

L'évaluation de la recherche : l'exemple de trois pays européens

Séverine Louvel, Stefan Lange

▶ **To cite this version:**

Séverine Louvel, Stefan Lange. L'évaluation de la recherche : l'exemple de trois pays européens. Sciences de la société : Les cahiers du LERASS, 2010, 79, pp.11-26. halshs-00576369

HAL Id: halshs-00576369
https://shs.hal.science/halshs-00576369

Submitted on 11 Apr 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.

L'évaluation de la recherche : l'exemple de trois pays européens

Séverine LOUVEL*
Stefan LANGE*

Article paru dans la revue *Sciences de la société*. Pour le citer :
Louvel, S. et S. Lange (2010). "L'évaluation de la recherche : l'exemple de trois pays européens." *Sciences de la société*, volume 79, pp. 11-26.

Résumés

Résumé

L'évaluation des systèmes de recherche est au cœur de nombreuses controverses en Europe. Entrée dans l'ère du quantitatif, formalisation des procédures et des critères, transformation du travail des pairs… Ces débats appellent à qualifier précisément les connaissances mobilisées et produites par les dispositifs d'évaluation. Cet article compare des dispositifs en vigueur au Royaume-Uni (le *Research Assessment Exercice* introduit en 1986), en Italie (le *Valutazione Triennale della Ricerca* conduit entre 2001 et 2003) et en Allemagne (*l'Excellenz Initiative* du gouvernement fédéral et le classement du Conseil de la Science, tous deux lancés en 2004). Il analyse ces réformes à l'aide de deux idéaux-types construits d'après la nature des connaissances mobilisées et leur place dans l'action publique (Gläser 2007 ; Whitley 2007 ; Gläser, Lange et al. 2010). Les systèmes d'évaluation nationaux peuvent ainsi être globalement qualifiés de « systèmes intrusifs » ou de « systèmes compétitifs », à conséquences fortes ou faibles sur la recherche (op. cit.). Les formes d'hybridation traduisent toutefois les tensions inhérentes à chaque système et les difficiles arbitrages sur les formes de connaissances à privilégier.

Abstract

Research evaluation is at the heart of numerous controversies in Europe: advent of the realm of metrics, formalisation of procedures and criteria, new expectations placed upon peers, etc. These debates call for a subtle examination of the knowledge used and produced during the evaluation. This article compares the evaluation procedures in use in the UK (the *Research Assessment Exercice* introduced in 1986), in Italy (the *Valutazione Triennale della Ricerca* conducted between 2001 and 2003) and in Germany (the *Excellenz Initiative* launched by the Federal Government; the Ranking of institutions and disciplines by the German Science Council – both started in 2004). It analyses these reforms with the help of two ideal-types defined after the knowledge base and the link between such knowledge and public action (Gläser 2007; Whitley 2007; Gläser, Lange et al. 2010). National evaluation systems can be predominantly qualified as "intrusive evaluation systems" or "competitive evaluation systems" with strong or weak consequences on research (op. cit.). However, various forms of hybridization reveal the tensions inherent to all research evaluation systems and some difficult trade-offs between the kinds of knowledge which should be given priority.

* Respectivement, maître de conférence en sociologie, Université de Grenoble CNRS PACTE (UMR 5194) BP 48, 38040 Grenoble Cedex 9, France, Severine.Louvel@iep-grenoble.fr ; et chaire subventionnée pour les sciences de l'organisation, l'enseignement supérieur et le management de la science, Ecole supérieure des sciences administratives de Speyer, P.O. Box 14 09, 67324 Speyer, Allemagne, stefanklauslange@aol.com.

L'évaluation est au cœur des politiques publiques de la recherche. Son expansion s'amorce dans les années 1980, corollaire direct de la stagnation des investissements publics et du renforcement des injonctions de « pertinence » (Rip, 1997), d'efficacité et de rentabilité qui s'exercent sur la recherche. La « République de la Science » (Dasgupta et David, 1994), bien que n'ayant jamais été autonome dans l'histoire (Pestre, 2003), entre alors clairement dans l'ère du soupçon. Evaluer suppose que les pouvoirs publics, les établissements de recherche, les universités, les agences, mettent en place des procédures de recueil et d'analyse de l'information, établissent des indicateurs de performance, d'efficacité, etc. Ces dispositifs souffrent souvent d'un déficit de légitimité, du fait notamment d'une tendance croissante à formaliser et à quantifier (Geuna et Martin, 2003 ; Feller, 2009). En France, les réformes de l'évaluation et notamment la création de l'AERES ont provoqué des controverses similaires (Cités, 2009). Pour éclairer ces débats internationaux, cet article compare les systèmes d'évaluation introduits au Royaume-Uni, en Italie et en Allemagne. Il les analyse à l'aide d'idéaux-types construits selon deux dimensions (Gläser, 2007 ; Whitley, 2007 ; Gläser, Lange et *al.* 2010) : les connaissances mobilisées et leur place dans l'action publique. Après avoir présenté finement cette typologie des systèmes d'évaluation, l'article situe alors les réformes nationales dans ce cadre, et se conclut par les enseignements qui peuvent être tirés quant aux liens entre connaissances, évaluation de la recherche et action publique.

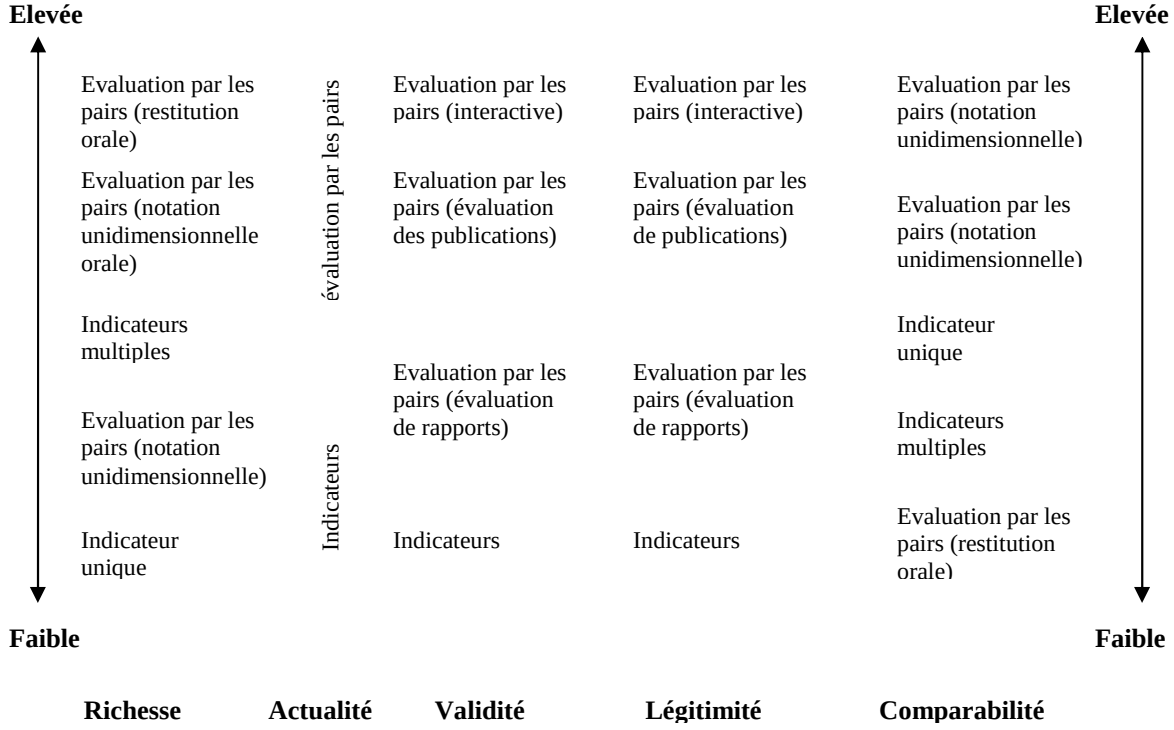
Un cadre d'analyse pour les systèmes d'évaluation de la recherche

