

## Les rankings : une mesure de la qualité de l'enseignement supérieur ?

4<sup>ème</sup> après-midi Qualité ImpeQes, lundi 9 novembre 2009

### Introduction :

*Shanghai, Times Higher Education Supplement, Times on line, Financial Times, CHE Excellence ranking, classement EQUIS des écoles de commerce, ... Les classements réalisés par des universités ou des médias se multiplient et inquiètent. De plus en plus d'établissements veulent en être mais, dans le même temps, les critiquent à l'envi ! Les critères utilisés sont à la fois très précis (performance de la recherche, nombre de prix Nobel, articles publiés dans les revues de recherche, qualité des infrastructures, niveau de satisfaction des étudiants, etc.) mais aussi, d'un système à l'autre, changeants et contradictoires... Parfois ces classements deviennent même les révélateurs de situations tendues comme le classement des universités espagnoles réalisés par l'IAIF (Instituto de Análisis Industrial y Financiero) de l'Université Complutense de Madrid (publié en mai 2009) qui a bouleversé (et agité !) le paysage universitaire espagnol. Pourtant, ce classement est un ranking sophistiqué incluant 32 variables regroupées en 10 facteurs, conformément à la méthode d'analyse factorielle utilisée au sein de l'OCDE !*

*Les résultats des rankings sont comparables au fantôme du manoir hanté, gage d'ancienneté de l'édifice : nous savons bien qu'il est là, nous en avons peur mais nous aspirons à le rencontrer, parfois même nous payons pour cela. Et dès que nous l'apercevons, nous nous enfuyons en hurlant, pour nous rassurer, que les fantômes n'existent pas ! Faut-il donc avoir peur des rankings ? Faut-il les fuir ? Ou au contraire chercher à les apprivoiser et à les comprendre ? Faut-il jouer leur jeu ? Sont-ils vraiment des indicateurs fiables de qualité ?*

---

Catherine Dehon (ULB) : Que mesure le ranking des institutions d'enseignement supérieur proposé par l'Université de Shanghai Jiao Tong ?

Catherine Dehon, en collaboration avec A. McCathie et V. Verardi, a rédigé un article (prochainement publié dans *Scientometrics*) qui consiste en un exercice statistique visant une analyse des facteurs sous-jacents du ranking de Shanghai.

Son choix de travailler sur les critères de classement du ranking de Shanghai et non sur ceux du ranking de *Times Higher Education* s'explique par les éléments suivants :

- le ranking de Shanghai, contrairement à celui de *Times Higher Education*, n'a pas de but lucratif. Les critères de classement semblent dès lors être moins opaques.
- Bien que les données brutes du ranking de Shanghai ne soient pas accessibles, il semble faire preuve de davantage de transparence que celui de *Times Higher Education* notamment car il publie sur son site internet les scores attribués aux établissements. Le ranking de *Times Higher Education*, quant à lui, se base pour 50 % de ses variables sur des enquêtes auprès d'académiques et de chercheurs sans qu'aucune information soit diffusée sur la méthodologie employée pour mener et traiter ces enquêtes.

L'objectif initial du ranking de Shanghai (qui existe depuis 2003) était de mesurer l'écart entre les universités chinoises et les universités considérées comme les universités les plus performantes à l'échelle mondiale.

Variables employées par le ranking de Shanghai :

- 1) 10 % *alumni* : nombre de Prix Nobel - en-dehors des Prix Nobel de littérature et de paix - et médailles *fields* obtenus par les anciens étudiants de l'établissement

- 2) 20 % *HICI* : nombres de chercheurs repris dans la liste des *highly cited researchers* de Thomson (qui décline la recherche en 21 domaines : 19 domaines relevant des sciences dites « dures » ; 2 domaines relevant des sciences humaines)
- 3) 20 % *N&S* : nombre d'articles publiés dans les périodiques *Nature* et *Sciences*
- 4) 20 % *PUB* : nombre d'articles indexés dans *Sciences Citation Index-Expanded* (SCIE) et *Social Science Citation Index* (SSCI)
- 5) 10 % *PCP* : moyenne pondérée des cinq scores précédents divisée par le nombre d'équivalents temps-plein (*academic staff*) par établissement

Quelques remarques sur le choix des variables :

- les variables 1) et 2) relèvent du prestige et sont fortement liées à l'ancienneté de l'établissement
- la définition de l'*academic staff* varie fortement d'un pays à l'autre (voire, à l'échelle de la Belgique, entre les différentes communautés)
- les variables 1) à 5) sont des variables dites de production alors que la variable 6 est une variable dite d'efficacité : il ne semble pas méthodologiquement robuste d'employer conjointement deux types de variables pour mesurer un même résultat
- les différentes variables ne sont pas connues pour les différentes universités classées. Par contre, pour chaque variable, l'on connaît l'université qui est considérée comme la plus performante. Cette dernière se verra attribuer le score maximum (100). Les autres universités se verront appliquer une règle de trois pour obtenir un score reporté sur 100. D'un point de vue statistique, ce choix méthodologique ne peut être considéré comme robuste.
- les auteurs du ranking estiment que la variable 6) est celle qui mesure la qualité de l'enseignement. De l'avis de Mme Dehon, cette variable mesure davantage la « production » de la recherche et le prestige de l'université classée.

