

REPÈRES

Peut-on mesurer l'excellence académique au plan mondial ?

Dominique Desbois*¹

*« Miroir, gentil miroir, dis-moi qui est la plus belle ? »
Blanche Neige et les sept nains d'après les Frères Grimm*

Le classement des universités publié par l'Université Jiao Tong de Shanghai, communément appelé « classement de Shanghai² », se présente comme une publication à vocation académique réalisée par une équipe universitaire indépendante sans aucun soutien financier extérieur. Il entend classer au plan mondial les institutions académiques à partir de données publiques et vérifiables.

Initiée en 2003, la quatrième édition du classement de Shanghai est parue en août 2006 : première université française, Paris VI (Pierre et Marie Curie) occupe la 45^e place du classement. Sa consœur, l'université Paris XI (Orsay) occupe le 64^e rang tandis que l'université Strasbourg I occupe le 96^e rang et que, dernière institution académique française à figurer dans le Top 100, l'École normale supérieure occupe une 99^e place peu glorieuse. Plus loin, dans le peloton, on retrouve les universités Paris VII (140^e rang), Grenoble 1 (175^e), l'École Polytechnique (207^e), les universités Lyon 1 (266^e), Méditerranée (271^e), Montpellier I (273^e), Paris V (280^e), Toulouse III (289^e), l'École des Mines (309^e), Supélec (310^e), les universités Bordeaux I (356^e), Nancy I (373^e), Paris IX (380^e). Puis, en s'enfonçant dans les profondeurs du classement, on trouve l'École normale de Lyon (406^e), les universités Aix-Marseille I (449^e), Bordeaux II (454^e), enfin, dernière française et non des moindres, l'université de Nice au 477^e rang.

Pour le moins, ces résultats n'apparaissent pas en contradiction avec l'édition initiale du classement qui, dans un contexte de crise vécue par la recherche française, avait suscité une floraison d'articles de presse titrant sur cette déconvenue. En écho à cet impact médiatique, surfant sur le mouve-

* Membre du Comité de rédaction de *Terminal*.

1. Une version de cet article a été publiée dans le numéro 67 de la revue *Recherche, technologie et société*, avril 2007.

2. Site ARWU : <http://ed.sjtu.edu.cn/>

ment des chercheurs de 2004, les responsables des universités françaises ainsi épinglées ont vivement réagi au classement de Shanghai, certains pour réclamer plus de moyens. Ainsi Gilbert Béréziat, président de l'Université Pierre et Marie Curie (Paris VI, premier rang français, 8^e rang européen et 46^e rang mondial dans le classement initial) juge dans une interview au *Figaro*³ que « le classement traduit une réalité : l'insuffisance des moyens consacrés à la recherche ».

À l'autre extrémité du spectre des prises de position, on trouve un « Manifeste », dont il vaut mieux par charité taire l'origine, réclamant comme signal politique fort aux plus hautes instances de l'État, la constitution « d'un Mit à la française », ignorant dans sa superbe le contexte législatif de préparation de la loi d'orientation de l'enseignement supérieur et de la recherche. On appréciera à sa juste mesure la conception essentiellement technocratique et jacobine du fonctionnement des institutions que révèle cette tentative de contournement des instances législatives nationales.

Le « classement de Shanghai » semble avoir eu, pour l'ensemble des institutions académiques françaises, un effet de catalyse socio-institutionnelle aboutissant à une prise de conscience des effets pervers que peuvent engendrer les spécificités françaises en matière d'organisation de la recherche, qu'il s'agisse des universités, des grandes écoles et des établissements publics scientifiques et techniques.

Ainsi, en réaction à l'impact médiatique de ce classement jugé défavorable, certains établissements de recherche ont tenté de modifier leurs pratiques institutionnelles, pour compenser les biais défavorables dont souffre l'évaluation de la recherche française en raison des règles qui président à l'élaboration des indicateurs bibliométriques, en particulier l'incontournable facteur d'impact de l'ISI⁴. L'Inserm a édité une charte des publications précisant les règles de rédaction des adresses et des affiliations à l'intention de ses chercheurs selon un principe hiérarchique supposé permettre une meilleure visibilité de l'Inserm. Une lettre du directeur général de l'Inserm vient appuyer le dispositif auprès de l'ensemble de la communauté de travail par un avertissement explicite⁵.

Dépassant l'impact médiatique initial et la revendication catégorielle en déplorant l'absence d'analyses de fonds dans les médias français⁶, François Orivel, chercheur à l'IREDU souligne le caractère stratégique de la taille des institutions académiques pour figurer honorablement dans le « classement de

3. *Le Figaro*, 20/10/2005 : « Gilbert Béréziat : le système universitaire français ne marche plus ».

