

Action publique éducative : mesurer l'essentiel ?

Paul Devin

page 1/5

L'usage d'indicateurs dans la gouvernance de l'action publique revendique une double perspective d'amélioration des services rendus aux usagers et de maîtrise des coûts. Il s'inscrit dans une culture du résultat qui, à en croire ses promoteurs, constituerait la garantie des finalités fondamentales de l'action publique en permettant de définir des priorités stratégiques et d'en mesurer les performances.

On pourrait être séduit par ces principes qui affirment le primat d'une mesure objective et d'une efficacité pragmatique, d'autant qu'ils prétendent se focaliser sur la qualité du service rendu aux citoyens et qu'ils entendent responsabiliser les décideurs politiques comme les cadres de l'administration sur les effets réels des politiques décidées et mises en œuvre. Pour beaucoup, cette logique est devenue incontournable et vouloir s'y opposer relèverait d'un conservatisme aveugle, inconscient des enjeux de maîtrise des coûts de l'action publique et résistant par principe à toute transformation qualitative.

Pourtant, on peut douter que le pilotage aux indicateurs soit réellement en capacité d'atteindre les finalités essentielles, celles de la démocratisation des savoirs et de la culture commune, celles de l'élévation générale du niveau de connaissances et de qualification. Et si les indicateurs n'étaient que des leurres, qui sous couvert d'objectivation de l'évaluation, ne servaient que la légitimation des politiques libérales et de leurs réductions budgétaires ?

Le modèle finlandais : corrélation n'est pas causalité !

Au vu de ses résultats aux évaluations PISA, l'école finlandaise a été pendant plus d'une dizaine d'années le modèle idéal dont il fallait s'inspirer pour résoudre les difficultés de l'école française. Xavier Darcos et Luc Chatel l'affirmèrent avec enthousiasme. On chercha donc à identifier les traits singuliers du système finlandais pour en faire des principes d'évolution nécessaire du système français : suppression des filières pour un enseignement transversal et modulaire au lycée, autodiscipline, bienveillance et confiance, autonomie des choix des élèves, autonomie des établissements... Au dire de bien des commentateurs, toutes ces caractéristiques du système finlandais étaient vertueuses et portaient en soi les prémices d'une amélioration qualitative.

Pourtant de telles conclusions procèdent de simples corrélations. Il est bien plus complexe de s'assurer de réelles relations causales car il reste très difficile de mesurer la part de l'un ou l'autre des facteurs dans les résultats de l'école finlandaise et de présumer les effets de leur transposition dans l'école française. C'est d'autant plus vrai que certaines corrélations ne sont même pas réelles. Par exemple, il n'y a pas un seul élément dans PISA qui administre la preuve que les résultats du système finlandais puissent reposer sur la modularité des enseignements en lycée, ce qui est pourtant régulièrement affirmé... Et pour cause : PISA évalue des élèves de 15 ans qui sont en fin « d'école fondamentale » et n'ont donc pas encore pu avoir des résultats scolaires influencés par l'organisation pédagogique spécifique du lycée finlandais.

Généralement les corrélations mises en valeur sont celles qui vont dans le sens de choix politiques pré-existant à l'analyse. D'autres caractéristiques du système finlandais auraient pu être considérées comme des éléments du modèle à suivre mais ont été écartés : des classes aux effectifs plus réduits, un volume de formation initiale et continue des enseignants largement supérieur, un nombre conséquent de conseillers accompagnant les élèves sur les méthodes de travail et l'orientation, des salaires enseignants plus élevés...

Ce tri des corrélations peut aller jusqu'au plus étonnant des paradoxes : le système finlandais se caractérise par l'importance donnée à l'aide aux élèves en difficulté à l'école primaire ; cela n'a pas empêché que Xavier Darcos, tout en portant le système finlandais aux nues, supprime une grande partie des RASED !

Et puis la référence récurrente à une objectivité du chiffre n'empêche pas qu'on puisse vouloir mettre de côté les chiffres qui relativisent ce qu'on aimerait

“ Les caractéristiques inégalitaires du système français, c'est ailleurs que dans la comparaison classante qu'il faut aller les chercher. ”

poser comme une évidence. C'est pourquoi les commentateurs se firent plus discrets sur TIMSS 1999⁽¹⁾, qui introduisait quelques doutes sur le miracle des résultats finlandais !

