



HAL
open science

Enjeux et limites du pilotage par les indicateurs en management public, l'exemple de la recherche en sciences de gestion

Dominique Bessire, Pascal Fabre

► **To cite this version:**

Dominique Bessire, Pascal Fabre. Enjeux et limites du pilotage par les indicateurs en management public, l'exemple de la recherche en sciences de gestion. Comptabilités, économie et société, May 2011, Montpellier, France. pp.cd-rom. hal-00646755

HAL Id: hal-00646755

<https://hal.science/hal-00646755>

Submitted on 30 Nov 2011

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

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

Enjeux et limites du pilotage par les indicateurs en management public, l'exemple de la recherche en sciences de gestion

Dominique BESSIRE, Professeur à l'Institut d'Administration des Entreprises d'Orléans

LOG - Laboratoire Orléanais de Gestion, Dominique.Bessire@univ-orleans.fr

Pascal FABRE, Maître de conférences HDR

LCT – Laboratoire des Collectivités Territoriales, Pascal.Fabre@univ-orleans.fr

Résumé : Le discours du *New Public Management* promeut un pilotage par les indicateurs et une logique de résultats. Nous nous appuyons sur le cas de la recherche en sciences de gestion pour mettre en évidence, en termes de contrôle de gestion, les errements conceptuels qui marquent l'adoption de ce mode de management, en nous fondant sur une modélisation du processus de « production » et du système d'activités. Nous soulignons en particulier la confusion entre indicateurs de résultat et indicateurs d'impact ainsi que la réduction du système complexe d'activités de recherche à un nombre réduit d'indicateurs. Plus généralement, nous nous efforçons de montrer que les activités de recherche ne remplissent pas les conditions pour qu'un contrôle de type cybernétique fonctionne correctement.

Mots clés : management public – indicateurs – recherche – évaluation

Abstract: Discourses on New Public Management promote control through indicators and management by results. We analyze, in terms of management control, the case of research in management science to point the mistaken ideas which underpin this type of management. Our demonstration is based on the modelization of the production process and of the system of activities. We underline the confusion made between outputs and outcomes as well the reduction of the complex system of activities which contribute to research to a limited set of indicators. More generally, we aim to demonstrate that research activities do not meet the conditions required for a cybernetic control.

Key words: public management – indicators – research – assessment

A partir des années soixante-dix s'affirme un discours qui promeut ce qu'il est convenu d'appeler le *New Public Management* (NPM dans la suite du texte, traduit en français par Nouvelle Gestion Publique (NGP) ; Hood, 1991 et 1995). La mise en œuvre des principes qu'énonce ce discours va de pair avec la volonté de passer d'une gestion par les moyens à une gestion par les résultats et la croyance qu'il est possible de mettre sous contrôle les organisations publiques sur la seule base d'un système d'indicateurs (Biondi et al., 2008). L'objectif de cet article est de mettre en évidence le caractère erroné et dangereux d'une telle croyance. Nous avons pris ici le parti, non pas de mener une critique globale du NPM et de ses sous-tendus idéologiques, mais de mener une critique en quelque sorte de l'« intérieur », sur la base d'un strict raisonnement en termes de contrôle de gestion, reposant sur une modélisation générique du processus de « production » et d'une modélisation spécifique du système d'activités de recherche.

Nous avons choisi comme champ d'analyse la recherche, parce que c'est un champ qui aujourd'hui en France connaît aujourd'hui une mutation sans précédent sous les effets notamment de ce pilotage¹ par les indicateurs. Nous avons aussi choisi ce champ parce que la recherche est au cœur d'un développement durable de notre société, en particulier parce qu'elle est centrée sur les potentialités offertes aux générations futures mentionnées dans la fameuse définition de Brundtland du développement durable (1987). Nous ne raisonnerons que

¹ Le terme pilotage utilisé ici ne préjuge en rien des multiples usages des indicateurs.

pour les sciences humaines et plus spécifiquement pour les sciences de gestion, même s'il nous semble que nombre de critiques ou de propositions peuvent s'appliquer *mutatis mutandis* à d'autres champs. Par ailleurs, nous ne prétendons nullement à l'exhaustivité : en effet, plus la réflexion avance, plus le phénomène apparaît complexe. Enfin nous nous concentrerons sur un niveau précis correspondant à l'unité opérationnelle de base : le laboratoire. Il faut noter par ailleurs que si la diffusion du NPM a donné lieu à une abondante littérature, il existe relativement peu d'écrits centrés sur la recherche : en français sur la base de données *Vente et gestion*, il n'y avait à la date du 28/03/2011 aucun article traitant spécifiquement du NPM appliqué à l'Université ; dans la base de données *Business Source Premier*, les références sont un peu plus nombreuses (18).

Dans une première section, nous évoquerons donc le contexte dans lequel les indicateurs de gestion se diffusent aujourd'hui dans les organisations publiques françaises. Dans une deuxième section nous proposerons une grille d'analyse permettant de mettre en évidence les conditions d'un pilotage efficient par les indicateurs. Dans une troisième section, nous appliquerons cette grille d'analyse à la recherche en sciences de gestion pour proposer une critique raisonnée du mode de pilotage qui tend à s'imposer sous l'égide de l'AERES (Agence de l'évaluation de la recherche et de l'enseignement supérieur).

1. *New Public Management* et indicateurs de gestion

Le développement des indicateurs de gestion est le corollaire de la diffusion des principes du NPM. Sans instrumentation de gestion, ceux-ci ne seraient restés qu'à l'état de discours. Cependant le pilotage par les indicateurs renvoie à une forme très spécifique de contrôle, le contrôle cybernétique, qui s'avère dépourvu de pertinence dans certains contextes.

1.1. Le discours du *New Public Management* et sa traduction managériale : le pilotage par les indicateurs

L'idée générale portée par le *New Public Management* est que les organisations publiques sont mal gérées et qu'en conséquence l'argent public est gaspillé. Les publications de l'OCDE sur le thème du NPM se rangent d'ailleurs sous la bannière « Value for money in Government »². Ce discours va avoir des conséquences très concrètes. Au niveau macroéconomique, il justifiera dans de nombreux pays les baisses générales d'impôts qui à leur tour entraîneront une stabilisation, voire une diminution des ressources allouées aux organisations publiques ; il entraînera la dévolution à des organisations privées de missions jusque là assumées par des organisations publiques, imposant une concurrence supposée bénéfique entre ces deux formes d'organisation ; il entraînera, directement ou indirectement, une redistribution des ressources entre organisations ; il ira de pair avec une exigence accrue sur la reddition des comptes (*accountability*)³ et l'imposition de méthodes qui ont fait leurs preuves dans les organisations privées (managérialisme) ; enfin il contribuera à une redistribution interne des moyens (humains, financiers).

² Pour une analyse de l'influence décisive exercée par l'OCDE sur la diffusion du NPM, voir Pal (2009).

³ Traduit dans la récente norme ISO 26 000 par le terme redevabilité.

Les pays anglo-saxons ont été parmi les premiers à avoir tenté de mettre en œuvre des réformes inspirées du NPM. L’Australie, la Nouvelle-Zélande, le Royaume-Uni et les Etats-Unis font figure de pionniers (Lapsley, 2009) ; ils ont été rejoints par un nombre grandissant de pays. Les études menées dans les pays qui disposent aujourd’hui d’un historique significatif mettent en évidence des résultats fréquemment décevants, quels que soient les domaines considérés (prisons, santé, éducation..., cf. par exemple Lapsley, 2009 ; Olson et al. 2001 ; Norman, 2007 et Gregory, 2000 pour la Nouvelle-Zélande ; Biondi et al., 2008). Cela n’a pas découragé d’autres pays, y compris la France, de vouloir eux aussi mettre en œuvre des réformes similaires. La Loi organique relative aux lois de finances (LOLF) apparaît en effet comme la traduction concrète des principes du NPM et impose dans son sillage la production de toute une batterie d’indicateurs dans toutes les organisations sous contrôle direct ou indirect de l’Etat ; c’est en particulier le cas dans les universités, les hôpitaux, les administrations... Les collectivités locales auxquelles s’applique un principe de libre administration (qu’elles défendent jalousement) échappent à cette obligation légale, mais sont menacées par le gouvernement d’une conditionnalité des dotations publiques à la production d’indicateurs attestant de leur « bonne gestion » (Roturier, 2010). Les associations, largement financées par la puissance publique (Etat ou collectivités territoriales), se voient progressivement imposer les mêmes contraintes par leurs bailleurs de fonds.

Dans les paragraphes qui suivent, nous tentons de mettre en évidence la relation pilotage par les indicateurs et modes de contrôle et les risques qu’entraîne le choix erroné d’un mode de contrôle.

1.2. La place des indicateurs dans les systèmes de contrôle

De nombreux auteurs se sont attachés à définir une typologie des modes de contrôle et les conditions de leur validité en fonction de différents paramètres. Nous nous référerons ici à trois catégories de travaux dont les conclusions sont assez communément acceptées dans la communauté scientifique en contrôle de gestion et ont été largement reprises dans diverses études ; il s’agit de ceux d’Ouchi (1979), d’Hofstede (1978, 1981) et de Burchell et al. (1980).

Ouchi (1979, p. 843) distingue trois mécanismes principaux de contrôle : mesure des comportements, mesure des extrants (*output*⁴) et contrôle par le clan. Il montre que l’efficacité de chacun de ces modes de contrôle dépend du degré de connaissance du processus de transformation et de la possibilité de mesurer avec une relative précision les extrants.

