

Evaluation et orientation

Rapport pour le Haut Conseil de l'Education

Pascal Bressoux
Université Pierre Mendès France, Grenoble
Laboratoire des Sciences de l'Education (EA n° 602)

Remerciements

Je remercie les membres du Haut Conseil de l'Education pour m'avoir fait confiance en me confiant la responsabilité de cette étude sur « évaluation et orientation ». J'espère que le rapport apportera une contribution à leur réflexion sur les questions qui m'ont été posées dans la commande.

Je tiens à remercier Philippe Tiquet pour son accompagnement dans ce travail et ses remarques sur une version provisoire du rapport.

Le présent rapport répond à une commande du Haut Conseil de l'Education à Pascal Bressoux d'une étude sur « évaluation et orientation ». La lettre de commande fait état des points suivants :

Comment se construit le jugement professoral, sur quoi repose-t-il ?

Le jugement professoral, l'évaluation professorale, l'évaluation scolaire vous paraissent-ils aptes à motiver une décision d'orientation ?

Si non, sur quels autres éléments se fonder ?

Comment évaluer dans de bonnes conditions pédagogiques et méthodologiques un projet d'évaluation d'élève ?

Avec quels acteurs et quelle place pour chacun d'eux ?

Du point de vue de votre recherche et de votre spécialité, quelle vous semble pouvoir être une orientation réussie ?

L'étude s'intéressera prioritairement au collège et au lycée.

Des préconisations sont attendues.

Sommaire

INTRODUCTION	5
DE QUELQUES PROCESSUS FONDAMENTAUX DANS LE JUGEMENT SOCIAL	7
Juger dans des situations d'incertitude : les heuristiques et les biais	8
La formation des impressions à l'égard d'autrui	11
A quoi attribuons-nous les causes des comportements observés ?	12
L'influence des stéréotypes.....	14
D'où viennent les catégories que nous utilisons pour classer, juger ?	16
LA CONSTRUCTION DU JUGEMENT DES ENSEIGNANTS.....	18
Les jugements des enseignants sont-ils « exacts » ?	18
Pourquoi certains enseignants sont-ils plus « exacts » dans leur jugement que d'autres : le dogmatisme comme facteur de biais	21
Capacité des enseignants à distinguer entre différents processus et habiletés spécifiques	24
Jugement sur performances « actuelles » et jugement sur réussite scolaire ultérieure	25
L'influence des normes sociales générales et des normes professionnelles	27
Les effets des attentes : les prophéties autoréalisatrices	28
Les études sur la notation	32
Synthèse concernant les qualités métrologiques du jugement comme mesure : fidélité et validité.....	35
Mais finalement, est-ce le véritable problème ? Le jugement scolaire est-il bien une mesure ?	37
Les tests ont-ils eux-mêmes des qualités métrologiques parfaites ?.....	39
Présentation synthétique de travaux français sur le jugement enseignant	40
Les choix d'orientation : sélection et autosélection	43
PRECONISATIONS	47
Reprise des résultats principaux sur la fabrication du jugement scolaire	47
Du bon usage des épreuves standardisées	50
La standardisation à l'aune des procédures de sélection et d'autosélection	51
Qu'est-ce qu'une orientation réussie ?	53
REFERENCES BIBLIOGRAPHIQUES	57

Introduction

Commençons par quelques données anciennes, mais édifiantes, pour nous convaincre que le problème qui constitue notre objet d'étude n'est pas nouveau et pour nous permettre d'entrevoir l'ampleur de la question. Dans une étude conduite par Sauvy et Girard, publiée en 1965 (citée par Prost, 1968, p. 430), 100 % des enfants de cadres supérieurs dont la réussite en primaire est jugée excellente entrent en 6^e. Ils sont 91 % parmi les enfants d'ouvriers. 92 % des enfants de cadres supérieurs dont la réussite en primaire est jugée moyenne entrent en 6^e. Ils sont 42 % parmi les enfants d'ouvriers. 50 % des enfants de cadres supérieurs dont la réussite en primaire est jugée mauvaise entrent en 6^e. Ils sont 3 % parmi les enfants d'ouvriers... Durant la même période, Girard et Bastide (cités par Duru-Bellat, 2002, p. 80) indiquent que, pour des élèves dont la réussite en primaire est jugée moyenne, l'avis des instituteurs, nécessaire à l'époque pour l'entrée en 6^e, est favorable dans 78 % des cas lorsqu'il s'agit d'un enfant de cadre supérieur, dans 30 % des cas lorsqu'il s'agit d'un enfant d'ouvrier.

Ces quelques données montrent l'ampleur d'un phénomène qu'il ne s'agit évidemment pas ici de cristalliser autour d'une période ou d'un niveau scolaire fixé dans le passage du primaire au secondaire, phénomène qui ne cesse d'être constaté par les sociologues, sous des formes partiellement renouvelées, à des étapes différentes du cursus, avec une translation due à la massification et à l'allongement de la scolarité, mais qui ne cesse de demeurer, fondamentalement, le même : les différenciations sociales d'orientation scolaire. La période contemporaine a vu le poids de l'Ecole augmenter considérablement dans la lutte pour le maintien ou l'amélioration de la position sociale d'origine. Elle devient de ce fait le centre de toutes les attentions, le lieu des stratégies les plus subtiles pour les uns, les moins informées pour les autres, le terrain d'élaboration de jugements lourds sur le devenir de chacun.

Les études sur l'orientation scolaire (Duru-Bellat, 2002 ; Duru-Bellat & Mingat, 1988) font état de différences dans l'orientation des élèves au niveau de l'enseignement secondaire, même pour des élèves de performances scolaires identiques. Ces études font en général apparaître des différences d'orientation qui sont associées à des facteurs individuels et à des caractéristiques environnementales. Concernant ce dernier point, certains collègues, par exemple, apparaissent plus sévères que d'autres dans leur notation et cette exigence accrue n'est que rarement comblée par des procédures d'orientation moins sélectives. Il existe de façon manifeste des politiques d'orientation scolaire qui sont pour partie liées à l'établissement scolaire. Ainsi, tel élève ne connaîtra pas nécessairement la même orientation selon qu'il est dans tel établissement ou dans tel autre. Ce type de différence joue surtout pour les élèves moyens et moyens-faibles, c'est-à-dire ceux pour lesquels l'orientation scolaire ne va pas de soi. Du point de vue des facteurs individuels, il a été montré que, à réussite scolaire identique, certaines catégories d'élèves connaissent une orientation moins sélective que d'autres. Les enfants de cadres supérieurs et professions libérales ont ainsi plus de chances que d'autres (et en particulier que les enfants d'ouvriers) d'être orientés vers les filières générales longues. Il en va de même de certaines caractéristiques liées au déroulement de la scolarité des élèves : il a pu être montré par exemple que les élèves en retard scolaire ont, toutes choses égales par ailleurs, une probabilité plus élevée que les autres d'être orientés vers des filières techniques courtes.

Bien que l'objet de cette étude ne réside pas dans l'étude des procédures d'orientation elles-mêmes, mais dans les liens que l'orientation entretient avec le jugement des enseignants, il

convient néanmoins de cadrer les différenciations sociales d'orientation scolaire par quelques chiffres. Prenons pour cela le différentiel de passage en classe de 4^e générale entre les enfants de cadres supérieurs et les enfants d'ouvriers, qui s'établit respectivement, au début des années 1980, à 90 % et 55 % (Duru-Bellat, 2002). Cet écart peut être décomposé en (au moins) quatre parties dont l'importance est à peu près égale : le premier quart revient aux inégalités de réussite scolaire ; le second revient aux différenciations sociales inscrites dans les vœux d'orientation ; le troisième revient aux inégalités sociales dans les décisions d'orientation ; le quatrième, enfin, revient aux écarts d'orientation entre établissements (Duru-Bellat, 2002). Au niveau du lycée, les inégalités sociales d'accès aux différentes filières s'estompent, là où le niveau scolaire et les choix d'option deviennent fondamentaux. Au niveau de l'enseignement supérieur, les filières étant (à l'université) d'accès libre, c'est l'autosélection qui devient le processus dominant. Dans cette autosélection, les différenciations sociales redeviennent importantes, même au sein de filières, ou d'écoles, qui se différencient de manière très subtile sur le marché de l'excellence scolaire (Bourdieu, 1989).

Ces phénomènes, loin d'être entièrement homogènes et déterminés, peuvent être vus comme étant, pour partie, le résultat de jugements portés sur les élèves, jugements dont la fabrication obéit elle-même à des processus complexes qui mêlent effets de contexte, attentes normatives générales, influence de stéréotypes, observations de comportements et de performances, etc. Le but de cette étude est de porter au jour les éléments qui concourent à la fabrication de ce jugement scolaire et de voir en quoi cela influe sur l'orientation des élèves. Nous tenterons de montrer que le jugement scolaire n'est pas le simple enregistrement de performances scolaires. Il s'agit en grande partie d'une activité d'inférence qui va au-delà de la simple observation et qui surajoute des éléments, leur donne une coloration, une interprétation particulières. Si l'on veut se donner une chance de comprendre complètement le jugement scolaire ou, plus modestement, d'en comprendre les déterminants fondamentaux, il faut le considérer, non comme une activité séparée des autres activités de jugement, comme si elle obéissait à des processus différents par nature, mais au contraire comme une activité d'inférence qui obéit aux mêmes lois que d'autres activités d'inférence, jugements quotidiens. Ces jugements quotidiens sont ceux formés par l'homme de tous les jours dans les situations les plus diverses, formelles comme informelles, fortuites ou anticipées, durables ou éphémères, à propos d'une multitude d'objets, personnes, événements, groupes sociaux, etc. Il faut donc appréhender le jugement scolaire de manière générale, en montrant ce qu'il a en commun avec le jugement quotidien, mais aussi étudier la spécificité de ce jugement scolaire, non pas en ce qu'il serait par nature différent, mais en ce qu'il s'établit dans des situations spécifiques, sur lesquelles s'exercent des contraintes spécifiques, qui fournissent des informations spécifiques. Ainsi, à la différence de bien d'autres situations sociales, la période d'interaction entre l'enseignant et l'élève est longue, asymétrique, elle s'établit dans un contexte qu'on pourrait qualifier d'hypernormatif, l'enseignant dispose d'informations pertinentes sur l'élève eu égard au jugement qu'il doit prononcer, etc. A la lumière de cette analyse, nous apporterons certaines préconisations sur les processus d'orientation, dans un but d'équité.

La première partie de ce rapport est consacrée à l'étude des processus fondamentaux dans la fabrication du jugement social. Cette partie nous est nécessaire pour comprendre pourquoi les jugements des enseignants peuvent, en certains cas, être biaisés. La deuxième partie est consacrée de façon spécifique à l'étude du jugement des enseignants. On y pointe la question de son « exactitude » mais on y développe aussi l'idée que le jugement est une attribution de valeur, qui peut avoir d'autres fonctions qu'une seule fonction descriptive des performances

des élèves. C'est dans cette partie qu'on répondra à la première question qui nous a été posée, qui est aussi la plus large : comment se construit le jugement professoral, sur quoi repose-t-il ? La troisième partie sera consacrée aux préconisations. Sur la base des constats établis dans les parties précédentes, nous nous risquerons à en tirer des conséquences en vue d'améliorer l'équité dans les pratiques d'orientation.

De quelques processus fondamentaux dans le jugement social

La formation d'un jugement consiste à produire des inférences à partir d'informations, ce qui implique plusieurs étapes de traitement de l'information qui sont autant de sources de biais potentiels (Dépret & Filisetti, 2001). Nous verrons que plusieurs sources de biais ont été mises au jour : effet de primauté, biais d'échantillonnage, illusion de contrôle, etc. (Fiske & Taylor, 1991). Certaines sources de biais tiennent dans le contexte : effet de contraste perceptif, effet de halo... D'autres tiennent dans les théories implicites des individus, qui piloteraient fortement l'inférence sociale. Tendance à sélectionner ou à accorder plus de poids aux informations susceptibles de confirmer nos attentes (biais de confirmation), interprétation des informations ambiguës dans le sens des attentes évitant ainsi la formation de croyances nouvelles et contre-intuitives sont autant de biais sociocognitifs potentiels. Darley et Gross (1983, cités par Yzerbyt & Schadron, 1996) évoquent plusieurs mécanismes qui peuvent conduire à des confirmations d'hypothèses : accorder davantage d'attention aux éléments cohérents avec l'hypothèse et, de ce fait, mieux mémoriser ces éléments ; accorder plus d'importance à ces éléments ; réinterpréter les informations contradictoires avec l'hypothèse ; attribuer les comportements contradictoires avec l'hypothèse à des facteurs situationnels (plutôt que dispositionnels). Par exemple, il a été montré empiriquement que le fait de recevoir *a priori* une information stéréotypique génère des attentes qui tendent à être confirmées par la perception des comportements observés, davantage en tout cas que si l'information stéréotypique a été donnée après que les comportements des agents visés par l'information ont été observés (Yzerbyt & Schadron, 1996).

Certains facteurs motivationnels ou socio-affectifs sont aussi aptes à influencer le jugement. Par exemple, la perception par l'individu de l'importance de sa tâche peut améliorer la validité du jugement (Yzerbyt & Schadron, 1996). De même, le degré de similarité entre la source et la cible peut affecter le jugement (les individus ont tendance à aimer davantage ceux qui leur ressemblent, à préférer ceux qui sont d'accord avec eux, ou ceux qui disent les aimer, etc.). Il en va de même du degré de gratification et de satisfaction associé à la cible, de l'attrance physique, ou encore du degré de proximité, de familiarité et d'intimité entre la source et la cible (Dépret & Filisetti, 2001).

Pourtant, nous essaierons aussi de montrer que, en matière de jugement, tout ne se joue pas dans une dialectique biais *versus* exactitude, qu'il faut aussi étudier les fonctions des jugements, que ceux-ci sont compris (aux deux sens du terme) dans un univers normatif et qu'ils ne sont pas seulement l'expression d'individus cognitivement limités dans leurs capacités à réaliser des inférences objectivement correctes. D'une part, les jugements sont aussi l'expression d'individus qui cherchent à interpréter leur environnement et les comportements d'autrui dans un but pratique (souvent très personnel, très contextualisé)

plutôt que dans un but de connaissance objective (transpersonnel, transcontextuel), ce qui peut orienter le jugement moins vers une exactitude globale que vers une exactitude locale (Swann, 1984). D'autre part, les agents sociaux, dans bien des cas, chercheront à attribuer de la valeur aux événements et aux individus qu'ils côtoient, plutôt qu'une connaissance descriptive et objective. Autrement dit, dans bien des cas, les jugements assertoriques sont aussi des jugements évaluatifs.

Juger dans des situations d'incertitude : les heuristiques et les biais

Les jugements des agents sociaux peuvent être abordés de divers points de vue, selon une orientation normative ou selon une orientation descriptive. Nous n'évoquerons ici que brièvement, dans un premier temps, l'orientation normative, puis nous nous arrêterons plus longuement sur l'orientation descriptive, qui est en rapport direct avec l'objet de cette note de synthèse. Il n'est cependant pas inutile d'évoquer l'orientation normative car, comme le précise Engel (1996), même l'orientation descriptive a besoin de références normatives, ne serait-ce que pour servir de point de comparaison, ou encore de modèle par lequel décrire les processus de jugement effectifs des agents sociaux (par exemple, suivent-ils les règles de la logique lorsqu'ils ont à raisonner sur un objet...).

Les jugements peuvent être étudiés du point de vue de la pensée « normale », non pas au sens de pensée « habituelle » ou « quotidienne » mais au sens de la manière dont on *doit* raisonner pour formuler un jugement. Il s'agit donc dans ce cas d'une question normative. En ce sens, les questions normatives traitent de savoir comment on doit appliquer des concepts et formuler des jugements, selon des théories normatives déterminées. La logique est une théorie normative par excellence du raisonnement mais on peut aussi penser à certains théorèmes statistiques par exemple. Dans ce cadre, il s'agit donc moins de se demander comment les individus raisonnent *en fait* que comment, *idéalement* (en droit pourrait-on dire), ils doivent penser. C'est de cette manière que se pose la question de la formation et de la validité des concepts, raisonnements et théories en philosophie : dans un raisonnement logique déductif, tout concept doit avoir des frontières précises, ne pas être vague, etc. (Engel, 1996). De nombreux auteurs, travaillant essentiellement dans une perspective rationaliste utilitariste, en économie ou en psychologie pour la plupart, ont développé des théories normatives sur le jugement et la prise de décision (comment maximiser ses gains dans un jeu, etc.). Ces théories introduisent en général un postulat de rationalité limitée (c'est un des points de changement par rapport aux questions que se posent les philosophes) et divers axiomes (tels que celui de la transitivité des préférences : si A est préféré à B et si B est préféré à C, alors A devrait être préféré à C). Elles proposent de décomposer de manière analytique, mais sous ce postulat de rationalité limitée, une situation complexe en des éléments plus simples et elles fournissent des règles sur la manière de combiner les conséquences supposées, qu'elles soient positives (gains) ou négatives (pertes), de différentes décisions possibles. Ces théories normatives de la prise de décision (*expected utility theory*, *multi-attribute utility theory*, *theory of reasoned action*, etc.) ont cependant un faible pouvoir descriptif. Manifestement, les agents sociaux ne semblent pas adopter, en tout cas pas de manière dominante, les théories normatives construites (cf. Van Der Pligt, 1996). À l'évidence, la rationalité des acteurs qui est engagée dans leurs opérations de jugement et de prise de décision, si rationalité il y a, n'est pas celle prônée par ces théories.

C'est pourquoi de nombreux auteurs se sont attachés à *décrire* les processus de raisonnement effectivement mis en place par les agents sociaux et, éventuellement, leurs écarts (et les raisons de ces écarts) par rapport à des règles normatives de raisonnement. Dans ce cadre, les travaux de Tversky et Kahneman (1974, 1981) ont grandement influencé la manière dont on conçoit la formation des jugements en situation d'incertitude, lorsqu'il s'agit d'estimer la probabilité d'occurrence d'un événement. Ils ont porté au jour que, dans de telles situations, les individus ont tendance à utiliser des sortes de *raccourcis de pensée*, qu'ils nomment des *heuristiques* (i.e. littéralement, ce qui sert à découvrir), qui ont l'avantage de conduire à des opérations de jugement simples, en présence de données pourtant complexes. Si ces heuristiques sont généralement efficaces, évitant d'avoir à se lancer dans des traitements trop compliqués et permettant notamment des prises de décision rapides (en tout cas pas trop coûteuses en termes cognitifs), elles peuvent aussi conduire à des erreurs systématiques. Dans un premier article important paru en 1974, Tversky et Kahneman ont porté au jour trois heuristiques.

L'heuristique de représentativité permet de juger du degré de correspondance entre un événement, une personne, etc. et un modèle mental. Souvent, il s'agit d'évaluer la similarité à un stéréotype (tel individu correspond-il au stéréotype de l'ingénieur par exemple ?). Ainsi, la probabilité que A fasse partie de B (B étant une catégorie, un processus...) et, par conséquent, possède telle qualité, tel trait, adopte tel comportement, etc., va être évaluée en fonction du degré selon lequel A est représentatif de B, c'est-à-dire en fonction du degré selon lequel A ressemble à B. Cependant, l'utilisation de cette heuristique conduit parfois à des erreurs grossières telles que la non-prise en compte d'indices pertinents.

L'heuristique de disponibilité énonce que l'évaluation de la fréquence d'une classe ou de la probabilité d'un événement va être affectée par la facilité avec laquelle certaines occurrences peuvent être accessibles à la pensée (accès en mémoire). Par exemple, estimer la probabilité d'une attaque cardiaque chez les adultes d'une certaine classe d'âge se fait en se remémorant de tels cas parmi nos connaissances. Cette heuristique peut être efficace pour juger des fréquences ou des probabilités car on se remémore mieux et plus vite des cas de grandes classes que des cas de classes plus petites. Toutefois, la disponibilité est affectée par d'autres éléments que la fréquence et la probabilité. Ces éléments peuvent par conséquent conduire à des biais. Dans cette heuristique, l'accès en mémoire de certains cas est très important. Cet accès peut être facilité par plusieurs phénomènes tels que la familiarité de certains cas : si on présente par exemple à un sujet des listes d'hommes et de femmes, en augmentant le nombre de personnes connues dans une classe donnée, le sujet aura tendance à surestimer le nombre d'individus appartenant à la classe en question. La saillance de l'événement entre aussi en compte (événements frappants ou récents par exemple). Des biais peuvent aussi apparaître dans l'efficacité de la recherche en mémoire (favorisant ainsi le rappel de certaines classes d'événements plutôt que d'autres). L'imagination joue aussi un grand rôle (le risque d'une expédition peut, par exemple, être affecté par le fait d'imaginer certains incidents). En général, les individus surestiment la probabilité d'un événement si des éléments concrets de cet événement sont accessibles en mémoire.

