

## Évaluer, noter : un repérage didactique

Yves Chevallard  
UMR ADEF & IUFM d'Aix-Marseille

*Ce repérage de quelques aspects cruciaux du problème didactique de la notation est sous-tendu par une idée qui affleure ça et là : d'une façon générale, nous avons en matière de notation des idées trop simples et, en consonance avec elles, des pratiques très ritualisées qui verrouillent fortement toute évolution significative. Une seconde idée est cachée derrière cette première idée : c'est que nos idées sont simplistes en grande partie parce que notre culture scientifique collective, et notamment notre culture mathématique, est encore trop peu évoluée. Cette seconde idée explique que les quelques références de nature mathématique dont ce court texte est parsemé ne sont pas une simple affaire d'intérêt personnel de l'auteur.*

Évaluation et notation ne se superposent pas. Toute évaluation ne se traduit pas par une note chiffrée. Toute assignation de note n'est pas une évaluation : elle doit conquérir son droit à l'être. On connaît ainsi le problème de la note zéro attribuée pour sanctionner un comportement inapproprié : en dépit d'une jurisprudence administrative ancienne et constante qui la désavoue et malgré une circulaire récente<sup>1</sup>, cette pratique perdure. On a là un cas particulier d'un phénomène plus général : la notation est une *structure* qui se trouve associée, par bouturage et provignement, si l'on peut dire, à une foule de *fonctions* dont les effets de réel sont souvent sans proportion avec l'humble acte évaluatif que le professeur assume dans la classe. La notation scolaire est une structure didactique à faibles défenses immunitaires, qui souffre de multiples pathologies opportunistes contre lesquelles la kyrielle de thérapeutiques envisagées depuis trois quarts de siècle semble durablement impuissante.

De l'empilement fonctionnel qui surcharge la notation scolaire et en occulte les fonctions didactiques témoigne par exemple la « cote *R* » introduite au Québec à partir de 1995 pour évaluer le dossier scolaire des candidats à l'entrée à l'université. Depuis 1977, chaque postulant était jugé sur ses résultats aux cours suivis dans l'enseignement collégial, la réussite à chacun de ces cours étant appréciée, non sur la base de la note *X* obtenue, mais à l'aide d'un indicateur numérique spécial, la « cote *Z* » définie par  $Z = \frac{X - m}{s}$ , où *m* est la moyenne arithmétique et *s* l'écart type de la série des notes obtenues par les élèves ayant suivi le même cours et subi la même procédure d'évaluation (collectif d'élèves dit « groupe à l'évaluation »). La distribution des cotes *Z* des membres du groupe est, par construction, *centrée* (sa moyenne est nulle) et *réduite* (son écart type est 1) : l'objet du passage de *X* à *Z* était motivé par un souci d'équité, un étudiant *e*<sub>1</sub> ayant obtenu avec un professeur exigeant une certaine note *X*(*e*<sub>1</sub>) pouvant en vérité surpasser tel autre postulant, *e*<sub>2</sub>, qui aurait reçu une note *X*(*e*<sub>2</sub>) > *X*(*e*<sub>1</sub>) d'un notateur moins sévère. Mais on voit le défaut de la cote *Z* : un étudiant « fort », inscrit dans un cours où *beaucoup* d'élèves sont « forts », est désavantagé face à un postulant d'égale force ayant suivi un cours réunissant des élèves de niveau *peu élevé*. D'où l'adoption de la « cote *R* », définie par  $R = (Z + \text{IFG} + 5) \times 5$ . Le sigle IFG désigne l'« indicateur de la force du groupe » – du groupe à l'évaluation, s'entend<sup>2</sup> ; le calcul de l'IFG se fonde sur les notes obtenues dans les deux dernières années du secondaire par les membres du groupe : il est défini, arithmétiquement, de façon que, si le groupe à l'évaluation se révèle ainsi plus faible que la moyenne, son IFG est négatif, en sorte qu'il vient diminuer la cote *Z*, tandis que, dans le cas contraire – si le groupe de l'élève dont le dossier est examiné est d'une « force » au-dessus de la moyenne –, son IFG est positif, si bien que la cote *Z* s'en trouve augmentée. Arrêtons-là ce bref récit : bien entendu, la polémique qui avait emporté la cote *Z* renaît avec la

<sup>1</sup> Voir le BOEN spécial n° 8 du 13 juillet 2000 (<http://www.education.gouv.fr/bo/2000/special8/default.htm>).