Mme Dehon estime qu'il serait davantage informatif de pouvoir disposer des scores attribués à chaque université pour chaque variable plutôt que le rang octroyé aux différentes universités.

De plus, deux sortes de classements se côtoient au sein du même ranking :

- les rangs 1 à 100 relèvent effectivement du ranking
- les rangs 100 à 500 sont divisés en sous-classements de 50 rangs au sein desquels les universités sont classées par ordre alphabétique

Cette façon de procéder implique différentes mécompréhensions de l'information transmise (cf. article du *Vif l'Express* de la semaine du 9 novembre).

Les critiques classiques adressées au ranking de Shanghai :

- les universités sont des organismes complexes : leurs missions sont diverses (enseignement, recherche et services à la collectivité) et elles ne peuvent se limiter à la recherche
- il semble que le ranking de Shanghai entretienne un certain biais en faveur des universités anglophones (au sein des variables liées à la publication, seuls les articles publiés en anglais sont comptabilisés)
- un autre biais qui semble entretenu par ce ranking est en défaveur des sciences sociales et des *Humanities*
- certaines mesures ont été réalisées de manière erronée et l'on note une absence d'analyse de sensibilité
- la non-disponibilité des données brutes par université
- l'utilisation conjointe de variables de production et d'efficacité
- une tendance à la normalisation illustrée par le choix méthodologique de mesurer l'ensemble des universités par rapport à l'université considérée comme étant la meilleure
- le choix arbitraire de pondérer la moyenne (variable 6)

- aucune variable d'input (telle que le financement accordé à l'université classée, l'autonomie octroyée aux universités dans leur contexte national, etc.) n'est prise en considération
- la nécessité d'un intervalle de confiance (au-delà du rang 150 approximativement, les différences de scores entre les universités sont infimes : dès lors quelle est la véritable différence entre l'université classée 170<sup>e</sup> et l'université classée 210<sup>e</sup> ?)

L'objet de la recherche menée par Catherine Dehon était de répondre à la question suivante : peut-on considérer qu'un seul indicateur composite est capable de résumer l'excellence de la recherche au sein des établissements d'enseignement supérieur ?

Pour tenter de répondre à cette question, elle a employé la méthode statistique dite d'analyse en composantes principales. L'objectif annoncé était d'établir par ce procédé une variable qui, à elle seule, reprendrait un maximum de l'information contenue dans les cinq premières variables employées par le ranking de Shanghai. Cette nouvelle variable consiste en la meilleure combinaison linéaire des cinq premières variables employées par le ranking de Shanghai. Elle reprend 64% de l'information globale du ranking. Malgré la faible homogénéité de l'échantillon, cette variable a été appliquée aux dix premiers rangs du ranking de Shanghai. L'application de cette variable a permis de mettre en évidence deux facteurs sous-jacents.

Par ailleurs, il est important de préciser que la pondération accordée à chacune des variables (exprimée ci-dessus en pourcentages, au début de chaque variable) influence fortement le résultat du classement. A titre d'exemple, en variant le poids accordé à la variable 1), l'École Normale sup de Paris peut passer de la 250<sup>e</sup> place à la 20<sup>e</sup> !

Catherine Dehon insiste sur la prudence nécessaire dont il faut se munir pour employer ce ranking. Selon elle, s'il ne faut pas avoir peur des rankings, il faut cependant avoir peur de ceux qui les utilisent !

### Denis Bouyssou :

Quelques remarques préalables :

- Au sein des 20 premiers rangs du ranking de Shanghai, Denis Bouyssou souligne que l'on retrouve 17 universités américaines (situées en Nouvelle Angleterre, en Californie et à Chicago), 2 universités anglaises et l'université de Tokyo.
- Si l'on ne reprend que les universités européennes, l'on considère que les 20 premières universités européennes sont principalement des universités anglaises.

Quels sont les impacts de ces rankings ?

- Un impact médiatique certain
- Un impact académique non négligeable : les universités américaines ne manquent pas d'afficher leur rang sur leur site internet et, dans certains cas, adaptent la rémunération des doyens d'universités en fonction du résultat obtenu au ranking de Shanghai.
- Des impacts futurs en augmentation : de plus en plus de décideurs politiques emploient les résultats de ce rankings à des fins de promotion de leurs propres stratégies politiques (ainsi le président Nicolas Sarkozy a fait savoir sa volonté de voir intégrer dix universités françaises dans le top 100 de rankings internationaux).