L'heuristique d'ajustement et d'ancrage énonce que, dans beaucoup de situations où les individus doivent fournir des estimations de probabilité ou de fréquence, ils commencent leur raisonnement à partir d'une valeur initiale (suggérée par la formulation du problème ou par un calcul partiel) qu'ils ajustent par la suite pour parvenir à une réponse finale. Or, différents points de départ donnent différentes réponses. Autrement dit, le choix du point de départ affecte la réponse finale et peut donc produire des biais. Dans une expérience célèbre, il était

demandé aux sujets d'estimer certaines quantités, par exemple le pourcentage de pays africains aux Nations Unies. Pour chaque estimation, un nombre de 0 à 100 était déterminé en faisant tourner une roue de la fortune en présence du sujet. Dans un premier temps, on demandait au sujet d'indiquer si ce nombre était supérieur ou inférieur à la valeur de la quantité demandée et, dans un second temps, d'estimer la quantité en s'éloignant vers des valeurs supérieures ou inférieures au nombre donné. Ce nombre, bien sûr totalement arbitraire, se révélait avoir un effet marqué sur l'estimation produite par le sujet. Ainsi, le pourcentage estimé médian de pays africains appartenant aux Nations Unies était de 25 et de 45 pour des groupes de sujets qui recevaient respectivement les valeurs de 10 et 65 comme valeurs de départ. Et pourtant, quel rapport entre le nombre déterminé par une roue de la fortune et la valeur recherchée ?

Par la suite, Kahneman et Tversky (1982) ont mis au jour une autre heuristique, nommée heuristique de simulation. Cette heuristique consiste dans la construction de scénarios hypothétiques pour tenter d'estimer ce qui va advenir. La facilité avec laquelle ce scénario hypothétique vient à l'esprit influe alors pour juger ce qui va vraisemblablement advenir dans la vie réelle ou pour réaliser des inférences causales sur les événements. Cette heuristique peut, en outre, exercer des effets sur certaines expériences affectives, en particulier dans les cas où un événement a été vécu comme frustrant et où l'on peut s'imaginer qu'une autre situation plus favorable aurait pu facilement advenir (du type « si seulement... »).

Ces heuristiques ne sont sans doute pas les seules à être utilisées par les agents sociaux dans la vie courante. La plupart des chercheurs s'accordent à dire que les heuristiques mises au jour par Tversky et Kahneman sont parmi les plus importantes et celles dont l'application est la plus générale, mais qu'il existe aussi d'autres heuristiques activées dans des classes de situation plus restreintes, voire des heuristiques quasi idiosyncrasiques, c'est-à-dire n'apparaissant qu'à l'occasion d'une situation particulière.

Tversky et Kahneman insistent sur le fait que, malgré leur efficacité générale, les biais qui peuvent être engendrés par ces heuristiques sont des biais cognitifs et qu'ils ne sont pas imputables à un quelconque facteur motivationnel. Ils avancent aussi que ces erreurs systématiques ne sont pas réservées seulement aux profanes. Mêmes des spécialistes peuvent se laisser abuser en fondant leurs estimations sur de telles heuristiques.

Est-ce à dire que les individus sont irrationnels ? Deux points méritent ici d'être soulignés. Premièrement, il existe selon toute vraisemblance des situations où les agents sociaux raisonnent davantage et utilisent donc moins ces raccourcis de pensée que sont les heuristiques. Il peut en être ainsi lorsque la prise de décision consécutive au jugement est très engageante, lorsque le temps disponible pour le raisonnement est assez long, lorsque les informations sont disponibles sous une forme qui se prête à l'utilisation de procédures mentales rationnelles, etc. Par exemple, Nisbett (cité par Fiske & Taylor, 1991, p. 393) a montré que, lorsqu'un problème repose sur une information « statistique » claire (problème lié au jeu, à la chance, mettant en évidence les aspects de variation aléatoire du phénomène), les agents sociaux utiliseraient une « heuristique statistique » qui met en œuvre des procédures proches de certaines règles statistiques abstraites (telles que la loi des grands nombres) et qui limite beaucoup certains biais de jugement. Les agents sociaux ne sont cependant pas toujours capables d'appliquer ces règles statistiques dans la vie de tous les jours car ils échouent parfois à reconnaître leurs situations d'application. En somme, la procédure utilisée serait toujours en partie liée à la situation et ne découlerait jamais entièrement d'une règle abstraite. Les processus mentaux que les agents sociaux engagent pour faire des inférences apparaissent

ainsi dans un entre-deux, ni totalement spécifiques à une situation réelle, ni totalement abstraits.

Deuxièmement, comme l'a montré Gigerenzer (cité par Engel, 1996), les individus semblent fonctionner bien davantage selon un mode de raisonnement fréquentiste que selon un mode de raisonnement bayésien (théorie subjectiviste des probabilités), qui a en grande partie servi de théorie normative à Tversky et Kahneman pour décrire les biais de raisonnement des agents. Lorsque les problèmes leur sont formulés en termes fréquentistes plutôt que sous le format de probabilités subjectives, cela conduit à beaucoup moins d'erreurs de leur part. Les agents pourraient en conséquence être rationnels, mais ne pas utiliser la bonne procédure de résolution, c'est-à-dire utiliser la mauvaise théorie normative. La notion de rationalité apparaît dès lors très relative. De plus, les agents sociaux pourraient avoir de bonnes raisons de se comporter ou de croire comme ils le font bien que ces raisons soient mauvaises au sens des théories normatives de raisonnement (cf. Boudon, 1995). En particulier, tous les comportements et toutes les croyances n'ont pas pour objectif unique ou principal de maximiser l'utilité (maximiser son intérêt propre). Et on peut douter que l'ensemble de ces raisons soit codifiable en termes de rationalité *versus* irrationalité.

La formation des impressions à l'égard d'autrui

Nombreux sont les travaux qui montrent que l'impression que nous nous faisons d'autrui ne s'opère pas sur la base d'un simple enregistrement de la réalité. D'abord parce que nous sélectionnons des éléments de cette réalité. Mais cela ne suffit pas. Notre connaissance d'autrui n'est pas une simple liste d'attributs. Elle suppose aussi une mise en cohérence de manière à parvenir à une image globale. Cela se réalise essentiellement par deux voies. D'une part, les éléments perçus sont organisés et reliés entre eux de manière à ce qu'ils correspondent les uns aux autres et soient subordonnés à l'ensemble (Yzerbyt & Schadron, 1996). D'autre part, des éléments non directement perçus sont inférés, qui vont permettre de compléter le tableau global et d'assurer certaines relations entre les éléments directement perçus. L'activité d'inférence est en général d'autant plus importante que les éléments directement perçus sont rares. Les éléments perçus et stockés en mémoire, de même que les éléments inférés (et en partie grâce à ces derniers) sont ainsi organisés et intégrés en une image globale. On sait par exemple que de nouveaux traits présentés comme caractérisant des personnes-cibles, après que d'autres traits de ces personnes-cibles sont connus de personnes évaluatrices, vont être interprétés en fonction d'une « direction évaluative » donnée par les premiers traits (Yzerbyt & Schadron, 1996). L'effet de primauté des premiers traits connus entraîne une évaluation des traits suivants en cohérence avec les premiers traits connus. On voit qu'il ne s'agit pas ici seulement d'une heuristique d'ancrage au sens de Tversky et Kahneman, mais aussi d'une mise en cohérence avec l'information initiale. Plusieurs approches sont en débat pour expliquer les mécanismes de mise en cohérence des divers traits, ce que l'on nomme les théories implicites de la personnalité. Certaines se fondent sur les couples d'oppositions entre traits en supposant l'existence d'une sorte de matrice de corrélations stockée en mémoire ; d'autres supposent l'organisation de l'ensemble des traits selon des grandes dimensions (notamment, une dimension intellectuelle et une dimension sociale) ; d'autres encore évoquent un fonctionnement par typologie, sorte d'idéal-types qui se distinguent entre eux par des conjonctions différentes de divers traits. Mais toutes s'accordent sur le fait que l'individu quotidien s'attache à mettre de l'ordre dans ses perceptions d'autrui

de manière à en avoir une appréhension globale, qui n'est pas synonyme de monolithique toutefois.

Les théories implicites de la personnalité interviennent dans la formation des impressions à l'égard d'autrui car elles créent des cadres d'interprétation qui affectent la perception d'autrui en accord avec ces cadres. Plusieurs théories implicites de la personnalité ont été mises au jour par les chercheurs. L'une des plus intéressantes pour notre propos est une théorie implicite relative à l'intelligence des personnes (e.g. Mueller & Dweck, 1998). Il a été montré que les individus pouvaient être distingués selon qu'ils considéraient que l'intelligence était un trait fixe et global (théorie de l'entité) ou bien selon qu'ils considéraient que les acquisitions étaient surtout le résultat du travail, des efforts fournis (théorie incrémentielle)¹. Dans le premier cas, les individus ont plutôt tendance à penser que l'intelligence est innée, stable, globale tandis que, dans le second cas, ils ont plutôt tendance à penser qu'il s'agit d'une habileté acquise, malléable et spécifique. Les comportements d'autrui et, pour ce qui nous intéresse en premier lieu dans cette étude, les comportements des élèves, peuvent ainsi être interprétés par les enseignants à l'aune de leurs théories implicites de la personnalité. Lorsqu'un élève démontre un bon niveau d'habileté dans un domaine spécifique, un enseignant qui adopte une théorie de l'entité aura ainsi tendance à attribuer également à cet élève un bon niveau de compétence dans des domaines connexes, voire peut-être dans des domaines éloignés, et à considérer que les habiletés révélées sont un bon indicateur du niveau de performances dans le futur. Il en ira bien différemment de la part d'un enseignant qui adopte une théorie incrémentielle.

Ces théories implicites de la personnalité ne se généralisent pas forcément à tous les attributs humains cependant. Un individu peut avoir une conception rigide (théorie de l'entité) pour un attribut et une conception malléable (théorie incrémentielle) pour un autre attribut tel que les caractéristiques morales ou bien certains traits (timidité, gentillesse, etc.).

A quoi attribuons-nous les causes des comportements observés ?

Une des raisons majeures pour lesquelles nous faisons des inférences à propos d'autrui, c'est que nous voulons connaître les causes de leurs comportements ou de ce qui leur arrive. Cette activité d'inférence causale a reçu beaucoup d'attention de la part des psychologues sociaux, qui se sont intéressés à la « psychologie quotidienne » des individus. Le pionnier en la matière a été Heider (1958). Pour lui, nous cherchons en permanence à savoir quels sont les déterminants de nos comportements et nous attribuons les causes des comportements d'autrui, et des nôtres à des forces situationnelles externes à la personne (l'environnement, la situation) ou à des forces internes (capacité et motivation, cette dernière incluant intention et effort.). Bien que le résultat d'une action puisse être le produit d'une combinaison de ces deux types de forces, c'est surtout la distinction entre causalité interne et causalité externe qui sera retenue. Pour Heider, l'homme quotidien se comporte un peu comme un scientifique naïf qui va chercher à connaître ce qui se passe autour de lui de manière aussi juste et valide que possible, et cela dans le but de mieux appréhender la réalité, la prévoir et la contrôler. On se doute par exemple que, selon qu'un enseignant va attribuer les causes des mauvaises notes

¹ Bien sûr, il s'agit plutôt d'une tendance marquée vers l'un ou l'autre de ces deux pôles. Certaines personnes développent des théories assez ambiguës de l'intelligence.

d'un élève à son manque de capacité ou bien à son environnement familial délétère, il en tirera des conclusions toutes différentes sur la manière de se comporter avec lui et sur la meilleure orientation pour lui.

Toutefois, et cela est important pour notre propos, avons-nous cette capacité d'inférence juste et valide ? Savons-nous distinguer de manière objective entre les forces situationnelles et les forces internes ? La réponse est que, en général, nous avons tendance à surestimer le poids des facteurs personnels dans l'explication des comportements et à sous-estimer le poids des facteurs situationnels. Ce phénomène a été nommé « erreur fondamentale d'attribution » par Ross (1977). Cette tendance peut connaître des variations, en particulier dans l'attribution par rapport à soi-même ; par exemple, il est assez fréquent d'expliquer nos propres échecs ou mauvais comportements par des forces externes et, inversement, nos propres réussites ou bons comportements par des forces internes, dans un but d'autocomplaisance. Globalement toutefois, les résultats sont tout à fait clairs et convergents pour montrer cette tendance à favoriser la causalité interne lorsque nous sommes confrontés à un événement. Divers types d'explication ont été avancés pour expliquer cette erreur fondamentale d'attribution. Il y a tout d'abord des interprétations cognitives ; elles renvoient à la saillance de l'explication interne (accessibilité en mémoire facilitée) ou à sa facilité de traitement : les explications internes seraient plus simples, donc moins coûteuses cognitivement (elles peuvent se réduire à l'invocation d'un trait de personnalité alors que les situations mettent en jeu des causalités complexes). Il y a aussi des interprétations motivationnelles : l'individu ayant un besoin de contrôle, il apparaît plus facile de contrôler des personnes que de contrôler des situations (qui peuvent en outre relever de la chance, du hasard, etc.). L'interprétation sacionormative postule quant à elle que certaines explications sont plus valorisées que d'autres. Les explications internes seraient choisies de manière préférentielle parce qu'il existerait, dans les sociétés occidentales, une norme de jugement : la norme d'internalité. Dans cette optique, les personnes qui, pour interpréter un événement, adoptent des explications internes, seraient mieux jugées que celles qui adoptent des explications externes et cela, indépendamment de la valence des événements (Beauvois & Dubois, 1988). Selon cette interprétation, si l'autocomplaisance peut sans doute permettre aux individus de maintenir un bon niveau d'estime de soi, en revanche, ils seraient moins valorisés que les individus qui donnent des explications internes de leurs bonnes comme de leurs mauvaises actions, de leurs réussites comme de leurs échecs. De nombreux résultats empiriques sont venus confirmer cette hypothèse dans des milieux très divers (entreprises, centres de formation, milieu scolaire, etc.). Cette norme d'internalité permettrait l'évaluation sociale dans nos sociétés occidentales individualistes et permettrait d'attribuer aux individus la responsabilité de leurs actes et de ce qui leur arrive. Dans bien des cas en effet, notre société enjoint d'évaluer des personnes (le cadre qui doit évaluer ses subordonnés, l'enseignant qui doit évaluer ses élèves, etc.), et l'attribution de causalité interne est tout à la fois ce qui va rendre cette exigence sociale possible et, dans le même temps, ce qui va permettre de définir la valeur des personnes, le rôle et la place qu'il doivent tenir dans la société.

Une autre interprétation mérite d'être avancée pour expliquer la plus grande valorisation sociale des explications internes : elles pourraient être perçues comme ayant un plus grand pouvoir d'explication causale. Ce sont en effet les individus qui agissent, non directement les situations, qui ne seraient pas considérées comme une causalité ultime mais comme des éléments modérateurs ou médiateurs dans un processus dont la cause reviendrait, *in fine*, à l'individu, non directement contraint par la situation (sauf situation particulière liée à l'imposition d'une force physique ou morale externe coercitive qui ne laisse aucun autre choix à l'individu que la soumission totale et qui, du coup, serait totalement privative de liberté).

Tant que l'individu serait vu comme ayant une marge d'initiative ou de liberté, il serait perçu comme recelant la causalité ultime de ses actes. Une telle hypothèse mériterait cependant d'être étayée empiriquement.

L'influence des stéréotypes

Un stéréotype est une connaissance socialement partagée à propos d'un groupe, qui tend à admettre que l'appartenance au groupe correspond à la possession d'une certaine propriété. Il peut toucher des attributs négatifs ou non, exacts ou faux². Un stéréotype a souvent prétention (en général implicite) à attribuer les attributs d'un groupe à sa « nature » (son essence). De là sans doute les éléments les plus fondamentaux pour discriminer, qui ont rapport avec l'indélébile, le corps : le port du corps, le maintien, l'élocution, l'accent, l'hexis corporelle, etc., c'est-à-dire les dispositions les plus fortement intériorisées, les plus fortement incorporées (cf. Bourdieu, 1979). Connaître la nature permettrait de prédire les comportements futurs des agents sociaux, leurs réactions avec autrui, etc. Plusieurs auteurs ont ainsi pu montrer que le simple fait de connaître la catégorie d'appartenance d'un individu conduit à des jugements plus affirmés sur cet individu.

Tout porte à penser que les stéréotypes jouent un rôle important dans le jugement d'autrui. Les stéréotypes ont fait l'objet d'un très grand nombre de travaux. Leur exploration laisse poindre une évolution dans la conception des stéréotypes et notamment de leurs fonctions. Dans les premiers travaux, l'agent social était vu comme un avare cognitif. Le stéréotype était une connaissance en permanence disponible, qui évitait de se lancer dans des traitements complexes. Les travaux suivants ont pourtant montré un agent social davantage perfectible dans ses processus de connaissance, d'attribution et de jugement, si on lui laisse du temps pour raisonner, s'il est persuadé de l'importance de sa tâche, s'il est conscient du risque d'erreur que lui font encourir les stéréotypes. Par exemple, le fait de disposer d'informations individuelles considérées comme pertinentes par rapport à l'objet du jugement (i.e. diagnosticité typique d'une information) produirait un « effet de dilution » qui conduirait les individus à délaisser l'information stéréotypique. Cela est important pour notre propos car les enseignants disposent d'informations pertinentes sur les élèves. Les résultats aux nombreux exercices proposés sont un bon indicateur pour juger de leur niveau et décider d'une orientation. Il faut cependant relever ici que les performances des élèves, même si elles révèlent en grande partie des habiletés cognitives, sont aussi le produit de facteurs situationnels souvent négligés dans leur ampleur, voire dans leur existence (ce qui nous ramène à l'erreur fondamentale d'attribution évoquée *supra*). On sait par exemple que la manière de présenter une tâche peut influencer de manière considérable sur les résultats obtenus en fonction des différentes catégories d'élèves. Les travaux de Monteil et de ses collaborateurs ont bien montré par exemple que l'habillage de la tâche (présenter une tâche comme étant du dessin *versus* de la géométrie), de même que le fait de rendre les performances visibles ou non (faire explicitement mention du niveau des élèves par exemple), pouvaient jouer de manière tout à fait frappante sur les performances des élèves et, en particulier, sur les écarts entre élèves faibles et forts (cf. la synthèse de ces travaux présentée dans Monteil & Huguet, 2002). Plus récemment, les travaux de Steele (e.g. 1997) ont montré que des personnes membres d'un groupe qui fait l'objet d'un stéréotype négatif (les noirs, les

² Certains auteurs réservent cependant le terme « stéréotype » à des croyances négatives et/ou erronées.

filles, etc.) pouvaient, dans les situations où elles devaient réaliser une performance intellectuelle, se sentir menacées par le stéréotype (les noirs qui seraient moins compétents, les filles qui seraient moins scientifiques, etc.), c'est-à-dire craindre, en cas de mauvaise performance, de confirmer les allégations négatives dont elles sont l'objet. La « menace du stéréotype » générée par la situation réduirait les performances (état de distraction vis-à-vis de la tâche, baisse de motivation, anxiété, conscience de soi augmentant la charge mentale et interférant avec la tâche à réaliser) lorsque l'échec peut être interprété comme la confirmation d'une infériorité groupale (Croizet, 2001). Les réputations d'infériorité intellectuelle mettraient en jeu une « menace du stéréotype » qui affecterait les performances cognitives... risquant ainsi de confirmer les allégations initiales.

