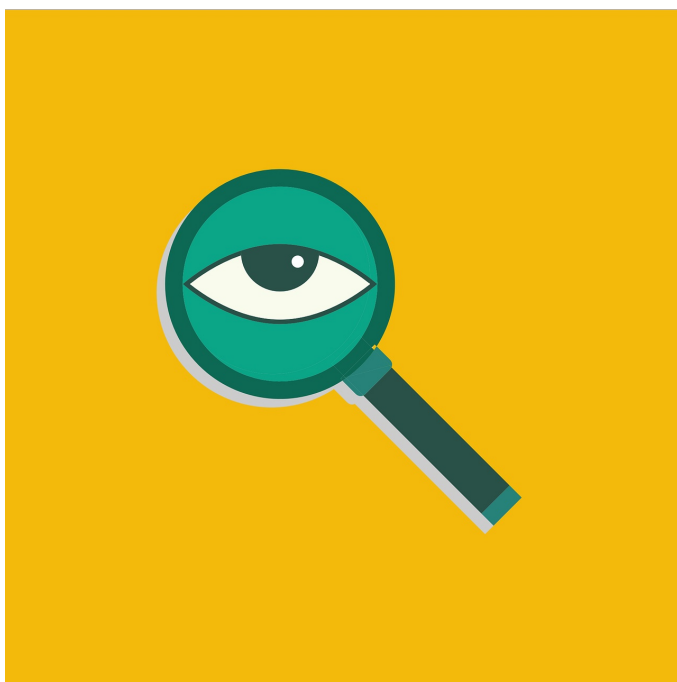


Vous êtes ici : [Accueil](#) > [Politiques Publiques](#) > Quelques limites et critiques de l'évaluation et de la quantification

QUELQUES LIMITES ET CRITIQUES DE L'ÉVALUATION ET DE LA QUANTIFICATION



ÉTUDE

Reconnue comme indispensable pour gérer les sociétés modernes, la quantification est également l'objet de nombreuses critiques...

De nombreux écrits ont ainsi été publiés au cours des dernières décennies sur ce que le mensuel Alternatives Economiques nommait récemment le « fétichisme des chiffres » (Jeanneau & de Ravignan, 2014). Car si tout le monde (ou presque) semble s'accorder sur l'idée que la quantification est indispensable pour comprendre et gérer les sociétés modernes, nombreux sont également ceux qui avertissent sur ses limites, ses dérives... voir ses dangers. On peut tenter de présenter ici quelques-unes de ces critiques en distinguant deux grandes tendances : les critiques d'ordre politique (le pouvoir des chiffres), et celles d'ordre technique (les limites intrinsèques à la quantification).

Tag(s) : Prospective, Service public, Évaluation

Date : 14/12/2015

Aurélien BOUTAUD

Environnementaliste, consultant-chercheur indépendant

Gouverner par les chiffres : la statistique au cœur du pouvoir...

L'outil statistique est un objet hybride, une construction à la fois politique et technique. Ce qui relativise sa dimension supposée objective...

La statistique est le fruit d'une médiation technique et politique. Un premier type d'analyse critique vient des sociologues et des politologues, qui ont développé depuis plusieurs décennies un travail de fond sur la question de la production statistique et son lien avec le pouvoir. Car derrière son apparence d'outil technique visant à l'objectivité, la statistique est en fait le fruit d'un croisement hybride entre ce que Max Weber appelait « le savant et le politique » (Weber, 1919). Alain Desrosières fut probablement en France l'un des premiers à montrer l'imbrication du politique et du technique dans la production de statistiques publiques, ce qui pose de nombreuses questions, notamment dans le domaine des sciences sociales. Car « l'idée de mesure, inspirée des sciences de la nature, suppose que quelque chose de réel peut être 'mesuré', selon une métrologie réaliste. Dans le cas des sciences sociales, l'emploi immodéré du mot mesurer induit en erreur, en laissant dans l'ombre les conventions de la quantification. Ce verbe quantifier, dans sa forme active, implique qu'il existe une série de conventions préalables, de négociations, de compromis, de traductions, d'inscriptions, de codages et de calculs conduisant à la mise en nombre. La quantification se décompose en deux moments : convenir et mesurer ». (Desrosières & Kott, 2005).

