



Les cafés de la statistique

"La statistique éclaire-t-elle les questions de société" ?

Soirée du 8 décembre 2009

Classements et palmarès

Synthèse des débats ^[*]

Depuis plusieurs années, des statistiques servent fréquemment à établir des classements, souvent publiés à grand bruit par la presse sous forme de palmarès : classement des pays, des régions ou des villes (du point de vue du parc immobilier, de la qualité de la vie, du marché du travail ...), classement des lycées, des universités, des hôpitaux, des valeurs mobilières, etc....

Ces classements peuvent être utilisés au sein d'organismes à des fins de bonne gestion ou d'amélioration du service rendu, directement, par une prise de conscience chez les responsables des entités classées ou via des mécanismes d'émulation. La presse a bien sûr d'autres buts avec ces publications et ce que les lecteurs en font est encore autre chose.

Comment sont établis ces classements ? Quelle en est la valeur ? Sont-ils déterminants pour les décisions privées ou publiques ? Leur publication sous forme de palmarès a-t-elle des effets vertueux ou pervers, et lesquels ?

Invité :

Claude Thélot,

Ancien directeur de l'évaluation au ministère de l'éducation.

Exposé introductif :

Les classements souvent présentés résultent d'indicateurs d'évaluation. Mais la visée de ceux-ci débordent largement cette utilisation. L'orateur s'appuiera principalement sur l'exemple des lycées, qu'il connaît bien, tout en soulignant que cela s'applique à l'évaluation de toutes unités localisées du service public : écoles, hôpitaux, tribunaux, etc.

^[*] Pour l'exposé liminaire, le présent texte est reformulé à partir des notes du secrétariat suivant le plan de l'orateur. En revanche, le contenu des échanges est structuré en quelques thèmes, sans suivre l'ordre chronologique. Par ailleurs, on a choisi de ne pas attribuer nominativement les propos échangés. Ceux-ci ont été reconstitués à partir des notes du secrétariat sans reprendre leur formulation détaillée. Lorsqu'un point est évoqué sous forme d'une question, ce qui vient ensuite ne retrace pas la seule réponse de l'invité, mais l'ensemble des contributions des participants.

Depuis 1994, le ministère de l'éducation nationale (direction de l'évaluation et de la prospective) établissait des "indicateurs pour le pilotage des établissements d'enseignement" (IPES). L'intention de fournir un instrument pour guider l'action répondait aussi à une demande forte. Ainsi, plusieurs années auparavant, *Le Monde de l'Education* faisait sa vente annuelle sur le palmarès des lycées selon leurs résultats au bac. Là naissaient des réputations, qui n'étaient pas toujours correctement fondées. Il ne fallait pas laisser la rumeur s'établir : souvent, peu fondée, elle fait du mal. Ou, si elle correspond à une réalité, celle-ci peut changer : assez vite même, tandis que la réputation est rémanente. Mais, ce désir de répondre à une demande externe s'ajoutait aussi au besoin, interne au système éducatif, d'améliorer son efficacité. Ceci ne s'obtient pas par des circulaires. Le pilotage doit se faire par l'évaluation. Or, l'évaluation suppose un regard externe à l'établissement ; mais elle fait retour aux acteurs et, par effet miroir, leur permet d'améliorer leurs pratiques. Elle leur offre des comparaisons à la fois dans le temps et dans l'espace, sans qu'il soit nécessaire d'avoir un palmarès (ou un pilori) général et public.

Pour répondre à ces deux besoins, interne et externe, il fallait donc que le service public construise des indicateurs ; sans pour autant en avoir le monopole. Et, cela supposait une information exacte. Or, celle-ci n'était pas à la base de qualité suffisante. D'où une première diffusion avant 1994, pendant 3 ans, interne, de qualité médiocre encore, mais pour inciter à améliorer les données.

Construction.

Il fallait couvrir quatre domaines :

- l'environnement (la "clientèle" des lycées),
- les coûts,
- les services offerts, le fonctionnement,
- les résultats.

Les indicateurs devaient avoir les caractéristiques suivantes :

- être en nombre restreint, quitte à se voir reprocher d'être simplistes,
- à la fois généraux et spécifiques, prenant en vue les objectifs généraux du système éducatif et les spécificités de chaque établissement,
- à la fois quantitatifs et qualitatifs : le quantitatif pur étant discrédité, ne prenant pas en compte, par exemple le "climat",
- en partie publics, malgré les réticences des responsables et donc les effets pervers qui peuvent en résulter, et malgré la possibilité d'être instrumentalisés. Il fallait en effet, notamment, nourrir le débat public, où l'on alléguait que "le niveau baisse", ou encore faire face à la rumeur sur tel ou tel établissement, évoquée plus haut. Une part de publicité permettait aussi une pression pour améliorer,
- permettre la comparaison temporelle de l'établissement avec lui-même auparavant (bien que l'environnement ou les objectifs soient évolutifs) et entre établissements : non pour pointer les bons et les mauvais, mais pour entretenir une certaine émulation sur les pratiques.