Denis Bouyssou souligne que la méthodologie employée par le ranking de Shanghai a beaucoup évolué durant ses six années d'existence. Dès lors, les modifications qui peuvent être notées au sein des rangs d'une année à l'autre ne peuvent être interprétées avec transparence puisque les critères méthodologiques employés ne sont pas équivalents.

### Une série de microdécisions non documentées

Il pointe également que les six variables ne sont pas employées pour l'ensemble des universités présentes dans le ranking. Lorsque le nombre d'ETP *academic staff* n'est pas connu pour une université (c'est un cas récurrent car il s'agit de données difficiles à obtenir), ce critère n'est pas pris en compte et l'université n'est classée que sur la base des cinq autres variables. De même, les universités offrant uniquement des programmes en sciences humaines et sociale ne publient pas de *Nature* ou *Sciences*. Dès lors, cette variable n'est pas prise en compte. Selon les cas donc, une université peut être analysée selon 4, 5 ou 6 variables.

Une autre critique émise par Denis Bouyssou à l'encontre de ce classement : le nombre de Prix Nobel attribués aux *alumni* de l'université. Le paysage de l'enseignement supérieur et des universités est en perpétuelle évolution : lorsqu'une université est englobée dans un plus grand ensemble ou qu'au contraire elle est subdivisée en plusieurs universités, quelle est l'université actuelle qui pourra comptabiliser le Prix Nobel ? Illustration : le Prix Nobel d'Albert Einstein. Il a étudié à l'Université de Berlin. Le paysage universitaire berlinois a subi de nombreuses modifications à l'issue de la seconde guerre mondiale (création de deux universités distinctes : Universität Humboldt et Freie Universität Berlin). Ces deux universités réclament la paternité du Prix Nobel d'Albert Einstein. Ne sachant les départager, le ranking de Shanghai ne les a attribué ni à l'une ni à l'autre et a retiré ces universités de son classement...

Dans le cas de la bibliométrie (*N&S*) : lorsqu'un article est cosigné, le score qu'il représente est multiplié car il est attribué à chacun des auteurs de l'article. Dès lors, un article cosigné par cinq chercheurs pèse cinq fois plus aux yeux du classement qu'un article rédigé par un seul auteur. A ce stade, Denis Bouyssou estime qu'il vaudrait mieux évaluer l'impact des articles (nombre de citations) que leur nombre.

Quant au choix des bases de données : le ranking de Shanghai comptabilise les articles publiés dans SCIE et SSCI. D'aucun estime que ce choix n'est pas le plus pertinent. D'autres bases d'indexation, telles que Scopers d'Elsevier, semblent employer des méthodologies plus robustes d'indexation. En effet, SCIE et SSCI n'harmonisent pas leurs listes d'indexation, puisqu'elles se basent uniquement sur la signature des chercheurs. Si un chercheur signe au nom de son laboratoire de recherche et non au nom de l'université à laquelle le laboratoire est relié, cet article ne sera pas attribué à l'université classée.

Lorsque l'on envisage le laboratoire identifié UMR XXX de l'Université de Paris VI (anciennement Université Pierre et Marie Curie), cela peut entraîner de nombreuses possibilités de signatures pour les articles publiés : UMR XXX, Université de Paris VI, Paris VI, Paris 6, Université de Paris 6, UPMC, etc. Toutes ces signatures seront autant d'entrées dans l'index de ces bases de données.

L'établissement du ranking de Shanghai est jalonné de microdécisions non documentées de ce type. Ce truchement d'éléments parasites induit un certain obscurantisme dans le traitement des données des établissements.

### Qu'est-ce qu'une bonne université ?

Qu'est-ce qu'une bonne université aux yeux du ranking de Shanghai : une grosse institution, vieille, ayant connu peu de changements institutionnels, portant un nom simple et unique (si possible en anglais ou à tout le moins dans un alphabet ne comportant aucun signe diacritique), englobant les laboratoires de recherche en son sein et non en sa périphérie, etc. Il s'agit en fait de la définition des universités de l'Ivy League !

Enfin, Denis Bouyssou rappelle que, comme dans toute évaluation, le comportement de l'évalué se modifie par le principe-même de l'évaluation. Puisque les universités savent qu'elles seront évaluées notamment sur le nombre de leurs publications, il ne leur reste plus qu'à multiplier la publication d'articles, quitte à en diminuer la qualité, la longueur ou l'impact... L'université doit aussi, pour satisfaire aux critères du ranking de Shanghai, diminuer les budgets octroyés aux sciences humaines et sociales et renforcer les budgets octroyés aux sciences biomédicales.