4. Cf. *Médecine/Sciences* 2004;20:909-15, « Indicateurs bibliométriques : Réalités, mythes et prospective », Lesya Baudoin, Nicole Haeffner-Cavaillon, Nicole Pinhas, Suzy Mouchet, Claude Kordon. *Nature*, vol. 438, 1 décembre 2005, « Automated grading of research performance clearly fails to measure up », N.Haeffner-Cavaillon, C.Graillot-Gak, C.Bréchet.

5. Christian Bréchet, « Je dois dire que dans le futur je serais attentif aux libellés des adresses dans les publications issues d'unités Inserm, en partenariat avec les universités et les hôpitaux », ref. CB/JGG n° 1765, octobre 2005.

6. « Pourquoi les universités françaises sont-elles si mal classées dans les palmarès internationaux ? » François Orivel, IREDU/CNRS, Université de Bourgogne, Notes de travail de l'IREDU, avril 2004.

Shanghai ». Si l'on considère la taille simplement en termes d'étudiants inscrits, on note qu'effectivement peu d'universités françaises sont susceptibles d'atteindre les seuils critiques. La seconde explication à cette déconvenue réside selon François Orivel dans le traitement uniforme des universités pour l'attribution de ressources. Cela tend à homogénéiser leurs performances sur un niveau moyen qui ne présente guère de visibilité dans un type de classement fondé quasi exclusivement sur des critères académiques élitistes.

Les autres remarques formulées par cet économiste de l'éducation s'attachent à mettre en lumière les maux dont souffre l'enseignement supérieur dans notre pays. Selon François Orivel, la partition de l'enseignement supérieur entre universités et grandes écoles est préjudiciable au recrutement des premières sans améliorer la visibilité des secondes. En effet, fait remarquer ce chercheur, les doctorants sont issus dans leur grande majorité d'une université. Cependant, ils ne font pas parti des « meilleurs » étudiants de leur classe d'âge recrutés par les grandes écoles dont le diplôme en cinq ans assure des rémunérations bien supérieures à une thèse sanctionnant huit années d'étude. Pour les meilleurs étudiants, il n'y a donc pas d'incitation financière à effectuer un doctorat.

Quant aux grandes écoles, si elles attirent les meilleurs étudiants et les meilleurs enseignants, elles n'ont malheureusement ni la taille critique ni la pluridisciplinarité nécessaire pour figurer dans le classement de Shanghai à un rang qui reflète la qualité de leur enseignement et de leur recherche. D'autre part, elles n'apparaissent pas, sauf de façon marginale (pour 10 % d'entre elles), comme des lieux centrés sur l'excellence de la recherche.

À ces prises de position, ont succédé des critiques, formulées par des chercheurs en bibliométrie, argumentant à partir des biais présentés par les indicateurs utilisés pour l'évaluation de l'activité scientifique. En effet, contrairement à d'autres classements internationaux voire nationaux dont la notoriété est plus ancienne (par exemple, le classement du *Times*), le classement de Shanghai est construit à partir des activités de recherche scientifique et, de ce fait, accorde une place prépondérante aux indicateurs bibliométriques. Certaines de ces critiques émanent d'universitaires et de chercheurs⁷ reconnus pour leur expertise dans le domaine de la mesure de l'activité scientifique.

Afin d'en saisir la portée, rappelons tout d'abord brièvement l'objectif affiché du classement de Shanghai et les critères pris en compte par ce classement. Le but affiché est de détecter les institutions universitaires de classe mondiale, en positionnant l'université de Shanghai et les autres universités chinoises par rapport à cette sélection, afin d'identifier d'éventuels partenaires pour des collaborations internationales. Les critères utilisés sont plutôt élitistes puisque la qualité de la formation est mesurée par le nombre d'anciens élèves (Alumni) ou d'équipes (Award) ayant eu le prix Nobel ou la

7. Voir notamment les articles publiés sur ce sujet en 2005 et 2006 par la revue spécialisée *Scientometrics* (ISSN 0138-9130).

médaille Fields⁸, et le nombre d'enseignants parmi les chercheurs les plus fréquemment cités dans une liste de vingt et une spécialités académiques (HiCi), tandis que l'activité de recherche s'apprécie à l'aune du nombre d'articles publiés dans la revue *Nature & Science* (N&S), et du nombre d'articles référencés dans le Science Citation Index (SCI) ou le Social Science Citation Index (SSCI).