Evaluation par les pairs *versus* évaluation par les indicateurs et nature des connaissances produites

Les dispositifs d'évaluation de la recherche peuvent être répartis en deux grandes « familles » : l'évaluation par les pairs (*peer review*) et l'évaluation par les indicateurs. Dans la première famille, des pairs réalisent un examen approfondi et contextuel des activités de recherche. Celui-ci peut prendre différentes formes : lectures critiques (d'articles, de rapports d'activité, de projets) ; échanges avec les évalués et entre évaluateurs (visite de laboratoires, auditions, délibérations des évaluateurs). L'évaluation par les pairs repose plutôt sur des connaissances qualitatives, ce qui n'exclut pas le recours à des indicateurs. Elle aboutit à des jugements qualitatifs (rédaction d'avis, de recommandations) ou synthétiques (notations, classements uni ou multidimensionnels). Dans la deuxième famille, l'évaluation se passe du recours direct aux pairs. La connaissance des activités de recherche repose uniquement sur des indicateurs, sur les moyens et les résultats (financements, contrats, recrutements, thèses, publications, facteur d'impact des revues, citation des articles, etc.). Les connaissances mobilisées par ces deux familles de systèmes d'évaluation ont été caractérisées selon cinq dimensions (Gläser, Lange et *al*., 2010) : la richesse, la validité, l'actualité, la comparabilité et la légitimité (tableau 1)[1].

1. Ces dimensions ont été établies à partir de 179 entretiens semi-directifs conduits dans 7 universités australiennes et de 60 entretiens conduits dans 2 universités allemandes. Les deux tiers des entretiens ont été réalisés avec des chercheurs, et un tiers avec des présidents d'université, doyens et responsables de département. L'objet de ces enquêtes était d'analyser les conséquences sur les activités de recherche de l'instauration d'un lien direct entre évaluation et financement : voir Laudel et Gläser (2007) pour la méthodologie et Gläser et Laudel (2007) ainsi que Lange (2007) pour les études de cas.

Tableau 1. Caractéristiques des connaissances mobilisées par les dispositifs d'évaluation



Les connaissances sont tout d'abord plus ou moins *riches*, selon leur niveau de détail, leur exhaustivité, etc. L'évaluation par les pairs mobilise généralement des connaissances plus riches que l'évaluation par les indicateurs. Ainsi, la lecture d'un rapport d'activité apporte une connaissance plus fine des programmes de recherche que le seul décompte de collaborations, de contrats ou de publications. Cette connaissance détaillée permet d'estimer le potentiel de certaines recherches, de réfléchir à l'adéquation entre des moyens et des objectifs, etc. Néanmoins, les évaluations par les pairs produisent parfois des connaissances assez pauvres, par exemple les notations unidimensionnelles. Ensuite, les connaissances sont plus ou moins *valides*, selon leur capacité à rendre compte de la qualité des recherches. Lors de l'évaluation par les pairs, les auditions sont généralement considérées comme plus valides que la seule lecture d'un bilan d'activité[2]. Les évaluations par indicateurs peuvent produire des connaissances peu valides car l'interprétation des mesures en terme de qualité reste délicate. La bibliométrie induit des artefacts désormais bien connus. Ainsi, le taux de citation d'un article renvoie à d'autres caractéristiques qu'à la qualité (auto-citation, citations critiques plutôt que laudatives, propension à citer davantage ses compatriotes, les auteurs connus, les revues prestigieuses, etc. (Rostaing, 1996). L'absence de discussions avec des pairs sur la pertinence des mesures est considérée comme une limite de la validité de l'évaluation par indicateurs[3]. *L'actualité* des connaissances dépend principalement de la périodicité des évaluations : une évaluation biannuelle mobilise *a priori* des connaissances plus actuelles qu'une évaluation quinquennale. Les connaissances mobilisées sont aussi plus ou moins *comparables*, à l'intérieur ou entre plusieurs vagues d'évaluation. Pour les deux familles d'évaluation, la comparabilité dépend du mode de restitution des résultats. Les

2. Voir cet écart à propos des publications Butler (2004) ou des financements contractuels Laudel (2006).
3. Dans les universités allemandes, même les universitaires aux scores bibliométriques très élevés restent réservés quant à la validité des indicateurs quantitatifs et privilégient les échanges entre pairs pour analyser les activités passées : Lange (2007).

recommandations et synthèses qualitatives offrent de plus faibles possibilités de comparaison que les classements ou les notes. Enfin, les connaissances bénéficient d'une plus ou moins grande *légitimité* auprès des collectivités scientifiques, selon leur adéquation avec leurs pratiques et leurs normes. L'évaluation par indicateurs est de ce fait globalement moins légitime, toutefois la légitimité de l'évaluation par les pairs varie fortement selon les dispositifs : les interactions entre évaluateurs et évalués sont plus légitimes que les évaluations à distance (fondées sur des lectures) ; les synthèses qualitatives davantage que les notations ou les classements (Gläser, Lange et *al.* 2010).

Systèmes d'évaluation intrusifs et systèmes d'évaluation compétitifs

Peu de systèmes d'évaluation s'appuient entièrement sur des indicateurs. Les universités australiennes sont évaluées de la sorte depuis le milieu des années 1990 : les principaux indicateurs portent sur les contrats, la formation doctorale et les publications (Gläser, Laudel, 2007 ; Geuna et Martin, 2003). En définitive, la plupart des pays implantent diverses configurations d'une évaluation par les pairs. C'est le cas de la création en France de l'Agence d'évaluation de la recherche et de l'enseignement supérieur (AERES). Des observateurs (Dodier, 2009) soulignent que l'AERES ne fait pas entrer l'évaluation dans le règne du quantitatif, mais qu'elle modifie considérablement le travail des pairs et les attentes à leur encontre. Cette remarque plaide pour une qualification précise des connaissances mobilisées et produites par l'évaluation par les pairs. L'évaluation de la qualité de la recherche s'inscrit certes dans des tendances communes (Sousa et Hendriks, 2007) : relâchement du contrôle centralisé des tutelles et autonomie croissante des opérateurs, passage d'un contrôle bureaucratique *ex ante* à une évaluation *ex post* des résultats. Toutefois, elle emprunte des voies diversifiées. Celles-ci peuvent être rattachées à deux grands systèmes (Kogan, 1989 cité par Geuna et Martin, 2003 ; Gläser, 2007) qui mobilisent des connaissances de nature très différente. Les systèmes d'évaluation intrusifs font appel à des évaluateurs qui acquièrent une connaissance approfondie des activités ; dans les systèmes dits compétitifs, ils compilent les informations nécessaires à une notation comparative des entités évaluées. À l'évidence, ces idéaux-types apparaissent rarement sous une forme pure, en particulier à l'échelle d'un pays. Toutefois, ils permettent de caractériser la conception de l'évaluation qui prime dans une politique nationale de recherche, et de comprendre les formes de connaissances des activités qui sont privilégiées. Par ailleurs, l'adoption d'un système ne présume pas de ses usages politiques et des conséquences de l'évaluation sur le système de recherches[4].