Ce tableau appelle plusieurs remarques. En premier lieu le pilotage par les indicateurs relève d’un contrôle par les extrants ; celui-ci n’est possible que pour autant que ceux-ci soient mesurables avec une certaine précision (nous devrions ajouter : et avec une certaine complétude) ; ceci réduit considérablement son champ d’application dans la mesure où dans de nombreuses situations, cette mesure est impossible. En second lieu, le contrôle d’un

⁴ Nous traduisons *outputs* par « extrants ». La traduction usuelle « résultats » nous semble trop restrictive et en outre source de confusion dans la mesure où comme nous le verrons ultérieurement, il est particulièrement important en matière de management public de distinguer résultats et impacts.

laboratoire de recherche dans une grande entreprise relèverait, selon Ouchi, d'un contrôle par le clan. Son argumentation (p. 844) mérite d'être reprise en détail.

« On ne dispose d'aucun moyen pour définir les règles qui, si elles étaient suivies, permettraient de conduire aux percées scientifiques désirées qui à leur tour conduiraient à de nouveaux produits susceptibles d'être mis sur le marché par l'entreprise. Il est possible de mesurer le succès ultime d'une découverte scientifique, mais le processus pour qu'une découverte apparemment obscure soit pleinement appréciée peut prendre 10, 20, voire 50 ans. Sans nul doute, nous ferions bien d'y réfléchir à deux fois avant d'utiliser un type de contrôle fort par les extrants pour encourager certains chercheurs dans notre laboratoire tout en décourageant d'autres. De fait, nous sommes dans l'impossibilité d'utiliser une mesure des comportements ou des extrants, ce qui ne nous laisse aucune forme « rationnelle » de contrôle. Dans ces circonstances, l'organisation fait largement appel à des formes de contrôle ritualisé et cérémoniel. Cela inclut le recrutement d'un nombre limité d'individus sélectionnés, dont chacun est passé par un processus de formation et de professionnalisation qui lui a appris à internaliser les valeurs désirées et à révéler les cérémonies appropriées. Les plus importantes de ces cérémonies telles que le « bizutage » de nouveaux membres dans des séminaires, la participation à des réunions d'associations professionnelles et l'écriture d'articles pour la publication dans des revues savantes, seront encouragées au sein du laboratoire. »⁵

Tableau 1
Conditions déterminant la mesure des comportements ou des extrants (*outputs*)
Source : Ouchi (1979, p. 843, libre traduction des auteurs)

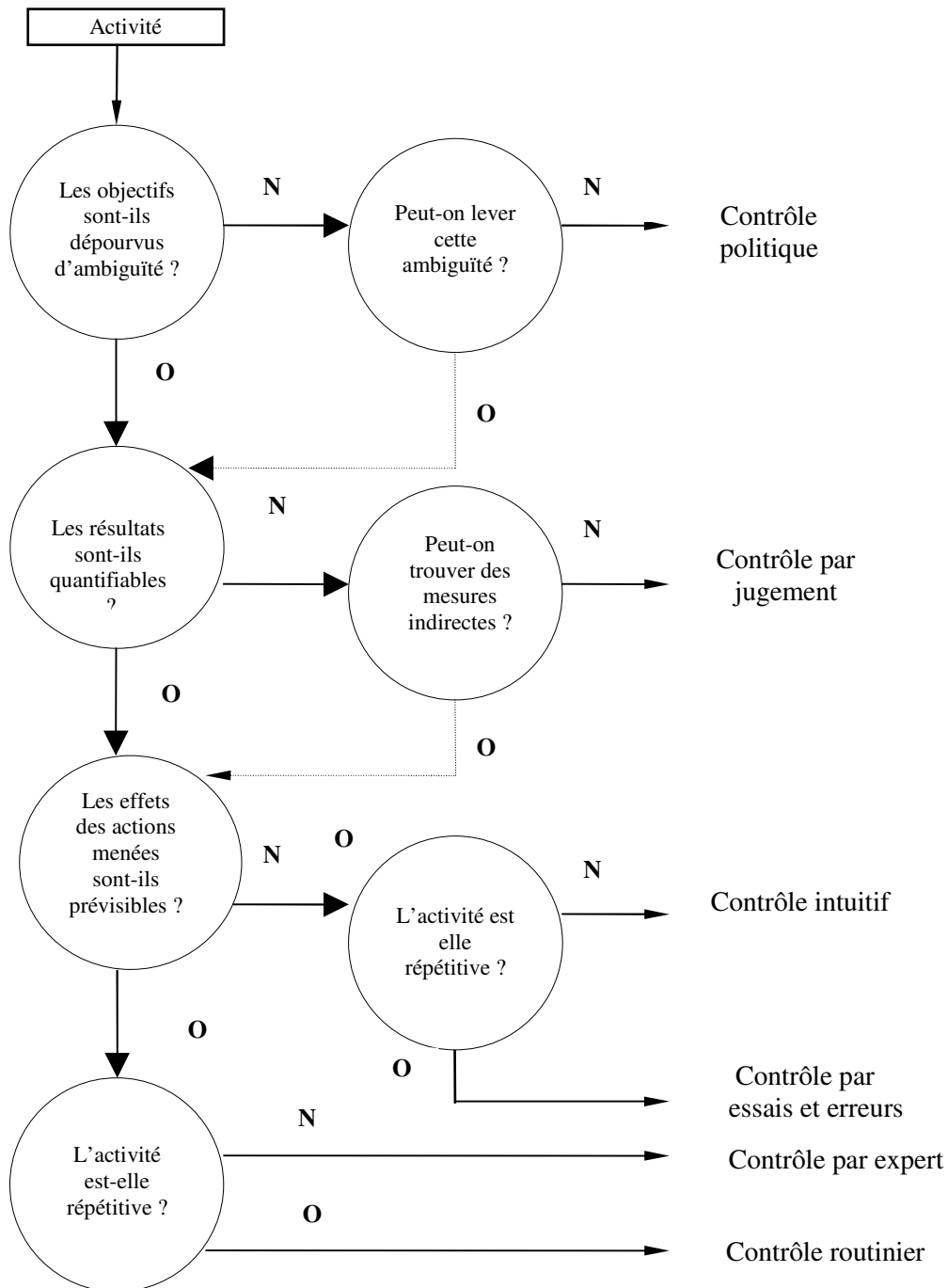
		Connaissance du processus de transformation	
		Parfaite	Imparfaite
Possibilité de mesurer les extrants (<i>outputs</i>)	Elevée	Mesure des comportements ou des extrants (Programme Apollo)	Mesure des extrants (Boutique de mode)
	Faible	Mesure des comportements (Usine de boîtes de conserves)	Rituel et cérémonial, contrôle par le « clan » (Laboratoire de recherche)

Hofstede (1981) propose un modèle plus complexe, spécifiquement adapté au contrôle de gestion des activités non marchandes (mais qui a également une validité pour les autres types d'activité). Il définit six modes de contrôle : contrôle routinier, contrôle par experts, contrôle par essai et par erreur, contrôle intuitif, contrôle par jugement et contrôle politique. Il propose de définir le mode de contrôle le plus approprié à partir de la réponse à quatre questions : les objectifs de l'activité sont-ils ou non ambigus, les extrants sont-ils ou non mesurables, les effets des interventions du management sont-ils ou non connus, l'activité est-elle répétitive ou non ? On notera que le questionnement d'Hofstede recouvre assez largement celui d'Ouchi (même s'il ne fait pas référence à ce dernier dans son article) : nous retrouvons dans les deux cas la mesure des extrants ; quant au caractère plus ou moins répétitif de l'activité et à la connaissance des effets des interventions du management, ils renvoient à la connaissance du

⁵ Sauf indication contraire, les traductions de l'anglais sont le fait des auteurs de cet article.

processus de transformation. Hofstede ajoute cependant un nouvel élément : le degré d'ambiguïté sur les objectifs. La réponse d'Hofstede ne se présente plus comme chez Ouchi dans une matrice 2x2 mais sous forme d'un arbre de décision, reproduit ci-après.

Figure 1
Conditions déterminant le choix d'un mode de contrôle
Source : Hofstede (1981, p. 196)



Pour Hofstede, les types 1 (contrôle routinier), 2 (contrôle par expert) et 3 (contrôle par essai et erreur) renvoient au modèle cybernétique de contrôle de gestion dont relève le pilotage par les indicateurs. Les types 4 (contrôle intuitif), 5 (contrôle par jugement) et 6 (contrôle politique) renvoient à un modèle non cybernétique. Ce qui fait que l'un ou l'autre modèle est applicable dépend du caractère mesurable des extrants : le modèle cybernétique dont relève le pilotage par les indicateurs ne peut être utilisé que lorsqu'il est possible de trouver une mesure des extrants ou un substitut acceptable à cette mesure. Dans le contrôle par jugement, selon la structure d'influence et de pouvoir, il pourra y avoir ou non un juge suprême (ou une coalition de juges) dont le jugement constituera la base de l'intervention, les jugements pourront être ou non négociés avant que l'intervention ne devienne possible, un jugement pourra se révéler impossible si bien que le contrôle n'arrivera que par accident ou pas du tout. Dans le contrôle politique, il existe cinq possibilités de passer outre l'ambiguïté des objectifs : l'imposition soit d'objectifs, soit de règles ou de politiques par la hiérarchie, la négociation des objectifs, le recours à des experts ou la crise comme élément fédérateur. Ce dernier nous indique que dans ce cas le facteur crucial est la perception des experts qu'ont les managers : « non s'il sait réellement, mais s'il apparaît suffisamment crédible aux yeux de l'organisation en tant que personne susceptible de résoudre l'ambiguïté ». Burchell et al. (1980) présentent comme Ouchi un modèle à quatre cases sur les usages de la comptabilité et comme Hofstede intègrent une dimension politique. Nous reprenons ci-après la présentation qu'en font Biondi et al. (2008, p. 115).