...dans le domaine social en particulier, les processus de conventions et de traductions jouent un rôle primordial qu'il ne faut pas oublier au moment d'analyser les chiffres

On peut prendre l'exemple des chiffres du chômage, qui restent très difficiles à comparer d'un pays à un autre car ils relèvent de conventions historiquement différentes, qui ont été longues à s'uniformiser... et qui ne sont toujours pas complètement similaires. Il en va de même des catégories socioprofessionnelles, dont la définition relève de nombreux arbitrages qui n'ont rien de naturel. Et on peut également citer le cas des statistiques ethniques, qui existent dans de nombreux pays et font régulièrement l'objet de débats en France (Roucaute, 2015), mais qui supposent de pouvoir catégoriser des individus en fonction de leur appartenance ethnique – ce qui là encore ne correspond pas à une catégorisation naturelle. Tout cela signifie bien entendu que les chiffres issus de la production statistique sont des constructions, faisant l'objet de médiations techniques et politiques : ils sont empreints de valeurs et se font le reflet de visions du monde parfois très différentes.

Zoom sur les indicateurs de développement durable : quand la quantification devient objet d'appropriation idéologique

On peut dire sans grand risque de se tromper que le caractère éminemment politique et idéologique de la statistique réapparaît aussitôt qu'un concept commence à s'imposer dans les débats publics. Par exemple, lorsqu'au tournant des années 2000 le développement durable est devenu un élément incontournable du discours politique, une impressionnante littérature sur les indicateurs de développement durable a soudainement vu jour. Il faut dire qu'au début des années 2000, ce concept fait l'objet d'interprétations très différentes. Les outils d'évaluation ont alors joué un rôle central dans ce combat idéologique, comme le montre la comparaison des résultats de l'*Environmental Sustainability Index* et de l'*Environmental - Human Development Index*. (Boutaud, 2005) Ces deux indicateurs synthétiques, qui cherchent tous deux à classer les pays en fonction de leurs performances de développement durable, fournissent des résultats à peu près en tout point opposés. Le premier, commandité par le Forum économique de Davos, reflète une vision du développement durable qui favorise les pays riches, en ignorant largement les externalités environnementales liées à leur niveau de vie. Le second, commandité par l'ONG *Friends of the Earth*, pondère au contraire fortement les pressions environnementales liées aux modes de vie, au détriment des résultats socio-économiques (qui ne comptent que pour un tiers du résultat final). Soit deux visions du monde (et du développement durable) radicalement opposées...

	ESI	E-HDI
		Total
Norvège	1	16
Finlande	2	17
Canada	3	6
Islande	4	20
Suède	5	12
Suisse	6	2
Australie	7	10
Danemark	8	19
Etats-Unis	9	14
Pays-Bas	10	18
...	...	
Algérie	11	11
Singapour	12	15
Sénégal	13	13
Malawi	14	9
Vietnam	15	3
Iran	16	4
Nigeria	17	7
Ouganda	18	1
Bangladesh	19	8
Madagascar	20	5

[note: 1 = "meilleur", 20 = "moins bon"]

Les outils statistiques changent au gré du pouvoir...