Tableau 2. Quatre types de systèmes d'évaluation de la recherche

		Conséquences	
		Faibles	**Fortes**
Type de connaissances	Examen approfondi d'une situation Connaissances locales et non transparentes faible publicité	**Systèmes d'évaluation intrusifs**	
		Connaissances sur les activités de recherche ; préconisations sans prescription	Utilisation des résultats pour allouer les ressources

4. C'est pourquoi cette terminologie parait plus satisfaisante que celle de Kogan (op. cit.) reprise par Geuna et Martin (op. cit.). Les auteurs distinguent une évaluation « formatrice » (*formative evaluation*) et une évaluation « sanction » (*summative evaluation*). Cette opposition sous-entend que les systèmes intrusifs ne conduisent à aucune sanction, et que les systèmes compétitifs ne permettent aucun apprentissage.

		Systèmes d'évaluation compétitifs	
	Comparaison des positions et performances respectives Connaissances publiques et transparentes forte publicité	Compétition pour la réputation (et pour les étudiants et pour les enseignants chercheurs dans l'enseignement supérieur)	Compétition pour la réputation (ainsi que pour les étudiants et les enseignants chercheurs dans l'enseignement supérieur) Compétition pour les financements

Source : d'après Gläser, 2007, notre traduction

Les pouvoirs publics, les universités ou les établissements de recherche, peuvent développer des systèmes d'évaluation qualifiés d'« intrusifs ». Ceux-ci s'appuient sur une connaissance approfondie de chaque entité évaluée et s'articulent étroitement à la mise en œuvre de réformes. La *richesse* et la *validité* des connaissances permettent ici de formuler un diagnostic précis et d'élaborer des scénarii d'évolution. Par ailleurs, une forte *légitimité* auprès des acteurs évalués facilite l'élaboration et la mise en œuvre des réformes[5]. Enfin, la priorité allant à la résolution de problèmes spécifiques, la *comparaison* de connaissances produites lors de plusieurs évaluations n'est pas primordiale.

Systèmes d'évaluation intrusifs

Dès lors, les systèmes intrusifs recourent prioritairement aux dispositifs qui produisent des connaissances riches, valides et légitimes : visites de sites, auditions, lecture extensive de certaines publications et de documents produits pour l'évaluation. Cela n'exclut pas l'usage de mesures synthétiques de l'activité, notamment en matière de publications. Les évaluations se concluent généralement par un bilan qualitatif des forces et des faiblesses, et par des propositions (restructurations organisationnelles, politique d'emploi scientifique, investissement dans des équipements, définition des programmes de recherche, etc.). Ces analyses sont généralement diffusées auprès des seuls évalués. Enfin, les résultats s'accompagnent d'incitations ou de sanctions[6].

Systèmes d'évaluation compétitifs

Ces systèmes d'évaluation s'inscrivent généralement dans une conception de l'action publique fondée sur le nouveau management public. Les mécanismes marchands ou quasi-marchands, et en particulier la concurrence, y jouent un rôle important (Ferlie, Musselin et *al.*, 2008). Par ailleurs, les tutelles « pilotent à distance » (De Boer, Enders et *al.*, 2006) les opérateurs, fixant une stratégie globale mais leur laissant une grande autonomie pour définir et mettre en œuvre les politiques concrètes. Évaluation et réformes constituent ici des processus distincts. Les évaluateurs diffusent des connaissances *comparables* sur les performances relatives, en vue d'informer la prise de décision des opérateurs de recherche, de leurs compétiteurs et de leurs financeurs. Corollaire au primat de la comparaison, les connaissances peuvent être *pauvres*, et les exigences de *validité* peu élevées. Enfin, destinées

5. Voir à ce propos le travail du comité consultatif universitaire (*Academic Advisory Council*) de Basse Saxonie Schiene et Schimank (2007).

6. Certains décideurs s'appuient sur des évaluations intrusives pour mettre en oeuvre des réformes radicales ou pour réaliser leurs arbitrages budgétaires : voir Schiene et Schimank (2007) à propos de la réforme de l'enseignement supérieur en Basse-Saxonie et Röbbecke et Simon (2001) au sujet des conséquences de l'évaluation des instituts de recherche Leibniz en Allemagne.

à plusieurs parties prenantes, les évaluations ne sont pas nécessairement très *légitimes* auprès des collectivités scientifiques[7]. Dès lors, les évaluations compétitives reposent principalement sur des mesures, ce qui n'exclut pas le recours à certaines connaissances qualitatives. Le cas échéant, le recueil et la restitution des données suivent des procédures standardisées, gage de comparabilité des évaluations. Le résultat à l'évaluation conditionne directement le financement de la recherche dans certains pays[8], tandis qu'il reste pratiquement sans conséquences dans d'autres[9]. Enfin, l'évaluation peut être indirectement prescriptive lorsqu'elle apporte des signaux à plusieurs parties prenantes (agences publiques, entreprises, étudiants, etc.). C'est le cas des notations et des classements réalisés par des agences, dont la prolifération structure par exemple le champ du management (Wedlin, 2006)[10].

Des réformes en cours

Trois exemples[11] européens illustrent l'opposition entre ces deux idéaux-types, ainsi que les emprunts qui conduisent éventuellement à des formes d'hybridation. Archétype d'un système compétitif, le Royaume-Uni s'est tout d'abord récemment rapproché d'un système intrusif. L'Italie a ensuite introduit un système compétitif inspiré de l'exemple britannique, mais aux conséquences plus limitées. Enfin, l'Allemagne présente la situation originale du développement parallèle de deux systèmes, l'un compétitif et l'autre intrusif.