Il faut reconnaître la fragilité d'une modélisation qui présume de relations causales là où il ne s'agit que de corrélations et qui choisit les corrélations qui servent son projet et écarte les autres. C'est d'autant plus préoccupant que ces corrélations sont souvent diffusées sans les analyses qui montreraient la complexité des systèmes et de leur relation à leurs environnements. Ces analyses amèneraient à devoir renoncer à l'idéalité des caractérisations. Par exemple, si les réformes des années 1970 ont fortement réduit les

inégalités dans le modèle finlandais, elles ne l'ont pas mis à l'abri de tris très discriminants notamment par une filialisation totalement cloisonnée de l'enseignement professionnel et par un recours croissant à un enseignement privé très sélectif. La fascination pour le miracle finlandais a produit des enthousiasmes irraisonnés et rejeté dans l'ombre les analyses critiques

qui tentaient de mettre en évidence une réalité plus complexe. Mais ces adhésions superficielles sont fragiles : il a suffi que PISA 2013 rétrograde la place de la Finlande pour que le modèle cesse de produire ses effets séducteurs, alors que ce recul s'explique davantage par l'arrivée de nouveaux pays en tête du classement que par une chute des résultats finlandais !

Les leurreurs de la classification

Les évaluations internationales ne produisent pas seulement des données : elles classent les systèmes éducatifs des pays. Dans le discours commun et tout particulièrement lors de la publication des scores par la presse, le rang et l'évolution du rang ont pris une part majeure. Pourtant on peut douter de la pertinence de ce classement lorsqu'il s'agit d'évaluer la qualité du système, d'autant qu'une bonne partie des commentaires ne s'embarrasse pas de détails et grossit largement le trait.

Le premier problème est que la validité du classement supposerait une totale comparabilité. C'est loin d'être le cas. Depuis PISA 2009, les excellents résultats des élèves de Shanghai conduisent certains commentateurs à considérer qu'il y aurait là un modèle à suivre. Mais si les chiffres bruts de cette métropole asiatique présentent ce haut niveau de réussite,

signifient-ils la pertinence de son système scolaire ? Shanghai constitue une exception sociologique en Chine, caractérisée pas sa richesse et le haut niveau de formation de sa population. Son système scolaire est loin de scolariser l'ensemble de la population des jeunes de 15 ans et s'avère fortement ségrégatif. Un universitaire américain, Tom Loveless, a alerté à plusieurs reprises sur ce biais qui empêche toute comparaison crédible. À cela, il faut ajouter que la réussite des élèves n'est pas seulement imputable au système scolaire car 80% des Shanghaiens bénéficient d'un important soutien extérieur à l'école, le soir comme le week-end. Dans ces conditions, quelle leçon pourrions-nous tirer de la comparaison avec Shanghai ?

Le Japon figure en tête des résultats mais pour autant son système éducatif doit-il nous inspirer ? Les ambitions de ses finalités scolaires sont davantage portées sur un conditionnement que sur le développement de l'esprit critique et de la capacité à raisonner. D'ailleurs bien des experts, y compris japonais, expriment des inquiétudes sur les incidences d'une telle conception de l'éducation sur l'avenir de la société nipponne. Quant aux conditions de vie des écoliers japonais, elles sont dramatiquement inquiétantes : au Japon, le taux de suicide chez les enfants et les adolescents est l'un des plus élevés au monde et continue à croître alors que le taux général de suicide diminue. Là encore, quel sens peut avoir le constat d'un rang élevé de classement ?

C'est d'autant plus vrai que le classement, notamment des pays médians comme la France, repose sur des écarts faibles qui ne sont pas forcément significatifs. D'autant que des facteurs extrinsèques aux résultats eux-mêmes interviennent et doivent amener à relativiser ces faibles écarts : par exemple, le fait que l'évaluation s'adresse à un âge précis (15 ans) et non à un niveau n'est pas sans effet suivant que le pays pratique un cursus avec passage automatique au niveau supérieur ou non.