Etant donné que la confiance dans le jugement augmente avec la quantité et la pertinence perçue des informations disponibles, il n'est pas impossible que les performances elles-mêmes entérinent des effets de stéréotype. Comme on le voit, la situation n'est pas simple. Si, à l'évidence, disposer des performances effectives est un des moyens les plus efficaces pour lutter contre les stéréotypes et parvenir ainsi à un jugement valide, il reste que, parfois, les performances elles-mêmes portent la trace des stéréotypes et, en particulier, de la menace qu'elles font peser sur les groupes stigmatisés. On ne saurait trop ici recommander aux enseignants de faire varier les conditions d'évaluation et de les rendre les moins menaçantes possible. Dédramatiser l'évaluation, faire porter les comparaisons sur les performances anciennes de l'élève de manière à raisonner en termes de progrès plutôt que faire porter les comparaisons sur les performances d'autrui, anonymiser le rendu des copies sont, par exemple, des moyens de réduire cette menace.

Toutefois, il faut encore aller au-delà de cette vision qui laisse voir un individu, certes de bonne volonté, mais tout de même pas très doué dans ses opérations de jugement. Il faut voir que les stéréotypes peuvent aussi être l'expression de nos capacités intellectuelles et de nos compétences sociales et que les facteurs motivationnels (objectifs, valeurs...) sont partie intégrante de l'activité de jugement. Il serait faux de penser que juger consiste simplement à récupérer une information en mémoire. Il s'agit au moins d'une interaction entre les informations en mémoire et l'interprétation de la situation de jugement (son sens social, ses objectifs...). Il est vrai toutefois que les agents sociaux sont peu conscients des processus qu'ils utilisent pour former leur jugement. Les stéréotypes peuvent être vus comme de véritables théories naïves qui revêtent une dimension explicative, apportant des informations nouvelles, donnant un sens aux informations disponibles et permettant de gérer des informations divergentes. En effet, les individus sont en permanence confrontés à des informations incohérentes voire contradictoires. Ces incohérences ou contradictions ne sont pas toujours niées ou ignorées. Elles font bien souvent l'objet d'un travail d'explication et de réconciliation. Dans cette optique, les stéréotypes seraient bien davantage utilisés pour faire face à un manque d'informations que pour alléger la charge cognitive d'un individu qui serait submergé par trop d'informations.

Les stéréotypes permettent aussi de justifier les rapports qui existent entre notre groupe et les autres groupes. Tout stéréotype de groupe met en effet en évidence la question de l'identité sociale. On sait par exemple que l'endogroupe (i.e. le groupe auquel on appartient) est perçu d'autant plus homogène qu'il est davantage minoritaire. Les stéréotypes obéiraient à des fonctions : ce serait des genres de « fictions » qui auraient une fonction de rationalisation, en naturalisant les différences sociales, ce qui permettrait aux individus de justifier l'ordre social établi (c'est ainsi, par exemple, que les groupes stigmatisés font pratiquement toujours, aussi, l'objet d'une infériorité intellectuelle, dans nos sociétés où l'idéologie veut que chacun puisse

réussir selon ses mérites). Cela permettrait d'assurer la reproduction sociale, notamment par l'intériorisation des images (positives ou négatives) de soi. On peut ainsi comprendre qu'un groupe dominé en vienne à adopter sur lui-même le point de vue du groupe dominant. « Dominées jusque dans la production de leur image du monde social et par conséquent de leur identité sociale, les classes dominées ne parlent pas, elles sont parlées » ; elles seraient une « classe objet » « dépossédée du pouvoir de définir sa propre identité », écrivait Bourdieu (2002, pp. 255-256). Les stéréotypes auraient donc une vertu « explicative » de l'ordre social établi.

Ainsi les stéréotypes ne sont-ils pas seulement le résultat d'une corrélation entre des comportements et des groupes sociaux. Encore faut-il que ces actes soient attribués à des facteurs dispositionnels. Les stéréotypes ont ainsi à voir avec l'attribution d'une identité sociale, une « nature » ou d'une « essence » sociale.

D'où viennent les catégories que nous utilisons pour classer, juger ?

Nous avons vu que, dans les derniers travaux de psychologie sociale, au-delà d'une fonction supplétive d'information, les stéréotypes sont conçus comme des catégories qui reflètent et ont pour fonction de justifier les rapports entre les différents groupes sociaux. Ils ont donc à voir avec un certain ordre social, avec l'attribution d'une identité sociale en naturalisant les rapports sociaux pour les transmuter en des caractéristiques personnologiques.

Certes, nous avons « en tête » des catégories qui nous permettent de classer, de juger les individus (entre autres). Mais d'où nous viennent ces catégories ? Un point important consiste à dire que, si nous pouvons raisonner sur la manière dont nous classons les individus, au moins partiellement, voire expliciter ce que revêt, pour nous, telle ou telle partition que nous adoptons et telle ou telle répartition des objets et des individus dans diverses catégories, si nous avons à cet égard une certaine lucidité, voire une certaine liberté, nous ne sommes néanmoins pas libres du choix de nos catégories fondamentales de perception et d'appréciation³, nous n'avons pas d'accès direct aux principes par lesquels nous les intériorisons. De même que, lorsque nous parlons, nous n'avons pas de représentation claire des différents phonèmes qui nous ferait choisir telle élocution, tel accent plutôt que tel autre, de même les catégories par lesquelles nous classons les individus ne résultent pas d'un choix conscient, ni d'une représentation claire des principes par lesquels nous classons et apprécions autrui. Ce qui ne signifie que nous ne soyons pas capables de moduler notre intonation, notre élocution, voire notre accent en fonction de la situation particulière dans laquelle nous nous trouvons, mais cela se réalise bien davantage de manière agie que clairement représentée.

En ce sens, un auteur comme Bourdieu a tenté de montrer que les catégories par lesquelles nous jugeons autrui sont des catégories incorporées (au sens propre, c'est-à-dire acquises par le corps), reflétant pour partie l'ordre social, qui deviennent des principes de vision pratique du monde : un système de schèmes de perception, d'appréciation et d'action (ou, autrement

³ Le langage, instrument de catégorisation, est à cet égard exemplaire car nous n'avons pas choisi le langage qui nous permet d'effectuer les activités de classement. Songeons à ce que seraient nos catégorisations si, comme les Chinois anciens, nous ne possédions pas le verbe « être ».

parlant, des structures mentales). Entre la réalité et la construction que s'en fait l'agent social, il y a une activité structurante dans notre perception de la réalité, qui a pour opérateur un système de schèmes de perception et d'appréciation. Cette activité structurante ne provient pas d'un système de catégories universelles mais d'un système de schèmes incorporés, constitués au cours de l'histoire collective et acquis au long de l'histoire individuelle, fonctionnant à l'état pratique et pour la pratique (c'est-à-dire sans représentation explicite). C'est pourquoi cette activité de classement n'a pas *stricto sensu*, une fonction de connaissance au sens scientifique ou scolastique du terme, de recherche d'une quelconque cohérence logique, ou de quête de l'objectivité. Du confus, de l'indistinct, il s'agit de créer de la distinction, de la différenciation. De grandeurs continues, il s'agit de tirer du discontinu, du discret, séparant par exemple les blancs des noirs (rendant de ce fait, au moins dans un premier temps⁴, tous les blancs identiques entre eux et différents de tous les noirs, eux-mêmes tous semblables). Les systèmes de classement opposent les groupes dans le même mouvement qu'ils les caractérisent. Nombre de travaux ont en effet montré que le seul fait de créer (même fictivement) des groupes, de les juxtaposer, conduit à des opérations de dissimilation et d'assimilation, créant de l'homogénéisation dans l'endogroupe et de la différenciation avec l'exo-groupe (Yzerbyt & Schadrin, 1996). Les systèmes de classement fixeraient ainsi, mieux encore que des limites, de véritables frontières à défendre (*versus* à détruire, à déplacer...) (Bourdieu, 1979, p. 556). Mais on voit que cette destruction ou ce déplacement des frontières ne peut se faire que si l'on a réussi à imposer un autre système de classement (surtout dans le premier cas) ou une nouvelle définition du groupe en question.

C'est pourquoi, pour Bourdieu, de manière très générale, il n'y a pas de perception sans intérêt à percevoir (pas de perception sans évaluation dirait Beauvois, 1976, dans un sens voisin). L'intérêt porté à un aspect (trait individuel...) est lié à l'intérêt que les individus ont à percevoir (apercevoir ce trait). L'intérêt est ici conçu, non pas dans un sens strictement économique, mais comme prise au jeu, reconnaissance des enjeux (*illusio*). Les « jugements d'attribution » par lesquels on assigne quelqu'un à une catégorie sociale ont à voir avec la définition d'une véritable identité, ont pour enjeu le fait d'en faire partie ou de ne pas en faire partie, avec tous les avantages et désavantages qui y sont liés, et c'est pourquoi les systèmes de catégories ne sont pas seulement des instruments de connaissance, mais aussi des instruments de pouvoir.

Par conséquent, la perception du monde social n'est pas un simple enregistrement, un simple reflet de la réalité. C'est un acte de *connaissance* (pratique) qui met en œuvre des principes extérieurs à l'objet. Mais c'est aussi un acte de *méconnaissance* car il ne renferme pas la maîtrise de ces principes et des relations qu'ils entretiennent avec l'ordre social établi. L'activité de jugement (comme d'autres activités sociales) est pour P. Bourdieu « une connaissance sans concept » (1979, p. 548), c'est-à-dire une connaissance pratique, agie plutôt que représentée.

⁴ Un système de sous-catégories étant toujours possible.

La construction du jugement des enseignants

Les jugements des enseignants sont-ils « exacts » ?

Plusieurs travaux ont tenté d'évaluer l'exactitude du jugement des enseignants. Il est important toutefois de distinguer de quel type de jugement il s'agit, selon que nous avons affaire à des mesures directes ou indirectes. On dit qu'un jugement constitue une mesure directe du niveau de performances des élèves lorsqu'on demande aux enseignants de juger du degré de réussite de leurs élèves à une épreuve ou un exercice qui est effectivement présenté à l'enseignant. Par exemple : à votre avis, quel sera le score obtenu par cet élève au test suivant ? Réussira-t-il à résoudre cet exercice ? Etc. Dans le cas de mesures indirectes, l'enseignant doit évaluer le niveau de ses élèves dans un domaine particulier mais sans qu'il lui ait été demandé de prononcer son jugement sur la base d'épreuves ou d'exercices particuliers qui lui auraient été présentés à voir. Par exemple : à votre avis, quel est le niveau de cet élève en arithmétique ? Que vaut cet élève par rapport aux autres élèves de la classe en grammaire ? Etc.

Une étude de Hoge et Butcher (1984) traite du jugement des enseignants comme mesure directe des performances des élèves. Elle donne une corrélation de 0,85 qui indique que le jugement des enseignants est par conséquent très prédictif de la réussite effective au test proposé. Les auteurs avaient aussi demandé aux enseignants de donner sur une échelle en 5 points leur degré de confiance dans le jugement qu'ils énoncent pour chaque élève. Ce degré de confiance ne s'est pas révélé prédictif des écarts entre les scores prédits et les scores réels obtenus. Autrement dit, les enseignants manquent de confiance dans leur jugement, ou bien ne sont guère perspicaces dans leur capacité à juger la performance des élèves, même si leur jugement est en général exact. On peut dès lors se demander quel serait le degré de consistance de leur propre jugement si on leur fournissait une information nouvelle, ou bien lors d'un débat ou d'une discussion en conseil de classe à propos des performances des élèves.

Quelques études ont demandé aux enseignants de prédire quel serait le résultat de leurs élèves à des items particuliers. Par exemple, une étude de Coladarci (1986) a montré qu'entre 70 et 77 % des prédictions des enseignants se révélaient exactes. Les prédictions étaient meilleures pour les élèves forts que pour les élèves faibles, et elles étaient meilleures lorsqu'il s'agissait de prédire le résultat à un item de calcul mathématique plutôt que de résolution d'un item impliquant la compréhension de concepts mathématiques. Toutefois, ces résultats dépendent beaucoup du niveau de difficulté de l'item. Si l'item est facile, il est aisé de prévoir que les forts vont réussir. Inversement, si l'item est difficile, il est sans doute aisé de prévoir que les faibles vont échouer. Mais quels seront les résultats des élèves faibles dans le premier cas et des élèves forts dans le second ? La prédiction apparaît en ce cas beaucoup moins assurée.

Une bonne synthèse des travaux sur le jugement des enseignants a été faite par Hoge et Coladarci (1989). Dans une revue de 17 études, les auteurs révèlent des corrélations jugement/performance qui varient de 0,28 à 0,92 avec une corrélation médiane de 0,66. Il est utile cependant de distinguer les résultats selon plusieurs critères :

- selon que les mesures de jugement sont directes ou indirectes (rappelons que dans le premier cas l'épreuve est présentée à l'enseignant et pas dans le second). La différence de résultats

n'est guère importante : la corrélation médiane vaut 0,69 dans le premier cas et 0,62 dans le second, montrant par conséquent une exactitude un peu plus élevée dans le cas de mesures directes. Les variations enregistrées dans les différentes études recensées s'établissent entre 0,48 et 0,92 dans le cas de mesures directes, entre 0,28 et 0,86 dans le cas de mesures indirectes. Cela confirme que le jugement est plus conforme aux performances quand les enseignants ont directement pris connaissance du contenu des épreuves. Les écarts sont cependant loin d'être dramatiques.

- selon le degré de spécificité du jugement. La corrélation est un peu plus élevée quand le jugement est plus explicite. Estimer des scores conduit en général à un jugement moins exact (corrélation médiane de 0,61) que s'il s'agit d'estimer des équivalents-années (corrélation médiane de 0,70), d'estimer la réussite à des exercices particuliers (corrélation médiane de 0,70), ou d'estimer des rangs (corrélation médiane de 0,76). Estimer des rangs revient en général à comparer les performances des élèves à celles de leurs pairs dans la classe. Nous y reviendrons.

- selon que le jugement est référencé à une norme ou référencé à un critère indépendant des performances des pairs⁵. Dans le premier cas, la corrélation médiane vaut 0,64. Dans le second cas, la corrélation médiane vaut 0,68. Il n'y a donc pas de différence flagrante d'exactitude selon que les performances des élèves sont ou non référencées aux performances de leurs pairs. Cependant, si ce constat vaut en moyenne, la variabilité des corrélations est beaucoup plus grande dans le premier cas que dans le second. En effet, les corrélations enregistrées dans les différentes études varient de 0,28 à 0,92 quand le jugement est référencé à une norme, alors qu'elles varient de 0,67 à 0,72 seulement quand le jugement est référencé à un critère. Il est cependant possible que l'exactitude d'un jugement référencé à une norme soit dépendante de la familiarité avec le groupe de référence (Perry & Meisels, 1996). Par exemple, selon que le groupe de référence sera l'ensemble de la population correspondante (faible familiarité) ou bien les autres élèves de la classe (grande familiarité) on peut comprendre que l'exactitude du jugement sera susceptible de beaucoup varier. La dernière distinction, *infra*, confirme ce fait.

- selon que l'appartenance à la classe est prise en compte ou non. Lorsque les classes sont indifférenciées, la corrélation médiane est de 0,64 (les corrélations variant de 0,28 à 0,86) mais la corrélation médiane intra-classe est de 0,70 (les corrélations variant de 0,48 à 0,92). Ces résultats sont importants pour nous dans le cadre de la présente étude. Ils montrent en effet que les corrélations intra-classes sont plus élevées que celles qui s'établissent sur l'ensemble de l'échantillon, c'est-à-dire lorsque l'on ne tient pas compte du fait que les élèves appartiennent à des classes différentes. Cela signifie que les enseignants établissent bien, dans leur classe, un jugement qui est conforme à ce que produisent les résultats aux épreuves standardisées, mais qu'ils ont davantage de difficultés à situer le niveau de leurs élèves dans l'ensemble de la population correspondante. Ce constat est important et nous y reviendrons pour expliquer les effets de contexte dans la construction du jugement.

Globalement, les jugements des enseignants sont donc relativement valides et la justesse des jugements ne dépend que dans des proportions modérées des différences de méthodes

⁵ L'évaluation peut être référencée à une norme ou à un critère. Dans le premier cas, on évalue les élèves en les comparant à une norme constituée par les performances de leurs pairs : donner des scores en équivalents-années, en percentiles d'une distribution, produire un rangement, etc. Dans le second cas, on évalue les élèves sur un critère qui n'implique pas de comparaison directe avec un groupe de référence : estimation du nombre d'items réussis, probabilité de réponse exacte à un exercice, etc.

employées par les chercheurs. En général cependant, l'exactitude du jugement augmente avec le degré de spécificité de la mesure de performances, probablement parce que cela donne moins lieu à des interprétations variées. Ce sont les rangs qui donnent la meilleure exactitude de jugement, sans doute en grande partie parce que cela revient à comparer les résultats à un groupe de référence très familier, celui des autres élèves de la classe (et aussi peut-être parce que cela conduit à s'abstraire de la prédiction d'un score somme toute arbitraire).

L'exactitude du jugement est aussi meilleure dans certaines circonstances que dans d'autres. Les jugements sont en général plus exacts pour les élèves forts que pour les élèves faibles. Les jugements sont en général plus exacts dans les disciplines telles que la lecture et les mathématiques que dans des disciplines telles que les sciences sociales. De même, les jugements semblent plus exacts, ce qui n'est pas étonnant, lorsqu'il s'agit d'évaluer un comportement observable (e.g. calcul en mathématiques) que des processus inobservables (e.g. raisonnement en mathématiques).

Si l'exactitude du jugement des enseignants apparaît en moyenne assez élevée, en revanche, il semblerait que les enseignants varient considérablement dans leur interprétation de la capacité de descripteurs fins à prédire la réussite scolaire (e.g., « se rappelle bien les détails des histoires », « oublie les instructions », « est maladroit et empoté »...). La précision sur les résultats est donc assez bonne mais les enseignants ne tombent guère d'accord sur les prédicteurs des performances.

Cela explique sans doute en partie pourquoi l'exactitude du jugement varie aussi d'un enseignant à l'autre. Hoge et Coladarci (1989) soutiennent que les enseignants diffèrent de manière parfois considérable dans l'exactitude de leur jugement. Par exemple, dans une étude de Hopkins *et al.* (citée par Hoge & Coladarci, 1989), les corrélations intra-classes variaient de 0,44 à 0,88. Les variations apparaissent beaucoup plus considérables encore dans une étude de Helmke et Shrader (citée par Hoge & Coladarci, 1989), où les corrélations intra-classes variaient de 0,03 à 0,90, c'est-à-dire d'une indépendance entre jugements et performances à une corrélation quasi parfaite.

Il se peut toutefois que ces grandes variations dans l'exactitude du jugement des enseignants proviennent d'un problème de variance d'échantillonnage. Dans nos propres études pour lesquelles, précisons-le, le jugement est référencé à un critère, nous avons trouvé des différences mais souvent assez modestes d'exactitude d'un enseignant à l'autre (Bressoux, à paraître). Par exemple, dans une de nos recherches, nous avons trouvé que certaines corrélations faibles pouvaient être attribuées à un petit nombre d'élèves évalués (petits effectifs évalués ou données manquantes). Si l'on exclut de l'analyse les deux classes à effectifs très faibles, les corrélations intra-classes varient de 0,42 à 0,92 pour le français (la corrélation totale valant 0,73) et de 0,50 à 0,88 pour les mathématiques (la corrélation totale valant 0,69). Soit des mesures somme toute assez proches de celles obtenues par Hopkins *et al.* Même si l'on maintient ces classes dans l'analyse, mais avec des procédures qui permettent de prendre en compte la variance d'échantillonnage (modèles multiniveaux), la variance de la relation entre le jugement et les scores aux épreuves standardisées, quoique significative, n'est pas très importante.