Deux systèmes compétitifs : le Royaume-Uni et l'Italie

Royaume-Uni : le Research Assessment Exercise (RAE), un système compétitif aux conséquences majeures sur la recherche

Le RAE[12] est introduit en 1986 dans un contexte de réduction des dépenses publiques et de développement du nouveau management public (Henkel, 1999). L'objectif explicite est de réallouer les budgets récurrents vers les entités les plus performantes. Le RAE est ainsi géré par l'instance gouvernementale chargée d'attribuer ces financements[13]. Six cycles d'évaluation ont été conduits (en 1986, 1989, 1992, 1996, 2001 et 2008) selon différentes méthodes. Les données évaluées ont changé : en 1986 les cinq meilleures publications des quatre dernières années ; en 1989 une à deux publications (au choix) par membre permanent ; en 1992 toutes les publications, avec possibilité d'inclure seulement les chercheurs considérés comme « actifs » ; enfin, depuis 1996, quatre publications maximum (au choix) par « chercheur actif ». Les évaluations sont conduites par environ 70 comités disciplinaires, dont

7. Voir Lucas (2006) à propos du RAE britannique.
8. La Grande-Bretagne et l'Australie sont ici exemplaires : Barker (2007) ; Gläser et Laudel (2007).
9. Voir l'exemple du financement « à la performance » en place dans les *Länder* allemands ces dix dernières années : Lange (2007).
10. Certains systèmes d'évaluation par indicateurs sont également des prescripteurs indirects. C'est le cas des classements publiés dans la presse (classement du *Times Higher Education*, de Shanghai, de l'Ecole des Mines, etc.), qui occupent une place considérable dans l'attribution de la réputation aux opérateurs de la recherche et de l'enseignement supérieur : Weingart et Maasen (2007) ; Wedlin (2006).
11. Cette étude comparative a été réalisée dans le cadre du contrat européen MEHEM (Mapping European Higher Education Models : www.mehem.org): Lange et Blümel (2009) et Louvel et Gonthier (2009).
12. Le RAE a été très étudié. Nous nous appuyons ici sur une publication récente (Martin et Whitley (2010)) qui comporte une synthèse des réformes du RAE et des controverses que celui-ci suscite. Pour des analyses plus détaillées voir Barker (2007), Henkel (1999 ; 2000) ainsi que les nombreuses publications du *Higher Education Funding Council for England* (HEFCE).
13. *University Grant Council*, puis *Universities Funding Council* (1988), enfin *Higher Education Funding Councils* ou HEFC (1992). Elle est intégrée au *Department of Education and Science.*

le nombre et la composition sont du ressort des HEFC[14]. Depuis 2001, ces comités ont pour consigne d'évaluer la qualité des publications indépendamment de leur support (ouvrages, revues, actes de conférences, etc.), ce qui supposerait de les lire intégralement. En pratique, les standards nationaux ou internationaux, et notamment le classement ou la réputation des revues, orientent largement les jugements. Le prochain cycle d'évaluation (dénommé *Research Evaluation Framework*) introduira un changement majeur : dans les sciences pour l'ingénieur et les sciences exactes, l'évaluation par les pairs (jugée trop coûteuse) sera remplacée par une évaluation par indicateurs. Les résultats ont été restitués sous plusieurs formes : en 1986, notation sur une échelle en 4 catégories ; en 1989, notation sur une échelle de 1 à 5 ; en 1996, notation sur une échelle en 7 catégories et possibilité de noter des sous-unités (par exemple, les équipes d'un département). Un changement important intervient en 2008 : la note est remplacée par un « profil de qualité » (*quality profile*), qui précise la répartition des activités entre les différents niveaux de qualité. Par exemple, un département aura 15% de la recherche de niveau « 4 étoiles » (« Qualité de niveau mondial en termes d'originalité, de pertinence et de rigueur »), 30% de niveau 3 étoiles (qualité de niveau international), etc. Enfin, les conséquences des évaluations s'alourdissent. En 1989, 50% des financements dédiés à la recherche sont attribués en fonction de la note obtenue, en 1992 ce pourcentage passe à 90%. Les exercices successifs du RAE se caractérisent aussi par une concentration croissante des financements sur les unités les mieux évaluées, les plus mal évaluées n'obtenant plus de financement (Barker, 2007; Lucas, 2006 ; Henkel, 2000).

Le RAE est emblématique des *systèmes d'évaluation compétitifs avec des conséquences fortes sur la recherche* (Martin et Whitley, 2010). Il instaure un jeu à somme nulle entre les universités, mais également entre les départements ou les équipes. Il produit des incitations très fortes aux restructurations organisationnelles (qui se sont traduites par des fermetures), à des stratégies de recrutement des meilleurs scientifiques, à la définition d'une politique scientifique claire, etc. *(ibid.)*. Ce système compétitif produit et restitue des connaissances hautement *comparables*, bien que le remplacement en 2008 de la note unique par des « profils de qualité » diminue cette comparabilité. Les connaissances sont assez *pauvres* puisque limitées à une sélection de publications (avec un primat aux revues internationales) et à des indicateurs de rayonnement scientifique (distinctions, activités éditoriales, etc.). À présent, les profils de qualité apportent aux départements évalués une connaissance plus fine de leurs atouts et de leurs faiblesses que la note unique. *L'actualité* des connaissances est également limitée par l'espacement croissant des cycles d'évaluation (à présent tous les sept ans). Enfin, leur degré de *validité*, ainsi que leur *légitimité* portent à débats. Des observateurs avancent que le RAE impose une définition étroite et standardisée de la qualité, plus pertinente dans certains domaines : les sciences exactes, les approches les plus conventionnelles, les revues généralistes (mieux « cotées » que les revues spécialisées), les travaux disciplinaires. Ils justifient leurs interrogations sur la validité des connaissances par la forte sensibilité de la mesure à l'instrument. L'apprentissage des règles (privilégier les thèmes ou les approches qui se publient le « mieux », calculer le nombre optimal de chercheurs « actifs »[15]) expliquerait ainsi l'amélioration d'ensemble des scores au RAE, mesurée par la hausse de la proportion des chercheurs rattachés aux départements les mieux notés (23% en 1992, 31% en 1996 et 55% en 2001 ; Martin et Whitley, 2010). La légitimité des données du RAE est enfin discutée. Si le principe d'évaluation par des comités disciplinaires est globalement légitime, la collecte et le traitement des données sont jugés très coûteux (pour les évaluateurs comme pour les unités évaluées). Surtout, le lien majeur entre résultats au RAE et financement de la recherche est très controversé. Certains observateurs soulignent que le RAE

14. Voir note précédente.

15. Comme une partie des financements dépend du nombre de chercheurs, exclure des « inactifs » augmente les chances d'être bien classé mais réduit les financements attribués en fonction des effectifs.

a amélioré la performance de la recherche britannique, mais d'autres insistent sur ses effets délétères : abandon des recherches moins rentables ; découragement des chercheurs jugés inactifs ; course à la publication aux dépens de l'enseignement ; renforcement du contrôle hiérarchique au mépris de la collégialité, etc. Plus globalement, le RAE est considéré comme emblématique d'une « '*marchandisation' du travail académique, le travail des académiques étant de plus en plus transformé en un bien dont le RAE détermine la valeur* » (*ibid.*, p. 20).