Une conséquence importante de ce mariage entre sphères techniques et sphères de pouvoir concerne la champ de l'évaluation : on ne mesure que ce que l'on veut bien mesurer ! Et encore le fait-on en fonction de moyens qui ne sont pas infinis. Alain Desrosières explique par exemple qu'il existe historiquement deux sources de la statistique publique : les enquêtes directes (recensements, sondages...) et les registres administratifs (fichiers de gestion, répertoires...). « Ces deux sources sont le produit de formes d'activités et d'enregistrements dont les buts sont très différents : pour dire vite, les enquêtes visent en principe à décrire la société, tandis que les registres, conçus à des fins de gestion, reflètent plutôt les rouages de l'État et des institutions ». Or les premières sont plus coûteuses et plus complexes à mettre en œuvre, tant et si bien qu'on tend à privilégier les sources administratives. Mais celles-ci sont « tributaires de définitions et de catégories réglementaires,

elles diffèrent de celles dont les utilisateurs des statistiques ont besoin, et elles peuvent rendre difficiles les comparaisons, entre pays ou dans le temps ». (Desrosières, 2005).

Les outils de la statistique publique dépendent des acteurs publics et des moyens que ceux-ci veulent bien y mettre : on ne voit que ce qu'on veut voir, avec les moyens dont on veut bien disposer. Au gré de la mondialisation économique, la statistique publique recule... au bénéfice des normes comptables des entreprises, qui ont su s'internationaliser et s'imposer au public.

...et nous passons du règne de la statistique publique à une contre-révolution comptable privée

Une autre évolution importante des dernières décennies touche précisément au fait que les normes statistiques ont eu bien du mal à s'internationaliser. L'hétérogénéité des moyens alloués à la statistique entre les différentes nations du monde, mais aussi les particularismes nationaux et historiques ont nuit à cette internationalisation. Dans un processus de mondialisation économique, la puissance statistique traditionnellement portée par les Etats semble donc sur le point de s'affaiblir. Or Fabrice Bardet (2014) montre que ce recul a largement profité à la montée en puissance d'une autre forme de quantification : la comptabilité.

Davantage portée par le monde de l'entreprise et de l'économie, la « contre-révolution comptable » est progressivement parvenue à réaliser ce que les statistiques publiques avaient échoué à faire : internationaliser les normes comptables... et les imposer aux acteurs publics. Béatrice Touchelay (2011) démontre par exemple comment les statistiques privées et publiques se sont rapprochées par le biais de la comptabilité. Au point que ce sont aujourd'hui les cabinets d'audit externes qui vérifient les comptes publics, et les agences de notation financière qui jugent la performance des Etats ! Un phénomène qui traduit un renversement historique du rapport de force entre le public et le privé : la légitimité du contrôle semble avoir changé de camp (Chiapello & Medjed, 2007).