La presse titre régulièrement à propos des résultats de PISA : « la France recule ». Mais si on procède à l'observation des résultats des pays qui participent au classement depuis 2000, on constate au contraire une hausse depuis 2006. Entendons-nous bien, cette hausse n'est pas plus signifiante que la prétendue baisse mais elle devrait au moins amener à davantage de mesure dans les constats. Il ne s'agit pas de faire la sourde oreille aux alertes que les évaluations internationales nous adressent. L'inégalité de l'école française est une réalité qu'il faut combattre mais ce n'est pas dans la dramatisation des qualifications du classement que nous trouverons des solutions ! La présentation de graphiques aux ordonnées tronquées laisse croire à de forts écarts, là où les évolutions sont peu marquées. Ce qui est évidemment

“ Les questions techniques sur la pertinence de l'indicateur, sur l'écart éventuel entre son chiffrage et la réalité qu'il cherche à transcrire sont submergées par les perspectives de sa mise en scène politique. ”

(1) TIMSS (*Trends in International Mathematics and Science Study*) évalue les mathématiques et les sciences

soutenu par le catastrophisme des titres. Parfois la baisse évoquée est explicitement considérée dans les données de PISA comme « statistiquement non significative » ce qui n'empêche pas des commentateurs de la qualifier !

Dans sa note 2013,30, la DEPP considérait que les pays se situant autour de la moyenne OCDE, dont la France, ne pouvaient pas être distingués par des écarts significatifs. À rebours des dramatisations régulières, nous pouvons donc considérer que la France se situe dans la moyenne et c'est le seul véritable enseignement que nous livre l'aspect classificatoire de PISA. Le reste, et notamment les caractéristiques inégalitaires du système français, c'est ailleurs que dans la comparaison classante qu'il faut aller le chercher.

Le chiffre, outil de communication politique... jusqu'au paradoxe

Alors que l'indicateur chiffré est valorisé par ses qualités objectives, il est devenu un instrument de communication témoignant de la réussite ou de l'échec d'une politique. Prenons un exemple en matière de sécurité publique : le nombre de voitures incendiées pendant le réveillon et son évolution d'une année sur l'autre. Il est évidemment loin de pouvoir signifier quoique ce soit des effets réels des actions mises en œuvre. Pourtant les répercussions du chiffre sont si risquées, qu'en janvier 2017, à la veille des élections présidentielles, craignant les effets de l'annonce d'une augmentation, l'indicateur a été modifié pour ne prendre en compte que les véhicules mis à feu en excluant ceux brûlés par voisinage². Si la manipulation est des plus critiquables, elle témoigne surtout de l'effet pervers du diktat du chiffre quand la réalité et l'analyse de sa complexité doivent céder le pas à l'affichage d'indicateurs.

Les questions techniques sur la pertinence de l'indicateur, sur l'écart éventuel entre son chiffrage et la réalité qu'il cherche à transcrire sont submergées par les perspectives de sa mise en scène politique. C'est d'autant plus inquiétant que bien des destinataires de ces éléments chiffrés ne disposent pas des informations qui permettraient de relativiser les résultats, d'en interroger les biais pour mieux prendre en compte la complexité de la réalité. La fonction d'objectivation, théoriquement permise par l'indicateur, est anéantie par sa fonction communicationnelle.

Un des exemples les plus problématiques est fourni par la question de l'apprentissage de la lecture. Pendant plusieurs années, les analyses des évaluations de lecture passées lors de la Journée d'appel de préparation à la défense ont constaté que près de 80% des jeunes appelés étaient des lecteurs efficaces. En

creux, le chiffre de 20% allait désormais dans les médias caractériser la part d'élèves sortant du système scolaire « sans savoir lire ». En réalité, ces 20% de « lecteurs inefficaces » se répartissent tout au long d'un continuum de compétences qui va du quasi-analphabétisme à une capacité de lecture réelle mais handicapée par la nécessité de compenser un faible niveau lexical ou des difficultés d'identification des mots. Le chiffre de 20% est devenu, à l'opposé d'un élément de mesure objective et d'analyse circonspecte, une invective idéologique.