Italie : le VTR (*Valutazione triennale della ricerca*)*, un système d'évaluation compétitif aux conséquences intermédiaires sur la recherche*

Le VTR[16] représente une innovation institutionnelle (Minelli, Rebora et *al*., 2008) : il rompt en effet avec une approche bureaucratique de l'évaluation, centrée sur le respect de règles administratives. Conduit entre 2001 et 2003, le premier cycle d'évaluation a porté sur 77 universités et 25 institutions de recherche : il est ainsi trop tôt pour évaluer avec certitude ses effets sur le système de recherches italien. Comme le RAE, le VTR est un système d'évaluation centralisé, géré par une instance ministérielle, le Comité national pour l'évaluation de la recherche (CIVR). Il s'agit aussi d'une évaluation par les pairs à visée compétitive, instaurée pour informer l'attribution d'une partie des financements récurrents. Dans le détail, le RAE et le VTR diffèrent toutefois largement. Tout d'abord, le VTR a porté sur un large ensemble de « productions de recherche » (articles, livres, actes de colloque, brevets, etc.), complétées par différentes informations (mobilité internationale, formation à la recherche, projets et contrats de recherche, transfert technologique). Le nombre de « productions de recherche » par entité évaluée s'élève au plus à la moitié du personnel titulaire (en équivalents temps plein, les enseignants chercheurs comptant pour un demi équivalent temps plein). Les données à évaluer (au total, environ 17000) ont ensuite été transmises à 20 comités disciplinaires composés de 5 à 17 membres (selon le volume à évaluer). Ceux-ci ont été nommés par le CIVR, après un travail conjoint avec les collectivités scientifiques (appel à propositions pour obtenir des listes de membres possibles, définition des critères de sélection des membres, élaboration des règles générales de fonctionnement du VTR). Les comités ont désigné à leur tour plus de 6000 experts italiens et internationaux chargés de l'évaluation. Pour chaque production de recherche, les experts ont rédigé une appréciation qualitative et attribué une note (excellent, bon, acceptable, insuffisant) en tenant compte de quatre dimensions (qualité, pertinence, originalité ou innovation, portée internationale). Les comités disciplinaires ont synthétisé les évaluations et produit un rapport en trois parties : un jugement final sur l'entité évaluée (*consensus report*) ; une note par entité évaluée et par domaine disciplinaire (*area ranking list* : une production de recherche « excellente » compte pour 1 point, une « bonne » pour 0,8, « acceptable » pour 0,6 et « insuffisante » pour 0,2) ; enfin un rapport (*aera final report*) indiquant les points forts et les points faibles. Par ailleurs, le CIVR a calculé un indice d'excellence (proportion de productions excellentes) et a établi des bilans par secteur disciplinaire. Tous les résultats par établissement et par discipline sont publiés sur son site. En revanche, les données individuelles ne sont pas publiques. Enfin, il est prévu d'attribuer 30% du budget recherche en fonction des résultats au VTR[17] : la notation de la qualité doit peser pour plus de la moitié dans le calcul ; viennent ensuite les données complémentaires rassemblées dans le *consensus report*. Toutefois, l'application de cette règle est encore incertaine.

Le VTR peut être qualifié de *système d'évaluation compétitif aux conséquences intermédiaires* : il encourage en effet, comme le RAE, la compétition entre établissements et

16. Cette partie s'appuie sur les principales publications en langue anglaise sur le VTR.
17. Source: http://vtr2006.cineca.it/ Final Report.

entre départements (Minelli et *al*., 2008). Par ailleurs, il se rapproche par certains aspects des systèmes d'évaluation intrusifs : en particulier, le *aera final report* fournit des recommandations détaillées. La notation des productions de recherche et l'attribution de notes globales assurent la forte *comparabilité* des connaissances produites. Les données évaluées sont par ailleurs plus *riches* que celles du RAE, et incluent une large gamme de productions de recherche et une connaissance fine des « entrants » à l'activité. Les observateurs attribuent une *validité* élevée aux connaissances mobilisées et produites. Ils soulignent notamment que le CIVR (comité national pour l'évaluation de la recherche) n'a pas imposé de définition standardisée de la qualité, « *la qualité scientifique étant ce que les comités [NB disciplinaires] décident qu'elle soit* » (Minelli et *al*., 2008). L'« excellence » est ainsi une appréciation relative, réservée aux productions de recherche se situant « *dans les 20% les mieux évalués selon une échelle de valeur partagée par la communauté scientifique internationale* »[18]. Des auteurs apprécient cette validité à l'absence d'effet Saint Matthieu (Merton, 1973), autrement dit à la faible sensibilité de l'évaluation à la réputation des universités ou des départements (Reale, Barbara et *al*., 2006). Des artefacts sont toutefois relevés, et notamment la sensibilité des résultats à certains choix stratégiques : en choisissant de soumettre à évaluation un petit nombre de productions de recherche, les universités accroissent leurs chances d'obtenir un fort pourcentage de productions excellentes. Enfin, les données du VTR sont assez *légitimes* auprès des collectivités scientifiques. Les chercheurs soutiennent l'absence de standardisation des critères de qualité, le choix des productions de recherche soumises à évaluation, enfin leur participation au choix des experts, à la définition des critères et des procédures. De ce fait, les scientifiques ont vu dans le VTR le prolongement aux exercices d'évaluation coutumiers dans la profession (Minelli, Rebora et *al*., 2008). L'absence de lien avéré entre le premier VTR et les décisions budgétaires a certainement accru la légitimité de l'exercice.

Deux systèmes dans un même pays : l'Allemagne

L'Allemagne 1 : l'initiative pour l'excellence (Exzellenzinitiative, EI), un système compétitif aux conséquences intermédiaires sur la recherche

Introduite en 2004 par le gouvernement fédéral, l'initiative pour l'excellence (EI) vise à lutter contre deux faiblesses apparentes de la recherche allemande : son éclatement entre universités et institutions non universitaires, la faible visibilité internationale de la recherche universitaire (reflétée par la place des universités allemandes dans les classements internationaux). L'EI est un outil pour ramener la recherche dans des universités considérées comme le « centre organisationnel de la science » (Strohschneider, 2009). Le dispositif repose sur l'allocation, entre 2006 et 2011, d'1,9 milliards d'euros vers des universités dites « d'élite ». L'EI instaure clairement un système d'évaluation compétitif, aux conséquences intermédiaires sur la recherche allemande. Les fonds s'ajoutent aux financements récurrents et sont attribués à l'issue d'une compétition entre les projets présentés par les universités[19]. Issu d'un compromis entre le gouvernement fédéral et les *Länder*, le dispositif respecte les prérogatives de ces derniers en matière de gouvernement des universités. En 2007 et 2008, les appels à projets ont concerné trois grands domaines : les écoles doctorales ; la recherche, avec comme cible la recherche reconnue à l'échelle internationale, les programmes pluridisciplinaires et la collaboration entre universités, institutions de recherche (instituts Max-Planck, Fraunhofer, Helmholtz, et Leibniz), écoles (*Fachhochschulen*) et entreprises ; enfin, les plans de développement. Particulièrement médiatisé, ce dernier domaine doit permettre aux universités de s'élever au rang des établissements les plus prestigieux à l'international. Les projets relatifs aux écoles doctorales et à la recherche ont été évalués par la