Une méta-analyse réalisée par Dusek et Joseph (1983) a permis de rassembler les résultats de plusieurs recherches ayant porté sur les informations qui influencent le jugement des enseignants. Certaines informations semblent produire des relations équivoques avec le jugement des enseignants : il en va ainsi de l'influence des performances des frères et sœurs

qui ont eu le même enseignant, des stéréotypes de prénoms (certains prénoms étant plus « désirables » que d'autres), ou encore des comportements liés à un rôle de genre (comportements féminins *versus* masculins). L'analyse de Dusek et Joseph ne montre pas de lien entre le sexe des élèves et le jugement des enseignants mais il est probable qu'en ce domaine la discipline scolaire évaluée joue un grand rôle : si les mathématiques par exemple sont plutôt vues comme une discipline masculine, la littérature est en revanche plutôt vue comme féminine. D'autres informations apparaissent clairement influencer le jugement des enseignants : il en est ainsi de la conduite de l'élève en classe, des informations transmises par dossier sur l'élève, de l'origine ethnique (blancs, noirs, hispaniques...), de la classe sociale.

De plus, certains facteurs inattendus s'invitent à la formation du jugement que d'autres forment sur autrui. Il en est ainsi de l'attractivité physique. Ross et Salvia (1975) ont montré qu'elle exerce un effet sur le jugement scolaire des enseignants. Leur expérience consiste à présenter aux enseignants des dossiers d'élèves, dont il est dit qu'ils ont un niveau scolaire faible, et dont la photographie révèle un physique tantôt attractif, tantôt peu attractif. Les élèves physiquement peu attractifs sont moins bien jugés : l'accord est plus fort pour eux sur un éventuel placement dans une classe spéciale pour élèves mentalement retardés, sur des problèmes psychologiques ultérieurs, sur des difficultés dans les rapports ultérieurs avec leurs pairs et sur des difficultés scolaires ultérieures.

Ritts, Patterson et Tubbs (1992) ont mené une synthèse de vingt-huit études (incluant celle de Ross et Salvia, 1975) sur les liens entre jugement et attractivité physique. Leur synthèse confirme que les élèves au physique attractif sont en général jugés plus favorablement sur plusieurs dimensions incluant l'intelligence, le potentiel scolaire, les notes et diverses habiletés sociales : les élèves physiquement attractifs sont jugés plus amicaux, plus prévenants (*attentive*), plus populaires et plus ouverts (*outgoing*). Ces résultats confirment l'impact sur le jugement des enseignants du stéréotype énoncé par Dion, Berscheid et Walster dans les termes : « Ce qui est beau est bien ». Toutefois, Ritts, Patterson et Tubbs concluent de leur synthèse que la taille d'effet est plus forte pour la relation entre l'attractivité physique et les jugements portant sur les habiletés sociales (0,48) que pour la relation entre l'attractivité physique et les jugements académiques tels que l'intelligence, le potentiel scolaire et les notes (0,36). Un des problèmes avec ces travaux expérimentaux, c'est qu'on peut se demander si, en milieu naturel, le jugement des enseignants est effectivement influencé par l'attractivité physique des élèves, et avec la même intensité, ou bien si l'influence de l'attractivité physique est due aux conditions expérimentales. On peut en effet décrire deux grandes caractéristiques de ces études, qui en limitent la validité écologique. D'une part, ce ne sont pas les élèves de l'enseignant lui-même qu'il doit juger ; son jugement n'est par conséquent pas engageant. D'autre part, l'enseignant n'a à sa disposition qu'un faible nombre d'informations sur les élèves. On ne sait donc pas si cette information serait utilisée si l'enseignant en possédait d'autres comme c'est le cas dans sa classe.

Pourquoi certains enseignants sont-ils plus « exacts » dans leur jugement que d'autres : le dogmatisme comme facteur de biais

Brophy et Good (1974) distinguent trois figures idéal-typiques d'enseignants. Les enseignants « sur-réactifs » (*overreactive*) tendent à percevoir les élèves de manière très rigide et stéréotypée. Ils sont en général très conventionnels, autoritaires et dogmatiques et, de ce fait,

traitent les élèves sur la base de stéréotypes plutôt que sur une base individuelle, exacerbant ainsi les différences initiales. Ces enseignants sont ceux chez lesquels on rencontre le plus systématiquement de comportements différenciateurs. Les enseignants « réactifs » (*reactive*) développent quant à eux des jugements peu rigides et les ajustent régulièrement en fonction des feedback qu'ils reçoivent de la part des élèves. Ces enseignants n'exacerbent pas les différences entre élèves et réagissent principalement en fonction des comportements des élèves. Les enseignants « pro-actifs » (*proactive*) cherchent quant à eux à compenser activement leurs jugements de manière à ce que les interactions maître-élèves ne soient pas déterminées par les élèves, par exemple en sollicitant plus souvent la participation des faibles.

Babad, Inbar et Rosenthal (1982) montrent que les enseignants très dogmatiques, qui sont ceux qui biaisent le plus l'information en exagérant les différences réelles entre élèves, sont aussi ceux qui manifestent les comportements les plus différenciés et les plus rigides vis-à-vis des élèves qu'ils jugent forts ou faibles. Leur étude mérite qu'on la présente de manière détaillée. Les auteurs veulent étudier les liens entre le jugement qu'émettent les enseignants et certaines caractéristiques de leur personnalité. Les auteurs se posent la question de savoir 1) qui est susceptible de biaiser l'information de manière stéréotypée, 2) quels sont les comportements différenciés des enseignants « biaisés » (*high-bias teachers*) et « non biaisés » (*no-bias teachers*), 3) quels sont les effets des jugements sur les performances effectives des élèves.

Afin de déterminer les enseignants « biaisés » et « non biaisés », il a été enseigné à des enseignants (stagiaires) d'éducation physique comment noter le Goodenough-Harris Draw-A-Person test. Puis il leur a été demandé de noter le dessin de deux enfants présentés comme ayant un statut différencié (haut statut et bas statut social, indiqué par le nom, l'ethnie, le niveau culturel et socioprofessionnel des parents, l'école fréquentée, le quartier de résidence, le nombre de frères et soeurs, etc.). En fait, ces dessins étaient tirés du manuel du test ; leur score « objectif » était donc connu. La différence entre les deux scores (corrigé de la différence « objective ») détermine un index de propension à biaiser l'information et permet de déterminer des groupes extrêmes d'individus. Sur les 86 enseignants qui ont passé le test, les auteurs extraient deux groupes contrastés d'enseignants : 15 *high-bias teachers* et 11 *no-bias teachers*. Ils demandent alors à ces enseignants de fournir les noms de 3 élèves forts et de 3 élèves faibles dans leur classe. De plus, les auteurs induisent des attentes élevées vis-à-vis de 2 élèves dans chaque classe (méthode type « *late blooming* » dans le fil de l'étude Pygmalion). Bien sûr, ces élèves ont en fait été choisis au hasard. Les attentes induites doivent permettre de tester si les éventuelles différences de comportements sont ou non justifiées par le niveau réel des élèves.

Des observations d'une séquence d'enseignement sont faites par deux observateurs sur une série de variables comportementales. Les deux observateurs ne savent pas qui sont les *no-bias teachers* et qui sont les *high-bias teachers*. Après synthèse des résultats, il est apparu 4 facteurs distincts et interprétables en matière de comportement des enseignants : le dogmatisme (rigidité, intolérance...), la distribution de l'attention et des renforcements (*responsive*), le comportement critique (blâme...), et le comportement amical. De plus, les auteurs ont extrait un facteur général de comportement censé refléter la qualité de l'enseignement (qualité des comportements, bonne écoute, faibles réprimandes, etc.). Pour la mesure des performances des élèves, il s'agit de mesurer en fin de séance des performances physiques (courir, sauter...) : 1) nombre de *sit-ups* (pour les filles) et de *push-ups* (pour les garçons), 2) saut en longueur, 3) temps de course entre deux points en évitant des blocs. Les

résultats révèlent plusieurs éléments intéressants, que nous déroulons en reprenant l'ordre des trois questions posées initialement par les auteurs.

1) on observe une corrélation négative entre le niveau culturel des parents et la propension à biaiser l'information de même qu'une corrélation négative entre l'âge et la propension à biaiser l'information.

2) Concernant les comportements en classe. Pour la variable de dogmatisme, on observe un effet fort et très significatif. Les *high-bias teachers* ont en moyenne un comportement plus dogmatique que les *no-bias teachers*. De plus, alors que les *no-bias teachers* se comportent de la même façon avec les 3 groupes d'élèves, les *high-bias teachers* se comportent de manière beaucoup plus dogmatique avec les élèves qu'ils jugent faibles.

En ce qui concerne la distribution de l'attention et des renforcements, on n'observe pas de différences entre les *no-bias* et les *high-bias teachers*. Les deux groupes d'enseignants accordent globalement moins d'attention et de renforcements aux élèves jugés faibles.

Globalement, les *no-bias* et les *high-bias teachers* ne se différencient pas sur la quantité de blâmes et critiques. Cependant, leur distribution varie. La distribution est équitable pour les *no-bias teachers* entre les trois groupes d'élèves, alors que les *high-bias teachers* critiquent beaucoup plus les élèves qu'ils jugent faibles.

Globalement, on n'observe pas non plus de différences entre les *no-bias teachers* et les *high-bias teachers* sur le nombre de comportement amicaux vis-à-vis des élèves. Là encore cependant la distribution varie. Les *no-bias teachers* se comportent équitablement avec les trois groupes d'élèves tandis que les *high-bias teachers* sont beaucoup moins amicaux envers les élèves qu'ils jugent faibles.

Les auteurs relèvent une tendance ($p < 0,06$) indiquant une différence entre les *no-bias teachers* et les *high-bias teachers* sur le facteur général de comportement censé refléter la qualité de l'enseignement. Surtout, les *no-bias teachers* ne traitent pas différemment les trois groupes d'élèves sur cette dimension, tandis que les *high-bias teachers* les traitent très différemment : ils démontrent plus de comportements négatifs envers les faibles, et davantage de comportements positifs en faveur des élèves pour qui des attentes positives ont été expérimentalement induites. En effet, le comportement vis-à-vis de ces élèves est identique à celui que les *high-bias teachers* adoptent vis-à-vis des élèves qu'il jugent d'eux-mêmes forts (sans attentes induites).

3) Les différences enregistrées entre les *no-bias* et les *high-bias teachers* sur les performances des élèves (en contrôlant l'habileté initiale) sont moins fortes qu'elles ne le sont sur le comportement des enseignants. On relève toutefois certains effets significatifs. Les performances des élèves faibles sont moins bonnes chez les *high-bias teachers* tandis qu'on n'observe pas de différences chez les *no-bias teachers*⁶.

En conclusion, cette étude est riche d'enseignements pour l'objet du présent rapport. Les résultats présentent une image intégrée des relations entre dogmatisme, propension à biaiser l'information et effets d'attentes. Les enseignants qui varient dans leur propension à biaiser

⁶ Sauf pour le saut en longueur, sans doute parce que cet exercice est, parmi ceux qui étaient proposés, celui qui met en jeu le moins d'efforts prolongés.

l'information varient dans la probabilité de manifester des comportements différenciés des élèves qu'ils jugent forts ou faibles, et le dogmatisme semble être le corrélât (l'antécédent ?) majeur de ces phénomènes. On rejoint ainsi la typologie de Brophy et Good (1974). Les *high-bias teachers* de l'étude de Babad, Inbad et Rosenthal apparaissent dogmatiques, leur jugement est très sensible aux stéréotypes et, sur cet ensemble de points, ils ressemblent étonnamment aux enseignants « sur-réactifs » (*overreactive*) de la typologie de Brophy et Good (1974).

Les interprétations psychologiques ont été souvent remises en cause dans les recherches récentes. Il faut se garder de voir le dogmatisme comme un trait trop intégré à la personne. Le fait que, dans l'étude de Babad, Inbad et Rosenthal, ce soient les plus jeunes qui aient le plus tendance à biaiser l'information, à reposer leur jugement sur des stéréotypes sociaux, peut aussi s'interpréter comme un manque d'assurance, ou de connaissances. A notre sens, le dogmatisme ne doit pas nécessairement être entendu comme un trait de personnalité, qui supposerait qu'il caractérise l'être de la personne, sorte de qualité immuable et inébranlable. Cette attitude (plutôt que ce trait) peut être aussi fonction des situations, faire l'objet d'un apprentissage social, varier selon les affects de la personne, etc. Gardons-nous donc de retomber, par un curieux travers, dans le paradigme des variables de présage qui affichait l'ambition démesurée de détecter la qualité des enseignants dans des variables psychologiques. Ce courant prétendait qu'il suffisait de repérer, grâce à des tests appropriés, certaines caractéristiques chez les candidats pour sélectionner les futurs « bons » enseignants. Le bilan de ce courant de recherche a été suffisamment cinglant pour qu'on ne répète pas ici les mêmes erreurs ni les mêmes inférences (signalons d'ailleurs au passage que cela reviendrait à agir de la même façon que l'homme quotidien, qui infère trop souvent des caractéristiques psychologiques sur la base de l'observation de comportements).

Capacité des enseignants à distinguer entre différents processus et habiletés spécifiques

On peut se demander dans quelle mesure le jugement des enseignants discrimine bien des habiletés spécifiques. Plusieurs auteurs ont relevé une tendance de la part des individus à évaluer de manière similaire des dimensions qu'ils perçoivent comme conceptuellement ou logiquement reliées. Cette tendance est souvent considérée comme un type spécifique d'effet de halo (Balzer & Sulsky, 1992 ; Cooper, 1981). Un tel effet se produit-il dans le jugement des enseignants ? Forment-ils un jugement d'habileté qui est effectivement distinct du jugement produit pour une autre habileté, ou bien l'un diffuse-t-il son influence sur l'autre ? Ces questions s'entendent à plusieurs niveaux :

- au niveau disciplinaire. Des jugements spécifiques à une discipline pourraient être sous l'influence des performances réalisées dans d'autres disciplines académiques. Une recherche de Dompnier, Pansu et Bressoux (2006) a montré que le jugement des enseignants sur le niveau de leurs élèves en français était dépendant de leurs performances en mathématiques. Plus le niveau de performances en mathématiques (mesuré par des épreuves standardisées) était élevé, meilleur était le jugement en français, même après avoir contrôlé le niveau des élèves en français (lui aussi mesuré par des épreuves standardisées). Pour le jugement en mathématiques, la même chose s'observait : le jugement de l'enseignant était influencé par les

performances des élèves en français. Cette recherche a permis de porter au jour un évident effet de halo.

- entre habiletés proches constitutives d'une habileté disciplinaire plus générale (e.g. calcul en mathématiques et résolution de problèmes mathématiques). Peu d'études ont abordé cette question. Certains auteurs ont considéré qu'un jugement sera d'autant plus exact qu'il est plus spécifique car cela réduirait le nombre d'influences potentielles que l'évaluateur doit considérer pour former son jugement (e.g. Swann, 1984). Dans cette optique, plus les habiletés à juger seraient spécifiques, plus la capacité des sujets à les distinguer devrait être forte. Cependant, cette interprétation a été largement mise en question, à juste titre pensons-nous. En effet, les influences potentielles relèvent sans doute moins des déterminants du monde réel que des hypothèses que forme la personne qui doit porter un jugement. Ainsi, une compétence très générale peut être l'objet de peu d'hypothèses alors qu'une habileté très spécifique pourrait être l'objet de nombreuses hypothèses, en fonction des « théories » accessibles et disponibles que l'évaluateur a pu se former sur l'objet à juger (Kruglanski, 1990).

- entre habiletés qui renvoient à des processus liés mais relevant d'une place différente dans la chaîne causale des acquisitions (résultats avérés et processus inférentiels supposés favoriser lesdits résultats : performances à une épreuve et niveau de motivation par exemple). Plusieurs études montrent que les enseignants éprouvent quelques difficultés à discriminer finement de tels processus. La question de savoir dans quelle mesure cette difficulté est effectivement liée à des difficultés de discrimination ou à une influence réelle des processus entre eux demeure assez largement posée (cf. Meisels *et al.*, 1996). Hoge et Butcher (1984) ont relevé, auprès d'un échantillon de 12 enseignants, les jugements sur leurs élèves dans trois domaines : motivation, acquisitions en lecture et niveau d'habileté générale. Les auteurs montrent des intercorrélations fortes ($r = 0,60$ entre motivation et acquisitions en lecture ; $r = 0,58$ entre motivation et niveau d'habileté générale), à très fortes ($r = 0,87$ entre acquisitions en lecture et niveau d'habileté générale). L'argument des auteurs disant que ces intercorrélations ne seraient pas dues à un effet de halo n'est guère convaincant puisqu'il repose seulement sur le constat que deux de ces corrélations sont significativement différentes entre elles. Or, cela n'efface en rien le fait que toutes les valeurs de ces corrélations sont élevées, ce qui pourrait bien être le produit d'un effet de halo.

L'effet de halo que mettent en évidence ces travaux est sans doute lié aux théories implicites de la personnalité des enseignants. On peut supposer que l'effet de halo est d'autant plus fort que les enseignants développent des conceptions de l'intelligence qui se rapprochent d'une théorie de l'entité. Une telle théorie implicite qui, comme nous l'avons vu *supra*, renvoie au fait de concevoir l'intelligence comme un trait immuable et général, entraîne que les performances réalisées dans un domaine sont vues comme révélatrices d'une capacité générale qui doit « régir » un ensemble de performances dans différents domaines.

Jugement sur performances « actuelles » et jugement sur réussite scolaire ultérieure

Plusieurs recherches ont établi que, selon que les enseignants ont à juger un niveau de performance actuel, ou bien un niveau de performance éloigné dans le temps, ils n'utilisent

pas tout à fait les mêmes informations et ne leur accordent pas le même poids (e.g. Hallinger & Murphy, 1986 ; Teddlie *et al.*, 1989). On constate notamment que plus on demande aux enseignants de prédire une réussite lointaine (entrée au lycée, entrée à l'université, etc.), plus ils fondent leur jugement sur des critères sociaux.

Dans une étude où l'on demandait aux enseignants d'évaluer les performances de leur classe (au niveau global), le niveau « objectif » des élèves (score moyen par classe à des épreuves standardisées de lecture) expliquait 20,33 % de la variance du jugement (Bressoux, 1993). Mais, lorsqu'il s'agissait d'émettre un pronostic sur l'entrée en seconde (émettre un pronostic sur la proportion d'élèves de la classe qui entreraient en seconde), le pourcentage de variance expliquée par le niveau « objectif » des élèves passait à 16,79 %. Lorsqu'il s'agissait d'émettre un pronostic encore plus lointain, sur l'entrée dans l'enseignement supérieur, le pourcentage de variance expliquée par le niveau « objectif » des élèves chutait à 10,28 %. Parallèlement, l'effet de la tonalité sociale de la classe augmentait : non significatif pour le niveau « actuel » de performances, l'effet était de 5,11 % sur le pronostic d'entrée en seconde et de 5,44% sur le pronostic d'entrée dans l'enseignement supérieur. Dans ce dernier cas, s'ajoutait un effet, modeste mais significatif, du pourcentage d'élèves étrangers dans la classe (contribution de 1,73 % à l'explication du pronostic d'entrée dans l'enseignement supérieur).

Tout se passe donc comme si le poids des facteurs sociaux augmentait au fur et à mesure que le jugement à émettre porte sur une performance plus éloignée dans le temps. Dans la situation de relative incertitude où se trouve la personne qui doit émettre le jugement, celle-ci se met alors, de manière très largement implicite bien sûr, en quête d'autres informations susceptibles d'assurer la qualité du jugement. En particulier, on voit apparaître des indices qui caractérisent la personne en tant que telle, son origine sociale, sa nationalité, qui sont des indicateurs stables, sortes de « marqueurs » personnologiques qui ne sont pas susceptibles de varier dans le temps. Cependant, lorsqu'il y a augmentation de la durée dans le temps de l'estimation à fournir, on voit qu'il ne s'agit pas seulement de surajouter des informations sociales aux performances « actuelles » dans le processus de formation du jugement. Les informations sociales tendent à se *substituer* aux performances. On relève en effet que le poids des performances tend à baisser dans le même temps que celui des caractéristiques sociales augmente. Tout se passe donc comme si les performances actuelles, sans doute jugées moins prédictives de la réussite scolaire dans le long terme, étaient alors partiellement remplacées par des facteurs stables qui caractérisent en propre les individus, qui constituent des « marques » indélébiles, origine sociale, nationalité qui, elles, sont supposées agir de manière permanente.

