette relation est plutôt entre le demandeur d'emploi et une agence d'intérim qui valorisera un CV plus polyvalent. Ces exemples offrent un aperçu des leçons applicables à des contextes très différents et pouvant éclairer les porteurs de projet ou les décideurs publics dans la programmation future d'interventions de ce type.

Ainsi, loin de se cantonner à une prise de recul théorique, l'évaluation basée sur la théorie propose une démarche itérative permanente d'allers-retours entre la théorie et le terrain. La théorie, qu'elle soit profane ou emprunte à des théories académiques, apporte une conceptualisation rendant intelligibles des situations complexes ou compliquées et, en retour, les données du terrain questionnent systématiquement la validité de la théorie. L'hétérogénéité de cette démarche dans sa conceptualisation et dans ses applications, participe à en faire sa force, en favorisant sa malléabilité. De fait, la démarche basée sur la théorie connaît encore des degrés de rigueur de démonstration très variables selon les outils de collectes de données employés, le degré de triangulation des éléments de preuve et l'attention portée à la démonstration de la causalité.

La tendance actuelle en évaluation est d'opposer, d'une part, les protocoles expérimentaux qui seraient capables de produire des connaissances rigoureuses, mais contrôlent très fortement l'action publique et, d'autre part, un ensemble de protocoles qui contrôlent moins l'action publique, mais sont souvent perçus comme incapables de produire des connaissances rigoureuses. L'évaluation basée sur la théorie cherche justement à dépasser cette opposition, mais reste encore peu présente dans les pratiques d'évaluation. Son utilisation reste anecdotique en France, se cantonnant à une poignée d'évaluations conduites aussi bien par des évaluateurs privés que par des chercheurs (Delahais et Toulemonde 2012; Guichard et Ridde 2008). Le défi pour cette démarche réside dans sa promotion auprès des commanditaires d'évaluations afin que ses promesses théoriques trouvent leur incarnation de manière significative dans une diversité de pratiques rigoureuses.



## Bibliographie

- AVENIER M-J., THOMAS C., (2011). Mixer Quali et Quanti pour quoi faire ? Méthodologie sans épistémologie n'est que ruine de la réflexion ! *Cahier de Recherche CERAG* (2011-066E4) (mars).
- BARON G., MONNIER E., (2003). Une approche pluraliste et participative : Coproduire l'évaluation avec la société civile. *Revue Informations sociales* 110.
- BEFANI B., (2012). *Models of Causality and Causal Inference. In Broadening the Range of Designs and Methods for Impact Evaluation*. DFID.
- BIRCKMAYER J., WEISS C., (2000). Theory-Based Evaluation in Practice, What Do We Learn? *Evaluation Review* 24 (4) : 407-431.
- BLAMEY A., MACKENZIE M., (2007). Theories of Change and Realistic Evaluation, Peas in a Pod or Apple and Oranges? *Evaluation* 13 (4), 439-455.
- CHEN H-T., (1994). *Theory-driven Evaluation*. Sage, London.
- DAVID A., (2003). Études de cas et généralisation scientifique en sciences de gestion. *Revue Sciences de gestion* 39, 139-166.
- DELAHAIS T., TOULEMONDE J., (2012). Applying Contribution Analysis: Lessons from Five Years of Practice. *Evaluation* 18 (3), 281-293.
- DENZIN N., (1978). *Sociological Methods*. McGraw-Hill, New York.
- DUFLO E., (2005). Évaluer l'impact des programmes d'aide au développement : le rôle des évaluations par assignation aléatoire. *Revue d'économie du développement* 19 (2), 185-226.
- EISENHARDT K. M., GRAEBNER M., (2007). Theory Building from Cases: Opportunities and Challenges. *The Academy of Management Journal* 50 (1), 25-32.
- FUNNEL S., ROGERS P., 2011. Purposeful Program Theory, Effective use of Theories of Change and Logic Models. Jossey-Bass, Chichester.
- GEORGE A., BENNETT A., (2005). *Case Studies and Theory Development in the Social Sciences*. MIT Press, Cambridge.
- GIBERT P., (2003). L'évaluation de politique : contrôle externe de la gestion publique? *Revue française de gestion* 147, 259-273.
- GUICHARD A., RIDDEV., (2008). Étude exploratoire des mécanismes de l'efficacité des interventions visant à réduire les inégalités sociales de santé. Étude pilote dans trois régions françaises. Institut National de Prévention et d'Éducation pour la Santé, Paris.
- HALL P., (2006). Systematic Process Analysis: when and how to use it. *European Management Review* 3 (1), 24-31.
- HASSENTEUFEL P., (2011). *Les acteurs intermédiaires des politiques publiques. In Sociologie politique: l'action publique*. Armand Colin, Paris.
- L'HORTY Y., PETIT P., (2011). *L'évaluation aléatoire : un succès qui ne doit rien au hasard*. La vie des idées, Paris.
- LINDBLOM C., (1959). The Science of "Muddling Through". *Public Administration Review* 19, 79-88.
- MAYNE J., (2001). Addressing Attribution through Contribution Analysis: Using Performance Measures Sensibly. *The Canadian Journal of Program Evaluation* 16 (1), 1-24.
- PATTON M., (2001). *Qualitative Evaluation and Research Methods*. Sage, London.
- PATTON M. Q., (1997). *Utilization-focused Evaluation: the New Century Text*. Sage, London.
- PAWSON R., (2002). Evidence-based Policy: The Promise of "Realist synthesis". *Evaluation* 8 (3), 340-358.
- PAWSON R., TILLEY N., (2004). *Realist Evaluation*. British Cabinet Office, London.
- PRESSMAN J., WILDAVSKY A., (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*. University of California, Berkeley.