18. Idem.
19. Pour des analyses de cette dimension compétitive voir Bleiklie et Lange (2010) ; Lange et Krücken (2010)

Deutsche Forschungsgemeinschaft (DFG, principale agence publique de financement de la recherche), tandis que le Conseil de la science (*Wissenschaftsrat*, WR, instance consultative sur la politique scientifique auprès du gouvernement fédéral, des Länder et de la communauté scientifique) a évalué les plans de développement. La DFG et le WR ont constitué des comités d'experts composés en moyenne de 60% de membres étrangers. Les décisions finales d'attribution des financements ont été prises par un jury composé de représentants du gouvernement fédéral et des *Länder*. Ce jury a largement entériné les décisions des comités d'experts relatifs aux écoles doctorales et à la recherche, et n'a exercé une marge de manœuvre que pour les plans de développement des universités. Les financements se sont élevés à 44,7 millions d'euros par an attribués à 39 écoles doctorales ; 235,9 millions d'euros par an à 37 « clusters d'excellence »; enfin 131,1 millions d'euros par an à neuf universités « d'élite ». Le gouvernement fédéral et les Länder ont récemment conclu un accord pour une seconde EI à conduire de 2012 à 2017 avec un budget de 2,7 milliards d'euros. Les résultats de la première campagne sont d'ores et déjà notables, en particulier le rapprochement des universités et des institutions de recherche au sein des écoles doctorales et des clusters d'excellence. Dans un cas, l'EI a même suscité la fusion d'une université et d'un centre de recherche (création de l'Institut technologique de Karlsruhe, issu de l'université technologique et du centre Helmholtz). Plus généralement, même les universités non sélectionnées ont cherché à renforcer leurs liens avec les centres de recherche, de telles collaborations devenant un atout décisif pour obtenir des financements.

L'EI instaure un *système d'évaluation compétitif avec des conséquences intermédiaires sur la recherche*. Comme le RAE et le VTR, il stimule explicitement la compétition entre les universités. À la différence de ceux-ci, l'EI n'a pas d'incidence directe sur les financements récurrents, bien que les gagnants puissent occuper une position plus favorable pour négocier leur budget dans leur *Land*. L'EI peut également encourager la compétition entre chercheurs et groupes de recherche. En effet, les projets sont soumis par les établissements, mais les fonds sont attribués à des chefs de projet. Par ailleurs, l'EI présente aussi certaines caractéristiques d'un *système d'évaluation intrusif*. En effet, et à la différence d'appels d'offres auxquels répondent directement des chercheurs ou des équipes, les universités connaissent très bien les dossiers qu'elles soumettent. L'EI apporte alors aux équipes dirigeantes des connaissances précises sur la recherche conduite dans chaque département. Les premières observations empiriques suggèrent que l'EI peut avoir des conséquences importantes sur la recherche, alimenter des réformes organisationnelles ou des décisions de financement dans les universités ou les départements (Lange et Gläser, 2009). De ce point de vue, l'EI peut contribuer, comme le RAE britannique, à consolider les hiérarchies en place en matière de recherche. Ses effets sur l'enseignement pourraient aussi s'accroître à l'avenir. Les pouvoirs publics ont en effet recommandé aux universités non sélectionnées de se constituer un profil d'excellence en enseignement, et le gouvernement fédéral a récemment annoncé le lancement d'une compétition similaire à l'EI dans le domaine de l'enseignement. On peut alors s'attendre à une différenciation croissante entre des universités de recherche (les *research universities* anglophones, centrées sur l'enseignement en master et doctorat) et des universités d'enseignement (gérant surtout l'enseignement de premier cycle).

Ce système compétitif repose sur des connaissances relativement *riches* : les comités de pairs examinent précisément l'originalité et le potentiel des projets, et ils recueillent des données détaillées sur le bilan des structures candidates. Ces connaissances sont également considérées comme *valides*, car orientées vers une appréciation contextuelle et documentée de la qualité des projets (*ibid*.). Certaines critiques ont toutefois souligné les limites possibles à cette validité et l'existence d'un « effet Saint Matthieu » (cf. *infra*). L'EI mobilise des connaissances moins *comparables* que d'autres systèmes compétitifs. Cette caractéristique

concentre les critiques contre l'EI : manque de base claire pour comparer la qualité des soumissions, absence de publicité sur les projets soumis et sur les expertises. Enfin, et malgré cette absence de comparabilité, l'EI bénéficie d'une grande *légitimité* auprès des collectivités scientifiques : la DFG et le WR sont en effet hautement légitimes et les scientifiques y occupent une place centrale ; par ailleurs les procédures d'évaluation sont alignées sur des pratiques usuelles en recherche.

L'Allemagne 2 : le classement du conseil de la science (Wissenschaftsrat, WR), un système intrusif aux conséquences faibles sur le système de recherche

En 2004, le WR lance le premier dispositif d'évaluation de toute la recherche publique, aujourd'hui encore expérimental. Les dispositifs précédents étaient dispersés : évaluation des 80 institutions non universitaires ; évaluation des universités par le *Land* dont elles dépendent, dans le sillage de l'adoption du nouveau management public (Schimank et Lange, 2009). Cette initiative répond directement à la multiplication d'évaluations par indicateurs souvent médiatisées (par l'hebdomadaire le *Spiegel*, par le *think tank* CHE – centre pour le développement de l'enseignement supérieur – publié dans le *Stern*, classements internationaux, etc.). Le WR rejette explicitement des classements qu'il juge peu rigoureux, qui sont élaborés sans les collectivités scientifiques, et qui excluent les institutions non universitaires. Il défend une évaluation intrusive, dégageant les forces et les faiblesses de la recherche dans différentes disciplines et institutions et indiquant des voies d'amélioration. En 2005, le WR a mené une étude pilote en chimie (57 universités et 20 institutions de recherche) et en sociologie (54 universités et 3 institutions de recherche). Les résultats ont été publiés fin 2007 (en chimie) et au printemps 2008 (en sociologie). Ces études ont été supervisées par un comité de pilotage, composé de membres du WR, de représentants du gouvernement fédéral et des Länder, de représentants des collectivités scientifiques. Elles ont été conduites par des comités d'évaluation disciplinaires d'une quinzaine de membres. Trois domaines ont été évalués : la recherche, la formation à la recherche et le transfert technologique, à partir de données qualitatives et quantitatives sur les moyens mis en œuvre et les résultats obtenus. La recherche a par exemple été évaluée selon trois dimensions : sa qualité, son impact et son efficacité. L'efficacité a été appréciée en rapportant la qualité à des indicateurs de moyens (nombre de chercheurs, financements). Les évaluateurs ont pu produire leurs propres indicateurs, à condition qu'ils respectent les pratiques et normes disciplinaires ; ils ont aussi pu échanger avec les groupes évalués et visiter les sites. Les résultats ont été restitués sous la forme de profils de qualité situant l'institution ou la discipline dans chaque domaine évalué. Initialement, la notation était relative, comme dans le VTR italien : ainsi, la plus haute note en recherche signifiait que 50% des activités atteignaient des standards d'excellence internationaux. Les évaluateurs ont finalement retenu une échelle absolue (excellent, très bon, bon, etc.).