Tableau 2
Conditions déterminant les usages de la comptabilité
Source : Burchell et al. (1980, adapté par Biondi et al., 2008)

		Dimension politique : incertitude sur les objectifs	
		Faible	Forte
Dimension organisationnelle : incertitude sur les relations entre actions et résultats	Faible	Cybernétique (A)	Luttes de pouvoir (C)
	Forte	Apprentissage (B)	Rhétoriques de rationalisation et de démocratisation (D)

« Une vision cybernétique (case A) n'est envisageable que lorsque les objectifs politiques à atteindre et les activités pour les accomplir sont simples. Lorsque la mise en œuvre des politiques publiques est complexe et incertaine, les instruments comptables et financiers peuvent être utilisés pour apprendre, c'est-à-dire pour se poser des questions, mais ils ne peuvent être utilisés de manière mécanique pour atteindre les objectifs préfixés (case B). En outre, lorsque les objectifs mêmes de ces politiques sont complexes et incertains, les représentations comptables et financières peuvent être utilisées par chaque acteur pour favoriser ses intérêts (case C) ou dans une optique de justification rhétorique (au nom de la rationalisation, ou de la démocratisation) des décisions prises (case D), ce qui peut masquer l'impact véritable des politiques adoptées. »

Au terme de cette brève analyse, plusieurs conclusions majeures se dessinent : le pilotage par les indicateurs (contrôle cybernétique ou contrôle par les extrants) n'a de pertinence que lorsqu'il est possible de mesurer avec une certaine précision et une certaine complétude les

extrants, que la relation entre intrants et extrants est connue et qu'enfin les objectifs sont simples.

1.3. Les conséquences d'une erreur sur le choix du mode de contrôle

Hofstede (1981, p. 207) identifie deux types d'erreur. L'erreur de type 1 consiste à ne pas utiliser un modèle cybernétique alors que la situation remplit les conditions pour ce faire ; l'erreur de type 2 consiste à essayer d'utiliser un modèle cybernétique alors que les conditions ne s'y prêtent pas. Tandis que les erreurs de type 1 sont relativement fréquentes, mais assez bénignes parce qu'elles ne concernent qu'une partie limitée de l'organisation, les erreurs de type 2 sont plus rares, mais d'une plus grande gravité. « Elles se produisent lorsque sont introduites à large échelle et de manière radicale des techniques destinées à améliorer les processus de contrôle de gestion par jugement ou de type politique, mais qui relèvent de fait après plus ample examen de programmes du paradigme cybernétique ». Il semblerait que, dans les activités de recherche, comme nous allons tenter de le démontrer ci-après, l'erreur de type 2 devient le lot commun.

Ouchi (1979), de son côté, souligne que dans les situations d'incertitude, une mesure fiable et précise est impossible et qu'un système fondé sur de telles mesures « très probablement récompensera un éventail étroit de comportements faiblement adaptatifs, conduisant finalement au déclin de l'organisation ». Comme nous allons le voir, l'utilisation pertinente des indicateurs (outil apparemment aisé à mettre en place et d'une grande souplesse) impose des conditions d'emploi relativement rigoureuses.

2. Eléments de repère sur la systémique des indicateurs

L'indicateur procède d'une systémique qui semble extrêmement mal connue des décideurs publics, à moins qu'elle ne soit délibérément occultée.

On retrouve ici un des phénomènes observé par Démil et al. (2007) dans l'articulation entre pratique et recherche: « les lecteurs de théories ont tendance à les lire rapidement et à en tirer des idées simples et réduites ». Ils ajoutent en s'appuyant sur D'Andrade (1995) : « dans de nombreux cas, ces lecteurs simplifient ou interprètent les idées jusqu'à ce qu'elles correspondent à des schémas cognitifs préexistants ». Ils citent également Beyer (1997, p. 18) : « les managers adoptent nos idées et nos résultats en sélectionnant ceux qu'ils estiment pertinents en fonction de leurs objectifs et des circonstances, ce qui les amène à redéfinir, modifier, combiner et réinterpréter ces résultats ». Il semble que dans la mise en œuvre des principes du NPM, les décideurs publics se soient saisis d'une sorte de vulgate, d'une vision très appauvrie des théories du contrôle de gestion qui ignore toutes les réflexions menées sur les systèmes de mesure de performance et plus spécifiquement les indicateurs. Mais, comme Démil et al. (2007, p. 37) nous le rappellent, le praticien (en l'occurrence le décideur public) est susceptible de mobiliser des théories non seulement pour agir ou pour penser, mais aussi pour légitimer son action.

Le terme systémique renvoie au fait que les indicateurs sont construits pour appréhender des phénomènes complexes et forment eux-mêmes système. Nous présentons ci-après les principaux aspects de cette systémique : inscription dans un système d'acteurs et inscription dans un processus de transformation.

2.1. L'inscription des indicateurs dans un système d'acteurs

L'indicateur, comme tout instrument de gestion, s'inscrit dans un système d'acteurs. Il est en principe la traduction d'une finalité, d'une politique, portées par des acteurs. Mal conçu, il peut donner lieu à des manipulations et à des comportements opportunistes, mais aussi à des interprétations erronées ou difficiles. Nous ne mentionnerons ici que les éléments qui nous seront utiles dans notre étude empirique.

2.1.1. La dimension politique de l'indicateur

C'est ici la question de la pertinence de l'indicateur qui est posée (Bessire, 1999). Celle-ci ne peut être appréciée que par rapport à la finalité poursuivie et à sa déclinaison en termes d'objectifs (ce que Hofstede (1981) a déjà fort justement souligné, au contraire d'Ouchi (1979) qui fait l'impasse sur cette dimension). Dans les organisations publiques, les finalités sont souvent floues et même lorsqu'elles sont explicitées, elles peuvent ne pas être considérées comme pertinentes par certains acteurs. Les objectifs peuvent être difficiles à définir ; ils peuvent être multiples, voire contradictoires. C'est dans le choix des indicateurs d'impact (voir infra) que cette dimension intrinsèquement politique de l'indicateur se manifeste avec le plus de force.

2.1.2. Les risques de manipulation et de comportements opportunistes

Quelle que soit la qualité d'un indicateur, il ne saisit jamais – et d'une manière imparfaite - qu'un nombre limité de facettes du phénomène considéré. Le risque est toujours présent d'une focalisation sur certaines dimensions de la performance au détriment d'autres dimensions, avec comme corollaire le risque de comportements opportunistes et la tentation de manipuler les indicateurs, surtout si leur est associé un système d'incitations financières ou autres et si le système d'activités est complexe. On retrouve ici un problème d'incitations classique dit de « multitâches » analysé par Holstrom et Milgrom (1991). Quelques exemples illustreront ces risques ; quelques précautions permettraient de les réduire.

Quelques exemples de risques encourus dans le pilotage par les indicateurs

Le risque de manipulation des indicateurs d'impact est relativement faible ; en revanche il est élevé pour les indicateurs portant sur les intrants, les moyens, l'activité et les extrants⁶.

La SNCF utilise l'indicateur « pourcentage de trains partant à l'heure » ; pour maintenir ce ratio à un niveau satisfaisant, en cas de retard d'un train, elle évite de retarder les trains assurant à la correspondance, sans égard pour les passagers du train en retard. Au final, la

⁶ Une définition de ces différents types d'indicateurs sera proposée au paragraphe suivant.

qualité globale du service n'est pas améliorée. L'indicateur le plus pertinent serait le nombre de voyageurs arrivés à l'heure, mais il est évidemment plus difficile à calculer.

Dans une zone donnée, le nombre de plaintes déposées auprès d'un commissariat de police peut diminuer même si la criminalité augmente, car les policiers peuvent être tentés, s'ils sont évalués sur cet indicateur, de décourager le dépôt de plaintes (temps d'attente excessif, encouragement à déposer une simple « main-courante » non prise en compte dans les statistiques...).

Quelques précautions à respecter

Les précautions principales à prendre portent sur l'utilisation d'une batterie cohérente d'indicateurs, une combinaison appropriée d'indicateurs individuels et collectifs et la forme sous laquelle l'indicateur se présente. Elles sont a priori connues de tout enseignant en contrôle de gestion normalement compétent, mais semblent en revanche méconnues des décideurs publics.

La focalisation sur un indicateur unique ou un nombre trop réduit d'indicateurs tend à générer des comportements opportunistes. L'utilisation d'une batterie cohérente d'indicateurs permet de façon générale de grandement limiter ces risques. En particulier, les indicateurs quantitatifs devraient être chaque fois que possible couplés avec des indicateurs qualitatifs. Dans les exemples mentionnés plus haut, le pourcentage de trains partant à l'heure devrait être mis en regard du taux de satisfaction des voyageurs ; le nombre de plaintes déposées devrait être analysé à la lumière du taux de satisfaction des personnes s'étant présentées au commissariat. Sans prise en compte de la dimension qualitative, le jugement peut être gravement faussé : le nombre de vols dans les transports en commun d'Ile-de-France a ainsi légèrement diminué au cours des onze premiers mois de l'année 2010, mais les vols avec violence ont augmenté de 40% ; ne pas prendre en compte le degré de gravité du délit donnerait alors une image faussée de la situation.

Il convient également de trouver une combinaison appropriée d'indicateurs individuels et d'indicateurs collectifs. La seule utilisation d'indicateurs collectifs peut provoquer des phénomènes de passager clandestin (à notre avis surestimés, compte tenu par ailleurs de l'existence de pressions normatives propres à tout groupe humain). Le ciblage sur des indicateurs individuels à l'inverse casse la communauté de travail et a des effets délétères sur la qualité des services rendus aux usagers. Ainsi si un employé est absent, il n'y a aucun intérêt pour ses collègues à assumer une partie de sa charge de travail.

2.1.3. Des indicateurs à interpréter dans toute leur complexité

L'indicateur doit avoir un mode de calcul relativement stable dans le temps et l'espace. Une absence de stabilité empêche ou fausse la comparaison et donc ne permet pas une évaluation correcte de l'évolution du phénomène mesuré. Mais symétriquement un indicateur figé, face à une transformation importante du contexte, donne une vision biaisée du phénomène observé. Notons que des changements dans l'espace ou le temps peuvent correspondre à des manipulations, destinées à améliorer les statistiques. La variation du mode de décompte des chômeurs et les débats qu'elle suscite illustre ces difficultés.