<sup>2</sup> Les paramètres pris ici égaux à 5 ont pour fonction de donner aux cotes *R* des valeurs qui, le plus souvent, sont comprises entre 15 et 35.

cote  $R$ , que ses promoteurs prétendent sans tache mais qui tord manifestement le réel qu'elle devait se contenter de révéler le plus fidèlement possible<sup>3</sup>.

Plus peut-être que d'autres quantifications qui trament notre quotidien, les notes scolaires n'en finissent pas de fournir à la construction de la réalité sociale une matière première qui paraît au premier abord simple et facile à travailler mais se révèle en bien des cas rétive, indisciplinable, captieuse. L'obsession des concepteurs des cotes  $Z$  et  $R$  était de distinguer un individu des collectifs à l'intérieur desquels il a étudié et appris, ce qui rappelle au passage que la notation a son assise, non dans l'individu, mais dans des groupes et des institutions. Arrêtons-nous ici sur un exemple qui illustre autrement la dialectique de l'individu et du groupe en montrant comment l'individu « fait » le groupe. Sur une épreuve attribuant à chaque élève une note  $X$ , une classe de  $n$  élèves se donne à voir à travers une série statistique  $X(e_1), X(e_2), \dots, X(e_n)$  qu'on peut décrire par sa moyenne  $m_X$  et son écart type  $s_X$ . Une « bonne » classe, c'est d'abord une classe dont la moyenne  $m_X$  est élevée ; mais une classe peut avoir une moyenne élevée et... n'être pas une classe, parce qu'elle se présente *éclatée* en plusieurs sous-groupes, ce que traduit une valeur élevée de l'écart type  $s_X$ . Dès lors qu'une série de notes disperse les élèves au point que la cohésion du groupe semble menacée, on voit ordinairement le professeur qui « tient » sa classe (au plan didactique, s'entend) engager des manœuvres qui auront pour effet de réduire l'écart type de la série de notes  $Y(e_1), Y(e_2), \dots, Y(e_n)$  de l'épreuve *suivante*. Plusieurs techniques s'offrent. Si le professeur propose une épreuve notablement plus facile, les notes s'agglutineront vers les valeurs élevées et la dispersion s'en trouvera réduite. Mais la chose est fâcheuse pour la réputation de l'enseignement prodigué. Une autre manœuvre consiste à sacrifier un peu de la moyenne : si le professeur choisit de proposer une épreuve *plus difficile* – y compris pour les « forts » –, la classe se regroupera vaille que vaille « vers le bas ». Là encore, la technique est effectivement utilisée – c'est même un réflexe chez certains professeurs : reprendre les choses en main ! Mais elle est parfois inutilement douloureuse.

Au-delà de ces tentatives, un mécanisme traditionnel vient réduire la dispersion de la classe par un phénomène *purement arithmétique*, qui affecte l'image chiffrée donnée par les *moyennes trimestrielles*. Supposons pour simplifier que la moyenne trimestrielle  $M(e)$  de chacun des élèves  $e$  soit la moyenne de *deux* notes  $X(e)$  et  $Y(e)$  :  $M(e) = \frac{X(e) + Y(e)}{2}$ . La classe se donne à voir, en fin de trimestre, à travers la série des moyennes  $M(e_1), M(e_2), \dots, M(e_n)$  : sa cohésion se lit donc dans la valeur de l'écart type  $s_M$ . Or cet écart type *n'est pas* égal à la *moyenne* des écarts types  $s_X$  et  $s_Y$  : il est en général *plus petit* que cette moyenne, puisqu'il suffit que le coefficient de corrélation  $r_{X,Y}$  des séries de notes  $X$  et  $Y$  soit strictement inférieur à l'unité pour qu'on ait  $s_M < \frac{s_X + s_Y}{2}$ <sup>4</sup>. Les moyennes trimestrielles font ainsi apparaître la classe comme *davantage groupée* que ne le montraient ses performances du trimestre ! Ce résultat rappelle que les jeux arithmétiques fabriquent du réel institutionnel et social et révèle en outre