Il fallait aussi pouvoir approcher des concepts complexes :

- 'efficacité financière' (coûts/résultats),
- 'efficacité fonctionnelle' ou 'pédagogique' ; résultat / méthode (par exemple, question de la taille des classes ou de la durée de l'enseignement),

- la “valeur ajoutée”, c’est à dire la situation des élèves à la sortie par rapport à leur arrivée dans l’établissement. Ceci est tout à fait différent d’un taux de réussite au bac ! La question se pose identiquement pour tous les domaines : par exemple, pour un hôpital, il est probable que les sortants sont en moins bonne santé que ceux qui n’ont pas été hospitalisés, mais ce qui compte, c’est qu’ils soient en meilleure santé qu’à leur entrée.

Réalisation actuelle

Pour la réussite au bac, on a calculé trois indicateurs :

- taux de réussite = reçus / présentés,
- taux d’accès = reçus / admis en seconde,
- proportion de bacheliers parmi les sortants du lycée de l’année = reçus / tous sortants (y compris abandons).

Et, les deux premiers de ces indicateurs sont comparés à des ‘*taux attendus*’ : ces derniers prennent en compte la structure de la population d’élèves reçus par l’établissement selon l’âge, le sexe et la catégorie sociale. Pour déterminer cela, on formalise un *modèle*, par types d’établissements comparables. Noter que l’effet “des pairs”, c’est-à-dire la stimulation ou le découragement induit par les autres élèves de la classe est aussi fort que l’effet de l’origine sociale. Aussi, récemment l’effet des pairs, ainsi que le niveau scolaire moyen des lycéens qui sont entrés en seconde figurent également dans le modèle.

On ne peut affirmer qu’on aurait ainsi à proprement parler estimé rigoureusement un résultat, mais on a approché la capacité de l’établissement à accomplir sa fonction, par rapport à la moyenne des établissements qui ont la même structure que lui. Les proviseurs et l’administration s’y sont intéressés. En revanche, on regrette que les enseignants ne s’en soient guère saisis.

Et, Claude Thélot de conclure par deux remarques :

- l’évaluation doit procéder d’un regard extérieur, mais pour nourrir l’autoévaluation ;
- des indicateurs sont une base pour l’évaluation, mais ne constituent par en eux-mêmes l’évaluation.

débat:

Les interventions sont reclassées en quatre grands thèmes :

- A. La construction d’indicateurs ;*
- B. Le champ couvert et le contexte ;*
- C. Publicité et perversion ;*
- D. Initiative et utilisation.*

A. La construction d’indicateurs

Un participant souligne que les indicateurs évoqués par l’intervenant cherchent à mesurer la qualité du service, à la différence, par exemple, de la comptabilité nationale qui se contente de les mesurer à l’aune de ce qu’ils coûtent. Il relève aussi qu’on cherche à atteindre l’effet propre du service en neutralisant l’influence d’autres facteurs, notamment d’environnement. L’invité précise néanmoins qu’on ne prétend pas véritablement mesurer la qualité, mais qu’on en a une approche suffisante pour pouvoir s’interroger à son sujet et tenter de l’améliorer.

Un point notable est le fait de produire plusieurs indicateurs, et non un seul. N'en donner qu'un conduirait à instituer une norme et donc des classements c'est à dire un palmarès ou un ban d'infamie. Tandis que là, une pluralité d'indicateurs – une configuration plutôt qu'une notation – invite à une interprétation plus nuancée de “ la situation de notre établissement”. Ainsi, par un effet de miroir, l'acteur s'interroge sur sa pratique. Dans la suite du débat, C. Thélot aura l'occasion d'insister à nouveau : un indicateur (ni même des indicateurs) ne constitue pas une évaluation. Celle-ci l'englobe, s'appuie dessus, mais va au-delà. Il mentionne le cas d'une académie (Lille) où avait été constituée une équipe d'évaluation itinérante, qui, certes en s'appuyant sur des indicateurs débordait ces derniers pour aller discuter avec les établissements, leur apportant un point de vue externe tandis que leur situation propre était prise en considération.

Mais, jusqu'où multiplier les indicateurs, alors qu'on ne dispose finalement que de peu de paramètres ? C'est un arbitrage et Paul Valéry est cité : « tout ce qui est simple est faux, mais tout ce qui ne l'est pas est inutilisable ! » Si l'on donne trop d'indicateurs, on noie l'interlocuteur. Les acteurs concernés sont les premiers à ne pas supporter la simplification. C'est donc en lien avec les utilisateurs qu'il faut déterminer la juste dose entre pluralité et simplification.