La forme sous laquelle se présentent les indicateurs mérite également une attention particulière. Les indicateurs peuvent d'abord être exprimés en termes d'unités physiques ou d'unités monétaires. Si les ressources par exemple sont homogènes, peu importe qu'elles soient appréhendées au travers d'indicateurs physiques ou d'indicateurs monétaires ; en revanche, si leur coût relatif diffère sensiblement, l'utilisation d'un indicateur physique synthétique masquera une partie du phénomène.

Les indicateurs peuvent se présenter sous une forme assimilable à une variable continue (nombre d'étudiants inscrits, nombre de chômeurs inscrits...), d'autres sous forme d'une variable discrète prenant un nombre limité de valeurs (échelle de A à C, de 0 à 5) ; dans le cas le plus extrême la variable est binaire (oui/non, 0/1). Les phénomènes organisationnels étant généralement d'une grande complexité, l'utilisation d'indicateurs binaires devrait être limitée et les échelles utilisées dans le cas de variables discrètes présenter un nombre de points raisonnables.

2.2. L'inscription de l'indicateur dans un processus de transformation

Si l'on admet que les indicateurs ne font sens que pris dans un système, il convient alors de les organiser de façon appropriée, en fonction d'un modèle. Le modèle que nous proposons ci-après inclut, au-delà des classiques indicateurs de moyens, d'activité et de résultats, des indicateurs sur les intrants et des indicateurs d'impact. C'est sur ce dernier type d'indicateur que nous concentrerons notre analyse.

2.2.1. Une modélisation pour appréhender les différents types d'indicateurs

Nous proposons de nous appuyer ci-après sur un modèle, largement inspiré de Malleret (2004), qui prend en compte les différentes étapes du processus de transformation afin de mettre en évidence les différents types d'indicateurs susceptibles d'être utilisés et d'alerter sur les difficultés liées à leur calcul et à leur utilisation.

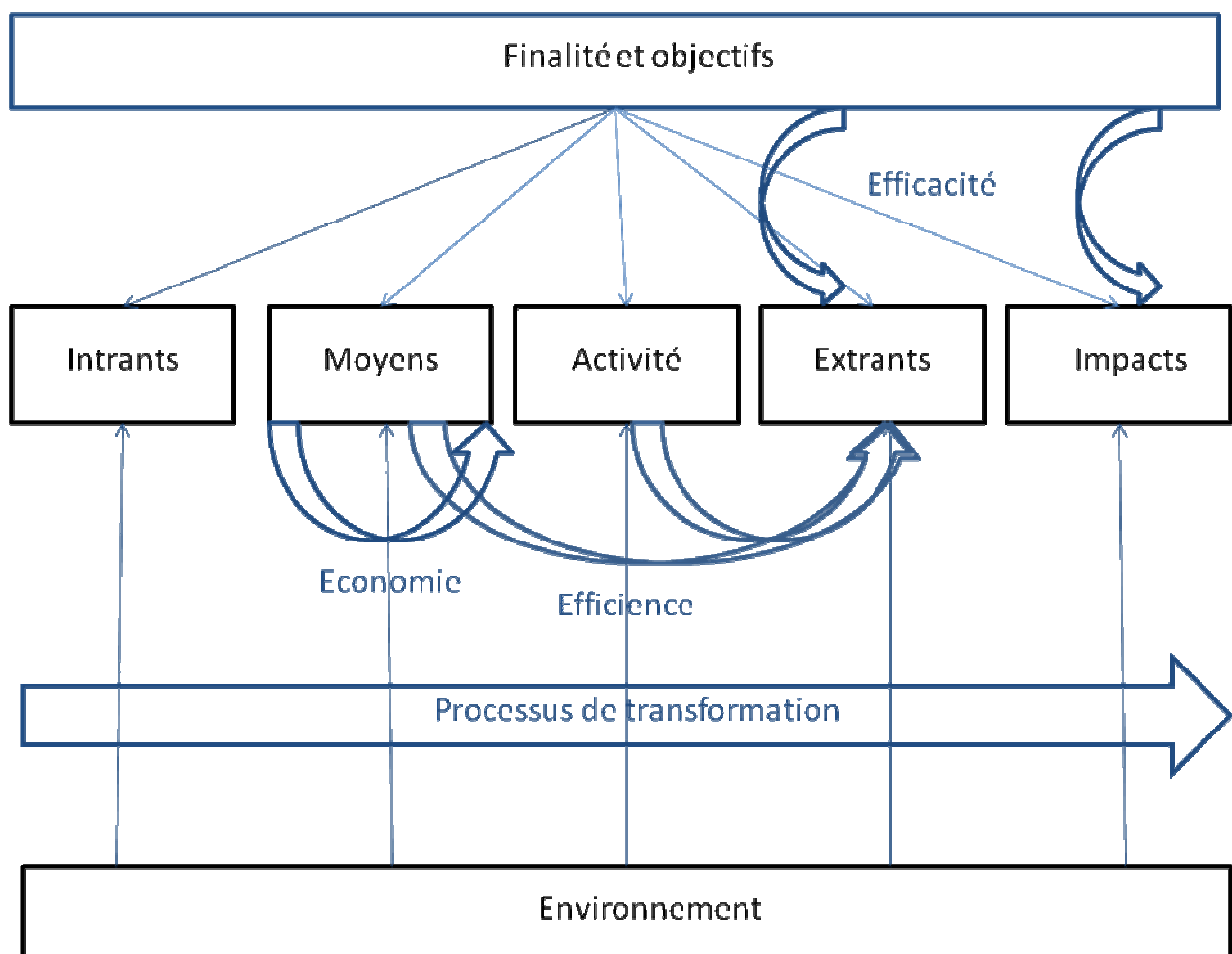
Notons en premier lieu que les finalités, autrement dit la politique poursuivie, influencent tous les maillons du processus de transformation. Elles se traduisent sous formes d'objectifs en termes d'extrants et d'impacts ; elles conditionnent le choix des intrants, des moyens utilisés et de leur mode de mise en œuvre. Notons également que l'environnement est susceptible d'affecter chacun des maillons du processus de transformation.

Pour donner une traduction concrète à ce schéma, prenons l'exemple d'un service départemental d'incendie et de secours (les pompiers). A côté d'autres missions moins connues, leur objectif majeur est la sauvegarde des biens et des personnes. Les intrants sont les demandes d'intervention qui peuvent varier en volume mais également en composition (fausses alertes et interventions mineures du type « nid de frelon » ou « chat dans un arbre » coexistant avec de véritables urgences), ceci pouvant profondément altérer la réalisation de leurs missions ; les moyens (ou ressources allouées) sont les effectifs de pompiers ainsi que les matériels spécialisés utilisés lors des interventions ; l'activité se traduit par un nombre d'interventions de nature variée (par exemple de type « lutte contre l'incendie »), les extrants expriment alors le résultat de l'action sur l'environnement (par exemple, le nombre

d'incendies éteints ou le nombre de personnes secourues). Les impacts sont plus diffus et moins aisés à saisir et représentent la réaction de l'environnement à l'action publique. Par exemple l'impact de la lutte contre les incendies de forêt peut être apprécié en examinant l'évolution du pourcentage de forêts détruites sur plusieurs années. Quant à l'environnement, il peut produire des effets favorables ou défavorables (par exemple la situation météorologique).

Ces indicateurs peuvent être complétés par les traditionnels indicateurs d'économie, d'efficience et d'efficacité. En termes de contrôle de gestion, l'économie renvoie à l'acquisition de ressources au prix le plus juste ; l'efficience (appelée aussi productivité ou rendement) fait référence au rapport entre d'une part les moyens mis en œuvre – les ressources consommées – et d'autre part l'activité ou les résultats mesurés sous leurs différentes formes (un nombre de vêtements produits par personne, le nombre d'heures pour fabriquer un vêtement...); l'efficacité enfin rapporte les résultats obtenus et/ou impacts produits aux objectifs initiaux qui sont eux même dépendants des finalités poursuivies.

Figure 2
Modélisation du processus de transformation et indicateurs



Note : chaque rectangle noir correspond à un type spécifique d'indicateurs.

2.2.2. Les indicateurs d'impact : des difficultés spécifiques

La plupart des indicateurs sont bien connus et largement utilisés par les décideurs publics. Cependant les indicateurs d'impact sont presque totalement absents du discours public qui, dans l'esprit du NPM, valorise le passage « d'une logique de moyens à une logique de résultats ». Or l'objectif ultime de l'action publique n'est pas un résultat (*output*), mais un impact (*outcome*), c'est-à-dire la création d'une externalité positive. Le nombre de kilomètres d'autoroute construits ne dit pas grand-chose de la performance d'une politique publique ; ce qui importe, c'est de savoir si cette nouvelle voie a permis de désenclaver une région et de stimuler son dynamisme économique, si elle a permis d'améliorer la sécurité routière. Les impacts négatifs (effets collatéraux) méritent également d'être appréhendés : la nouvelle autoroute et l'augmentation du trafic qu'elle a générée ont-ils contribué à une augmentation des gaz à effet de serre, des espèces rares ont-elles été fragilisées ? Augmenter le pourcentage d'une tranche d'âge ayant obtenu le baccalauréat est peut-être un objectif louable, mais ce taux ne nous renseigne guère sur l'évolution des compétences de la population en général et sa capacité à trouver un emploi à la mesure de ses ambitions. En se concentrant sur les résultats et non sur les impacts, le discours public se trompe donc de cible.

Il est vrai que les indicateurs d'impact, au contraire des autres indicateurs, sont particulièrement difficiles à calculer car ils mesurent des phénomènes dont la dimension qualitative est forte et qui ne prennent leur sens que sur le long terme. Une des particularités des organisations publiques est que généralement l'impact ne se fait sentir qu'à long terme (c'est d'ailleurs ce qui justifie que les élus le soient pour plusieurs années).