Les régulières volontés des politiques conservatrices d'imposer un retour à des méthodes syllabiques cherchent désespérément à se fonder dans les résultats des grandes évaluations internationales... en vain, car jusqu'à maintenant aucune n'est venue accréditer le simplisme d'un choix méthodologique univoque. La récente publication de PIRLS 2016 le confirme et si elle évoque plusieurs facteurs pour expliquer la faiblesse des résultats français, elle les situe dans le champ d'une insuffisante maîtrise des compétences de compréhension des textes et non dans celui d'une insuffisante maîtrise des relations graphophonologiques. Si l'usage des grandes enquêtes internationales s'inscrivait bien dans une volonté politique d'objectiver l'analyse préalable aux décisions, il ne fait aucun doute que la publication des résultats de PIRLS 2016 aurait dû infléchir les choix de la politique Blanquer. Il n'en est rien... au contraire, elle est l'occasion de les réaffirmer.

Mais que mesure-t-on réellement ?

La LOLF a introduit la mesure chiffrée comme un élément central de l'évaluation des politiques publiques. Désormais un projet annuel de performances définit des objectifs à atteindre et se dote d'indicateurs mesurant leurs effets. Mais l'hypothèse de la rationalisation induite vient se heurter à la manière avec laquelle l'indicateur reste trop souvent une construction superficielle incapable de se centrer sur les facteurs essentiels et donc de servir les finalités fondamentales.

Les études ne manquent pas pour montrer la faible efficacité du redoublement sur la réussite des élèves rencontrant des difficultés. On peut aujourd'hui considérer qu'un avis quasi unanime de la recherche sur la question a conduit les politiques éducatives

“ Il s'agirait seulement de renoncer aux fantasmes d'une rationalité binaire qui isole une méthode, une opération, une action et laisse croire qu'elle serait capable à elle seule de transformer positivement l'action publique. ”

(2) *Le Monde, Les décodeurs*, 2 janvier 2017

“ Les indicateurs agissent comme un écran déplaçant l'objet essentiel de progrès vers un effet illusoire et sans conséquence réelle sur les finalités. ”

depuis plusieurs années à réduire les redoublements. Pour autant, la suppression du redoublement en soi ne peut être considérée comme suffisante à produire une amélioration qualitative qui nécessite, en substitution au redoublement, des organisations pédagogiques capables de soutenir les élèves. La logique des indicateurs et la pression faite sur les écoles et les établissements pour que puisse être constatée une diminution du redoublement ont conduit à centrer l'action sur la seule baisse du taux de redoublement. Alors qu'ils seraient censés guider une reconfiguration des parcours scolaires permettant une meilleure démocratisation de la réussite, les indicateurs agissent comme un écran déplaçant l'objet essentiel de progrès vers un effet illusoire et sans conséquence réelle sur les finalités. Ainsi pilotée par des indicateurs superficiels, la politique de réduction du redoublement ne parvient qu'à son but formel et aux économies budgétaires qu'il engendre. Mais elle ne contribue en rien à l'amélioration qualitative du service public d'éducation.

Souvent l'indicateur isolé est incapable de guider les choix politiques car il n'incite pas aux croisements nécessaires à la réorientation d'une action pour qu'elle devienne capable de produire des effets réels. Prenons l'exemple de la scolarisation des élèves en situation de handicap qui, dans le premier cycle, constitue un autre indicateur de performance. Sont mesurés le taux de satisfaction des affectations en ULIS³ et le taux de certification spécialisée des enseignants qui y exercent. Entre 2014 et 2015, le taux de satisfaction a légèrement diminué du fait de la croissance du nombre de notifications. La logique aurait voulu qu'on puisse mesurer si les moyens d'enseignements affectés aux ULIS croissaient à la hauteur de l'augmentation régulière des affectations. Mais la nature isolée de l'indicateur ne le permet pas et le rend donc incapable de mesurer un élément essentiel de l'application de la loi de 2005 : l'État consacre-t-il les moyens nécessaires à permettre un développement de l'inclusion scolaire à la hauteur des besoins ? À défaut de cette analyse croisée, l'usage des indicateurs est des plus risqués car on pourrait être tenté d'améliorer l'indicateur sans qu'une amélioration réelle des résultats soit engagée... par exemple, en cherchant à influencer sur une diminution du nombre de notifications.