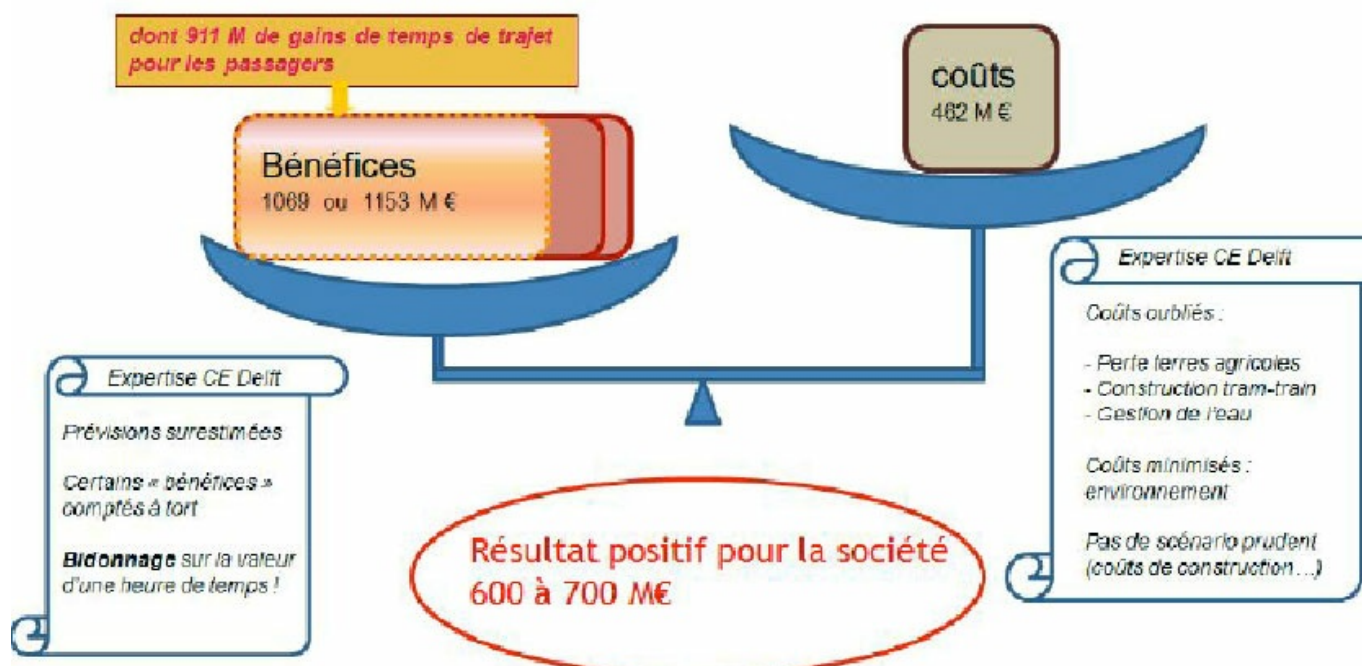
Enfin, l'évaluation peut outrepasser son rôle d'aide à la décision en imposant un point de vue technique. Ce qui amène parfois à manipuler les chiffres pour les rendre conformes aux volontés politiques !

Aider à la décision... ou décider ? Le risque d'expertocratie

Toujours en lien avec ces enjeux de pouvoir et de prise de décision, le risque est souvent pointé d'une dérive technocratique. Censée fournir une « aide à la décision », l'évaluation peut parfois se transformer en acte de décision pur et simple, lorsqu'elle impose une solution supposée optimale et objective. Un bon exemple de ces dérives est fourni par les évaluations monétaires de type « coûts bénéfiques », que Julien Milanesi analyse dans le domaine des projets d'aménagement : « en complétant les chiffres déjà disponibles (comme le coût de construction d'un équipement) ces évaluations sont censées permettre de juger du bien fondé d'une infrastructure ou de tout type de projet (...). Tout étant réductible à une évaluation chiffrée (le marchand comme le non-marchand) la décision publique se limiterait ainsi à une simple soustraction entre les coûts et les bénéfices attendus ». (Milanesi, 2009). Sous son aspect rationnel, l'évaluation risque alors de devenir l'élément d'arbitrage ultime que les instances de décision n'auront pas forcément le courage – ni la volonté – de contourner.

La conséquence logique de cette « dictature du chiffre » dénoncée par Julien Milanesi est que les évaluations sont parfois biaisées afin que leurs résultats correspondent à l'attente des commanditaires – le cas le plus médiatisé de ce type de dérive étant sans doute le projet d'aéroport de Notre-Dame-des-Landes (Cf. encart).

Zoom sur les controverses autour de l'analyse coûts bénéfiques du projet d'aéroport de Notre-Dame-des-Landes



Parmi les nombreuses controverses qui entourent le projet de construction de l'aéroport de Notre-Dame-des-Landes, celle sur l'analyse coûts-bénéfices occupe une place centrale (Weiler, 2011 ; Russell, 2014). On le comprend d'autant mieux que, conformément à la loi d'orientation sur les transports intérieurs, c'est sur la base de cette étude que la déclaration d'utilité publique du projet est fondée. En effet, l'étude commanditée par la Direction Générale de l'Aviation Civile en 2006 concluait à un intérêt global du projet, estimant que l'ensemble des gains seraient supérieurs aux coûts.