L'impact est également multiforme et n'est pas apprécié de la même façon par les différentes parties prenantes. Là où la construction d'une centrale nucléaire sera vue par l'Etat comme un moyen de limiter la dépendance énergétique, par une collectivité territoriale comme une opportunité pour la création de nouveaux emplois, elle sera perçue par les défenseurs de l'environnement comme un risque inacceptable et par les riverains comme une menace sur leur santé. L'impact renvoie à des finalités, donc à des choix politiques nécessairement conflictuels.

3. Des errements d'un pilotage par les indicateurs : l'exemple de la recherche en sciences de gestion

Avant d'analyser en détail les indicateurs mis en œuvre ou susceptibles d'être mis en œuvre dans le pilotage des activités de recherche, nous présentons ci-après nos choix méthodologiques.

3.1. Choix méthodologiques

La recherche en sciences de gestion offre aux chercheurs en contrôle de gestion un accès au « terrain » privilégié. On aurait donc pu penser que ceux-ci auraient saisi l'opportunité qui

leur était offerte d'analyser de manière concrète un phénomène qui s'est développé de manière accélérée dans les dernières années : la production et l'utilisation d'indicateurs de gestion. Force est de reconnaître qu'il n'y a cependant guère de travaux menés dans ce domaine, même si la question des indicateurs revient inévitablement dans nombre de conversations ou sert d'illustration rapide à des enseignements en contrôle de gestion. Un travail plus systématique nous semble nécessaire. A cette fin, nous nous appuyerons sur quelques travaux antérieurs menés dans ce champ, mais surtout sur notre position privilégiée d'observateurs participants, impliqués fortement dans l'animation de la recherche, la direction de diplômés, l'audit de formations et la production scientifique⁷. Comme nous l'avons dit plus haut, nous restreindrons notre champ d'analyse aux sciences de gestion, même si par endroits il sera nécessaire de faire référence à d'autres disciplines pour mettre en perspective nos propos. Par ailleurs, nombre de points de notre analyse nous semblent généralisables. Le champ des sciences de gestion se caractérise enfin par la coexistence de deux types de structures : des départements universitaires et des écoles de commerce ; nous concentrerons notre analyse sur les premiers, mais ceux-ci entrant en concurrence avec les écoles de commerce, il sera aussi nécessaire à certains moments d'inclure ces dernières dans notre champ d'analyse.

Au sein de l'Université, comme dans toute organisation, les entités à mettre sous contrôle se situent à différents niveaux : au niveau individuel (l'enseignant-chercheur), à un niveau intermédiaire (le laboratoire, l'école doctorale), à un niveau organisationnel (l'Université prise dans son ensemble), à un niveau inter-organisationnel (une fédération, l'ensemble des Universités, l'ensemble des laboratoires d'une même discipline ou d'un sous-ensemble de disciplines...). Nous avons choisi dans notre étude de privilégier en termes de niveau d'analyse un niveau intermédiaire, celui du laboratoire parce que c'est le niveau qui correspond aujourd'hui à la maille centrale du dispositif d'évaluation et de contractualisation. Nous verrons cependant qu'en certains endroits, il existe une certaine intrication des niveaux et qu'il est également nécessaire de raisonner au niveau du chercheur.

Nous raisonnerons sur les indicateurs les plus usuels, ceux qui notamment apparaissent dans les documents de contractualisation demandés par l'AERES. Nous montrerons, à la lumière du modèle présenté plus haut, que faute d'une réflexion, d'une part sur les modes de contrôle, d'autre part sur la systémique des indicateurs, le pilotage par les indicateurs tel qu'il est pratiqué aujourd'hui souffre de sérieuses déficiences.

3.2. La mise en œuvre forcée d'un contrôle cybernétique

Si l'on suit les raisonnements d'Ouchi, d'Hofstede ou de Burchell et al., les activités de recherche ne peuvent a priori relever d'un pilotage de type cybernétique. Aucune des conditions identifiées – pouvoir mesurer avec une certaine précision et une certaine complétude les extrants, identifier une relation entre intrants et extrants, définir des objectifs simples – ne semble remplie pour la recherche. Un contrôle par les extrants, de type cybernétique, et donc un pilotage par les indicateurs semble donc *a priori* inapproprié ou tout au moins insuffisant.

⁷ Ce travail a un caractère exploratoire. Nous envisageons dans un second temps d'interroger des membres des organismes de tutelle et d'évaluation.

En effet, au-delà de la définition d'une finalité en termes généraux et relativement consensuels, la définition des objectifs est particulièrement ardue. Le processus de transformation est multiforme (quoi de commun, pour ne prendre qu'un exemple de cette hétérogénéité, entre la recherche qui se fonde sur l'utilisation de bases de données et celle qui implique une plus ou moins longue immersion dans des organisations ?) ; il reste par ailleurs impossible à décrire de manière suffisamment fine et complète. L'activité et les extrants, comme nous le verrons ci-après, sont d'une grande hétérogénéité, ce qui rend toute mesure agrégée problématique. Si l'on suit Ouchi, un contrôle par le clan semble le mieux adapté à ce type d'activités. Cet auteur prend d'ailleurs comme nous l'avons vu le cas d'un laboratoire de recherche pour illustrer ce mode de contrôle.

Jusqu'à une époque relativement récente, l'évaluation des chercheurs relevait principalement du Conseil national des universités et le contrôle externe sur les laboratoires était relativement faible. Le pilotage relevait donc principalement d'une combinaison de contrôle politique et de contrôle par jugement (au sens d'Hofstede). La création de l'Agence d'évaluation de la recherche et de l'enseignement supérieur (AERES)⁸, a profondément bouleversé la situation. On peut analyser cette création comme l'utilisation d'un contrôle par expert ; nous verrons ci-après comment la référence au classement de Shanghai a permis de lever dans une large mesure l'ambiguïté des objectifs propres aux activités de recherche. L'introduction par l'AERES de la notion de « chercheur publiant », devenu ultérieurement « chercheur produisant », a eu pour effet de faire évoluer le contrôle vers un contrôle de plus en plus routinier. Naturellement, il s'agit de tendances générales et les différents types de contrôle continuent à coexister. Cependant, il est indéniable que le pilotage des activités de recherche, autrefois dominé par un modèle non-cybernétique, est aujourd'hui largement soumis à un modèle cybernétique.

Ce glissement a été rendu possible par deux étapes successives de « réduction » : sur la réduction de l'ambiguïté sur les objectifs permise notamment par la référence au classement de Shanghai, s'est articulé le désencastrement au sein d'un système d'activités complexes de l'activité de publication dans des revues scientifiques.

3.2.1. Réduction arbitraire de l'ambiguïté sur les objectifs : la référence au classement de Shanghai

La mission principale des chercheurs et des équipes de recherche est de produire et diffuser de la recherche qui contribue à l'amélioration du bien-être collectif. Au-delà de cette finalité relativement consensuelle, la définition d'une politique de recherche et sa déclinaison en objectifs pose de redoutables défis. Faut-il privilégier certaines disciplines au détriment d'autres ? Faut-il concentrer la recherche dans un nombre limité de grands établissements⁹ ? Faut-il l'encadrer par des programmes plus ou moins contraignants ? Ce ne sont là que quelques questions parmi d'autres.

Les choix qu'entraînerait une réponse explicite à ces questions sont politiquement périlleux. C'est pourquoi jusqu'à une époque récente, la politique de recherche, tout au moins française, restait relativement vague. Le fameux classement de Shanghai a fourni aux politiques un alibi

⁸ La création d'agences est typiquement un produit du NPM.

⁹ Dans le domaine de la santé, il apparaît régulièrement que les petits laboratoires sont plus innovants que les grands laboratoires.

idéal pour mener les réformes exigées par le NPM, sans être obligés d'afficher de manière explicite des objectifs potentiellement conflictuels. Rappelons que le classement de Shanghai repose sur six indicateurs (les quatre premiers ont un poids respectif de 20%, les deux derniers un poids respectif de 10%) : le nombre des membres du corps universitaire ayant reçu un Nobel ou une médaille Fields, le nombre de chercheurs de l'établissement parmi la liste des « plus cités » de Thomson Reuters, le nombre d'articles de l'établissement publiés dans les revues *Nature* et *Science*, le nombre total d'articles recensés dans le Web of Science de Thomson Reuters (*Science citation index* d'une part et *Arts & Humanities citation index* d'autre part), le nombre d'anciens étudiants ayant reçu un Nobel ou une médaille Fields, et enfin l'ajustement des résultats précédents par la taille de l'établissement (la somme pondérée des cinq indicateurs est divisée par le nombre de chercheurs équivalent temps plein).

Il existe une littérature scientifique abondante qui critique ce classement (voir par exemple Gingras, sd). Néanmoins, il a une influence directe sur (ou permet de justifier) la politique actuelle de recherche en France. En 2007 devant le Sénat, Valérie Pécresse, Ministre de l'Enseignement supérieur et de la Recherche, admettait « que le classement de Shanghai était certes critiquable mais que puisqu'on ne pouvait changer les indicateurs dont nous n'étions pas maîtres, il valait mieux les retourner en notre faveur. [...] C'est la mondialisation. On ne peut s'en abstraire et nous devons donc gagner des places, ce qui n'est pas contraire à l'exigence d'excellence de l'université française ». Le classement de Shanghai est d'ailleurs le premier argument avancé pour la constitution des Pôles de Recherche et d'Enseignement Supérieur (PRES). Ainsi en 2008, Valérie Pécresse déclarait que « les résultats pour la France du classement de Shanghai [...] plaident pour un regroupement de nos forces »¹⁰. On notera que la référence au poids de la recherche et développement dans le produit intérieur brut, qui mesure l'effort de recherche, n'est plus guère présente dans les discours politiques.