Plus essentiel encore, à moins de considérer l'inclusion comme une finalité en soi, indépendante de ses effets réels, la véritable évaluation de cette politique demanderait à ce que soient mesurés les effets de l'inclusion sur les compétences des élèves en situation de handicap. Ce n'est pas le cas. Nous disposons donc

d'indicateurs incapables de nous renseigner sur l'essentiel : l'inclusion scolaire dispose-t-elle des conditions nécessaires pour permettre ses finalités essentielles à savoir l'accès aux droits fondamentaux reconnus à tous les citoyens et tout particulièrement celui de la maîtrise des savoirs et de la culture commune ?

Un autre effet pervers des indicateurs est perceptible dans l'appréciation de l'action propre de chaque lycée par les indicateurs de résultats. La communication institutionnelle affirme que leur fonction n'est pas le classement des établissements. D'ailleurs les indicateurs de valeur ajoutée (IVAL) cherchent à relativiser les résultats bruts en mettant en valeur les effets de l'action pédagogique de l'établissement sur les résultats des élèves, au-delà des caractéristiques socio-économiques de la population scolaire accueillie. L'intention est clairement de relativiser des réussites qui ne tiennent qu'à la sélectivité de l'environnement social et de valoriser celles produites par l'action de l'établissement. Mais la diffusion publique permet à la presse de publier des « palmarès » qui organisent leur classement essentiellement sur le taux de réussite au bac et induisent les choix parentaux vers les établissements les plus performants sans prendre en compte la valeur ajoutée. Un tel traitement favorise les établissements privés usant d'une forte sélectivité pour garantir leurs résultats et joue le jeu d'un entre-soi social probablement satisfaisant pour les familles qui en profitent mais bien loin d'être capable de répondre à la nécessité d'une préoccupation politique globale agissant dans les perspectives de l'intérêt général.

On pourrait multiplier les exemples.

Pour mesurer le climat scolaire et l'apprentissage de la scolarité, trois indicateurs sont retenus : le taux d'absentéisme des élèves, le nombre de signalements de violences graves et le taux de participation des lycéens aux élections au conseil des délégués pour la vie lycéenne. Au-delà de la capacité de ces trois indicateurs à rendre compte d'une problématique aussi complexe, on sait comment le nombre de signalements de violences est largement dépendant de la perception subjective de la gravité de l'acte. Les tentatives de définition des critères d'appréciation sont loin de suffire à corriger les effets de la subjectivité ou de la volonté d'un établissement à atténuer ou accentuer sa situation propre. L'analyse des signalements d'incidents par la DEPP constitue d'évidence un travail appréciable et un élément d'appréhension de l'évolution des phénomènes violents. Mais pour autant, ce taux constitue-t-il un indicateur pertinent dans l'évaluation d'une politique ?

(3) Unités locales d'inclusion scolaire : dispositifs de scolarisation à effectifs restreints permettant la scolarisation des élèves handicapés en milieu ordinaire

Les progrès de l'indicateur mesurant le taux de remplacement dans le premier degré sont analysés comme ayant été produits par la mutualisation des moyens entre circonscriptions et la généralisation de l'usage d'un logiciel de gestion des remplaçants. Mais nul indicateur ne se préoccupe de mesurer les effets de l'augmentation de la tâche gestionnaire liée à cette rentabilisation du remplacement et de savoir si elle ne se fait pas aux dépens de missions d'accompagnement et de formation pourtant bien plus essentielles à l'amélioration qualitative du service public d'éducation. Cette centration du pilotage sur l'indicateur est d'autant plus paradoxale qu'il est évident que, parallèlement aux progrès faits à la marge, l'essentiel de l'amélioration du taux de remplacement entre 2014 et 2016 a été produite par la politique d'augmentation des recrutements qui a permis de reconstituer les viviers de remplaçants.

Mesurer l'essentiel ?