Enfin, la performance académique globale de l'institution, mesurée par la somme des scores obtenus pour les indicateurs précédents, est rapportée à la taille de l'institution exprimée par le nombre de scientifiques statutaires à plein-temps. Pour les institutions spécialisées dans les sciences sociales, comme la London School of Economics, le poids du critère non utilisable (N&S) est réparti sur les autres critères.

Dès 1988, Moravcsik⁹ pourfendait l'illusion entretenue par certains travaux en quête du Graal de la mesure de l'activité scientifique : disposer d'un outil générique pouvant s'adapter à différents contextes scientifiques à travers une procédure paramétrée et des batteries préfabriquées d'indicateurs. Depuis les premières publications effectuées dans le champ de la mesure de l'activité scientifique, les bibliométriciens n'ont cessé d'émettre des avertissements concernant les limites de validité des indicateurs bibliométriques dans les activités d'évaluation, tout en cherchant des moyens de dépasser ces limites.

De par le caractère plutôt fruste des indicateurs choisis, le classement de Shanghai constitue une excellente opportunité pour ces spécialistes pour discuter de l'état de l'art en matière d'indicateurs bibliométriques et rappeler les règles de bonne conduite en matière d'évaluations et de classement.

Résumons les critiques émises sur le classement de Shanghai, rassemblées soit par des revues traitant des institutions éducatives de l'enseignement supérieur (e.g. *Education Policy Analysis Archives* ou encore *Higher Education in Europe*) soit par des revues bibliométriques. On a bien entendu suggéré d'élargir l'éventail des distinctions scientifiques au-delà du prix Nobel et de la médaille Fields afin d'obtenir un panel plus diversifié. L'utilisation exclusive de *Science & Nature* altère la pertinence du classement pour les sciences humaines et sociales.

La signification et la robustesse d'un certain nombre d'autres critères ont été également mises en question. Les ambiguïtés d'identification des institutions académiques dans les bases de données ISI conduisent également à relativiser certains classements.

Compte tenu de la nature de l'information prise en compte par les études comparatives en matière de production scientifique, les problèmes d'échelle posés par les différences de taille entre les institutions concernées constituent un thème central dans les discussions relatives à la mesure de

8. Distinction décernée tous les quatre ans, équivalente au prix Nobel pour les mathématiciens.

9. Moravcsik M.J. (1988) « Some Contextual Problems of Science Indicators », *Handbook of Quantitative Studies of Science and Technology*, p. 11-30.

l'excellence académique. C'est pourquoi dans leur contribution visant à discuter du bien-fondé du classement de Shanghai¹⁰, Michel Zitt et Ghislaine Filliatreau de l'Observatoire des sciences et des techniques (OST)¹¹ mettent le doigt sur un aspect particulier du classement de Shanghai : les indicateurs utilisés sont tous des fonctions de taille. En effet, les mesures d'excellence utilisées, contribuant pour 70 % au classement, sont définies en termes d'occurrences (articles, auteurs...) atteignant un certain niveau académique. La mesure d'impact, qu'il s'agisse de l'indice SCI ou de l'indice SSCI, s'exprime également en termes de dénombrement. Ainsi, 90 % des contributions au classement sont constituées par des mesures dépendantes de la taille de l'opérateur scientifique ainsi classé.

En dépit de l'émoi suscité par le classement de Shanghai, il convient de rappeler que le classement des institutions académiques n'est pas un phénomène nouveau. En effet, dès 1983, *U.S. News & World Report*, magazine hebdomadaire de référence aux États-Unis, publie le premier classement annuel des meilleurs collèges universitaires aux USA (America's Best Colleges) et des meilleures universités (America's Best Graduate Schools).

Depuis, classements et évaluation d'universités ont fait florès dans la plupart des continents ou pays, qu'il s'agisse de l'Europe, du Japon, de la Russie, de l'Amérique latine, mais aussi de l'Asie : l'estimation faite, antérieurement à la parution du classement de Shanghai, par Kozminski et Sadlak en 2002 dénombre plus de vingt classements de ce type.