L'évaluation pilote du WR peut être qualifiée de *système intrusif à conséquences faibles* sur la recherche allemande. Elle analyse en effet en contexte des résultats, des moyens et modes d'organisation, pour proposer des recommandations. On peut remarquer que les « profils de qualité » peuvent lui donner une visée compétitive : toutefois, la comparaison des profils de qualité vise jusqu'ici moins à stimuler la compétition qu'à dégager des pistes de réformes. La comparaison de la formation à la recherche en chimie et en sociologie a par exemple alimenté une réflexion sur des voies d'amélioration pour celle-ci. Ce système intrusif repose sur des connaissances *riches*, mais moyennement *comparables*. Les profils de qualité constituent ainsi une solution de compromis pour disposer de repères d'évaluation clairs sans simplifier à l'excès les résultats. Les connaissances mobilisées ont été considérées comme relativement *valides*. En particulier, le WR a respecté un choix pragmatique de critères disciplinaires.

Confrontés à la faible pertinence des indicateurs bibliométriques en sociologie, les évaluateurs ont ainsi opté pour une lecture intégrale de certaines publications, au prix d'un travail considérable. Ce souci de la validité transparaît aussi dans la réflexion que le WR lance en 2009 pour adapter les dispositifs d'évaluation aux humanités, en réponse aux critiques des historiens concernés par l'extension de l'évaluation à leur discipline (cf. *infra*). Enfin, les connaissances sont très *légitimes* auprès des collectivités scientifiques, comme le montre l'approbation de l'extension des études pilotes au génie électrique et à l'histoire. Le retrait de l'association des historiens allemands apporte toutefois un bémol pour les humanités. Les historiens ont avancé que les seules connaissances légitimes à leurs yeux (la lecture approfondie des publications) restaient incompatibles avec le cadre et le calendrier de l'évaluation. Par ailleurs ils se sont élevés contre les effets possibles d'un tel exercice (atteintes à la collégialité, exacerbation de la concurrence entre les chercheurs) et ont fait part de leurs craintes quant à l'utilisation des résultats dans une optique de financement sélectif de la recherche.

Les systèmes d'évaluation constituent aujourd'hui une base de connaissance essentielle au management de la recherche. Cela étant, leurs modes d'institutionnalisation et les types de connaissances qu'ils mobilisent, diffèrent fortement selon les pays. La distinction d'ensemble entre évaluation par indicateurs et évaluation par les pairs rend notamment peu compte de la diversité des instruments de connaissance engagés. L'évaluation par les pairs peut ainsi être rattachés à deux formes idéales typiques: les systèmes intrusifs, qui mobilisent pour chaque entité évaluée des connaissances riches, valides et plutôt légitimes auprès des collectivités scientifiques ; des systèmes compétitifs, qui produisent pour toutes les entités évaluées des connaissances fortement comparables, souvent moins légitimes. Par ailleurs, chaque système a des conséquences plus ou moins fortes selon le lien entre les évaluations et l'attribution des financements.

Le système d'évaluation français parait doublement hybride. D'une part, l'opération « investissements d'avenir » lancée en 2010 par le Ministère de l'enseignement supérieur et de la recherche, inspirée en particulier de *l'Exzellenzinitiative* allemande, instaure un système compétitif qui pourra avoir des conséquences majeures. La sélection organisée dans le cadre des appels à projets (laboratoires d'excellence, d'équipements d'excellence, sociétés d'accélération du transfert de technologies, instituts hospitalo-universitaires, etc.), est en effet susceptible de provoquer des regroupements, de renforcer certains pôles et d'accentuer la différentiation entre sites. D'autre part, l'AERES constitue en lui-même un système d'évaluation hybride, aux conséquences intermédiaires. La logique intrusive réside dans l'analyse détaillée et contextuelle des activités et dans la production de recommandations. La logique compétitive se manifeste dans les repères simples permettant de comparer les entités évaluées. Enfin, les résultats ne conditionnent pas explicitement les financements à la manière du RAE britannique.

Les exemples européens présentés ici montrent que l'hybridation renvoie à des tensions inhérentes aux systèmes nationaux d'évaluation. Aucun n'échappe en effet aux arbitrages quant aux connaissances à privilégier. Les connaissances mobilisées par l'AERES sont ainsi relativement riches (évaluation exhaustive des activités – contrairement au RAE et au VTR –, visites de site, rapports détaillés, recours à des critères qualitatifs et quantitatifs) et assez valides (ce qui n'exclut pas des craintes quant à une standardisation de l'excellence, particulièrement pour les sciences humaines et sociales). Richesse et validité ont pour

corollaire des possibilités intermédiaires de comparaison (notation multidimensionnelle, notes globales comparables à l'intérieur d'une discipline seulement).

Surtout, les exemples étrangers indiquent que les arbitrages sur les connaissances fluctuent. Jusqu'en 2008, le RAE privilégie ainsi des connaissances très comparables. L'exercice de 2008 modifie cet arbitrage : les profils de qualité diminuent la comparabilité, mais accroissent la richesse et la validité des connaissances. Le VTR italien pourrait prendre le chemin inverse, notamment si sa dimension compétitive s'accentue et si les liens entre évaluation et financement de la recherche se durcissent. Enfin, en Allemagne, l'évaluation du WR et l'initiative pour l'excellence privilégient la richesse et la validité. Dans le second cas, cet arbitrage a très récemment évolué vers davantage de comparabilité, en réponse aux critiques exercées à l'encontre d'une évaluation qualifiée de « boîte noire ». Dans la seconde campagne de l'EI lancée en mars 2010, les évaluateurs disposent ainsi d'une liste de critères par domaine évalué. Cet exemple montre que certains systèmes d'évaluation aux conséquences fortes subissent des pressions considérables pour légitimer leurs décisions, ce qui peut les conduire à arbitrer en faveur de la comparabilité des connaissances. Mais la réforme du RAE signale que d'autres facteurs peuvent jouer en sens inverse, comme les traditions institutionnelles, la culture académique ou les rapports de force entre collectivités scientifiques, Etat central et équipes dirigeantes des universités et établissements de recherche.

Références bibliographiques

BARKER K., 2007, «The UK Research Assessment Exercise : The Evolution of a National Research Evaluation System », *Research Evaluation*, n°16, 3-12.

BLEIKLIE E., LANGE S., 2010, « Competition and Leadership as Drivers in German and Norwegian University Reforms », *Higher Education Policy*, n°23, 173-193.

CITES, 2009, *L'idéologie de l'évaluation : la grande imposture*, n° 37.

DASGUPTA P., DAVID P., 1994, «Toward a new economics of science », *Research Policy*, 23(5) : 487-521.

BUTLER L., 2004, «What Happens when Funding is Linked to Publication Counts? », *in* MOED H.F., GLÄNZEL W. et SCHMOCH U., *Handbook of Quantitative Science and Technology Research : The Use of Publication and Patent Statistics in Studies of S&T Systems*, Dordrecht, Kluwer, 89-405.