3.2.2. « Désencastrement » de l'activité de publication dans des revues classées

Les politiques ministérielles n'ont qu'une influence limitée sur le nombre de prix Nobel ou de médailles Field (d'autant plus que ce nombre est pris en compte, certes avec des pondérations décroissantes dans le temps, depuis 1901 pour les étudiants et 1911 pour les membres de l'établissement) ; accroître rapidement le nombre de publications dans les revues *Nature* et *Science* est un objectif difficile à atteindre. Il est donc assez logique qu'une fois admis l'objectif de progression dans le classement de Shanghai, les efforts soient orientés vers l'augmentation du nombre de chercheurs d'un établissement donné parmi la liste des « plus cités » de Thomson Reuters et vers le nombre total d'articles recensés dans le Web of Science de Thomson Reuters.

Dans la conception des dossiers de contractualisation des équipes de recherche et l'évaluation de ces mêmes équipes par l'AERES, ces priorités se traduisent par l'accent mis sur un indicateur, « une estimation de la proportion des chercheurs et enseignants-chercheurs « produisant en recherche et valorisation » (AERES, 2009, p. 1). Pour être considéré comme enseignant-chercheur « produisant », il faut avoir deux publications de rang A en quatre ans, mais « une publication dans *Nature*, *Cell* ou *Science* va suffire à cataloguer le chercheur ou enseignant-chercheur comme « produisant en recherche et valorisation » (id, p. 3.). Rappelons

¹⁰ Les Cahiers de la compétitivité, spécial enseignement supérieur, p II, encart dans *Le Monde* du 21 mai 2008.

que le nombre d'articles publiés dans *Nature* et *Science* compte pour 20% dans le classement de Shangai ; les raisons de l'inclusion de *Cell* dans ce trio nous restent inconnues (lobbying scientifique ?).

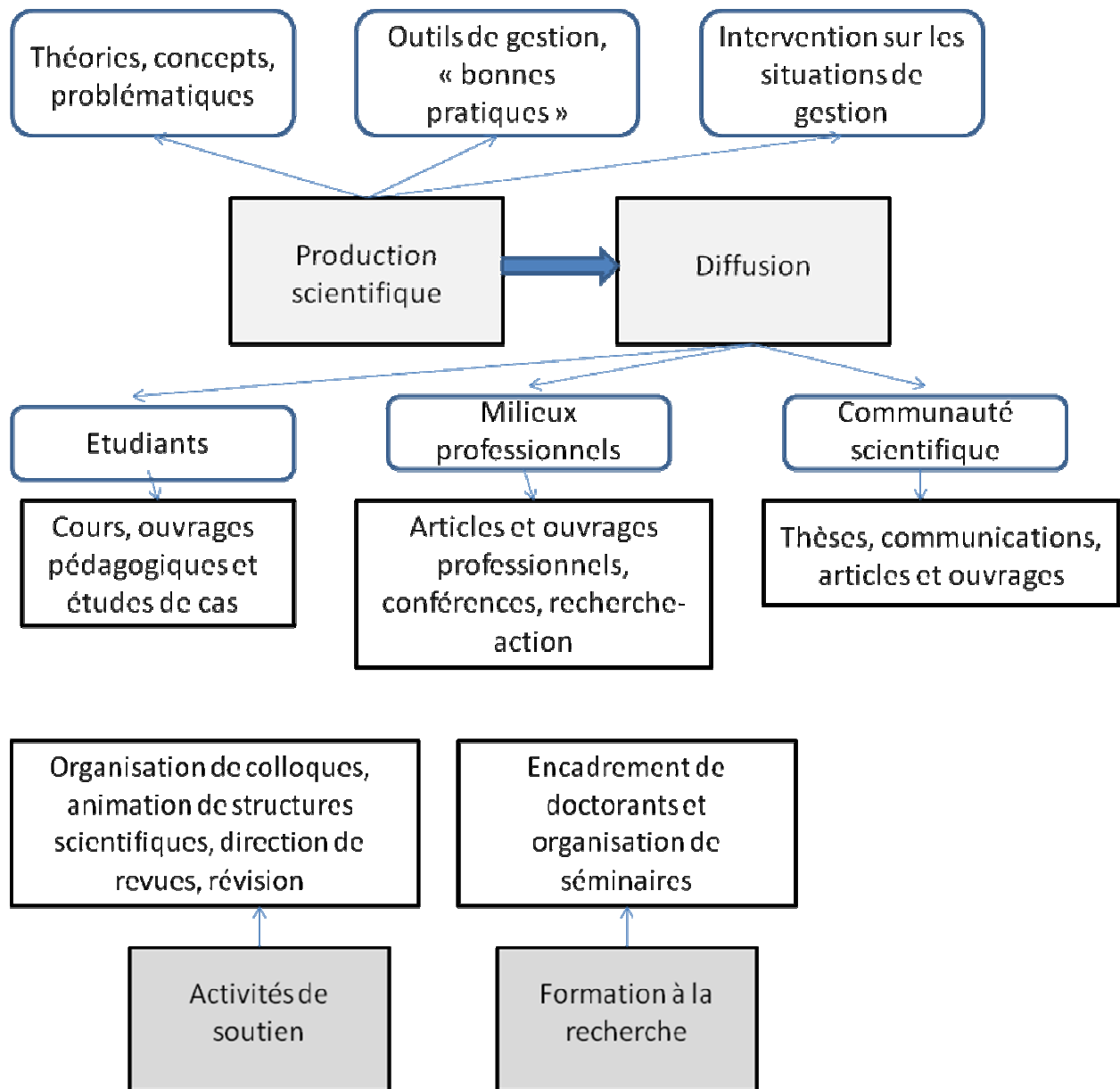
Une publication de rang A est définie, tous secteurs confondus, « comme un article dans une revue à comité de lecture [figurant sur une liste] « établie avec le concours de la communauté scientifique » (id., p. 3). La définition exclut la participation à des colloques (« la très grande variabilité qualitative de l'évaluation des actes des colloques ne permet pas de les retenir comme élément pertinent d'appréciation », id., p. 1). Dans les « remarques permettant de tenir compte du domaine de recherche, il est indiqué en revanche que dans le secteur SHS, « un bon ouvrage scientifique de recherche ou d'érudition est une publication de rang A (id., p.3). Il apparaît d'ailleurs ici une contradiction puisqu'il est également mentionné pour ce secteur qu'« une communication à un colloque ou à un congrès ne peut être comptée deux fois si elle est publiée par ailleurs » (ibid.), ce qui laisserait entendre qu'une production de ce type peut être prise en compte alors que l'inverse est affirmé p. 2. Cependant en sciences de gestion, le Professeur Patrick Rousseau, délégué scientifique à l'AERES en charge des sciences économiques et de gestion, nous a indiqué oralement, lors de la visite de notre laboratoire, que cette définition plus large ne s'appliquait pas à l'économie et à la gestion et qu'en conséquence le nombre de chercheurs « produisant » était déterminé uniquement en relation avec le nombre de publications dans des revues classées.

En mettant l'accent sur la publication dans des revues classées, le processus d'évaluation isole et « désencastre » cette activité du système complexe des activités de recherche que nous avons tenté de représenter dans le schéma ci-après (figure 3).

L'examen de ce schéma permet de dresser un certain nombre de constats : le discours officiel confond production scientifique et diffusion ; la diffusion elle-même prend des formes très diverses qui ne se réduisent évidemment pas à la publication dans des revues classées et ne se limitent pas au public constitué par les chercheurs et enseignants-chercheurs ; l'activité centrale de production scientifique ne peut se dérouler et se pérenniser indépendamment de toute une gamme d'activités qui présentent deux particularités : elles ne peuvent être correctement assurées que par des chercheurs ou sous leur supervision étroite ; elles sont étroitement interdépendantes (si plus aucun chercheur n'accepte de réviser d'article, l'existence de revues classées, dans leur modèle actuel, est menacée).

Il est par ailleurs étonnant de voir un système d'activités aussi complexe réduit à un indicateur binaire : chercheur « produisant » ou chercheur « non produisant ». Il y a là, nous semble-t-il, un effet désincitatif sur le chercheur débutant ou celui qui voudrait se remettre à la recherche après une période d'inactivité scientifique (due par exemple à l'acceptation de responsabilités administratives importantes). Une exigence minimale serait d'avoir une gradation dans l'évaluation de l'activité scientifique, par exemple sur une échelle à 5 points en fonction du degré d'implication dans la recherche appréciée de manière globale et contradictoire par au moins deux personnes de la discipline à partir de fiches d'activité détaillées ; il s'agit ici de prendre en compte toutes les dimensions de la recherche, ainsi d'ailleurs que tente de le faire le CNU lorsqu'il se prononce sur des promotions ou des primes d'excellence scientifique.

Figure 3
La recherche, un système d'activités complexe



La systématisation de l'évaluation des chercheurs par le CNU qui démarrera en 2012, si elle est bien conduite, pourrait faciliter l'exercice. C'est un indicateur certes imparfait, mais il a le mérite de prendre en compte les différentes facettes de l'activité des membres de l'équipe de recherche ; il est de toute façon préférable à l'indicateur binaire nombre de « producteurs ».

Certes, le dossier de contractualisation de l'AERES comprend toute une série d'indicateurs de résultats. En dehors du nombre d'articles publiés dans des revues classées, il faut fournir la liste des articles publiés dans d'autres revues, des communications dans des colloques, d'ouvrages et chapitres d'ouvrages, et de cahiers de recherche, des contrats de recherche pilotés par le laboratoire, des colloques qu'il a organisés et des thèses encadrées en son sein

(nombre de doctorants, nombre de thèses soutenues, qualité de leur insertion), mais, dans la pratique, ces indicateurs semblent jouer un rôle très secondaire.

3.2.3. Un « désencastrement » source d'opportunisme

La série de réductions opérées permet évidemment un contrôle de type cybernétique, mais a de nombreux effets pervers.