Parmi les classements européens les plus connus, signalons le Good University Guide du Sunday Times¹², un des principaux classements effectués au Royaume-Uni, le CHE HochschulRanking¹³ publié par le magazine *Die Zeit* du groupe Stern en Allemagne; les nouveaux états membres de l'Europe ne sont pas en reste, puisqu'en Pologne des classements universitaires sont publiés par *Perspektywy* (mensuel pour les lycéens), *Polityka* (hebdomadaire pour les cols blancs), et *Wprost* (l'équivalent polonais de *Newsweek*). Malgré les critiques et réserves émises par certains chercheurs universitaires se focalisant sur les problèmes méthodologiques¹⁴, la popularité indéniable de ces évaluations comparatives montre qu'elles répondent à un besoin d'information du public concerné : parents d'élèves, étudiants, mais aussi personnels de l'enseignement supérieur et de

10. Zitt M., Filliatreau G. (2007), « *Big is (made) beautiful, some comments about the Shanghai ranking of world-class universities* », à paraître dans Sadlack J. et Liu N.C. (éd.) *The World-Class University and Ranking: Aiming beyond Status*, UNESCO-CEPES, Institute of Higher Education, Shanghai Jiao Tong University, Cluj University Press (communication présentée initialement sous le titre « Being a World-Class University », à la First International Conference on World-Class Universities, Shanghai, 2005).

11. Pour de plus amples informations sur l'OST, voir le site www.obs-ost.fr

12. cf. Good University Guide, Timesonline (www.timesonline.co.uk/).

13. Classement publié par le Centrum für Hochschulentwicklung, (cf. www.daad.de/deutschland/studium/hochschulranking).

14. Cf. concernant en particulier le classement de l'U.S. News & World Report, l'article de Marguerite Clarck « Quantifying Quality: What Can the U.S. News and World Report Rankings Tell us About the Quality of Higher Education ? » *Education Policy Analysis Archives*, Vol. 10 n°16, 20 mars 2002.

la recherche en quête d'une affectation. La massification de l'enseignement supérieur et son orientation progressive vers le marché du travail par le développement de formations à vocation professionnelle incitent étudiants, parents, personnels, institutions académiques et universitaires, employeurs et gouvernements à vouloir préciser le « statut » des différentes universités, grandes écoles, instituts universitaires et plus généralement de l'ensemble des institutions d'enseignement supérieur. Pour les usagers de l'enseignement supérieur que sont les parents d'élèves et les étudiants, cette demande est soutenue par une logique de retour sur investissements, en raison non seulement du coût direct des études, mais également de leur coût d'opportunité.

S'engouffrant dans le marché ouvert par l'*U.S. News and World Report*, les éditeurs de magazine d'autres pays ont rapidement publié leurs propres évaluations des institutions nationales de l'enseignement supérieur. Depuis ces deux dernières décennies, les institutions d'enseignement supérieur ont donc été soumises non seulement aux évaluations du secteur privé, commanditées par les médias, mais également à différents audits commandités par des gouvernements ou des institutions de financement.

Ces pratiques d'évaluation, fondées en partie sinon en totalité sur des méthodologies de classement et des indicateurs statistiques, peuvent avoir une influence notable sur la qualité des formations et de la recherche, amenant des évolutions souhaitées car jugées conformes à l'intérêt public, mais également porteuses d'effets pervers en ce qu'elles sont susceptibles d'influencer significativement les politiques d'accréditation des départements universitaires et les stratégies d'accès des étudiants et des personnels de la recherche et de l'enseignement supérieur.

En rétroaction, parmi les effets les plus patents attribuables à cette culture de l'évaluation, on enregistre une compétition nationale mais aussi internationale entre institutions académiques pour attirer les meilleurs chercheurs, les meilleurs enseignants et les meilleurs étudiants. En raison de la massification de l'enseignement supérieur, il apparaît donc crucial pour l'information du public et pour la qualité du débat démocratique sur les politiques de l'éducation que de telles évaluations soient réalisées de manière rigoureuse, que leurs limites de validité soient précisées, et que leurs méthodologies et résultats soient librement accessibles afin qu'ils puissent être soumis à des processus de validation contradictoires. Dès 2002, des recommandations ont été émises par la première conférence internationale consacrée à ce sujet tenue du 13 au 15 juin à Varsovie¹⁵. En juin 2005, Anthony van Raan dans son adresse à la première Conférence sur les Universités de rang mondial¹⁶, malgré les

15. « Summary Report of the Invitational Roundtable on Statistical Indicators for the Quality Assessment of Higher/Tertiary Education Institutions: Ranking and League Table Methodologies » Jamie P. Merisotis, *Higher Education in Europe*, Vol. 27, n° 4, 2002.