- RIDDE V., ROBERT E., GUICHARD A., BLAISE P., OLMEN J., (2012). L'approche Réaliste à l'épreuve du réel de l'évaluation des programmes. *Revue canadienne d'évaluation de programme* 26 (3), 37-59.
- ROGERS P., PETROSINO A. *et al.*, 2000. Program Theory Evaluation: Practice, Promise, and Problems. *New directions for evaluation* 87, 5-13.
- SHADISH W., (1987). Program Micro-and Macro Theories: a Guide for Social Change. *New Directions for Evaluation* 33, 93-109.
- STAME N., (2004). Theory-based Evaluation and Types of Complexity. *Evaluation* 10 (1), 58-76.
- STAME N., (2009). Avantages et inconvénients des différentes méthodes d'évaluation : comment choisir? In S Trosa Évaluer les politiques publiques pour améliorer l'action publique une perspective internationale, Comité pour l'histoire économique et financière de la France, Paris.
- SUCHMAN E., (1967). *Evaluative Research: Principles and Practice in Public Service and Social Action Programs*. Sage, London.
- TOULEMONDE J. (2009). Voisinages : de la confusion à la complémentarité - Évaluation, audit de performance, pilotage de la performance, management par la qualité. *Les cahiers de la SFE* 4.
- TSANG E., KWAN K-M., (1999). Replication and Theory Development in Organizational Science: A Critical Realist Perspective. *The Academy of Management Review* 24 (4), 759-780.
- VANDERKNAAP P., (2004). Theory-based Evaluation and learning: Possibilities and Challenges. *Evaluation* 10 (1), 16-34.
- WEISS C., (2001). *What Kind of Evidence in Evidence-Based Policy? In Third International, Inter-disciplinary Evidence-based policies and Indicator system Conference*. <http://cem.dur.ac.uk>.
- WEISS C., (1997). Theory-based Evaluation: Past, Present, and Future. *New directions for evaluation* 76, 41-55.
- WEISS C., (1990). Evaluation Research for Decision: Is Anybody There? Does Anybody Care? In M. C. Alkin, *Debates on Evaluation*, Sage, London.
- WHITE H., (2009). *Theory-Based Impact Evaluation: Principles and Practice*. International Initiative for Impact Evaluation. [http://www.3ieimpact.org/admin/pdfs\\_papers/Working\\_Paper\\_3.pdf](http://www.3ieimpact.org/admin/pdfs_papers/Working_Paper_3.pdf).