Perplexe face à certains résultats, un collectif d'élus a demandé en 2011 une contre-expertise à CE Deft, un cabinet d'études européen spécialisé dans ce genre d'évaluation. Ce dernier a alors relevé de nombreuses irrégularités dans l'étude officielle, dont la plus importante concerne le temps gagné par les passagers pour

se rendre à l'aéroport, chiffré à 911 millions d'euros, soit plus de 80 % de l'ensemble des bénéfices du projet. Pour atteindre ce chiffre, l'étude officielle a notamment élargi le bassin de chalandise de l'aéroport, surestimé les prévisions de trafic et, surtout, accordé une valeur monétaire au temps passé dans les trajets qui s'avère près de 5 fois supérieure à la valeur actualisée proposée par l'instruction cadre du Ministère en charge des transports. Comme le résumait Hervé Kempf dans le Monde en 2011, cette « *"manipulation" a une lourde conséquence : elle permet de présenter comme bénéfique un projet qui, si l'on avait suivi la méthode recommandée, serait apparu comme déficitaire* ». (Kempf, 2011). En corrigeant ce qu'il considère comme des écarts par rapport à la méthode recommandée, le cabinet CE Deft aboutit en effet à des coûts supérieurs aux bénéfices escomptés.

Des chiffres qu'il faut apprendre à... déchiffrer !

La quantification comporte également des limites que l'on pourrait qualifier d'intrinsèques. Pour y faire face, certains auteurs comme Nico Hirt dans son « contre-manuel de statistiques pour citoyens militants » (Hirtt, 2007), appellent à développer une culture de la statistique parmi les citoyens.

Gare à la production des chiffres !

Parmi ces écueils, certains tiennent aux conditions de production des chiffres. Par exemple, le caractère biaisé de certains sondages fait régulièrement l'objet de polémiques. Identifiés de très longue date dans la littérature (par ex. Mucchielli, 1967), les différents biais cognitifs susceptibles d'orienter les réponses seraient ainsi parfois volontairement introduits dans les enquêtes. Mais la plupart du temps, ces biais sont présents de manière involontaire – en particulier dans les statistiques publiques. Un exemple très intéressant nous est fourni par l'Indice de Masse Corporelle des Etats-Unis, déjà présenté dans les pages précédentes. A l'échelle nationale, l'organisme public en charge de collecter les données (le Center for Disease Control and prevention - CDC) mène une enquête basée sur des rencontres durant lesquelles les personnes rencontrées sont pesées et mesurées.

Mais pour obtenir un volume suffisamment représentatif de données à l'échelle des 50 Etats, le CDC doit disposer d'un échantillon beaucoup plus large, ce qui l'amène à réaliser une enquête de grande envergure sur la base cette fois-ci d'entretiens téléphoniques. Or les personnes répondant à ces enquêtes ont généralement tendance à fournir des réponses erronées : en particulier, les femmes annoncent souvent un poids inférieur à leur poids réel tandis que les hommes ont tendance à se déclarer plus grands qu'ils ne le sont ! Il en résulte évidemment une sous-estimation de l'IMC, que les statistiques font clairement apparaître : la valeur nationale (réalisée avec la meilleure méthode) annonce 35% d'obèses parmi la population adulte en 2012... alors que les valeurs régionales indiquent qu'en 2013 seuls deux Etats des Etats-Unis (sur 50) franchissent le seuil de 35% d'obèses parmi la population adulte ! (Trust for America's Health, 2009 ; CDC, 2015). Les protocoles de collecte et de production des données peuvent avoir une influence considérable sur les résultats obtenus.

Attention à la présentation des résultats...