DE BOER H., ENDERS J. et *al*., 2006, « The Netherlands - Steering from a Distance », *in* B. M. KEHM et U. LANZENDORF, *Reforming University Governance. Changing Conditions for Research in Four European Countries*, Bonn, Lemmens, 53-89.

FELLER I., 2009, « Performance Measurement and the Governance of American Academic Science », *Minerva* n°47, 323-344.

FERLIE E., MUSSELIN C. et *al*., 2008, «The steering of higher education systems: a public management perspective », *Higher Education*, n°56, 325-348.

GEUNA A., Martin B., 2003, « University Research Evaluation and Funding : An International comparison. », *Minerva*, n°41, 277-304.

GLÄSER J., 2007, « The social orders of research evaluation systems », *in* R. WHITLEY et J. GLÄSER, *The Changing Governance of the Sciences : The Advent of Research Evaluation Systems*, *Sociology of the Sciences Yearbook*, n°26, Dordrecht, Springer, 245-266.

GLÄSER J., LAUDEL G., 2007, « Evaluation without Evaluators : The Impact of Funding Formulae on Australian University Research », *in* R. WHITLEY et J. GLÄSER, *The Changing Governance of the Sciences : The Advent of Research Evaluation Systems*, Dordrecht, Springer, 127-151.

GLÄSER, J., LANGE S. et *al.*, 2010, « Informed Authority? The Limited Use of Research Evaluation Systems for Managerial Control in Universities », *in* R. WHITLEY J. GLÄSER et L. ENGWALL *Reconfiguring Knowledge Production : Changing Authority Relations in the Sciences and Their Consequences for Intellectual Innovation*, Oxford University Press, à paraître.
HENKEL M., 1999, « The modernisation of research evaluation : The case of the UK », *Higher Education*, n° 38, 105-122.
HENKEL M., 2000, *Academic Identities and Policy Change in Higher Education*, London, Jessica Kingsley.
KOGAN M., 1989, « The Evaluation of Higher Education : An Introductory Note », *Evaluating Higher Education*, London, Jessica Kingsley Publishers, 11-25.
LANGE S., 2007, « The Basic State of Research in Germany: Conditions of Knowledge Production Pre-Evaluation » *in The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*, Dordrecht, Springer, 153-170.
LANGE S., BLÜMEL A., 2009, *Comparative Study of Research Performance Evaluation Systems in European Higher Education: Anglo-Saxon and Teutonic Models*, Rapport pour le projet européen MEHEM (Mapping European Higher Education Models), 40 pages.
LANGE S., GLÄSER J., 2009, « Performanzsteigerung durch Selektivität ? Erwartbare Effekte von Forschungsevaluationen an deutschen Universitäten im Lichte internationaler Erfahrungen », *in* R. WHITLEY et J. GLÄSER, *Der moderne Staat. Zeitschrift für Public Policy, Recht und Management*, 2. Jg., S. 355-374.
LANGE S., KRÜCKEN G., 2010, « German Universities in the New Knowledge Ecology. Current Changes in Research Conditions and University-Industry-Relations », *in* C. CALHOUN et D. RHOTEN, Knowledge Matters : The Public Mission of the Research University, New York, Columbia University Press, à paraître.
LAUDEL G., 2006, « The 'Quality Myth': Promoting and Hindering Conditions for Acquiring Research Funds », Higher Education, n° 52, 375-403.
LAUDEL G., GLÄSER J., 2007, « Interviewing Scientists », *Science, Technology & Innovation Studies* 3 : http://www.sti-studies.de/fileadmin/articles/laudelglaeserstivol3no2-2007.pdf.
LOUVEL S., GONTHIER A., 2009, *Performance evaluation regime : France*, Rapport pour le projet européen MEHEM, 27 pages.
LUCAS L., 2006, *The Research Game in Academic Life*, Maidenhead : SRHE/Open University Press.
MARTIN B., WHITLEY R., 2010, « The UK Research Assessment Exercise : A Case of Regulatory Capture? », *in* R. WHITLEY J. GLÄSER et L. ENGWALL, *Reconfiguring Knowledge Production: Changing Authority Relations in the Sciences and Their Consequences for Intellectual Innovation*, Oxford University Press, à paraître.
MERTON R. K., 1973, *The Sociology of Science*, Chicago, University Press of Chicago.
MINELLI E., G. REBORA, et *al.*, 2008, « The structure and significance of the Italian research assessment exercise (VTR) », *in* C. MAZZA P. QUATTRONE et A. RICCABONI, *European Universities in Transition: Issues, models and cases*, Cheltenham, Edward Elgar: 221-236.
PESTRE D., 2003, *Science, argent et politique. Un essai d'interprétation*, Sciences en questions, Paris, INRA éditions.
REALE E., BARBARA, A. et *al.*, 2006, « Peer review for the evaluation of the academic research. The Italian experience », *Working paper CERIS-CNR*, n°15/2006, 29.
RIP A., 1997, « A cognitive approach to relevance of science », *Social Science Information*, 36(4) : 615-640.
RÖBBECKE M., SIMON D., 2001, *Reflexive Evaluation. Ziele, Verfahren und Instrumente der Bewertung von Forschungsinstituten*, Berlin: Edition Sigma.
ROSTAING H., 1996, *La bibliométrie et ses techniques*, Collection Sciences de la société, Co-édition : Sciences de la société (Toulouse) et Centre de recherche rétrospective de Marseille (Marseille).
SCHIENE C., SCHIMANK U., 2007, « Research Evaluation as Organisational Development: The Work of the Academic Advisory Council in Lower Saxony (FRG) », *in* R. WHITLEY et J. GLÄSER, *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*, Dordrecht, Springer, 171-190.
SCHIMANK U., LANGE S., 2009, « The German University System : A Late-Comer in New Public Management », *in* C. PARADEISE, E. REALE, E. BLEIKLIE et E. FERLIE, *University Governance. Western European Comparative Perspectives*, Dordrecht, Springer, 65-92.

SOUSA C. A. A. et HENDRIKS P. H. J., 2007, « That obscure object of desire : the management of academic knowledge », *Minerva*, 45: 259-274.
STROHSCHNEIDER P., 2009, « Über Voraussetzungen und Konzeption der Exzellenzinitiative, *Beiträge zur Hochschulforschung* », 31: 8-24.
WEDLIN L., 2006, *Ranking Business Schools: Forming fields, identities and boundaries in international management education,* Cheltenham, Edward Elgar.
WEINGART P., MAASEN S., 2007, « Elite Through Rankings – The Emergence of the Enterprising University », *in* R. WHITLEY et J. GLÄSER, *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*, Dordrecht, Springer, 75-99.
WHITLEY R., 2007, « Changing Governance of the Public Sciences : The consequences of establishing research evaluation systems for knowledge production in different countries and scientific fields », *in* R. WHITLEY et J. GLÄSER, *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*, Dordrecht, Springer, 3-27.