Il n'entre pas dans notre propos de dire si les enseignants ont tort ou raison d'agir de la sorte. Il faut constater que, à l'évidence, des « théories » implicites sont ici à l'œuvre dans la formation du jugement de réussite à long terme, qui ne se fondent pas tout à fait sur les mêmes éléments, et ne leur accordent pas tout à fait le même poids que pour un jugement de performance « actuelle ». Même si, à l'évidence, certains stéréotypes sont à l'œuvre, on ne peut évacuer aussi que des connaissances (fondées) soient aussi en jeu ici. Toutes les enquêtes sociologiques montrent en effet que, à chaque palier d'orientation, à réussite scolaire identique, les élèves les plus favorisés choisissent les filières les plus longues et les plus prestigieuses.

L'influence des normes sociales générales et des normes professionnelles

Plusieurs travaux ont montré que le jugement des enseignants est sensible à des normes sociales très générales. Il en va ainsi de la politesse par exemple ou de certains comportements sociaux particulièrement valorisés (altruisme...) (Filisetti, 2005). Il en va de même de la norme d'internalité, présentée *supra* (Beauvois & Dubois, 1988). Plusieurs études ont en effet montré que les élèves qui adoptent un comportement normatif, au sens de la norme d'internalité, c'est-à-dire qui expliquent ce qu'ils font ou ce qui leur arrive par une causalité essentiellement interne (i.e. qui s'attribuent les causes de leurs réussites et échecs, de leurs bonnes et mauvaises actions) sont mieux jugés que les élèves qui expliquent ce qu'ils font ou ce qui leur arrive par une causalité externe (i.e. qui attribuent à autrui, à la chance, au hasard, à la situation, etc. les causes de leurs réussites et échecs, de leurs bonnes et mauvaises actions) (Dompnier, Bressoux & Pansu, 2006 ; Dubois & Le Poulitier, 1991 ; Bressoux & Pansu, 2003).

Gilly (1992) a montré que le jugement des enseignants dépend certes de normes sociales générales (valeur morales, rôle de l'homme dans la société, etc.), mais aussi de normes scolaires institutionnelles (règles de fonctionnement de l'École, objectifs de l'éducation, etc.). Les représentations des enseignants se situeraient à l'articulation entre ces normes générales et professionnelles et expliqueraient pour une bonne part le syncrétisme de ce jugement, qui combine parfois dans la description d'un élève ses qualités scolaires, morales, voire physiques (beauté).

L'influence des normes professionnelles ne se décline pas seulement au niveau « lointain » de la profession (en particulier de la formation professionnelle) ou de l'institution scolaire globale telle qu'elle s'impose à tous les enseignants. Elle se fait aussi plus proximale dans la définition par l'établissement scolaire lui-même d'un ensemble de règles et normes. Les travaux de Brookover *et al* (1979) ont montré que l'établissement scolaire doit être considéré comme une organisation sociale qui développe sa propre culture : normes, attentes, valeurs. Les auteurs ont montré que le niveau d'attentes des enseignants concernant la réussite des élèves variait d'une école à l'autre dans des proportions que le niveau scolaire des élèves ne peut pas entièrement expliquer. Il semble bien que des écoles « développent » un système d'attentes élevées tandis que d'autres « développent » un système d'attentes peu élevées (les premières seraient d'ailleurs plus efficaces que les secondes).

Ainsi peut-on montrer l'existence d'un effet-établissement sur la notation et sur l'orientation, fruit de normes souvent implicites et d'une politique de sévérité/indulgence par laquelle l'établissement redéfinit localement les grandes missions nationales d'éducation. Un établissement ne se réduit pas aux contraintes qui pèsent sur lui. Il dispose toujours d'une certaine marge de manœuvre par laquelle il redéfinit ses missions et les modes de relations entre ses membres.

Les effets des attentes : les prophéties autoréalisatrices

Il serait sans doute plus aisé de décrire les jugements en termes de biais, de décalage par rapport à la réalité, s'ils n'avaient aussi cette capacité étonnante à influencer la réalité et, par là, à s'autoréaliser. Parvenir à un jugement est ainsi facilité lorsque l'enseignant influence systématiquement l'élève en le conformant à son jugement.

Lorsque deux personnes entrent en contact pour la première fois, elles ne sont pas pour autant totalement inconnues l'une à l'autre. Cette interaction va en effet s'opérer dans un contexte social qui va activer une série d'indications pour chacun des partenaires. Par exemple, à l'école, l'enseignant sait qu'il aura en face de lui des élèves qui sont là pour apprendre (mais qui risquent de se désengager s'il ne les « tient » pas comme il faut, etc.), qui ont un certain niveau d'acquisitions, etc. Chaque partenaire aura aussi d'emblée des informations sur autrui au simple vu de la personne (âge, habillement, port de tête, tenue du corps, expression du visage, beauté, etc.) ce qui fait que, d'emblée, des inférences vont être formées sur l'identité du partenaire. La « négociation d'identité » va être un enjeu crucial de la relation qui s'établit alors (qui suis-je pour l'autre, « pour qui me prend-il ? ») et qui, dans le cas de l'école, va durer longtemps (en général une année scolaire au moins). Parmi la multitude de comportements de transaction qui vont s'opérer, deux ont pu être identifiés comme apparaissant très tôt, dès le début de l'interaction. L'un consiste à se comporter avec notre partenaire d'interaction comme si notre jugement initial était correct. L'autre consiste à rechercher très rapidement des informations diagnostiques sur la personne compte tenu de l'environnement social dans lequel va se dérouler l'interaction (est-il bon ou mauvais élève, se tient-il bien ou mal, etc.). Ces deux comportements, distincts, sont cependant corrélés car nous avons tendance à penser que les informations qui confirment notre jugement initial sont plus diagnostiques que les autres (Swann, 1984). Même si, nous l'avons dit, une multitude de comportements vont petit à petit se mettre en place, si d'autres informations vont émerger des diverses interactions, remettant en jeu la négociation d'identité, il y a là deux processus fondamentaux pour comprendre les prophéties autoréalisatrices. On peut voir aussi que ces processus ont partie liée avec l'heuristique d'ancrage et d'ajustement présentée par Tversky et Kahneman (1974) : la difficulté à ajuster totalement un jugement à une réalité sur la base de nouvelles informations quand une estimation initiale, éventuellement éloignée de la réalité, a été produite.

Du côté de la cible du jugement, deux types de comportements (au moins) peuvent apparaître. La vérification de soi et la confirmation du jugement d'autrui. La vérification de soi conduit à se comporter avec autrui de manière à confirmer notre concept de soi. Il s'agit donc d'amener autrui, parfois dans une véritable lutte, à nous juger tel que nous nous voyons, à obtenir d'autrui qu'il nous accorde l'identité que nous revendiquons. Sous l'hypothèse que le concept de soi est stable et transcontextuel, le jugement d'autrui risque d'avoir une validité assez générale quand il s'accorde au processus de vérification de soi. La confirmation conduit quant à elle à se comporter avec autrui de manière à se conformer à son jugement. Dans ce cas, la validité du jugement risque de n'être que contextuelle car il y a toutes chances que le comportement de la cible du jugement change dès lors que la source du jugement n'est plus la même. Le second comportement est toutefois plus à même de générer des effets de prophéties autoréalisatrices et ce d'autant plus qu'il y aurait une certaine congruence entre les diverses sources de jugement, qui pourront éventuellement conduire à une internalisation des étiquettes accordées à la cible (Swann, 1984).

Nulla étude en éducation n'est aussi célèbre que celle qui a été réalisée par Rosenthal et Jacobson (1968) sur ce qu'ils ont nommé l'effet Pygmalion. Point n'est sans doute besoin de s'attarder trop longtemps sur cette étude si ce n'est pour bien préciser et remettre dans leur contexte les résultats obtenus. Il faut rappeler que cette étude induisait des attentes erronées chez les enseignants puisqu'on leur donnait de fausses informations sur certains de leurs élèves. Il était dit aux enseignants, sur la base d'un pseudo-test dit « d'inflexion des courbes d'acquisition », que certains de leurs élèves progresseraient beaucoup en cours d'année. Or, lesdits élèves avaient en fait été choisis aléatoirement. A la fin de l'année, les élèves désignés comme allant beaucoup progresser avaient effectivement plus progressé que les autres. Rappelons cependant que le résultat n'était significatif que pour les élèves les plus jeunes (1^{er} et 2^e grades). De cette expérience de Rosenthal et Jacobson, et malgré de nombreuses polémiques, beaucoup de chercheurs ont été tentés de voir un impact très puissant des jugements des enseignants. D'où la conception d'un constructivisme social fort : l'idée que la réalité sociale serait dans une large mesure construite par la perception des individus. Ainsi, le fait de juger un élève « bon », même s'il ne l'est pas, conduirait un enseignant à se comporter de telle manière avec lui qu'il en viendrait petit à petit à se conformer au jugement de l'enseignant et qu'il deviendrait effectivement « bon ».

Dans une telle conception, on peut se demander si la question même qui est posée par ce rapport (le jugement professoral paraît-il apte à motiver une décision d'orientation ?) se poserait réellement, en tout cas avec une telle acuité. Si les enfants ne font que se conformer au jugement de l'enseignant, alors celui-ci est nécessairement adéquat, tout au moins en fin d'année⁷. Or, les travaux ultérieurs ont largement mis à mal cette conception d'un socioconstructivisme fort.

Une série de travaux sont apparus dans le sillage de cette étude, et aussi une série de critiques. L'une des plus importantes porte sur le fait que, dans la réalité, les enseignants disposent d'informations fiables sur les élèves et que leurs jugements ne sont pas, en situation naturelle, aussi biaisés que pouvaient l'être ceux des enseignants de l'expérience de Rosenthal et Jacobson. C'est dans cette optique que se situent les travaux de Lee Jussim et de ses collègues, qui conduisent actuellement à un réexamen des effets des prophéties autoréalisatrices (Jussim, 1989 ; Jussim & Eccles, 1992 ; Madon, Jussim & Eccles, 1997 ; Madon, Smith, Jussim, Russell, Eccles, Palumbo & Walkiewicz, 2001 ; Smith, Jussim & Eccles, 1999 ; Smith, Jussim, Eccles, VanNoy, Madon & Palumbo, 1998). Il s'agit ainsi, dans le cadre de ces recherches, non pas d'induire de faux jugements chez les enseignants, mais de se contenter d'enregistrer ceux qu'ils se forgent eux-mêmes naturellement et d'essayer de mesurer la part du jugement qui influence réellement les performances des élèves. Jussim a élaboré pour ce faire un modèle de « reflet-construction » grâce auquel il tente de faire la part de ce qui, dans la perception des enseignants, participe à la construction de la valeur des élèves (et donc de la réalité) de ce qui n'est que le reflet de leur valeur réelle (et qui ne l'influence donc pas).

Les résultats de Jussim et de ses collègues révèlent un effet Pygmalion bel et bien significatif mais faible. Globalement, 80 % de la corrélation entre le jugement des enseignants (en début d'année) et les performances des élèves (en fin d'année) seraient dus au fait que le jugement de l'enseignant n'est que le reflet de la valeur réelle des élèves (l'exactitude du jugement)

⁷ Ce qui ne veut pas dire que le caractère initialement erroné du jugement soit, lui, hors de question. Rappelons qu'il ne peut y avoir de prophétie autoréalisatrice que si le jugement est initialement erroné. Sinon, il s'agit d'un simple reflet de la réalité, non d'une construction de celle-ci.

tandis que seulement 20 % de cette corrélation pourraient être dus à un effet de prophétie autoréalisatrice. En termes de coefficient standardisé, la taille de l'effet des prophéties autoréalisatrices sur les acquis scolaires des élèves oscillerait entre 0,10 et 0,30. Jussim et ses collègues aboutissent donc à la conclusion d'un constructivisme social faible concernant la valeur scolaire des élèves. Il ne suffit pas d'avoir foi en un élève pour que cette foi crée des miracles comme on a pu le croire naguère dans l'enthousiasme suscité par les premières études sur l'effet Pygmalion.

Les recherches récentes ne posent plus la question de savoir s'il existe des effets de prophéties autoréalisatrices, ce qui est maintenant admis, mais d'étudier dans quelles conditions et avec quels élèves ces effets pourraient être atténués ou au contraire accentués. On en vient donc à étudier les modérateurs et les médiateurs des effets des prophéties autoréalisatrices. Il sort du cadre de ce rapport de faire un compte-rendu précis de cette question. Signalons simplement quelques résultats. Certaines pratiques pédagogiques, telles que les groupes de niveau, accentuent les effets des prophéties autoréalisatrices (Smith, Jussim, Eccles, VanNoy, Madon et Palumbo, 1998). Une étude de Brattesani, Weinstein et Marshall (1984) a montré que les effets des attentes des enseignants sur le sentiment de compétence des élèves sont plus forts dans les classes où ces derniers perçoivent un fort traitement différencié entre bons et mauvais élèves que dans celles où le traitement est perçu comme faiblement différencié. Les effets de prophéties autoréalisatrices sont plus forts dans les climats de classe contrôlants, plutôt que favorisant l'autonomie (Trouilloud, Sarrazin, Bressoux & Bois, 2006). Certains travaux avancent que les jugements des enseignants influencent davantage la réussite des élèves faibles que celle des élèves forts (Madon, Jussim & Eccles, 1997). De même, certains auteurs pensent que les élèves les plus « vulnérables » (groupes stigmatisés) sont plus enclins que les autres à se conformer aux jugements des enseignants (Good & Brophy, 2000). Certains ont émis l'idée que des jugements sous-estimés affecteraient davantage les élèves que des jugements surestimés (*e.g.* Brophy & Good, 1974 ; Eccles & Wigfield, 1985). Toutefois, une étude récente de Madon, Jussim et Eccles (1997) montre au contraire que les prophéties autoréalisatrices seraient plus susceptibles de favoriser que de pénaliser. Une raison pourrait être que les attentes négatives ne sont pas toujours néfastes, notamment lorsqu'elles conduisent un enseignant à œuvrer pour un élève en difficulté. Ainsi, Madon, Jussim et Eccles (1997) ont montré un effet des prophéties autoréalisatrices particulièrement fort ($\beta = 0,41$) quand les enseignants surestimaient la valeur des élèves faibles.

Les travaux sur les prophéties autoréalisatrices nous sont précieux aussi dans le cadre cette étude parce qu'ils ont posé la question des effets dans le long terme... pour peu qu'on les lise un peu différemment. On peut en effet examiner les résultats de ces travaux pour y étudier la capacité des enseignants à prédire la réussite dans le long terme (et pas seulement, ou pas nécessairement, à l'influencer). La plupart des études ont conclu à une dissipation des effets de prophéties autoréalisatrices dans le temps, mais cette dissipation peut être totale ou seulement partielle.

Dans la célèbre étude de Rosenthal et Jacobson (1968), les trois re-tests ont été administrés aux élèves quatre mois, huit mois et vingt mois après le démarrage du programme expérimental. Globalement, les résultats montrent que les effets des prophéties autoréalisatrices sont faibles après quatre mois, plus forts après huit mois mais que, après vingt mois, ils sont plus faibles tout en restant à un niveau légèrement plus élevé qu'après quatre mois. Il est important de signaler que, pour les élèves des petites classes, c'est-à-dire ceux pour lesquels les effets des prophéties autoréalisatrices avaient été les plus forts à la fin de l'année (après huit mois), les effets de prophéties autoréalisatrices disparaissent quasiment

après vingt mois. Ces résultats indiquent que les effets des prophéties autoréalisatrices tendent à se dissiper dans le long terme.

Dans une étude longitudinale dans laquelle ils ont suivi 3000 élèves au cours de leurs trois années de lycée, West et Anderson (1976) rapportent que les effets des attentes des enseignants de la 1^{ère} année tendent à se dissiper dans le temps puisqu'ils prédisent la réussite en 2^e année plus fortement que la réussite en 3^e année. Les coefficients de piste causale sont respectivement de 0,12 et de 0,06 ce qui indique bien, d'une part, que l'effet d'une année sur l'autre est faible et que, d'autre part, l'effet tend à se dissiper quand la durée s'allonge.

Toutefois, dans une étude réalisée sur une très longue période allant du 6^e grade au 12^e grade, Smith, Jussim et Eccles (1999) obtiennent des résultats qui ne confirment que partiellement cette conclusion. Après avoir contrôlé le niveau initial des élèves, ils ont montré que les perceptions des enseignants de mathématiques concernant la performance, le talent et l'effort de leurs élèves, enregistrés au cours du 6^e grade et du 7^e grade, exercent des effets de prophéties autoréalisatrices qui sont encore observables sur les scores à des épreuves standardisées de mathématiques au terme du 12^e grade. Bien que déclinant en intensité, les effets des prophéties autoréalisatrices ne disparaissent pas complètement ; plus précisément, après une lente et régulière baisse de l'intensité des effets, ceux-ci demeurent stables à partir du 10^e grade. Les auteurs en concluent que, bien qu'ils tendent à se dissiper, les effets des prophéties autoréalisatrices sont susceptibles de s'exercer, avec une faible intensité, sur le long terme.

Alvidrez et Weinstein (1999) ont montré que le jugement des enseignants de préélémentaire (*preschool teachers*), c'est-à-dire avant même le *kindergarten*, concernant l'habileté cognitive d'élèves de 4 ans prédisait les scores obtenus à des épreuves quatorze années plus tard (*grade point average, GPA, et Scholastic Aptitude Test, SAT*) après avoir contrôlé les scores de QI et l'origine socio-économique des élèves.

Au-delà des résultats enregistrés, c'est l'interprétation qu'en donnent les auteurs qui nous semble digne d'intérêt dans le cadre de ce rapport. Alvidrez et Weinstein (1999) insistent sur le fait que l'interprétation de ces résultats est délicate. Bien qu'elles contrôlent le niveau initial des élèves, des études conduites en milieu naturel telles que la leur (ou celle de Smith, Jussim & Eccles, 1999), ne permettent pas de déterminer si ces relations à long terme révèlent une réelle influence exercée par le jugement de l'enseignant sur la réussite ultérieure des élèves ou bien si elles révèlent un jugement prescient, c'est-à-dire une simple capacité à prévoir de manière pertinente ce qui adviendra.

Si une telle hypothèse se révèle exacte, cela est d'une grande conséquence pour l'objet de ce rapport. Cela signifierait en effet que les enseignants auraient repéré (très tôt dans l'étude d'Alvidrez et Weinstein, au début du secondaire pour l'étude de Smith, Jussim et Eccles) des éléments prédictifs des performances scolaires futures *non détectés par les épreuves initiales* et, par conséquent, non contenus en tant que tels dans les performances réalisées par les élèves au moment où le jugement des enseignants était émis. Dans le fil de cette hypothèse, cela conduit à avancer que, dans le long terme, le jugement des enseignants pourrait être plus valide que la réussite constatée à des épreuves standardisées. Il faut cependant accueillir cette hypothèse avec une grande prudence dans la mesure où d'autres hypothèses concurrentes ont été formulées, et qu'il est difficile de les départager sur le plan empirique.

Signalons une autre hypothèse, que nous ajoutons, qui est que les jugements des enseignants pourraient être relayés par les enseignants ultérieurs. Les enseignants ont sans doute tendance à être sensibles aux mêmes comportements des élèves et à réaliser des inférences similaires à partir de ces comportements (e.g. Stevenson *et al.*, 1976). Parce qu'ils ont reçu une socialisation générale proche du fait de leur appartenance à une même société (mêmes normes générales, mêmes stéréotypes généraux), parce qu'ils ont reçu une socialisation secondaire commune (professionnalisation), parce qu'ils sont exposés aux mêmes contraintes institutionnelles (hiérarchie, collègues, parents...), les jugements des enseignants sont, en partie, fondés sur les mêmes informations. Notamment, la capacité des élèves à jouer le jeu de l'excellence scolaire (élève docile – au sens étymologique, c'est-à-dire qui se laisse instruire –, attentif, ayant un comportement et des connaissances issus d'un fort capital culturel, respectant les normes générales telles que la politesse, etc.) n'est pas sans effet sur le jugement de l'enseignant. Ainsi, il est tout à fait plausible d'imaginer qu'un même élève va avoir tendance, au long de sa scolarité, à être jugé, au-delà de ses strictes performances, de manière globalement similaire d'un enseignant à l'autre. Bien sûr, cela ne vaut qu'en tendance et en moyenne. Les jugements, éventuellement jouant comme prophéties autoréalisatrices d'un enseignant donné, seraient ainsi relayés dans les années ultérieures par d'autres enseignants, en partie sensibles aux mêmes informations et comportements.