Les pièges potentiels ne se limitent évidemment pas à la production des chiffres, ils peuvent également

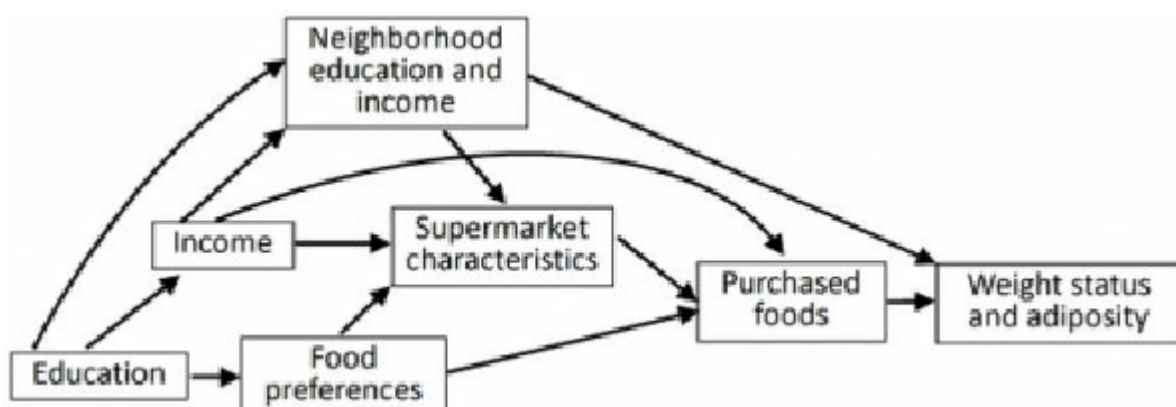
concerner la présentation des résultats. Un biais classique consiste à privilégier les valeurs moyennes par rapport aux valeurs médianes ou aux écarts types. Nico Hirtt développe cet exemple sur le cas de l'évolution des revenus en Belgique, montrant que l'accroissement de la moyenne des revenus présentée dans certaines publications masque en réalité un accroissement des inégalités que la stagnation de la médiane ne manquerait pas de révéler (Hirtt, 2007). En matière de comparaison trompeuse, Jean Gadrey nous rappelle quant à lui qu'il faut également se méfier des références : ainsi, dans la comparaison des chiffres de création d'emploi en France et en Allemagne, il est souvent fait référence dans l'actualité aux évolutions entre 2003 et 2014 : +11% pour l'Allemagne, contre +4,4% pour la France. Or, « pour comparer les variations d'une grandeur dans deux pays il faut à tout prix éviter de prendre comme année de départ une année qui, pour l'un des deux pays, est exceptionnellement basse, ou exceptionnellement haute » (Gadrey, 2015) En prenant comme année de référence 2000 au lieu de 2003, on observe alors un taux de croissance de l'emploi à peu près équivalent pour les deux pays. Et si on prend l'année 1994 comme référence, le taux de croissance de l'emploi constaté sur la période est alors beaucoup plus fort en France qu'en Allemagne. Une fois produits, les résultats peuvent eux-mêmes être présentés de manière trompeuse et/ou biaisée.

...et méfiance à l'égard des interprétations et autres conclusions hâtives

On trouve également des risques de confusion du côté de l'interprétation des résultats statistiques. C'est particulièrement vrai en ce qui concerne les corrélations et les liens de cause à effet, qui ne vont pas toujours de paire. Par exemple, Marc Tertre nous dit qu'il y a bien « corrélation (et relation de cause à effet) entre la consommation de tabac et certains cancers. Mais cette relation peut aussi être trompeuse. Il y a une statistique qui montre que les gens qui chaussent des souliers d'une taille supérieure à 45 commettent trois fois plus de meurtres que ceux qui chaussent entre 40 et 42 ». Corrélation, donc... mais pas causalité. Car cette dernière est en fait à chercher du côté du genre : les hommes ont en moyenne les pieds plus grands que les femmes, et « il se trouve que les comportements meurtriers se retrouvent principalement chez les individus de sexe masculin » (Tertre, 2013).