Elle encourage des comportements opportunistes. Certains chercheurs font le choix délibéré de la publication dans des revues classées, se désintéressant de toutes les autres activités, non seulement de recherche, mais aussi d'enseignement et refusant de s'investir dans des responsabilités scientifiques ou administratives. Une charge accrue pèsera sur les épaules de leurs collègues. L'accent mis sur les publications dans des revues classées tend en outre à multiplier le nombre de co-signataires d'articles : c'est positif si c'est une incitation pour les chercheurs à renforcer la coopération, c'est négatif si ce sont des signatures de complaisance. Le seul frein à cet opportunisme effréné est pour le moment constitué par les évaluations du Conseil national des universités : en règle générale, lorsque cette instance se prononce sur des promotions, elle tend à avoir une appréciation plus globale. On retrouve ici le problème mentionné par Holstrom et Milgrom (1991) et évoqué dans la deuxième partie. Comme le souligne en effet Cohen (2009) :

« que la performance des enseignants-chercheurs soit multidimensionnelle, contrairement à celle des professeurs du secondaire, complexifie considérablement l'organisation du système d'éducation supérieure. L'enseignement ou la contribution de chaque enseignant-chercheur au « capital social » s'avèrent ainsi difficiles à mesurer et à récompenser au moyen des systèmes incitatifs simples adoptés par exemple dans le cas d'une entreprise produisant un seul bien standardisé. Si l'on récompense uniquement l'une des fonctions des enseignants-chercheurs (par exemple la recherche fondamentale), on les incite à négliger les autres ».

Enfin, dans un contexte où la proportion de chercheurs « produisant » devient un élément central de l'évaluation d'une équipe de recherche, la tentation est grande d'exclure ceux qui ne sont pas « produisant » au sens de l'AERES. Cette pratique est malheureusement de moins en moins rare. Les effets sur la cohésion des équipes sont délétères.

3.3. Une logique de gestion par les résultats trompeuse

Nous avons vu dans la première partie que la mesure de l'impact est un élément essentiel de l'évaluation des activités relevant du management public. Dans les activités de recherche, trois indicateurs sont couramment utilisés : le facteur d'impact, le nombre d'articles dans des revues classées et l'indice Harzing (appelés également PoP pour *Publish or Perish*). Le facteur d'impact est une mesure de l'importance scientifique d'une revue. Il est calculé sur une période de trois ans. Le facteur d'impact est calculé sur une période de trois ans. Soit A le nombre de citations apparaissant en 2010 sur le Web of Science de Thomson Reuters se référant à des articles publiés dans une revue X en 2008 et 2009. Soit B le nombre total d'articles publiés dans cette même revue en 2008 et 2009. Le facteur d'impact de la revue X pour 2010 est égal au rapport A/B. En sciences de gestion, cet indicateur est peu utilisé. Nous

nous concentrerons donc sur les deux autres indicateurs, même si nombre des critiques formulées valent *mutatis mutandis* pour les facteurs d'impact. Les principaux écueils portent sur la confusion entre résultat et impact, les biais liés aux citations, et le caractère autoréférentiel de ces indicateurs.

3.3.1. La confusion entre indicateur d'impact et indicateur relatifs aux extrants

Nous avons vu dans la première partie que la mesure de performance des organisations publiques devrait privilégier non des résultats, mais des impacts. Les indicateurs utilisés dans le domaine de la recherche ne répondent que très partiellement à cette exigence. Le nombre de publications est un indicateur relatif à un extrant, le classement tentant lui de mesurer, de façon très discutable, un impact.

Le classement en sciences de gestion fait principalement référence à la liste établie par la section 37 (économie et gestion) du CNRS. Il suscite bien des controverses : quels critères doivent présider à l'inclusion d'une revue dans cette liste, comment classer les revues les unes par rapport aux autres ? L'éditorial publié par Charreaux et Gervais dans la revue *Finance Contrôle Stratégie* met en évidence les distorsions entre le classement opéré par la section 37 du CNRS et d'autres classements, ainsi que les variations difficilement explicables d'une version à une autre. La publication par l'AERES en juin 2010 d'une liste de revues classées qui abandonne tout critère de hiérarchisation traduit bien la difficulté d'un classement comparatif des revues. D'autres critiques ont été émises : ce classement privilégierait l'économie au détriment des sciences de gestion, il ferait la part belle aux revues américaines et tendrait à privilégier les courants de pensée dominants au détriment de l'innovation.

Mais quel que soit le classement utilisé, la référence au nombre de publications dans des revues classées a un effet pervers immédiat : elle rend difficile le lancement de nouvelles revues. La revue *Accounting, Organizations and Society*, créée dans les années soixante-dix, qui a renouvelé de façon considérable la recherche en comptabilité contrôle et audit et qui dans la plupart des classements est notée comme excellente, n'aurait probablement jamais vu le jour¹¹. Quel chercheur aujourd'hui prend le risque de publier dans une revue non classée s'il peut le faire dans une revue classée ? A terme, cette insistance sur la publication dans des revues classées peut générer une certaine sclérose de la recherche.

L'Australie a appris à ses dépens les limites de l'indicateur « nombre d'articles publiés dans des revues classées » : le système mis en place pour inciter les chercheurs à publier dans des revues classées a produit ses effets (le nombre de publications a explosé), mais le rang du pays apprécié en termes de citations a chuté de façon spectaculaire. Les chercheurs publiaient plus, mais l'intérêt de ces publications s'était amoindri.

3.3.2. Les biais liés aux citations

Les indices PoP développés par Harzing font référence au nombre de citations sous différentes formes. Ils permettent ainsi de prendre en compte une palette plus diversifiée d'activités scientifiques. L'utilisation des indices Pop fait apparaître généralement quatre

¹¹ Son instigateur principal, Anthony Hopwood, exprimait la plus grande méfiance vis-à-vis des classements en tous genres.

catégories de chercheurs : ceux qui ont des indices proches de zéro, ceux qui ont des indices faibles, ceux qui ont des indices relativement élevés et enfin ceux dont les indices « crèvent le plafond ». La confrontation entre l'évaluation d'un chercheur par le nombre d'articles publiés dans des revues classées et celles résultant de l'utilisation des indices PoP fait apparaître une contradiction immédiate : les ouvrages, qui aujourd'hui ne sont pas pris en compte dans l'évaluation des laboratoires en sciences de gestion, peuvent donner lieu paradoxalement à des indices PoP très élevés ; c'est le cas par exemple avec l'ouvrage de Boltanski et Chiapello (1999) intitulé *Le nouvel esprit du capitalisme*.

Les indicateurs liés aux citations se prêtent moins à des comportements opportunistes que le seul nombre d'articles publiés dans des revues classées ; malgré tout, les citations de complaisance ne sont pas à exclure. Un autre problème relatif à l'utilisation d'indices de citation réside dans la définition de l'horizon temporel. Comme le soulignent Démil et al. (2007, p. 42), « puisqu'elles sont parfois produites alors qu'il n'y a pas de problèmes à résoudre, les théories n'ont pas forcément une application immédiate mais peuvent avoir un écho dans le système PR [pratique-recherche] beaucoup plus tard. De nombreux décalages dans le temps dans la diffusion des théories existent. Il faut des dizaines d'années pour que des théories portées par des auteurs comme Coase ou Penrose soient relues et intéressent à nouveau des chercheurs (respectivement, dans le cadre de la théorie des coûts de transaction et de l'approche ressources) ».

3.3.3. Des indicateurs autoréférentiels

Qu'il s'agisse de facteurs d'impact, de classement ou d'indices tels que calculés sur PoP, force est de reconnaître que ces indicateurs proposent une mesure de l'impact par rapport à la seule communauté universitaire. Il s'agit donc d'indicateurs autoréférentiels qui ignorent les autres parties prenantes que sont les décideurs d'aujourd'hui, c'est-à-dire les managers, et les décideurs de demain, c'est-à-dire les étudiants. Ces indicateurs encouragent une forme d'autisme des chercheurs qui finissent par ne plus écrire que pour d'autres chercheurs. C'est une situation paradoxale au regard du discours dominant qui stigmatise le supposé fossé entre l'Université et le « monde réel » (Démil et al. 2007). La publication d'articles dans des revues professionnelles, la rédaction d'ouvrages pédagogiques, la participation à des contrats de recherche-intervention qui se valorise plus difficilement dans des revues classées que la recherche « standard », devraient d'une manière ou une autre être valorisées.

3.4. Une faible contextualisation des indicateurs

Les indicateurs relatifs aux extrants et aux impacts s'inscrivent, nous l'avons vu dans la première partie, dans une chaîne complexe d'indicateurs. Force est de reconnaître que les autres indicateurs se voient accorder peu d'intérêt ; c'est le cas en particulier des indicateurs de moyens et d'efficience.

Le coût d'une équipe de recherche comprend des coûts individuels liés aux chercheurs (rémunérations, participation à des congrès...) et des coûts collectifs de structure (équipements, logiciels, bases de données...). En sciences de gestion, en l'absence d'équipements lourds, l'essentiel des coûts est constitué par des coûts salariaux au sens large

(rémunération de base des chercheurs, augmentée des primes¹², du financement de la participation à des colloques, de travaux de traduction et de révision...). Avant la dernière vague de contractualisation, les moyens alloués étaient principalement saisis au travers d'un indicateur non-monnaire, les effectifs, détaillés par catégorie (chercheurs, enseignants-chercheurs, ATER, PRAG et PRCE, administratifs...) et d'un indicateur monétaire, le financements reçu par le laboratoire sous forme de subventions reçues du ministère, de subventions obtenues dans le cadre de projets de recherche gérés par diverses instances (collectivités territoriales, administrations publiques, Agence nationale de la recherche et plus rarement instances européennes), de contrats de recherche-intervention passés avec différentes organisations publiques ou privées... Les nouveaux dossiers AERES exigent à présent une évaluation de la masse salariale.