16. « Challenges in Ranking of Universities », First International Conference on World Class Universities, Shanghai Jiao Tong University, Shanghai, June 16-18, 2005. Pour en savoir plus, cf. l'article d'Anthony van Raan intitulé « Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods », *Scientometrics*, vol. 62, n°1, pp. 133-143.

difficultés inhérentes à la mesure de l'excellence, semble considérer le classement des deux cents premières universités de rang mondial comme un nouveau défi au plan méthodologique et pratique : de nouvelles méthodes¹⁷, des indicateurs plus précis peuvent-ils permettre d'aboutir dans cette quête d'une métrique de l'excellence dont il conviendrait également d'élucider la dynamique ?

Les problèmes les plus ardues relèvent de la nécessaire distinction entre mesure de l'activité scientifique, mesure de la productivité scientifique et mesure du progrès scientifique. Nonobstant les difficultés rencontrées, la première mesure s'effectue sans trop d'états d'âme à partir d'indicateurs tirés des bases de données de l'ISI. Cette source quasi monopolistique a donné lieu à une littérature pléthorique. Par contre, la seconde mesure demeure un problème ouvert en particulier au niveau macroéconomique.

En effet, la mesure de la productivité basée sur des ratios d'input/output requiert une connaissance fine du profil des dépenses publiques et du type d'activités scientifiques mesurées.

À propos d'une comparaison France/Royaume-Uni, Rémi Barré¹⁸ illustre quelques-unes des chausse-trappes posées par des ratios rapportant les résultats aux inputs pour la mesure de la productivité dans le contexte de notre espace européen de la recherche. Si les dépenses sont équivalentes entre les deux pays pour les structures universitaires, le montant de la dépense publique pour la recherche non universitaire est différent entre la France et le Royaume-Uni, en particulier pour le budget civil de recherche appliquée et celui de la recherche militaire.

Or, pour des raisons de confidentialité bien légitimes, ces secteurs scientifiques alimentent assez peu le Science Citation Index. Si l'on corrige les différences de structures par le poids relatif des publications académiques dans chacun de ces secteurs, universitaire et non universitaire, on en déduit un budget de recherche en France supérieur de 12 % à celui du Royaume-Uni, estimation beaucoup plus réaliste que les 28 % conduisant à une estimation surprenante (deux fois plus élevée au Royaume-Uni qu'en France) de la productivité scientifique en 1996¹⁹. Au numérateur du ratio de productivité, le dénombrement des citations de publications scientifiques par l'indice SCI comme mesure de l'output permet-il d'effectuer des comparaisons entre la France et le Royaume-Uni ? Les sciences médicales représentent 30 % des publications couvertes par l'indice SCI, le Royaume-Uni est fortement spécialisé en recherche clinique (12 % des articles mondiaux) alors que la France l'est beaucoup moins (5 %). Il convient alors de corriger le biais

17. Citons en particulier, l'apport des méthodes DEA (Data Envelopment Analysis) dans l'analyse des questions d'efficacité productive en termes de « frontière de production », cf. « Data envelopment analysis as a tool for constructing scientometric indicators », S. Rousseau et R. Rousseau, *Scientometrics*, vol. 40, n° 1, 1997.

18. « Sense and nonsense of S&T productivity indicators », Barré R., *Science and Public Policy*, vol.28, n°4, 259-266, 2001.

19. « The Scientific Investments of Nations », May R.M., *Science* n° 281, pp.48-51, 1998.

induit par la différence entre ces deux résultats par une pondération tenant compte des profils de publication. Ainsi, en dépit des améliorations apportées par les tentatives de normalisation internationale²⁰, la mesure de la productivité scientifique suppose une connaissance fine du profil des dépenses publiques et du type d'activités scientifiques mesurées. Quant à prendre la mesure du progrès scientifique... qui oserait aujourd'hui se confronter à une telle entreprise ?

En attendant, une question d'actualité demeure posée à la recherche française : comment, à partir d'une étude dont le choix des critères et du processus de classement a fait l'objet de critiques scientifiques convergentes exprimant des doutes sérieux sur sa validité, peut-il émerger un fait saillant aussi manifeste et persistant que la très médiocre visibilité des institutions académiques françaises ?

Cette question nous ramène aux interrogations plus fondamentales sur la crise que traverse depuis plusieurs années le système français de recherche et d'enseignement supérieur. Doivent y apporter des réponses non seulement les scientifiques, bien entendu les autres acteurs professionnels de ce système national de recherche et d'enseignement mais, plus largement, l'ensemble des groupes sociaux concernés afin de garantir la pluralité des points de vue indispensable au fonctionnement démocratique de notre société. ■

20. Devenu une référence en la matière, le manuel de Frascati en est à sa 6^e édition, publiée par l'OCDE.