Au vu de ces effets de leurres, d'illusions, d'erreurs, l'usage du chiffre ne peut se concevoir que dans des analyses vigilantes, méfiantes et exigeantes. Il ne s'agit pas, bien sûr, de renoncer à appréhender la réalité des effets d'une politique mais de refuser le diktat néomanagérial qui voudrait que cette évaluation n'ait de légitimité que circonscrite à la production d'une mesure chiffrée et dévaloriserait de ce fait toute autre conception de pilotage de l'action publique.

Toute personne qui accorde une attention un peu soutenue à l'évaluation de l'action publique éducative constate que la nature multifactorielle des actions qui contribuent aux buts porte, intrinsèquement, le risque de ne pouvoir faire fonctionner efficacement la mise en relation entre une action et ses effets. C'est d'autant plus vrai que les effets ne sont véritablement mesurables qu'à terme et que les conséquences immédiates des actions sont loin de prédire leurs conséquences pérennes. C'est à cette question que renvoie l'habituel constat d'un décalage entre « temps éducatif » et « temps politique ». Les indicateurs survalorisent ce qui est facilement et immédiatement mesurable et centrent le pilotage sur l'atteinte de chiffres qui jouent en définitive à la marge des progrès qualitatifs essentiels.

Une prétendue rigueur scientifique conduit à écarter tout ce qui procèderait d'une autre philosophie évaluative et à celui qui doute des impératifs du pilotage à la performance, on oppose immédiatement les valeurs de la modernité. Critiquer l'usage des indicateurs ne pourrait se fonder que dans une résistance primaire à tout changement ou dans un attachement obsessionnel aux intérêts corporatistes qui visent à protéger quelque avantage acquis ou quelque confort statutaire !

Pourtant le modèle du pilotage stratégique et opérationnel de la performance est loin d'avoir imposé sa pertinence à conduire les actions publiques. Au-delà de ses effets gestionnaires et des réductions des dépenses publiques qu'il permet d'engager à court terme, on peut interroger ce que recouvre dans ses effets réels l'amélioration de performance visée. L'amélioration du service public promise par le recours à cette nouvelle rationalité est loin d'imposer son évidence !

Les principes plus traditionnels d'un pilotage centré sur l'énoncé des buts et l'examen de la cohérence entre les actions et les finalités étaient-ils en définitive moins pertinents pour conduire l'action publique ? Il ne s'agit évidemment pas d'imaginer qu'une vertu téléologique conduirait par principe le système vers ses finalités : le pilotage reste une nécessité. Il s'agirait seulement de redonner à la mesure chiffrée sa juste place, celle d'un outil de repère et d'analyse capable de nous aider aux ajustements nécessaires, aux réorientations utiles, voire aux renoncements lorsque l'orientation apparaîtrait comme incapable d'engager les effets prévus. Il s'agirait seulement de renoncer aux fantasmes d'une rationalité binaire qui isole une méthode, une opération, une action et laisse croire qu'elle serait capable à elle seule de transformer positivement l'action publique.

Tout cela demande évidemment une condition que le contexte politique actuel est loin d'offrir : disposer d'un projet éducatif dont les buts transcendent les variations politiques consécutives aux alternances, considérer les enjeux de l'école comme un investissement essentiel de l'action publique, distinguer les perspectives de l'action publique de l'État des préférences idéologiques particulières du ministre... Cela supposerait que les buts de la démocratisation soient recherchés dans leur effectivité.

Tout cela demande de revenir à une question politique plus fondamentale encore, qui interroge le paradoxe d'un énoncé unanimiste à vouloir satisfaire la demande sociale d'une réussite scolaire pour toutes et tous sans pour autant vouloir porter un projet scolaire réellement émancipateur et égalitaire. Il en résulte une tension entre la réalité et les idéaux égalitaires de notre démocratie que les leurres de l'indicateur cherchent désespérément à masquer.

“ Cela supposerait que les buts de la démocratisation soient recherchés dans leur effectivité. ”

PAUL DEVIN
Inspecteur de l'Éducation nationale,
Secrétaire général du SNPI-FSU

(5) Un tiers de service maximum en master 2 pour rendre possible la dynamique de retour réflexif sur la pratique.