Les études sur la notation

Bien que l'étude docimologique de la notation ne soit pas au cœur de cette synthèse de questions, on ne peut totalement la passer sous silence dans la mesure où la note constitue un jugement objectivé. C'est le jugement des enseignants tel qu'il se donne à voir aux élèves et à leur entourage, notamment familial, avec toute la puissance que l'institution scolaire lui confère. Cette visibilité externe impose certaines contraintes et fonctions supplémentaires par rapport au jugement que l'on peut garder par devers soi.

Une caractéristique importante qui a été mise en évidence, il y a bien longtemps, est la tendance des enseignants à normaliser et centrer la distribution de leurs notes, ce qui se traduit par une distribution en forme de cloche (i.e. distribution gaussienne), avec une moyenne qui est proche de la valeur centrale de la distribution des notes (le fameux 10/20). Cette tendance, nommée effet Posthumus, a pour conséquence de produire de la différenciation à l'intérieur des classes, parce qu'on recrée ainsi tout un continuum de différences de niveaux d'habileté entre élèves, allant des plus forts aux plus faibles, et d'atténuer les différences entre les classes, parce que les écarts de notation entre classes sont moins grands que les écarts d'acquisitions réelles (Crahay, 1996).

Plusieurs mécanismes sont à l'œuvre dans ce phénomène. L'un a trait à la nécessité pour l'enseignant d'évaluer précisément les niveaux d'habileté de ses élèves. Si la distribution des notes est trop asymétrique (révélant des effets-plafonds ou des effets-planchers), l'enseignant ne pourra plus discriminer finement les élèves. En cas d'effets-plafonds prononcés, il ne pourra pas distinguer entre les élèves forts et très forts. En cas d'effets-planchers prononcés, il ne pourra pas distinguer entre les élèves faibles et très faibles. Par ailleurs, la note obéit à des fonctions sociales et l'une de ses caractéristiques les plus évidentes est qu'elle se donne à voir. Il ne semble guère « acceptable » socialement que les notes soient très élevées : n'est-ce pas trop facile ? L'enseignant n'est-il pas laxiste ? Inversement, il n'est guère « acceptable »

socialement non plus que les notes soient très basses : n'est-ce pas trop difficile ? L'enseignant n'est-il pas incapable de se faire comprendre ? N'est-il pas injustement sévère ? La « paix sociale » dans la classe risque de n'être pas obtenue dans le cas d'une sévérité perçue comme trop grande. A l'inverse, si les notes sont très indulgentes, comment obtenir des élèves qu'ils travaillent beaucoup ? Il faut donc négocier des notes ni trop élevées, ni trop basses. Nous aurons l'occasion de revenir plus loin sur ces fonctions de négociation de la note en classe. Toutes ces contraintes contribuent à créer l'effet Posthumus et à façonner le niveau de performances « visible » des élèves. Quand deux enfants d'établissements différents s'exposent leur moyenne trimestrielle pour savoir lequel des deux est le plus fort, lequel perçoit le côté fabriqué et contextualisé de sa note ? Le plus souvent, aucun des deux.

Il y a de bonnes raisons de penser que ces contraintes qui pèsent sur la notation ne sont pas toutes perçues, en tout cas pas toutes dans leur pleine mesure, pour pouvoir être prises en compte par les enseignants et corriger en conséquence un jugement libéré de ces contraintes propres à la notation. De sorte qu'on doit avancer que, *si le jugement fait la note, la note fait aussi le jugement*. En effet, nous avons constaté des effets similaires à l'effet Posthumus sur le jugement des enseignants même lorsque celui-ci est émis de manière informelle et anonyme (Bressoux & Pansu, 2003)⁸. En particulier, on constate un effet de contexte qui fait que les écarts inter-élèves sont maximisés au sein de la classe et minimisés entre les classes. Cela se traduit par une plus grande sévérité de jugement dans les classes fortes et une plus grande indulgence dans les classes faibles.

Les travaux de docimologie qui ont été rendus célèbres par Piéron (1969), en particulier l'enquête sur le baccalauréat de la Commission Carnegie, ont révélé les très forts écarts de notation de copies d'un correcteur à l'autre : l'écart moyen entre deux correcteurs était de 2,05 points sur une même copie de mathématiques, avec un écart maximum de 9 points, l'écart moyen entre deux correcteurs était de 3,29 points sur une même copie de composition française, avec un écart maximum de 13 points. Pour obtenir la « note vraie », entendue comme la moyenne d'un nombre assez grand de notations indépendantes (obtenir un coefficient de fidélité de 0,99), il aurait fallu 13 correcteurs pour une copie de mathématiques, 16 correcteurs pour une copie de physique, 78 correcteurs pour une composition française et 127 correcteurs pour une dissertation philosophique ! Et l'incohérence est encore plus grande dans les examens oraux. Ces écarts considérables ont conduit Piéron (1969) à lancer sa fameuse formule : « Pour prédire la note d'un candidat, il vaut mieux connaître son examinateur que lui-même ! » (p. 24). Cela a conduit Piéron à alerter sur les limites des examens et à préconiser « comme politique éducative générale, de substituer dans toute la mesure du possible aux sélections par examens une orientation continue fondée sur une connaissance approfondie de chaque personnalité avec des méthodes objectives » (p. 176). Notons qu'il s'agit toutefois dans ces travaux de traiter de jugements ponctuels sur des productions particulières en mettant au jour les problèmes « techniques » que pose la notation pour parvenir à un accord inter-correcteurs mais sans considérer les fonctions sociales de la notation. Les fluctuations autour de la « note vraie » sont censées se compenser parce qu'elles sont supposées provenir du hasard (i.e. suivre un processus stochastique), ignorant en cela les déterminants sociaux de la notation.

⁸ Ainsi, concernant le jugement des enseignants, la part de variance inter-classes s'élevait à 8,3 % de la variance totale en français et à 11,6 % en mathématiques. Concernant les scores des élèves aux épreuves standardisées, les parts de variance inter-classes s'élevaient à 29,9 % en français et à 21,8 % en mathématiques. La réduction des différences inter-classes produite par le jugement est ici évidente.

Les travaux sur la notation menés par Noizet et ses collègues ont porté au jour plusieurs phénomènes qui influencent la notation des copies (ordre de passage des copies...) (Noizet & Caverni, 1978). Si certains de ces facteurs peuvent être vus comme aléatoires (un élève donné verra, au long de l'année, sa copie passer tantôt après une bonne copie, tantôt après une faible et bénéficiera ainsi, en moyenne, tantôt d'un effet négatif tantôt d'un effet positif), d'autres nous intéressent bien davantage dans le cadre de cette revue de questions. Il en est ainsi en particulier de l'influence de la connaissance de la personne qui a réalisé la copie. Noizet et ses collègues ont mis en évidence ce qu'ils nomment un « effet de source ». Cet effet traduit un « effet d'assimilation de l'évaluation actuelle aux évaluations attribuées antérieurement au même élève » (Noizet & Caverni, 1983, p. 11). Caverni, Fabre et Noizet (1975) ont par exemple montré que le niveau scolaire de l'élève influait sur la note qui lui était attribuée. A copie identique, les élèves qui étaient supposés avoir réalisé antérieurement une bonne performance obtenaient une meilleure note. Une étude de Bonniol, Caverni et Noizet (1972) a également montré que les notes attribuées à des copies d'élèves de 6^e type 1 (censées accueillir les élèves les plus forts) obtenaient de meilleures notes que des copies, pourtant équivalentes, d'élèves de 6^e type 3 (censées accueillir les élèves les plus faibles).

Une étude de Felouzis (1997) a tenté d'analyser la notation des enseignants au niveau du lycée. L'un des buts de l'étude était de rendre compte des différences de sévérité entre enseignants. Pour ce faire, l'auteur a cherché à rapporter ces différences de sévérité à des différences de conceptions du métier que se forgeraient les enseignants. Ces conceptions seraient en lien avec l'évolution du système éducatif telle qu'ils la perçoivent. Felouzis a conduit son étude auprès d'enseignants de seconde (36 enseignants de mathématiques et 25 enseignants de français). L'auteur montre que la notation varie d'une classe à l'autre ; même pour des élèves comparables, certains enseignants sont globalement plus sévères que d'autres. Par des entretiens réalisés auprès des enseignants, Felouzis montre que le degré de sévérité est lié à une conception générale de l'évolution du métier d'enseignant et du rapport aux élèves. L'auteur met au jour deux grandes catégories d'enseignants. Certains restent attachés à un état antérieur du système éducatif, perçu comme une sorte d'âge d'or de ce système et de la profession d'enseignant. Ils sont porteurs (et défenseurs dans bien des cas) de ce que l'auteur appelle un « ritualisme académique ». Ces enseignants adoptent un discours très centré sur leur discipline et ils se révèlent plutôt nostalgiques d'un temps ancien où les lycées n'accueillaient qu'une frange favorisée de la population. D'autres enseignants sont porteurs de ce que l'auteur appelle un « pragmatisme pédagogique ». Le discours qu'ils adoptent se centre bien davantage sur leurs élèves que sur la discipline qu'ils enseignent, et il est beaucoup moins nostalgique. En somme, ils seraient plus proches des lycéens « tels qu'ils sont aujourd'hui » et refuseraient de s'arc-bouter sur un passé glorieux concernant le niveau des élèves. Les enseignants qui sont porteurs d'un « ritualisme académique » sont les plus sévères dans leur notation. Ils déclarent conserver à la note son rôle strict de mesure (affirmant une objectivité pourtant largement mise à mal par les analyses). Au contraire, les enseignants qui sont porteurs d'un « pragmatisme pédagogique » sont plus indulgents. Ils relativisent la note et s'en servent comme d'un outil pédagogique (pour récompenser les efforts accomplis par exemple). Il y aurait donc, chez les pragmatiques, une « indulgence calculée », qui ne serait en rien l'expression d'un quelconque laxisme (ceux-ci se définissant d'ailleurs plutôt comme des enseignants exigeants), mais bien une véritable stratégie de la note, considérée comme moyen d'information pour l'élève, voire d'incitation à le pousser à mieux faire. Une vision passéiste et négative des élèves serait donc liée à une notation plus sévère. Ces résultats mettent en évidence le rapport personnel de l'enseignant à la note, qui est lui-même le produit d'une histoire singulière : le propre passé scolaire de l'enseignant, son origine sociale, son expérience professionnelle, etc.

Synthèse concernant les qualités métrologiques du jugement comme mesure : fidélité et validité

Toute mesure a pour vocation de fournir une information précise et objective sur une réalité. La qualité de la mesure se fonde sur deux critères essentiels qui sont sa validité et sa fidélité. La validité se définit comme la capacité d'un procédé de mesure à mesurer réellement ce qu'il est censé mesurer. La fidélité se définit comme la capacité d'un procédé de mesure à produire les mêmes résultats lors d'observations répétées. La fidélité apparaît donc comme condition nécessaire mais non suffisante de la validité.

Plusieurs études permettent d'étudier les qualités métrologiques du jugement des enseignants *en tant que mesure* (nous verrons ultérieurement tout ce que cette conception peut avoir de réducteur). L'analyse psychométrique distingue plusieurs types de validité.

La validité critérielle est liée à la valeur pronostique des mesures. Il s'agit de comparer la mesure à un critère externe. Dans la recherche sur le jugement scolaire, l'exemple canonique de l'investigation de ce type de validité consiste à comparer le jugement que les enseignants portent sur le niveau d'acquisitions de leurs élèves aux résultats obtenus à des épreuves standardisées censées mesurer la même chose. Comme nous l'avons vu *supra*, la validité critérielle est en général considérée comme bonne. Elle varie toutefois en fonction des enseignants, en fonction des disciplines et des types d'élèves.

La validité de contenu concerne la définition de l'univers d'items censés représenter le concept. La validité de contenu n'a quasiment jamais été étudiée en tant que telle par les chercheurs (Meisels *et al.*, 1996). Il ne nous est par conséquent pas possible d'apporter d'informations solides sur ce sujet.

La validité de construit consiste à apprécier l'adéquation d'une mesure à un construit théorique. On va ainsi prédire qu'il existe des différences entre des groupes (par exemple, une mesure de l'estime de soi devrait différencier les personnes dépressives des autres personnes), établir des prédictions concernant les effets du concept construit sur des comportements ou attitudes, ou encore prédire les interrelations entre plusieurs mesures proches mais distinctes (capacité à distinguer des phénomènes proches). En général, les enseignants distinguent bien les grandes catégories d'élèves (élèves forts, faibles, moyens, élèves « à risque » et autres élèves, etc.) mais il arrive qu'ils ne détectent pas certaines lacunes très spécifiques, en particulier quand les résultats scolaires de l'élève sont globalement satisfaisants. Par ailleurs, Meisels *et al.* (1996) relèvent un assez faible niveau de pouvoir discriminant du jugement des enseignants. Bien que ceux-ci soient capables de distinguer leurs élèves sur de multiples dimensions, le jugement des enseignants est entaché d'un effet de halo : un jugement dans un domaine irise le jugement dans un domaine proche. La capacité discriminante du jugement est en tout cas moins bonne que celle liée à l'usage de tests standardisés, qui fournissent des interrelations plus précises entre des scores de lecture, maths, sciences, etc., ou bien entre diverses habiletés d'une même discipline (e.g. vocabulaire, grammaire, orthographe). Ainsi, une réanalyse des données présentées dans Bressoux et Pansu (2003), faite spécialement pour les besoins de la présente étude, révèle que la corrélation entre les jugements des enseignants sur le niveau des élèves en français et en mathématiques ($r = 0,81$) est plus élevée que la

corrélation entre leurs scores obtenus à des épreuves standardisées pour ces mêmes disciplines ($r = 0,72$). Cet effet de halo est sans doute à rapprocher des théories implicites de l'intelligence des enseignants. Plus ils adhèrent à une théorie de l'entité (l'intelligence comme trait général, stable...) plus il y a de chances que l'effet de halo soit fort. Hoge et Butcher (1984) montrent en effet que les enseignants ont tendance à surestimer la relation entre le niveau d'habileté générale des élèves (tel que mesuré par le score de QI obtenu aux échelles verbales d'un test d'intelligence) et leur niveau de réussite à des épreuves d'acquisitions particulières (en lecture). Les auteurs montrent que certains enseignants (mais pas tous) surestiment la performance des élèves ayant un niveau élevé d'habileté générale et sous-estiment la performance des élèves ayant un niveau faible d'habileté générale, cette surestimation ou cette sous-estimation se révélant assez fortes pour quelques enseignants.

Ce dernier point est sans doute à rapprocher de la fonction du jugement, qui ne sert pas à mesurer les élèves *in abstracto*, mais qui est aussi tourné vers l'action. Il a été montré que, pour piloter leur classe (présenter une leçon, choisir un exercice, fournir tel ou tel type d'aide, etc.), les enseignants se servent bien davantage de prédictions sur les résultats des élèves (Qui réussira ? Qui échouera ?) que de prédictions fines sur les procédures qui seraient mobilisées par les élèves pour parvenir à tel ou tel résultat (cf. Maurice, 1996). La classe se piloterait aux résultats plutôt qu'aux procédures et on peut supposer que cela est lié à la non-distinction de certaines lacunes spécifiques.

La fidélité du jugement scolaire a reçu beaucoup moins d'attention de la part des chercheurs que la question de sa validité. En général, la fidélité test-retest est élevée. On constate une bonne stabilité du jugement des enseignants dans le temps. Par exemple, les corrélations entre deux jugements émis à 7 mois d'écart peuvent atteindre des valeurs très élevées ($r = 0,88$). Cependant, dans le temps long, on sait finalement assez mal ce qui, de cette forte stabilité, revient à une perception adéquate de ce qui relève d'une fixation initiale (phénomène qu'on peut relier à l'heuristique d'ancrage et d'ajustement). De manière sans doute plus convaincante pour le propos qui est le nôtre dans cette étude, on relève une très bonne stabilité sur une durée plus courte de quelques semaines (les acquis des élèves étant supposés avoir peu varié dans un laps de temps aussi court).

On manque d'études vraiment probantes pour étudier la fidélité des jugements inter-enseignants⁹. Une étude rapportée par Meisels *et al.* (1996) indique une corrélation élevée entre les jugements des enseignants des *kindergarten* (jardins d'enfants) et les jugements d'autres enseignants réalisés deux ou trois ans plus tard (à l'école élémentaire). Il semble toutefois que tous les enseignants n'interprètent pas de la même manière les différentes caractéristiques des élèves (« chinois » par exemple peut évoquer des difficultés pour l'apprentissage de la langue ou bien une solidarité familiale favorable à la réussite scolaire). Bien souvent, le caractère plus ou moins prédictif de certaines caractéristiques des élèves varie selon la signification qu'ils prennent dans telle ou telle situation. Il apparaît toutefois qu'une homogénéité de l'expérience (ancienneté, formation reçue) favorise l'accord entre les enseignants. On obtient le même type de résultats en docimologie. Former les enseignants, s'accorder sur un barème, augmente la fidélité inter-correcteurs.

⁹ Rappelons que, si les travaux concernant la cohérence inter-enseignants en matière de jugement de la valeur scolaire des élèves sont peu nombreux, ceux concernant la notation sont nombreux, probants et montrent une assez forte incohérence inter-correcteurs. Nous avons signalé *supra* que les travaux de Piéron (1969) avaient montré que, sur une même copie, les notes des examinateurs pouvaient grandement varier. Les variations étant beaucoup plus grandes pour les copies littéraires que pour les copies scientifiques, beaucoup plus grandes aussi pour les examens oraux que pour les examens écrits.

Il n'existe pas à notre connaissance de travaux probants qui aient traité la question de la consistance interne du jugement, sans doute parce qu'il est souvent appréhendé comme une variable monolithique et non pas comme pas comme la synthèse d'un ensemble de micro-jugements qui seraient tous censés mesurer la même chose.

A ce stade de l'étude des qualités métrologiques du jugement comme mesure, il apparaît que le jugement des enseignants a une bonne validité générale. Il faut cependant sur ce point opérer une double distinction. « Validité générale » signifie ici « globalement valide pour la population des élèves ». Nous n'avons pas distingué une validité générale au sens de « valide en toute situation » d'une validité plus circonstanciée au sens de « valide dans une certaine situation ». Par ailleurs, il faudrait aussi s'interroger sur la validité individuelle du jugement, au sens de « valide pour *chaque* élève ».

Commençons par le second point. Ce n'est pas parce que le jugement des enseignants est globalement valide pour une population d'élèves qu'il est valide pour chaque élève pris individuellement. Il peut se produire des erreurs assez élevées pour un élève donné. Là où le chercheur en quête de mesure d'un phénomène se satisfera d'une validité globale, renvoyant les écarts individuels à des erreurs aléatoires qui se compensent, l'institution scolaire ne pourra se satisfaire d'un tel point de vue. On voudrait bien sûr que la décision d'orientation ne soit pas affectée par une erreur de ce type. Même des cas rares peuvent apparaître insupportables au vu des règles de l'équité scolaire *due à chacun*. Swann (1984) a par exemple montré que les individus dont le concept de soi est mal assuré sont peu armés pour contrer les mauvais jugements prodigués à leur encontre et peuvent en venir à se conformer (au moins en apparence) au jugement, fût-il erroné, confirmant ainsi le juge dans son erreur.

L'autre point que nous avons signalé est celui de la validité globale *versus* circonstanciée. Nombre de traits et d'activités jugés par l'Ecole ne relèvent pas que du domaine scolaire : sérieux, travailleur, motivé, enthousiaste, apathique, turbulent, etc. sont des traits que l'on retrouve évoqués dans bien des domaines de la vie sociale. On sait qu'un individu peut se comporter différemment selon les situations dans lesquelles il se trouve. Quel est par conséquent le type de validité qui est recherchée ? Dans bien des situations sociales, Swann (1984) a montré que les individus sont préoccupés par la situation d'interaction dans laquelle ils se trouvent et que le processus de « négociation d'identité » qui s'instaure est profondément dépendant de cette situation particulière. Si la cible du jugement ne lutte pas pour imposer une certaine vue de lui-même (vérification de soi), qui éventuellement peut provenir de situations externes à la situation d'interaction en question, il y a un plus fort risque que le jugement ait une validité circonstanciée plutôt que globale.