Ce que montre ce dernier exemple, c'est qu'une corrélation ne signifie pas forcément un lien de causalité... et qu'il faut parfois aller chercher plus loin les mécanismes susceptibles d'expliquer une corrélation. Et ce travail est parfois beaucoup plus difficile à réaliser que dans le cas évoqué par Marc Tertre, comme l'illustre le travail controversé de Kate Pickett et Richard Wilkinson sur les inégalités, cité précédemment. L'interprétation des résultats peut parfois elle-même être trompeuse. Notamment lorsque corrélations et causalités sont confondues.

Zoom sur une corrélation étrange : pourquoi les clients des Monoprix sont-ils plus minces que les clients des autres supermarchés ?



La question peut surprendre, mais c'est pourtant le constat qu'ont réalisé des chercheurs européens dans une étude publiée en 2012 (Chaix et coll., 2012). A l'origine, ces derniers cherchaient à vérifier si le type de supermarchés dans lequel les ménages font leurs courses a un impact sur leur consommation finale – notamment en termes de quantité de calories absorbées. Et leurs résultats font effectivement apparaître une différence d'Indice de Masse Corporelle entre les clients de différents types de magasins, avec des écarts particulièrement significatifs entre, d'un côté, les clients des supermarchés de proximité (dont l'IMC moyen est le plus faible) et, à l'autre opposé, ceux des hypermarchés et des hard discount des périphéries urbaines. Explorant les raisons d'un tel écart, les auteurs notent d'abord que le contenu en matière grasse ou en sucre est très similaire entre les produits vendus dans les différents supermarchés, ce qui ne saurait donc expliquer les différences. Comme le note ironiquement Colette Roos (2012) « il y a si peu de différence que lorsque vous transplantez un type riche et diplômé qui habite les beaux quartiers dans un Lidl ou un Aldi de Trappes, ses courses ne lui feront pas prendre un gramme. En revanche, placez un homme pauvre titulaire d'un brevet des collèges dans ces mêmes grandes surfaces alimentaires, et cela aura un impact sur son IMC ».

Ce que mettent en évidence les auteurs de l'étude, c'est en réalité qu'il faut chercher les explications de cette étrange corrélation dans les parcours de vie des individus. Le niveau d'éducation est ainsi le facteur le plus déterminant car il va avoir une influence sur les revenus, sur les préférences alimentaires, mais aussi sur le lieu de résidence. Colette Roos résume à merveille la mécanique qui s'installe alors : « Géomarketing aidant, les produits disponibles dans le supermarché le plus proche de chez vous (un Monoprix, bien sûr, puisque vous vivrez en centre-ville) seront plus propices à vous faire garder la ligne (...). Et puis vos goûts étant modelés par votre classe sociale, (...) vous serez enclin à acheter des baies de goji déshydratées à 155 €/kg, pas des barres chocolatées pleines d'acides gras trans. Conclusion : il ne faut pas confondre causalité (Monoprix rend mince) et corrélation (les gens qui gagnent bien leur vie sont minces, et ils fréquentent les Monoprix). Ce que l'étude prouve, (...) c'est que l'obésité est une maladie sociale ». (Roos, 2012).

Une dernière limite : la politique du chiffre

Enfin, une autre limite touche cette fois-ci à l'utilisation des chiffres, qui peut parfois devenir contreproductive. Le mensuel Alternatives Economiques dénonce par exemple le fait que, trop souvent, « au lieu de nous appuyer sur les statistiques pour construire des politiques appropriées et audacieuses, nous les adorons comme des dieux auxquels nous offrons notre jugement en sacrifice. » (Jeanneau & de Ravignan, 2014) Se concentrer sur un indicateur phare au point de perdre de vue tout autre enjeu est en effet une dérive récurrente, que les médias dénoncent régulièrement sous l'intitulé de « politique du chiffre ». Ainsi, malgré les avertissements du Conseil National de l'Information Statistique qui invitent à bien mesurer les limites d'une focalisation sur le taux de chômage (CNIS, 2008), l'actualité nous montre qu'une telle obsession est parfois difficile à éviter. Le risque est alors de mettre en œuvre des politiques dont les effets secondaires peuvent être redoutables : par exemple, en accroissant le taux de retour à l'embauche par le développement de la précarité de l'emploi, dont les indicateurs sont moins médiatisés.