Plus tard, un autre participant s'interrogera sur la pertinence du modèle par lequel on estime les résultats « attendus ». Il lui est répondu qu'en effet, on ne peut affirmer la qualité du modèle dans l'absolu, mais qu'il prend tout de même en compte les facteurs que l'on sait avoir l'influence prépondérante sur les résultats des élèves (sexe, âge, niveaux scolaire et social, ...) tandis que la nationalité ne joue guère en soi. L'âge est un facteur important dans la mesure où le fait d'être “en avance” ou “en retard” résume la carrière scolaire antérieure et révèle une aptitude ou une difficulté propre à l'élève qui facilitera ou freinera sa progression.

B – Le champ couvert et le contexte

Puisqu'on prend l'exemple de l'éducation nationale, quelques questions portent sur l'inclusion de certains types d'établissements : par exemple, l'enseignement privé sous contrat. Il est précisé que, intégré dans le service public par la “loi Debré”, il fournit les mêmes informations et donne lieu aux mêmes indicateurs que pour les établissements publics.

De même, plusieurs propos évoqueront les établissements situés dans des ZEP (zones d'éducation prioritaire). Ceci concerne les écoles primaires et les collèges, plutôt que le lycée et le bac : les indicateurs commentés pour celui-ci doivent être remplacés ici par d'autres appréciant l'acquisition des compétences de base. Toutefois que, le monde évoluant, on demande aussi à l'école de donner de nouvelles compétences. Or, on constate que les progrès dans les ZEP sont finalement assez parallèles à ceux des autres établissements ; c'est-à-dire qu'il n'y a pas de rattrapage. Est-ce à dire que la “politique de la ville”, notamment ce qu'on fait pour les établissements en ZEP serait inutile ? Pas nécessairement : il faut comparer non avec le hors-ZEP mais avec ce qui se passerait s'il n'y avait pas eu de dispositions propres aux ZEP. Ceci n'est pas facile techniquement mais cependant possible ; le plus malaisé étant de le faire comprendre. En fait, de telles dispositions spécifiques aux ZEP sont assez ‘contre-culturelles’, ce qui explique qu'elles soient d'ampleur finalement assez modeste (d'où leur difficulté à être efficace) et doivent en permanence être relégitimées.

Le système scolaire subit directement la polarisation de l'espace social. A cet égard, une politique de carte scolaire, si elle consiste à installer des écoles, des collèges ou des lycées dans les “cités”, aggrave l'assignation des élèves qui y résident.

C – Publicité et perversion

De nombreuses interventions portent sur l'opportunité de rendre publiques les évaluations. Car cela peut conduire à diverses perversions. Un participant en distingue de deux types :

- soit le détournement des indicateurs à des fins autres que celles qui présidaient à leur construction : l'exemple est celui des indicateurs de réussite au bac, conçus pour aider les lycées à prendre la mesure de leur performance et à s'améliorer, et qui a servi aux journaux pour établir des palmarès, dont les parents s'inspirent pour inscrire leurs enfants ;
- soit que cela compromette la signification, voire même l'existence, des résultats. En effet, les responsables auront tendance à communiquer ou non les informations en fonction de l'effet produit par les résultats : utilisés par l'administration pour l'allocation de leurs moyens (on l'a vu dans le domaine de la santé avec le PMSI¹ pour les budgets hospitaliers), soit aboutissant à une stigmatisation de l'établissement.

Ce dernier cas est illustré par l'observation des violences scolaires : le Ministère avait une centralisation d'information (dénommée SIGNA). C'était assez délicat, car il fallait cerner ce qui constituait un incident susceptible d'être signalé ; un rodage était en cours afin de rendre les données suffisamment homogènes dans l'espace et comparables dans le temps ; un hebdomadaire a obtenu les données par établissement et a publié un palmarès. Les chefs d'établissements, mécontents d'être ainsi montrés du doigt, ont refusé alors de communiquer les données. La publicité a abouti là à tuer l'instrument. Par la suite, un autre a dû être monté, avec d'autres dispositions.

Dans un autre domaine encore, celui de l'action sociale, les départements, qui en ont la responsabilité, sont très intéressés, chacun, à savoir ce que les autres font. Mais leur possibilité de modifier la situation n'est de toute façon que fort limitée. Aussi n'ont-ils aucune envie que l'on impute les différences de situations sociales à leur performance respective et qu'on en tire des classements.