Mais finalement, est-ce le véritable problème ? Le jugement scolaire est-il bien une mesure ?

On manquerait sans doute un aspect essentiel de la question si on ne voyait dans les imperfections métrologiques du jugement scolaire que des erreurs, des biais dans la mesure effectuée par des enseignants aveuglés par leurs préjugés, limités dans leur capacité de traitement de l'information, qui auraient par conséquent des difficultés à décrire objectivement la réalité. Il ne suffit pas de décrire le jugement en termes d'erreurs, de lacunes,

de manques parce que, au quotidien, les agents sociaux recherchent sans doute moins une connaissance objective des autres agents avec lesquels ils interagissent, qu'une connaissance évaluative : « Ce qui s'offre au regard du sujet, ce sont d'abord les conduites, et l'attitude première de toute personne impliquée sera d'attribuer à ces conduites une valeur plutôt que d'en rendre compte » (Beauvois, 1984, p. 173).

Le jugement est un construit complexe qui intègre de nombreux éléments et dont on pourrait dire qu'il vise moins la connaissance objective et descriptive que la recherche d'éléments qui permettent de signifier quelle est la valeur des personnes évaluées, de leur attribuer une « place » pourrait-on dire dans un environnement social donné (il ou elle est bien à sa place dans cette classe, il ou elle ne sera pas à sa place dans cette filière, il ou elle n'a rien à faire ici, etc.). Le jugement scolaire n'est pas une activité intellectuelle *in abstracto*. Il s'agit d'une activité pratique qui sert l'action. Il guide l'enseignant dans ses interactions et ses prises de décision : savoir comment se comporter avec tel ou tel élève, décider du passage ou non dans la classe supérieure, fournir une aide supplémentaire, valoriser un effort, etc.

Une critique radicale du jugement des enseignants a été fournie par les travaux de Bourdieu. Pour lui, les classements des élèves expriment « la hiérarchie véritable des propriétés à reproduire » (Bourdieu & de Saint Martin, 1975, p. 69). Les auteurs ont montré que les appréciations d'une enseignante de philosophie se faisaient d'autant plus sévères, à note égale, que les élèves (toutes des filles) étaient provinciales plutôt que parisiennes et d'origine sociale moyenne plutôt qu'aisée. Les jugements professoraux seraient donc des « jugements d'attribution » (Bourdieu, 1979, p. 550) par lesquels on assigne quelqu'un à une position sociale. Les critères de reconnaissance – ou de distinction – seraient d'autant plus forts qu'ils marqueraient la personne de manière indélébile, touchant son appartenance sociale d'origine, celle qu'on ne peut changer. D'où l'importance de l'accent, de l'élocution, du style, des manières, du maintien, etc. dans le jugement formulé à l'égard – et à l'encontre parfois – d'autrui. D'où aussi des jugements qui se réalisent dans des adjectifs dont la plupart dénotent des qualités personnologiques allant des plus élogieuses (« fin », « ingénieux », « intelligent », « subtil »...) aux plus péjorantes (« simplet », « niais », « servile », « vulgaire »...). Paradoxalement, entre ces deux extrêmes, les adjectifs qui traduisent surtout des qualités revendiquées par l'Ecole (« scolaire », « soigneux », « attentif », « sérieux »...) ne sont pas les plus valorisées. En dévalorisant les manières trop scolaires, l'Ecole privilégierait des manières acquises en dehors d'elle et trahirait, par là même, sa fonction de reproduction sociale.

Au niveau de la notation proprement dite, le même doute peut être émis quant à son assimilation à une mesure qui viserait à décrire objectivement la réalité. Pour Chevallard (1992), la note a une fonction didactique de transaction : « L'attribution d'une note n'est pas un acte de mesurage, ni même une tentative [...] de parvenir à une telle mesure. Elle participe d'une transaction, et constitue un moment particulier – mais essentiel – d'un processus plus large, celui de la négociation didactique » (p. 36).

La situation d'enseignement est une situation dissymétrique où s'exerce un rapport de force entre l'enseignant et les élèves. L'enseignant tente de conduire sa classe vers une progression dans le savoir en essayant d'obtenir la reconnaissance de cette nécessité de la part des élèves. Cependant, ceux-ci peuvent recourir à certaines stratégies de retardement, tenter de freiner l'avancée dans le savoir. C'est dans cette négociation à propos du savoir à acquérir, sans cesse en passe d'être rompue, que se comprend la fonction de la note comme élément de transaction, de régulation.

Chevallard montre ainsi que l'enseignant est contraint à transiger, à réguler, en tentant de trouver le meilleur compromis entre des exigences souvent contradictoires. Il faut maintenir un certain niveau d'exigence sans pour autant laisser se noyer un trop grand nombre d'élèves. Il faut donc que les notes ne soient ni trop élevées, ni trop basses. Il faut que le groupe classe demeure homogène de manière à ne pas se retrouver face à des groupes distincts qui exigeraient des contrats didactiques différents, mais il faut malgré tout discriminer les élèves. Il faut donc que les notes ne soient ni trop dispersées, ni trop concentrées.

On montre ainsi qu'il y a une véritable stratégie derrière la notation dont l'enjeu n'est pas seulement (pas essentiellement dirait Chevallard) la mesure, mais bien aussi (surtout ?) le maintien d'une interaction vivable en classe. Les travaux de Merle (1996) montrent que ces transactions prennent même parfois la forme de négociations explicites autour de la notation. La note devient alors un moyen régulateur du bon fonctionnement de cette « relation obligée » entre élèves et enseignant.

Les tests ont-ils eux-mêmes des qualités métrologiques parfaites ?

Comme nous l'avons montré, les travaux qui ont traité la question de l'exactitude du jugement reposent sur une comparaison avec les performances des élèves à des épreuves standardisées. Ils s'inscrivent dans ce que Kruglanski (1990) a nommé le paradigme réaliste, qui pose la question de l'exactitude du jugement sur la base d'un critère externe supposé valide et qui sert d'étalon de mesure.

Mais les épreuves standardisées garantissent-elles elles-mêmes un critère valide de comparaison ? De la même manière que nous avons étudié les qualités métrologiques du jugement (avec toutes les précautions que nous avons prises dans ce que cela implique de considérer le jugement comme une mesure) il nous faudrait aussi étudier les qualités métrologiques des épreuves standardisées. Nous ne creuserons pas cette question dans le détail mais il faut rappeler qu'il y a une imperfection dans tout instrument de mesure et que des épreuves standardisées d'acquisitions, si bien construites soient-elles, n'échappent pas à cette règle.

Même si, globalement, leur validité, tant critérielle que de contenu et de construit est en général bonne (elles sont construites en fonction de cela), de même que leur fidélité, celles-ci ne sont pas parfaites. La consistance interne des épreuves par exemple, mesurée par le coefficient *alpha* de Cronbach n'atteint jamais la valeur maximale de 1, bien qu'étant en générale supérieure à 0,80. Or, la corrélation entre deux variables est affectée par les erreurs de mesure : plus elles sont élevées plus la corrélation sera atténuée.

Cela permet de poser un regard complémentaire concernant le fait que les jugements sur les niveaux en mathématiques et en français sont plus fortement corrélés entre eux que ne le sont les scores de deux épreuves d'acquisitions dans ces mêmes disciplines. C'est exactement ce que l'on rencontrerait si l'on comparait la corrélation entre deux épreuves entachées par des erreurs de mesure faibles et la corrélation entre deux épreuves mesurant la même habileté mais entachées d'erreurs de mesure plus élevées (i.e. dont la fidélité serait moins bonne). La première corrélation serait plus forte que la seconde. Dès lors, la corrélation plus élevée entre

les jugements pourrait-elle être due au fait que ces derniers seraient entachés de moins d'erreurs de mesure que les épreuves standardisées ? La fidélité du jugement n'ayant guère fait l'objet de recherches précises, on ne peut que relever ici cette question. Il ne s'agit cependant pas par là de mettre en doute l'existence de l'effet de halo noté *supra*, et qui a été montré à maintes reprises avec des protocoles différents, mais de montrer que la comparaison directe des deux corrélations repose sur l'hypothèse (implicite) que les erreurs de mesure sont identiques d'un type de variable à l'autre. Si tel est le cas, alors l'effet de halo du jugement enseignant correspond bien à ce qui a été énoncé *supra*. Si les erreurs de mesure du jugement sont plus fortes que celles des épreuves, alors l'effet de halo est plus important encore. En revanche, si les erreurs de mesure du jugement sont plus faibles que celles des épreuves, alors l'effet de halo est sans doute plus faible.

Nombreux sont les auteurs qui pensent aussi que les épreuves d'acquisitions (de même que les tests d'intelligence) reflètent les effets du contexte de passation sur la définition de la tâche à résoudre et qu'ils ne renseignent pas seulement sur le niveau d'habileté intrinsèque des individus mais aussi sur leur statut social (Croizet, 2001).

Tout cela n'explique cependant pas pourquoi, à performances données (donc entérinant éventuellement des biais sociaux), le jugement des enseignants surajoute des effets sociaux (e.g. à performances identiques, les enfants de cadre supérieur sont mieux jugés que les enfants d'ouvrier). Nos propres données réanalysées pour la présente étude (Bressoux & Pansu, 2003) montrent d'ailleurs que les caractéristiques sociales (profession du père) et scolaires (en retard ou non) des élèves exercent un poids dans la fabrication du jugement scolaire qui est plus fort que dans la réalisation des scores aux épreuves standardisées (en mathématiques par exemple, la profession du père et le retard scolaire expliquent 9,39 % de la variance des scores aux épreuves standardisées, mais 16,75 % de la variance des jugements des enseignants).

Ces remarques sur les qualités métrologiques des épreuves standardisées montrent tout autant les forces que les limites des épreuves standardisées et doivent nous amener à considérer avec une certaine prudence les propositions d'aucuns qui seraient tentés de substituer à la subjectivité des enseignants l'objectivité d'épreuves standardisées.

Présentation synthétique de travaux français sur le jugement enseignant

Nous présentons maintenant dans le tableau 1, de manière synthétique, les résultats de plusieurs recherches françaises qui ont porté sur les liens qu'entretiennent différents facteurs contextuels (établissement, classe), individuels (sexe, origine sociale, retard scolaire) et normatifs avec le jugement scolaire. Le tableau 1 n'a pas prétention à explorer de manière exhaustive toutes les recherches françaises conduites sur ce thème, mais à réunir un certain nombre d'études qui ont pris en compte des variables communes et qui utilisent une méthodologie commune.

Tableau 1 – Six variables susceptibles d’être liées aux jugements scolaires : synthèse d’études françaises intégrant au moins l’une de ces six variables

Etude	Echantillon	Méthode	Résultats					
			Etablissement	Classe	Sexe	Origine sociale	Retard scolaire	Influence normative
Bressoux & Pansu (1998)	18 classes de CE2 (343 élèves).	Jugements sur une échelle (0-10) en français et en maths.	—	Plus la classe est forte, plus le jugement est sévère.	Les garçons sont mieux jugés en maths. Pas de différences en français.	Les CSP favorisées sont mieux jugées en français et en maths.	Effet négatif du retard scolaire en français et en maths.*	Les élèves qui expriment le plus d’internalité sont mieux jugés en maths.
Bressoux & Pansu (2001a)	19 classes de CE2 (406 élèves).	Jugements sur une échelle (0-10) en français et en maths.	—	Plus la classe est forte, plus le jugement est sévère.	Pas de différences.	Les CSP favorisées sont mieux jugées en maths.	Effet négatif du retard scolaire en français.	Les élèves « clairvoyants » sur la valorisation de l’internalité sont mieux jugés en français et en maths.
Dompnier, Pansu & Bressoux (2006)	38 classes de CE2 (663 élèves)	Jugements sur une échelle (0-10) en français et en maths	—	Plus la classe est forte, plus le jugement est sévère.	—	—	Effet négatif du retard scolaire en français et en maths.	L’internalité de l’élève perçue par l’enseignant est valorisée en français et en maths.
Duru-Bellat & Mingat (1993)	17 collèges (2500 élèves).	Moyenne annuelle global des notes de français, maths et 1 ^{er} langue vivante.	Plus l’établissement accueille un public fort, plus la notation est sévère.	Plus la classe est forte, plus la notation est sévère.	Les filles sont mieux notées.	Les CSP favorisées sont mieux notées.	Effet négatif du retard scolaire.	—
Felouzis (1997)	25 classes de seconde en français (685 élèves). 36 classes de maths (927 élèves).	Notes moyennes trimestrielles en français et en mathématiques.	—	Plus la classe est forte, plus la notation est sévère.	Les filles sont mieux notées en français.	Pas d’effet de la CSP, ni en français, ni en maths.	Pas d’effet du retard scolaire, ni en français, ni en maths.	—
Jarlegan (1999)	21 classes de CE1 (472 élèves). 22 classes de CM2 (519 élèves).	Jugements sur une échelle (1-5) portant sur les résultats actuels, le potentiel, l’intérêt, les efforts, l’attention.	—	Pas d’analyse sur les CE1. En CM2, plus la classe est forte, plus le jugement est sévère sur les résultats actuels et sur le potentiel (pas d’analyse sur les 3 autres dimensions).	Aucune différence en CE1. En CM2, pas de différence sur l’évaluation des résultats actuels. Les garçons sont vus comme ayant un meilleur potentiel alors	En CE1, les CSP favorisées sont mieux jugées sur les 5 dimensions. En CM2, les CSP favorisées sont mieux jugées sur les résultats actuels (tendance : $p < 0,10$), sur l’intérêt	En CE1, effet négatif du retard scolaire sur les 5 jugements. En CM2, effet négatif du retard sur les résultats actuels (tendance : $p < 0,10$), sur le potentiel et sur l’intérêt.	—

					que les filles sont vues comme faisant plus d'efforts et étant plus attentives.	(tendance : $p < 0,10$), et sur l'attention. Une « prime » est accordée surtout aux enfants d'enseignants.		
Kaufmann (1987)	44 classes de 3 ^e en français (959 élèves). 49 classes de 3 ^e en maths (1039 élèves).	Notes moyennes annuelles en français et en mathématiques.	—	En maths, plus la classe est forte, plus la notation est indulgente. Pas de relation significative en français.	—	—	—	—
* Ce résultat provient d'une réanalyse des données, présentée dans Bressoux (2000), la variable « retard scolaire » n'ayant pas été prise en compte dans Bressoux & Pansu (1998)								

On note que, au-delà de certaines divergences d'une étude à l'autre, des régularités apparaissent. Ainsi, même si toutes les études ne décèlent pas d'effet significatif de l'origine sociale sur le jugement, aucune ne trouve d'effet négatif qui indiquerait un jugement meilleur quand la CSP est plus défavorisée. Ces résultats vont donc bien dans le sens d'autres travaux, anglo-saxons pour la plupart (cf. Dusek & Joseph, 1983) qui montrent que les élèves d'origine aisée sont souvent jugés plus favorablement que les élèves modestes. Toutes les études qui concluent à un effet significatif du retard scolaire concluent que les élèves en retard scolaire sont moins bien jugés. Il semble donc bien y avoir un effet d'étiquetage en ce domaine. Toutes les études qui ont pris en compte des effets de contexte concluent que plus le public accueilli par un établissement ou une classe est fort, plus le jugement est sévère, sauf celle de Kaufmann (1987). Cette étude est la seule à ne pas montrer que les classes fortes conduisent généralement à des jugements et des notations plus sévères de la part des enseignants. Faut-il y voir là une conséquence des caractéristiques de son échantillon (issu d'un seul district) ? Concernant l'étude de Felouzis (1997), qui ne montre pas d'effet du retard scolaire ni de l'origine sociale sur la notation, l'explication réside sans doute dans le niveau d'étude considéré, le lycée, car la sélection qui a été opérée préalablement aux différents paliers, et notamment en fin de 3^e, tend à réduire la différenciation sociale de réussite scolaire.

Les résultats consignés dans le tableau 1 accèdent l'idée que des informations telles que le contexte scolaire (établissement ou classe), le sexe, l'origine sociale et le retard scolaire des élèves sont bien généralement des informations sur lesquelles les enseignants fondent leur jugement concernant la valeur scolaire de leurs élèves. De plus, ils conduisent à avancer que le jugement des enseignants est affecté par des normes sociales générales, non spécifiques au milieu scolaire. Les travaux ayant pris en compte l'influence normative montrent en effet que les élèves qui se montrent les plus porteurs de cette norme sociale, qu'ils l'aient effectivement intériorisée ou bien qu'ils sachent simplement s'en montrer porteurs à bon escient (clairvoyance normative), sont mieux jugés que les autres.

Les choix d'orientation : sélection et autosélection

Tous les travaux sociologiques ont montré, depuis les années 1950 en France avec l'INED, le caractère socialement marqué des carrières scolaires. Il est inutile d'y revenir ici dans le détail. Rappelons toutefois que cette différenciation sociale s'opère surtout, aux premiers niveaux de la scolarité, sur la base de différences culturelles et que, par la suite, les différences économiques prennent de plus en plus d'importance.

Les carrières scolaires sont évidemment en grande partie le résultat de la sélection opérée par l'Ecole qui a inventé et mis en place, pour ce faire, tout un ensemble de dispositifs : redoublement, examens, concours, orientation, *numerus clausus*, etc. Pour l'instant, en abordant les phénomènes de jugement des enseignants, nous n'avons fait, d'une certaine manière, que traiter la question de la sélection à l'Ecole. Nous avons, au moins implicitement, considéré que l'orientation d'un élève était le résultat d'un processus de jugement de ses performances et, nous l'avons vu, probablement d'autres caractéristiques supposées refléter sa valeur scolaire et sociale. Nous avons par conséquent supposé que ce jugement fondait en grande partie la décision d'orientation elle-même.

Or, ce phénomène de sélection à l'École, fondamental pour la commande qui nous a été adressée dans cette étude, ne devrait cependant pas occulter que d'autres partenaires interviennent dans les choix d'orientation. L'élève et son milieu familial (parents essentiellement) sont bien sûr partie prenante dans le choix d'orientation. Cela ouvre à d'autres processus de gestion des carrières scolaires.

Il a été montré en effet que, à côté de la sélection proprement dite, opérée par l'École, un autre processus régissait les carrières scolaires des élèves et des étudiants : il s'agit de l'autosélection opérée par les individus eux-mêmes. L'autosélection peut se définir comme le fait que, à réussite scolaire identique, certains élèves vont s'interdire l'accès à certaines filières tandis que d'autres vont au contraire les choisir. Ces choix sont socialement marqués, (ce qui n'empêche pas les élèves de les vivre sous le mode de la liberté, de goûts individuels, voire de la vocation). Ils traduisent des préférences systématiques pour certaines filières en fonction de caractéristiques sociales déterminées.

Des données récentes issues du panel 1995 (élèves entrés en 6^e cette année-là) montrent que, une fois parvenus en 3^e, les élèves font des vœux d'orientation en seconde qui sont très loin de ne correspondre qu'à leur valeur scolaire effective (mesurée par les notes obtenues au contrôle continu du brevet), et que plusieurs facteurs sociodémographiques entrent en compte de manière substantielle (Caille, 2004). On relève ainsi que, à notes comparables au contrôle continu du brevet, la probabilité d'émettre un vœu d'orientation en seconde générale et technologique augmente quand la PCS du chef de famille est plus favorisée, quand le diplôme du père et celui de la mère sont plus élevés. La probabilité diminue quand les élèves sont plus âgés, et quand ils sont français (plutôt qu'étrangers). A l'évidence, les familles les plus défavorisées socio-économiquement et culturellement¹⁰ s'autosélectionnent. Duru-Bellat et Mingat (1988) avaient trouvé, par exemple, que pour des élèves ayant une moyenne annuelle comprise entre 9 et 11, 97,5 % des enfants dont le chef de famille était cadre supérieur ou de profession libérale demandaient la 4^e générale, alors que seulement 76,7 % des enfants d'ouvrier émettaient le même vœu. De plus, le passé scolaire de l'élève intervient aussi dans ses vœux d'orientation : le niveau d'ambition est plus faible pour les élèves qui ont redoublé, même si ce redoublement n'est pas récent. Il semble bien que cet événement scolaire, issu d'un jugement scolaire précédent, ait marqué profondément l'élève au point de réduire ses ambitions scolaires.