Les gouvernements, quant à eux, ont intérêt à fusionner leurs universités entre elles. A titre d'illustration : si le gouvernement français décidait de fusionner Paris 6 et Paris 11, cette nouvelle institution serait classée entre le MIT et Caltech... Si on ajoutait à ce nouvel organisme Paris 5, cette institution fictive se retrouverait entre Harvard et Stanford ! Enfin, si l'on y ajoutait Paris 7, cette « super institution » équivaldrait à Harvard !!! Si on ajoutait à cette nouvelle entité un ensemble d'instituts de recherche, cet ensemble dépasserait Harvard et prendrait la tête du classement... Que mesure-t-on dès lors ? La grosseur d'une institution ? Son prestige ? La qualité de sa recherche ?

### Harvard est-elle la référence absolue et constante ?

Harvard bénéficie chaque année du maximum (100) et que les autres universités sont évaluées par rapport à elle. Les performances d'Harvard évoluent (même de manière infime) chaque année et, dès lors, le référentiel par rapport auquel sera évalué l'ensemble des universités évolue également...

---

### Séance de questions/réponses

Existe-t-il de bons indicateurs permettant de classer les établissements d'enseignement supérieur ?

La première question à se poser n'est pas celle des indicateurs mais celle du sens : que veut-on classer ? Pour qui ? Pour quoi ? Ensuite il faut se donner les moyens d'apprécier l'impact des différents choix méthodologiques et des indicateurs sur le résultat final voulu.

Dans leur forme actuelle, les grands rankings internationaux visent comme destinataires non les étudiants mais bien les gouvernements qui s'en emparent.

D'autres classifications visent davantage la diffusion d'une information complète à destination des étudiants. En Allemagne, ils ont mis au point une classification intitulée CHE qui décline, par discipline, un très grand nombre de variables. Le consortium responsable de ce classement a été consulté pour évaluer la faisabilité de ce type d'outil au niveau européen.

Ces classifications et/ou rankings doivent-ils être pris en compte pour attribuer le financement des universités ?

En Australie, le financement de la recherche au sein des universités se basait auparavant sur la qualité de recherche (mesurée notamment via les citations des articles). Ensuite, les critères ont été modifiés et se sont davantage basés sur la quantité d'articles produits par les universités. A la suite de cette modification, on a pu observer une augmentation massive du nombre d'articles publiés mais également une diminution significative de la qualité de ces articles...

Serait-il envisageable d'insérer au sein des variables du ranking de Shanghai des données relatives aux « inputs » contextuels des universités (financement, autonomie, etc.) ?

Techniquement, cela pourrait se faire sans trop de difficultés. Mais d'autres difficultés sont à présager : le financement doit-il être considéré comme un input ou comme un output ? Fournit-on des prestations de qualité parce qu'on bénéficie d'un financement adéquat ou reçoit-on un financement adéquat parce que l'on fournit des prestations de qualité ?

Malgré tout, ces données sont particulièrement pertinentes à prendre en compte. Peut-on évaluer avec les mêmes variables Harvard et une université du Laos en sachant le budget annuel d'Harvard équivaut à 1,2 fois le PNB du Laos ?

Actuellement, un projet européen vise l'élaboration d'une classification européenne des universités contextualisée, par pôles de spécificités pour tout établissement d'enseignement supérieur accrédité/évalué par une agence nationale d'évaluation de la qualité.

Ne faut-il pas dénoncer officiellement le caractère non-scientifique de ces classements ? Denis Bouyssou et certains de ses collègues ont écrit des articles scientifiques et vulgarisés qui ont été soumis à *Scientometrics*. Ces articles ont également été envoyés au service chargé d'élaborer le ranking de Shanghai. Il leur a été répondu qu'ils pouvaient, s'ils le souhaitaient, élaborer leur propre classement avec leurs propres critères.

L'Ecole des Mines de Paris a, pour sa part, proposé un autre type de classement en 2005 : la variable employée pour élaborer ce classement était le nombre d'*alumni* étant aujourd'hui P.D.G. d'une des 500 plus grandes entreprises mondiales (conformément au classement établi par *Fortune*). Bien sûr, ce classement n'était pas non plus un classement fiable, mais ses résultats étaient particulièrement contrastés par rapport à ceux du ranking de Shanghai : la France était particulièrement bien classée (ce qui s'explique notamment par les robustes réseaux d'*alumni* qui sont développés au sein des universités françaises).

Qui finance le ranking de Shanghai ?

Selon le discours officiel, l'équipe qui établit le ranking annuel n'est composée que des 4 personnes et que son objectif est purement académique. A l'origine, le ranking de Shanghai a bénéficié d'une impulsion du gouvernement chinois qui souhaitait réformer son système d'enseignement supérieur. Cependant, à l'heure actuelle, le contexte de rédaction de ce ranking reste opaque...

---

Compte-rendu rédigé par Eva Jaroszewski, attachée à la Cellule exécutive de l'AEQES.