Nous ignorons l'usage qui est fait de ces indicateurs. Nous pouvons cependant formuler un certain nombre d'observations : ils suscitent peu de débats au sein de la communauté des chercheurs et ne semblent pas orienter leurs comportements ; dans la préparation des dossiers de contractualisation, ces indicateurs se voient accorder peu d'attention ; enfin les comités de visite de l'AERES à notre connaissance posent rarement des questions sur ce type d'indicateurs. Tous ces éléments indiquent, tout au moins pour l'instant, un relatif désintérêt pour les indicateurs de moyens, assez paradoxal dans un processus de réforme qui prétend faire passer les organisations publiques d'une gestion par les moyens à une gestion par les résultats.

Ces indicateurs de moyens serviront-ils un jour à calculer une efficience ? Si la logique actuelle continue à prévaloir (accent mis sur le nombre d'articles publiés dans des revues classées) et que l'évaluation des équipes prend en compte leur efficience, l'incitation à exclure les chercheurs les moins « produisant » sera renforcée. En revanche, il serait possible, en utilisant l'information sur la masses salariale d'avoir une comparaison plus pertinente sur les « performances » respectives des laboratoires publics et des laboratoires privés. Les Universités recrutent en effet leurs chercheurs jusqu'à présent à un salaire imposé par le Ministère et relativement faible ; les écoles de commerce, et notamment les plus grandes, offrent des rémunérations (directes et indirectes) beaucoup plus attractives (il est aujourd'hui courant qu'un jeune docteur soit recruté en école de commerce à un salaire comparable à celui d'un professeur de 1^{ère} classe à l'Université). Les moyens financiers qu'elles mobilisent à cet effet permettent aux écoles de commerce de recruter des chercheurs anglo-saxons qui publieront plus facilement dans des revues dites internationales (en réalité pour la plupart américaines). La « qualité » des ressources (le terme qualité est pris ici dans son acception la plus générale, sans jugement de valeur) a ici un effet important sur les extrants mesurés en termes de publications dans des revues classées. Si l'on prend en compte la variable coût, rien ne nous permet en revanche de préjuger dans un sens ou dans un autre de l'efficience relative des équipes de recherche en Université et en école de commerce. En outre, rien ne garantit que l'impact sociétal des chercheurs recrutés à des niveaux de salaire plus importants soit plus élevé que celui de chercheurs moins bien rémunérés. Si l'on veut avoir une représentation convenable du paysage français en matière de recherche en sciences de gestion, les laboratoires d'écoles de commerce devraient en tout cas faire l'objet d'une évaluation, selon

¹² A l'Université : prime d'excellence scientifique, en école de commerce : primes à la publication.

les mêmes normes que les laboratoires universitaires sans omettre de prendre en compte les moyens alloués.

Conclusion

Cette communication est le fruit d'une recherche exploratoire et comme telle présente un certain nombre de limites. Outre celles que nous avons déjà mentionnées, il faut noter que nous n'avons pour le moment raisonné que dans une perspective statique, parce que la plupart des indicateurs utilisés aujourd'hui n'intègrent aucune dynamique. Il serait pourtant nécessaire de mesurer des évolutions plutôt que des situations. L'évaluation telle qu'elle est conduite aujourd'hui pour noter des équipes de recherche porte une ambiguïté essentielle : est-ce un bilan qui est évalué, la qualité d'un projet, une combinaison des deux et dans ce cas dans quelles proportions respectives ? Par ailleurs, le processus de production de la note attribuée à chaque équipe de recherche (A+, A, B, C) demeure une « boîte noire » ; il reste à compléter cette étude et à la nuancer par des entretiens avec différents acteurs de la communauté scientifique et en particulier avec des membres de l'AERES et des organismes de certification qui œuvrent dans ce domaine (Equis par exemple).

Nous espérons cependant au travers de cette communication avoir contribué à une meilleure compréhension des limites du pilotage par les indicateurs, notamment en management public et plus précisément pour les activités de recherche en sciences de gestion, une question aujourd'hui centrale. En termes théoriques, la modélisation du processus de transformation et le repérage des indicateurs qui lui sont associés nous semble être de nature à éclairer les recherches qui portent sur les indicateurs, les tableaux de bord, le contrôle de gestion et à un niveau encore plus général la gouvernance. Elle devrait également permettre aux enseignants de mieux former les décideurs du futur qui auront nécessairement à se saisir d'indicateurs. En termes managériaux, il nous semble que si les décideurs publics d'aujourd'hui pouvaient s'approprier les quelques repères simples que nous avons tenté de poser, la question de l'évaluation des organisations et des politiques publiques serait posée dans des termes plus pertinents qu'elle ne l'est aujourd'hui, quelle que soit ensuite la difficulté des réponses à y apporter : la sagesse populaire nous enseigne en effet qu'il ne peut y avoir de bonnes réponses qu'à de bonnes questions. En termes sociétaux, nous ne prétendons pas apporter de réponses définitives mais susciter un questionnement et alerter sur des dérives dangereuses du système d'évaluation de la recherche. Notons à cet égard que pilotage par les indicateurs a le caractère d'une prophétie auto-réalisatrice : là où le contrôle des activités de recherche faisait une large place à la socialisation et au contrôle par le clan, l'imposition d'un pilotage par les indicateurs oriente les comportements dans le sens d'un fonctionnement cybernétique, provoquant des comportements opportunistes et une fragmentation de la communauté scientifique.

Nous souhaitons plus généralement avec cette communication ouvrir un débat qui s'est déjà amorcé dans la société civile (voir par exemple le mouvement Appel des appels), mais qui n'a pas encore suffisamment pénétré l'univers de la recherche en sciences de gestion. Il faut noter que cette réflexion sur les indicateurs prend place dans un pays où la culture de l'évaluation publique se diffuse difficilement (Perret, 2008)). Puisque nos lecteurs potentiels font aussi partie de notre terrain, nous voudrions qu'ils soient nombreux à réagir à cette communication

(notamment en nous envoyant des commentaires par courriel) afin d'être en mesure de formuler des propositions les plus constructives possibles.

Bibliographie

- AERES (2009). Critères d'évaluation des chercheurs et enseignants-chercheurs « produisant en recherche et valorisation, disponible sur le site www.aeres-evaluation.fr, consulté le 04/04/2011.
- Bessire, D. (1999). Définir la performance. *Comptabilité Contrôle Audit*, tome. 5, vol. 2, septembre, 127-150.
- Beyer, J.M. (1997). Research utilization. Bridging a cultural gap between communities. *Journal of Management Inquiry*, vol. 6, n° 1, 17-26.
- Boltanski, L. et Chiapello, E. (1999). *Le nouvel esprit du capitalisme*. Gallimard.
- Burchell, S., Clubb, C., Hopwood, A., Hughes, J. et Nahapiet, J. (1980). The roles of accounting in organizations and society. *Accounting, Organizations and Society*, vol. 5, n° 1, 5-27.
- Charreaux, G. et Gervais, M. (2007). La « piste aux étoiles » - un commentaire sur dernier classement des revues élaboré par la section 37 du CNR, éditorial. *Finance Contrôle Stratégie*, vol. 10, n° 4, décembre, 5-16.
- Cohen, E. (2009). La réforme de la gouvernance des universités, *Académie des sciences morales et politiques*, séance du 6 juillet.
- D'Andrade, R. (1995). *The development of cognitive anthropology*. Cambridge University Press, New York.
- Démil, B., Lecocq, X. et Warnier, V. (2007). Le couple pratique-recherche. Divorce, mariage ou union libre ? *Revue Française de Gestion*, n° 171, pp. 31-48.
- Gingras Y. (sd). Le classement de Shangai n'est pas scientifique, www.larecherche.fr, consulté le 04/01 2011.
- Gregory, R. (2000). Getting better but feeling worse? Public sector reform in New Zealand. *International Public Management Journal*, vol. 3, issue 1, 107-123.
- Hofstede, G. (1978), The poverty of management control philosophy. *Academy of Management Review*, July, 450-461.
- Hofstede, G. (1981). Management control of public and not-for-profit activities. *Accounting, Organizations and Society*, vol. 6, n° 3, 193-211.
- Holstrom, B and Milgrom, P (1991). Multitask principal-agent analyses: incentive contracts, asset ownership, and job design. *The Journal of Law, Economics and Organization*, 7, 24-52.
- Hood, C. (1991). A public management for all seasons? *Public administration*, 69: 3-19.
- Hood, C. (1995). The "New Public Management" in the 1980s: variation on a theme. *Accounting, Organizations and Society*, 20 (2/3): 95-109.
- Lapsley, I. (2009). New Public Management: the cruellest invention of human spirit? *Abacus*, vol. 45, n° 1, 1- 21.
- Malleret, V. (2004). La mesure des performances dans les activités de service. L'art du management édition 2004 - e-learning HEC - *Les Echos*, disponible sur www.lesechos.fr
- Norman, R. (2007). Entre finalités de service public et production mesurable : la redéfinition de la « valeur publique » dans le modèle néo-zélandais d'administration, *Revue Française d'Administration Publique*, n° 123, 323-336.
- Olson, O., Humphrey, C et Guthrie, J. (2001). Caught in an evaluatory trap : a dilemma for public services. *The European Accounting Review*, 10:3, 505-522.
- Ouchi, W.G. (1979). A conceptual framework for the design of organizational control mechanisms. *Management Science*, vol. 25, n° 9, September, 833-848.
- Ouchi, W.G. (1980). Markets, bureaucracies and clans. *Administrative Science Quarterly*, 25 :129, 129-141.
- Pal L.A. (2009). The OECD and global public management reform. *NISPAcee 17th Annual Conference*, Montenegro, May, 14-16.

- Perret, B. (2008). *L'évaluation des politiques publiques*, La Découverte, collection Repères (réédition).
- Porter, M. (1982). *Choix stratégiques et concurrence, Techniques d'analyse des secteurs et de la concurrence dans l'industrie*, Economica.
- Roturier, L. (2010). Pour en finir avec le « nouveau management public. *La Lettre du Cadre Territorial*, n° 408, octobre, 14-17.