<sup>3</sup> Dès janvier 1997, des conseillers et conseillères d'orientation du Québec écrivaient : « La cote R crée aussi un climat d'incertitude propice à la recherche de stratégies douteuses de la part des clientèles pour obtenir une place convoitée. » Et de rappeler ce qui se dit ici où là : « les étudiants et leurs parents recherchaient les écoles et les collèges où il y a des groupes forts ; des écoles secondaires auraient bonifié les résultats dans certaines matières non sanctionnées par des épreuves provinciales ; des collèges auraient formé spécialement des groupes forts ou, mieux, des groupes virtuels pour avantager des étudiants ; des professeurs, afin d'augmenter l'indice de force de leur groupe, feraient des pressions pour que les plus faibles quittent les cours avant la date limite d'abandon », etc. ([http://www.orientation.qc.ca/site/pdf/grand\\_dossier/cote.R.janv.97.pdf](http://www.orientation.qc.ca/site/pdf/grand_dossier/cote.R.janv.97.pdf)).

<sup>4</sup> Ce coefficient est, par définition, compris entre  $-1$  et  $1$ . On a d'une part  $s_M = \frac{1}{2} \sqrt{s_X^2 + s_Y^2 + 2s_X s_Y r_{X,Y}}$  et d'autre part  $\frac{s_X + s_Y}{2} = \frac{1}{2} \sqrt{(s_X + s_Y)^2} = \frac{1}{2} \sqrt{s_X^2 + s_Y^2 + 2s_X s_Y}$ , expressions qui montrent que  $s_M \leq \frac{s_X + s_Y}{2}$ .

un paradoxe. Car le phénomène constaté s'explique par le fait que certains élèves nomadisent à travers l'échelle des notes, obtenant tantôt de bonnes notes, tantôt des notes faibles<sup>5</sup>. C'est en fin de compte à ses élèves instables que la classe doit son apparente unité !

Le problème central, ici, est celui de la *lisibilité* de la note assignée. La créance donnée autrefois à la note comme *mesure* de la valeur supposée de la prestation de l'élève avait un mérite certain et un désavantage sérieux : elle permettait d'invoquer un ensemble de métaphores métrologiques fournissant une grille de lecture toute faite de la notation, en même temps qu'elle figeait cet acte dans une imperfection essentielle, toujours à maîtriser – celle d'une opération de mesurage réalisée avec un instrument peu fiable en vue d'estimer la mesure d'une grandeur elle-même mal définie. Constatons qu'il n'est pas facile de sortir de cette utopie métrologique : depuis Laugier et Weinberg (1936) concluant qu'il faudrait 127 correcteurs pour parvenir à la « vraie note » d'une copie de philosophie du bac, on a tenté avec une persévérance admirable de construire solidement sur des fondations incertaines. En bien des cas, la clarification recherchée conduit à accroître l'opacité d'une note issue de manœuvres complexes et vouée à être exploitée en des constructions sociales non moins absconses<sup>6</sup>. La pression de la société pour savoir ce que savent ceux que l'école a formés et combien se situent en deçà ou au-delà d'un certain seuil de compétence est on ne peut plus légitime dans son principe. Mais en rendant obligatoire une certaine forme de quantification, elle pèse sur le processus même dont elle voudrait connaître le produit.

L'acte de notation, il est vrai, n'est pas soumis qu'à cette pression. En lui deux universaux anthropologiques s'entrecroisent : une dialectique du silence et de la véridiction tente de s'y articuler à l'intérieur d'une dialectique du don et du contre-don. Ce dernier aspect est déterminant<sup>7</sup>. Le professeur fait don d'un cours et offre à la classe un sujet de travail. L'élève lui remet sa copie. Le professeur répond à son tour : pour nombre d'élèves, aujourd'hui, la *note* qu'il assigne le libère pour l'essentiel de la dette contractée envers l'élève au moment où il en a accepté la copie. La dialectique du don et du contre-don, ici, tourne court : le marché se conclut trop vite, au prix d'un peu de verroterie numérique que chacun feint de regarder comme réglant la transaction adéquatement. La contre-prestation du professeur, il est vrai, ne se limite pas usuellement à cela : en nombre de disciplines, par delà l'attribution de la note et la restitution de la copie « annotée », elle comporte une correction que les élèves reçoivent souvent sans enthousiasme, suivie de la remise d'un corrigé pour solde de tout compte.