C'est ce qui est arrivé en Grande-Bretagne avec l'explosion des contrats « zéro heure » et en Allemagne avec les « mini-jobs », que Guillaume Duval décrit comme une « bombe sociale » représentant « cinq millions de personnes qui gagnent moins de 400 euros par mois et qui n'ont pas cotisé pour la retraite pendant dix ans. » (Duval, 2014) Un constat que l'OCDE confirme cette fois-ci sur la question des temps partiels, qui représentent en France 14,3% de salariés en 2013, contre 22,8% des Allemands et 23,9% des Britanniques (cité par Gadrey, 2015).

Enfin, un dernier danger guette les décideurs : la tentation de la « politique du chiffre ». Une obsession qui mène parfois à des politiques contreproductives.

Mais c'est sans doute lorsque l'évaluation s'assimile à un contrôle que les aspects les plus improductifs de la politique du chiffre apparaissent. Mis sous pression, les agents sont alors parfois amenés à adapter leurs pratiques pour « produire du chiffre ». La montée en puissance du mouvement de la nouvelle gestion publique (new public management), qui prétend mettre la quantification des objectifs au cœur de l'action publique, fournit malheureusement de nombreux exemples de cette contre-productivité. Un rapport de l'IGA dénonce par exemple les changements constatés à partir de 2007 dans l'enregistrement des plaintes de police et de gendarmerie en France. Selon l'IGA, « le management par objectifs de la délinquance, connu sous la dénomination de "politique du chiffre", a largement contribué à cette absence ou ce désengagement du contrôle de l'enregistrement. » Pour répondre aux objectifs de baisse de la délinquance fixés par le ministère, les fonctionnaires de police et de gendarmerie auraient donc artificiellement réduit le nombre d'enregistrement des plaintes, encouragés en cela par l'administration centrale, dont « certaines directives (...) ont pu contribuer à minorer fortement les statistiques de la délinquance en généralisant des pratiques d'enregistrement non conformes » (AFP, 2013).

Et on trouve des exemples de dérive similaires dans de nombreuses autres politiques soumises à la pression des indicateurs tels qu'inspirés de la nouvelle gestion publique, comme les services hospitaliers (Belorgey, 2010) ou encore la recherche universitaire. Dans ce dernier domaine, Yves Gingras (2014) montre par exemple les dérives de la « bibliométrie » qui amène les chercheurs à publier en nombre des articles de faible valeur ajoutée, publiés dans des revues mainstream et demandant peu (et si possible pas du tout) de travail de terrain. Avec comme conséquence de dévaloriser la recherche in situ et l'innovation. Pour « Peter Lawrence, professeur à Cambridge, 'le principal résultat de la bibliométrie, c'est que l'objectif principal des savants n'est plus de faire des découvertes mais de publier autant que possible', de sorte que 'l'utilité, la qualité et l'objectivité des articles se sont dégradées » (cité par Jourde, 2008).

Le management par objectifs prôné par la nouvelle gestion publique amène parfois les agents à des comportements contreproductifs qui n'ont pour seul but que d'atteindre les objectifs... au détriment du reste. Le genre de constat qui amène Laurent Jeanneau et Antoine de Ravignan à conclure qu'« il en va finalement des chiffres comme de la musique : déchiffrer une partition, ce n'est pas l'interpréter ». (Jeanneau & de Ravignan, 2014).

Bibliographie sur l'évaluation des politiques publiques (PDF 94.42 ko)