Pour l'intervenant, ces usages ou dévoiements dépendent de ce que les utilisateurs de tels résultats peuvent en faire, de leur marge de liberté. Par exemple, dans le cas d'une carte scolaire stricte, si l'on n'a pas le choix de l'établissement, un classement n'a pas beaucoup d'intérêt. Mais il estime que, même s'il y a une marge de liberté et même si des effets pervers sont possibles, mieux vaut quand même diffuser les résultats. Cela vaut mieux que de laisser libre cours à l'interprétation de résultats bruts. Car les médias, en tout état de cause, publieront et commenteront ces résultats bruts. Si un service public d'évaluation ne diffuse rien, et en particulier rien qui permette d'approcher la valeur ajoutée de l'unité évaluée, il est à craindre que le discrédit soit au moins aussi fort.

Un participant souligne qu'il faut néanmoins que les écarts, notamment par rapport aux résultats attendus, soient établis de façon suffisamment robuste pour apaiser les appréhensions des personnels.

D – Initiative et utilisation

Qui prend l'initiative de vouloir améliorer la performance d'un établissement et donc d'établir des indicateurs (on a fait remarquer que l'amélioration ne s'obtient pas en promulguant une circulaire) ? Est-ce l'administration centrale ? Dès lors, ressentie comme mettant en cause les habitudes, cette

¹ PMSI, programme de médicalisation des systèmes d'information : recueil de données sur les hospitalisations.

initiative ne va-t-elle pas susciter des méfiances, des résistances ? Il faut associer les destinataires : ceux à qui on demande les données de base et dont on espère qu'ils changeront leur façon de faire. C'est bien ce qui a été fait pour les lycées : divers groupes de travail, où des chefs d'établissements en particulier étaient nombreux, ont été réunis pour définir les indicateurs. Il ne fallait pas que ce soit un travail technocratique, imposé d'en haut : juste introduire un « extérieur », à bonne distance, de manière à réfléchir aux différents cas.

On peut avoir deux stratégies :

- on décide d'implanter un système de compte rendu et d'indicateurs en le justifiant : « ça va servir ». Le risque est alors de pervertir les données ;
- ou l'on pense que c'est difficile et on implante progressivement quelque chose pour s'en servir peut-être dix ans plus tard. Et il sera alors difficile d'en faire prendre les résultats en compte.

Mais alors, ce qui a été entrepris a-t-il eu des effets ? Ces indicateurs ont-ils permis d'améliorer le fonctionnement du système ? Il est répondu que cela a du moins développé et légitimé une culture de l'évaluation qui n'existait pas il y a vingt ans, le souci de voir l'efficacité de ce qu'on fait. Mais, c'est une difficile pédagogie dans un système où le ministre change tous les deux ans². L'administration (centrale comme celle des établissements) a commencé à s'en saisir. Mais, la démarche n'a pas eu l'impact espéré auprès des enseignants.

Les lois de décentralisation (1982-85) prévoient que, chaque année, chaque collège et chaque lycée (qui sont devenus des « établissements publics locaux de l'enseignement ») doit établir et voter un rapport d'activité annuel, qu'il doit transmettre au recteur, au préfet et au président soit du département (collège) soit de la région (lycée).. Est-ce fait ? et, pour quoi en faire ? Si ou quand c'est fait, est-ce lu ? Voilà un support théoriquement adapté à une démarche d'évaluation et d'autoévaluation, et qui n'a pas en fait depuis 20 ans été utilisé. Ce serait utile de s'en servir.]

D'autre part, si cette évaluation est utile en interne, qu'en est-il au niveau rectoral ou central : si un établissement est mal noté, à quoi l'attribue-t-on et qu'en fait-on ? Quelle diversification des moyens, par exemple ? L'introduit-on dans un processus de contractualisation avec l'établissement ?

Un participant s'interroge aussi sur la signification d'un critère d'accès et de succès au bac pour évaluer l'efficacité d'ensemble du système éducatif : dans la mesure où le bac est un construit du système même.

Au total, demande un participant, sur quinze ans, voit-on des évolutions, existe-t-il des profils d'évolution ? – Ce qui existe permet de répondre à certaines questions : Cela permet en particulier l'analyse de la variabilité des résultats en fonction du contexte (environnement de l'établissement, population recrutée, ...) ou en fonction des caractéristiques du système d'enseignement (taille des classes, durée des enseignements, ...) Par exemple, on avait pu montrer (mais ce serait à reprendre et confirmer) qu'environ la moitié de la variance des taux de réussite au bac entre lycées était imputable au contexte, et l'autre moitié aux établissements eux-mêmes. Ce genre de conclusion est capital pour faire réfléchir les acteurs à leurs pratiques (pédagogiques, éducatives, de pilotage de l'établissement, etc.) et aux résultats qui en découlent.



² Et, où chaque ministre s'attache à faire sa propre réforme !