C'est l'organisation et le contenu même de ces jeux de l'échange – dont professeur et élèves sont les acteurs plutôt que les auteurs – qu'il faut mettre en cause pour poser le problème de la note. Les formulations précédentes peuvent être généralisées : le professeur offre à la classe une question  $Q$  et une partie des matériaux utiles pour apporter à  $Q$  une réponse  $R$ . Ensemble, sous la direction du professeur, on examine des réponses  $R^\diamond$  : on les observe, on les analyse, on les évalue (que *valent-elles* ?). Du même mouvement, en autonomie didactique relative, on développe sa réponse  $R^\heartsuit$ , on la « publie », on la défend – dans la copie qu'on remet au professeur, quand copie il y a. Quel contre-don appelle alors le don de  $R^\heartsuit$  ? Question cruciale. Car s'il est vrai que, pendant tout un temps d'étude de  $Q$ , *noter*

---

<sup>5</sup> Ces élèves font en effet chuter le coefficient de corrélation  $r_{X,Y}$ , et donc aussi l'écart type  $s_M$ .

<sup>6</sup> Aux États-Unis, une majorité d'États imposent aux élèves de passer un test standardisé, qu'il s'agisse d'un test « normatif », d'un test « critérié », ou des deux. Dans un test « normatif », le score de l'élève est situé par rapport à la distribution des scores d'un groupe jugé représentatif de la population de ses congénères (« *norm group* »). Par contraste, un test « critérié » examine « *what test takers can do and what they know, not how they compare to others* » (Anastasi, 1988, cité in Linda A. Bond, *Norm- and Criterion-referenced Testing*, décembre 1996 : <http://www.mcgt.net/Articles/norm.htm>).

<sup>7</sup> Sur la dialectique du don et du contre-don, dont l'étude par Marcel Mauss (1872-1950) a fait époque, et qui se caractérise par la triple obligation de donner, de recevoir et de rendre, voir par exemple Mark R. Anspach, *À charge de revanche* (Seuil, 2002).

serait agressif à l'endroit du processus d'étude même, il ne convient pas davantage d'évaluer trop vite ! L'état d'*epochè*, de suspension de jugement, doit présider à l'observation et à l'*analyse* que celle-ci permet : analyser *n'est pas* évaluer<sup>8</sup>. Or, dans une certaine culture scolaire, la pression à l'évaluation – que l'attribution d'une note a surtout pour effet de diminuer fortement – tend à réduire l'observation et l'analyse des réponses  $R^\heartsuit$  à la portion congrue, alors même qu'il y a là, au plan didactique, une contre-prestation essentielle, au reste indispensable pour une évaluation non formelle des réponses  $R^\heartsuit$ .

La dialectique du silence et de la véridiction trouve là son lieu : d'abord se taire, faire, regarder faire ; puis analyser ce qu'on a fait, et pour cela *dire* (et éprouver) ce qu'on croit *vrai*, sans pour autant assigner encore à la réponse  $R^\heartsuit$  une valeur quelconque ; enfin, bien sûr, *évaluer* ce qu'on a construit, dire ce que, de tel point de vue, cela « vaut » – avec l'idée de poursuivre ou de reprendre le travail en cas d'insuffisance. L'évaluation gagnera souvent à s'expliciter sur des échelles de scores où noter certains aspects de la réponse évaluée – intelligibilité, économie de moyens, efficacité, fiabilité, « portabilité », etc. Une telle évaluation *vectorielle* pourrait-elle être résumée par un *scalaire* unique ? La chose se pratique couramment, d'autorité. Mais ne soyons pas dupes : si bien combiné soit-il, un indice synthétique et sa florissante descendance numérique éventuelle ne servent en général que bien imparfaitement les fonctions pour lesquelles on prétend les mobiliser ! Nos sociétés doivent accepter de renoncer à une croyance infantile aussi gratuite que paresseuse : toute qualité ne saurait se quantifier adéquatement sous la forme d'une grandeur scalaire (à l'instar des longueurs, des volumes, des masses, etc.), ni même se modéliser par une relation d'ordre total (comme il en va pour les températures). S'il est légitime qu'on veuille savoir, on ne saurait attendre que le savoir se présente à nous, complaisamment, sous une forme unidimensionnelle – qui ne conduit guère, en règle générale, qu'à méconnaître la réalité dont on prétend s'instruire.

---

<sup>8</sup> L'analyse ne saurait en outre se réduire à une traque des « erreurs ». Car, sauf à procéder par imposition dogmatique, le faux ne se montre pas de lui-même. Il doit être construit en tant que tel.