La décision d'orientation, comme l'avaient déjà bien montré Duru-Bellat et Mingat (1988) au niveau de l'orientation en 5^e, entérine en bien des cas le vœu d'orientation et, par conséquent, les différences sociales qu'elle incorpore : « Or, dans le déroulement de la procédure d'orientation de fin de troisième, le conseil de classe a plus souvent tendance à entériner cette forme d'autosélection de l'élève qu'à tenter d'y remédier » (Caille, 2004, p. 86). Les conseils de classe surajoutent même certains effets car, à notes et caractéristiques socio-économiques et culturelles identiques, les élèves en retard scolaire qui ont néanmoins formulé le vœu d'entrer en seconde générale et technologique obtiennent moins souvent satisfaction que les autres. Ce résultat est sans doute à rapprocher d'autres résultats qui montrent que les enseignants ont tendance à sous-évaluer le niveau des élèves en retard scolaire (Bressoux & Pansu, 2003). Tout se passe comme si le retard scolaire étiquetait l'élève comme faible et que les performances avérées de ces élèves ne parvenaient pas à compenser entièrement le pronostic qui leur est défavorable. Il faut aussi relever que Caille (2004) a trouvé des effets

¹⁰ Le cas des élèves étrangers mériterait un examen approfondi tant il est vrai que cette catégorie peut recouvrir des situations socioculturelles fort différentes.

sociaux dans les décisions d'orientation en seconde générale et technologique : à notes, parcours scolaire, vœux... comparables, les enfants dont le père est ouvrier non qualifié, dont la mère a un diplôme « inconnu » (cette absence d'information n'étant sans doute pas neutre socialement), ou dont la mère est inactive ont une probabilité plus faible d'entrer en seconde générale et technologique.

Ainsi, à réussite identique, les filières prestigieuses sont majoritairement choisies par les individus socio-économiquement favorisés et de moins en moins choisies au fur et à mesure que les individus sont socio-économiquement moins favorisés. Cela s'opère aussi bien pour les choix d'orientation dans l'enseignement secondaire (à réussite scolaire égale, les filières générales, et en particulier la filière scientifique voient une sur-représentation des classes sociales favorisées), que pour les choix d'orientation dans l'enseignement supérieur (les classes préparatoires aux grandes écoles, les grandes écoles elles-mêmes, les filières universitaires à concours telles médecine et pharmacie voient elles aussi une surreprésentation des classes sociales favorisées). Des différenciations fines apparaissent au fur et à mesure que l'on distingue plus précisément les diverses filières : ainsi, parmi les grandes écoles les plus prestigieuses (ENS, ENA, HEC, Polytechnique...), des différenciations sociales demeurent, qui vont voir par exemple les enfants d'enseignants se regrouper plutôt à l'ENS, les enfants de patrons de l'industrie et du commerce à HEC, etc. (Bourdieu, 1989, p. 191 et tableau 9, p. 192).

Les carrières scolaires sont ainsi faites d'innombrables petites choses, sélections successives, orientations successives, mais aussi toute une série d'incitations, de rappels à l'ordre émanant de l'Ecole, des pairs, de la famille, qui vont petit à petit « orienter », non pas tous azimuts car le monde social n'est pas chaotique, mais dans une direction plutôt qu'une autre.

Les mécanismes de l'autosélection ont fait l'objet de deux grands types d'interprétations. Pour Bourdieu (1974), l'autosélection obéit aux lois de la « causalité du probable ». Les agents sociaux, transformant ce qui est le plus probable pour eux en système de préférences (la pente qui devient penchant) vont en venir à choisir ce qui leur est offert et à s'interdire ce qui leur est interdit. Par un genre d'amour de leur destin (*amor fati*), les individus en viendraient ainsi à faire des « choix » qui, tout en étant raisonnables (ne pas se lancer dans une filière dont on ne pourrait assumer les coûts, ni dans une filière où on n'a aucune chance de réussir), c'est-à-dire globalement adaptés aux contraintes sociales qu'ils subissent, seraient guidés par des prédispositions, elles-mêmes inconscientes. Ces « stratégies » scolaires feraient partie d'un ensemble plus vaste de stratégies de reproduction (stratégies matrimoniales, sociales, économiques...), qui visent à maintenir ou améliorer la position sociale d'origine. Les stratégies scolaires dépendraient alors de la relation qui s'établit entre le patrimoine dans son volume et sa structure (part du capital économique, du capital social, du capital culturel) et les différents instruments de reproduction disponibles. Ainsi, les individus seraient d'autant plus enclins à investir dans l'école que le poids de la réussite scolaire dans la reproduction de la position sociale serait plus fort. L'investissement dans l'école favoriserait à son tour la réussite scolaire ce qui, en retour (rendant la réussite escomptée plus probable), augmenterait l'investissement scolaire. Ainsi peut-on comprendre pourquoi ceux qui n'ont rien d'autre à léguer que leur capital culturel réalisent un investissement dans l'école à hauteur de l'ambition du maintien de leur position sociale d'origine.

Pour Boudon (1973), les choix d'orientation doivent être compris comme un arbitrage rationnel opéré au sein d'un « espace de décision » défini par des coûts, des risques et des bénéfices, lesquels varient en fonction de la position socio-économique des acteurs sociaux.

Ceux-ci réaliseraient un arbitrage qui viserait à optimiser la combinaison coûts, risques, bénéfices. Or, cet arbitrage aura toutes chances d'être différent en fonction de la position socio-économique car cela va conduire à pondérer très différemment chacun de ces facteurs. Pour un individu socio-économiquement défavorisé, les coûts relatifs seront plus élevés, les risques plus élevés (perte de bourse en cas d'échec, risque de devoir renoncer aux études...), tandis que les bénéfices relatifs seront eux-mêmes plus vite attractifs (devenir technicien supérieur représente une ascension sociale pour l'enfant d'ouvrier, tandis que cela représente une régression sociale pour l'enfant d'ingénieur). On conçoit dès lors que les individus socio-économiquement défavorisés ont « intérêt » à se lancer dans des études moins longues et moins sélectives, desquelles ils pourront néanmoins escompter des bénéfices relatifs importants, tout en limitant les coûts et les risques. Aux individus socio-économiquement défavorisés, il ne reste guère que l'excellence scolaire pour diminuer les risques d'échec et accepter ainsi un arbitrage ambitieux mais néanmoins acceptable sur le plan des risques encourus.

Ces deux théories, différentes, voire antagonistes, ont cependant pour ambition d'expliquer les mêmes faits : l'autosélection scolaire socialement marquée. Par des voies interprétatives différentes, elles expliquent pourquoi, à réussite scolaire égale, les classes sociales défavorisées fréquentent moins que d'autres, proportionnellement parlant, les filières scolaires d'excellence. La sélection scolaire et l'autosélection, toutes deux socialement typées, ajoutent leurs effets pour produire les fortes différenciations de carrières scolaires que l'on observe.

Si le lien entre sélection scolaire et jugement des enseignants est d'emblée évident, il n'en va sans doute pas de même avec l'autosélection. Pourtant, nous voudrions ici montrer certains mécanismes par lesquels des liens peuvent s'opérer.

Le jugement que les enseignants portent sur la valeur scolaire de leurs élèves a toutes chances d'exercer des effets sur la perception que l'élève a de lui-même. Plusieurs travaux montrent que le jugement des enseignants affecte le sentiment de compétence scolaire des élèves, par-delà leurs performances avérées. Certains résultats montrent même que l'influence de ce jugement ne se limiterait pas au seul domaine scolaire mais qu'elle se diffuserait aussi sur d'autres domaines de la perception de soi tels que le domaine des relations sociales (facilité à se faire des amis, se sentir populaire, etc.) et celui du comportement (sentiment d'autocontrôle, bien se comporter, etc.) (Bressoux & Pansu, 2003). Tel un miroir, le jugement de l'enseignant indique à l'élève qui il est, jusqu'à influencer son propre jugement sur lui-même. Il lui confère une partie de son identité. Cela ne signifie évidemment pas que l'élève soit passif, qu'il soit le simple réceptacle du jugement d'autrui. Il y a toutes chances que s'opère une négociation, voire une lutte autour de la définition de l'identité. Les psychologues sociaux ont mis en évidence des phénomènes tels que la vérification de soi par exemple (tendance à ne percevoir et ne retenir que les éléments qui vérifient le concept de soi) qui montrent des stratégies (souvent inconscientes) par lesquelles les individus participent à la définition d'eux-mêmes. Il n'en demeure pas moins que la relation enseignant-enseigné est asymétrique, qu'elle se réalise dans une institution hautement normative et qu'elle confère de fait un poids important au jugement de l'enseignant, qui dispose d'une force sociale pour produire une « vérité sociale » sur les individus.

Le jugement des enseignants a toutes chances d'être repris et relayé par les pairs, de même que par les parents. Il sera même probablement d'autant plus fortement relayé par les parents qu'ils disposent de peu de ressources pour imposer une définition concurrente de leur enfant dans le cas où la définition scolaire serait peu valorisante. Il faut insister ici sur le fait qu'il

s'agit de fournir une définition autre, mais sur le même domaine, et non pas de renvoyer à l'institution scolaire une définition de soi décalée, voire antinomique avec les valeurs scolaires. Il s'agit de rester dans le jeu scolaire, non de s'en exclure. Ainsi peut-on comprendre par exemple le dilemme de la rébellion scolaire de certains élèves, contraints de fuir la définition d'eux-mêmes que leur impose l'institution scolaire, mais jouant du coup un autre jeu qui va précipiter leur rejet (symbolique puis concret) de l'Ecole.

Le jugement porté par l'Ecole, pour partie intégré par l'élève, pour partie relayé par les pairs, et par la famille, peut en conséquence influencer les choix d'orientation et, par là, alimenter l'autosélection.

Préconisations

Reprise des résultats principaux sur la fabrication du jugement scolaire

Nous avons tenté de montrer dans cette étude que la fabrication du jugement scolaire est complexe, loin de n'être que l'enregistrement objectif de performances avérées, et que ses relations avec l'orientation s'inscrivent elles-mêmes dans une dynamique aux multiples ramifications.

Rappelons tout d'abord quelles sont les voies par lesquelles le jugement des enseignants peut intervenir. Pour appréhender clairement les choses, reprenons les quatre éléments qui sont associés aux différenciations sociales d'orientation scolaire, tels que nous les avons présentés dans l'introduction.

- Les inégalités sociales de réussite scolaire. Le jugement de l'enseignant intervient ici selon deux modalités. Il peut effectivement transformer la valeur scolaire des élèves ; dans cet « effet Pygmalion », le jugement des enseignants, erroné au départ, devient vrai par un mécanisme de prophétie autoréalisatrice. Le jugement peut aussi traduire une simple erreur de perception qui, cependant, n'influence pas la réalité : la valeur scolaire de l'élève n'est pas affectée, mais les notes reçues par l'élève peuvent entériner ces biais de perception. Il faut ici faire intervenir aussi l'effet du contexte classe. De nombreux travaux ont montré que le jugement des enseignants dépend du niveau des autres élèves dans la classe : le jugement est d'autant plus sévère que la classe est plus forte. Toutefois, cet effet devrait jouer à l'encontre des enfants aisés puisqu'ils sont plus fréquemment dans des classes fortes.

- Les différenciations sociales inscrites dans les vœux d'orientation. Le rôle du jugement des enseignants dans la formulation des vœux d'orientation a été très peu étudié. Tout se passe comme si la majorité des chercheurs avaient considéré que les vœux d'orientation sont l'apanage de la famille. Un constat réalisé sur les vœux d'orientation au lycée peut cependant être rapproché de nos préoccupations. Si, en début d'année, dans les vœux d'orientation provisoire, on ne distingue guère de différenciations sociales à niveau scolaire comparable, on s'aperçoit, à la fin de la phase de concertation, que les différenciations sociales apparaissent. Les élèves faibles (mais pas trop) abandonnent plus fréquemment le projet d'intégrer la

filière S quand ils sont d'un milieu modeste que ceux de milieu aisé. Si l'on retient, comme le fait Duru-Bellat (2002), que les élèves de milieu modeste se montreraient par là « plus sensibles aux avis des enseignants » (p. 85), on peut aussi se dire que l'avis n'est peut-être pas tout à fait le même selon que l'élève est de milieu modeste ou aisé.

- Les inégalités sociales dans les décisions d'orientation. Le jugement intervient ici car les décisions d'orientation ne sont pas toujours les mêmes, à valeur scolaire comparable et à vœux d'orientation identique. Certaines catégories d'élèves, ceux qui sont en retard, ou ceux qui sont de milieu modeste obtiennent moins fréquemment satisfaction. Il est donc erroné de penser que la décision d'orientation serait le simple reflet du niveau scolaire et des vœux d'orientation.

- les écarts d'orientation entre établissements. Le jugement des enseignants intervient ici au niveau d'une sévérité globale. Celle-ci est sans doute, le plus souvent, le fruit d'une politique d'établissement qui vise à promouvoir *versus* à sélectionner les élèves (évidemment, il ne s'agit pas de choix dichotomiques mais d'une question de placement du curseur sur le continuum promotion-sélection). Plusieurs études ont montré que, à cet égard, les politiques des établissements sont fort diverses. Ceux-ci cherchent à gérer leur image, mais subissent aussi le poids de leur morphologie (ont-ils des filières techniques, ont-ils des classes préparatoires...), de leur environnement proche (les familles par exemple) ou plus éloigné (où sont les établissements concurrents et comment se « placent-ils »...). Certains établissements, sévères dans leur notation, compensent cette sévérité accrue par des orientations moins sélectives, mais d'autres maintiennent un haut degré de sélection dans les procédures d'orientation. Ces différences de sévérité entre établissements ne sont que pour une part accessibles aux acteurs eux-mêmes. Toute politique n'est pas complètement explicite. Elle est aussi véhiculée par une norme locale plus ou moins consciemment perçue. De plus, il est très difficile pour les acteurs de placer objectivement leur établissement par rapport aux autres établissements et, ainsi, d'avoir une vision objective des effets objectifs de cette politique.

Dans l'ensemble de ces processus, quel est le poids des jugements des enseignants ? Au global, sur chacun de ces processus, le poids du jugement n'est pas prédominant. Dans chacun des cas, la valeur scolaire avérée est déterminante. Et même dans la fabrication de la valeur scolaire, le processus de prophétie autoréalisatrice montre des effets relativement limités, l'enseignant se révélant, à cet égard, plutôt « Pygmalion à la marge ».

L'une des raisons essentielles à ce poids modéré est que le jugement des enseignants est globalement valide. Autrement dit, il traduit bien la valeur scolaire des élèves et ne joue, à cet égard, dans les processus d'orientation, qu'à la marge de la valeur scolaire. On constate, pour le jugement scolaire, une plus forte prévention des biais potentiels énoncés en première partie que pour d'autres formes de jugement quotidien. La raison tient à plusieurs éléments liés à la spécificité du jugement scolaire. Il y a tout d'abord la durée de l'interaction, qui laisse un long temps de connaissance mutuelle entre les acteurs, et laisse aussi un temps de réflexion et de maturation du jugement. De plus, les enseignants sont engagés dans la relation et sont conscients de l'importance de leur jugement, qui a des conséquences lourdes pour les élèves. Cela les conduit à éviter des procédures trop rapides et expéditives de jugement. Ils sont souvent conduits à porter un regard réflexif sur leur jugement. Suis-je objectif ? Quel poids dois-je attribuer à tel élément favorable et à tel élément défavorable ? Etc. De plus et surtout, ils disposent de tout un ensemble d'informations pertinentes par rapport à l'objet de leur jugement. Pour juger de la valeur en mathématiques d'un élève par exemple, les résultats qu'il obtient aux exercices de mathématiques proposés constituent une information en général

valide (quoique, nous l'avons vu, non nécessairement parfaite). A l'évidence, le jugement scolaire se fonde sur des informations beaucoup plus pertinentes que celles dont on dispose dans les situations courantes du jugement quotidien (juger qu'un individu est courageux, que cette personne est quelqu'un de « bien », quelqu'un de confiance, un bon professionnel, etc.).

Est-ce à dire que le jugement des enseignants est parfait ? Nous l'avons vu, à l'évidence non. On ne peut nier l'influence de grandes normes sociales et de stéréotypes. A performances identiques, les élèves qui savent montrer qu'ils ont intégré certaines grandes normes sociales (politesse, internalité, hygiène, altruisme...), mais aussi les élèves les plus favorisés socialement, les élèves au physique attractif, etc. sont en général mieux jugés que les autres. Si ces effets ont été observés à de maintes reprises, ils ne jouent cependant, en moyenne, qu'assez marginalement par rapport au poids des performances avérées des élèves¹¹.

Cette conclusion n'est cependant valable *qu'en moyenne*. Nous avons montré que certains enseignants ont des jugements plus « exacts » que d'autres, que certains se révèlent beaucoup plus sensibles que d'autres aux stéréotypes. Les enseignants dogmatiques sont en particulier plus sensibles à l'influence des stéréotypes et, par conséquent, davantage susceptibles d'émettre des jugements « biaisés ». Nous insistons ici toutefois sur le fait que le dogmatisme ne s'entend pas pour nous nécessairement comme un trait de personnalité. Il peut s'agir d'une attitude, plus ou moins stable, et qui surtout peut se forger sur des caractéristiques situationnelles : le manque d'expérience, le manque d'assurance, un sentiment d'insécurité, etc. peuvent favoriser le dogmatisme, en augmentant la fixité des points de repère dans un monde qui risque d'apparaître, sans cela, bien inquiétant.

Ce n'est parce qu'un jugement est *en moyenne* exact qu'on peut admettre, dans la pratique, qu'il s'éloigne fortement de la réalité pour certains élèves et la question doit rester posée de savoir comment réduire les erreurs sur ces élèves.

L'un des problèmes qui nous apparaissent les plus cruciaux pour la question de l'équité des jugements et de l'orientation scolaire est celui des effets de contexte dans la fabrication du jugement. Pour juger un élève, les enseignants disposent nécessairement d'un cadre de référence. En l'absence d'un critère absolu de mesure de performance, ce cadre de référence est notamment composé des performances des autres élèves de la classe. Cela implique que les enseignants se fondent sur des comparaisons qui s'opèrent surtout par rapport aux élèves de leur classe, éventuellement par rapport aux élèves des diverses classes auxquelles ils enseignent (ou ont enseigné). Comme les classes et les établissements accueillent des publics fort variés du point de vue de leur niveau social et de leur niveau de performances scolaires, il s'ensuit que les cadres de référence des enseignants sont différents les uns des autres. Lorsque que le public accueilli est fort, les références sont élevées et, inversement, lorsque que le public accueilli est faible, les références sont basses.

Il ne s'agit pas de dire que les enseignants sont totalement aveugles à ce phénomène. Les enseignants en ZEP savent que le niveau de leurs élèves a toutes chances d'être moins élevé que celui des élèves du lycée bourgeois de centre ville. Mais, d'une part, dans les situations moins marquées socialement, l'évidence de l'effet de contexte est sans doute moins forte et, d'autre part, les enseignants ne disposent guère de points de repère fixes et valides pour situer

¹¹ Pour un facteur comme l'attractivité physique, ce sont souvent des travaux expérimentaux qui l'ont mis en évidence. En rendant cette information saillante et en neutralisant d'autres, il est possible que les conditions expérimentales accentuent quelque peu l'effet de ce facteur.

exactement le niveau de leurs élèves par rapport à l'ensemble de la population nationale correspondante.

Cela se traduit par des effets de contexte marqués sur le jugement des enseignants et la notation. Les enseignants sont plus sévères dans les établissements qui accueillent un public fort et moins sévères dans les établissements qui accueillent un public faible. De sorte qu'un même élève verra son évaluation et, du coup, son orientation, varier selon l'établissement où il se trouve. Cela est surtout sensible pour les élèves qui sont « objectivement » de niveau moyen car ils pourront passer du statut de fort dans une classe faible à celui de faible dans une classe forte.

Du bon usage des épreuves standardisées

Dans un but d'équité de jugement et d'orientation, l'une des conséquences évidentes de ce qui vient d'être dit est que l'on aurait tout intérêt à harmoniser autant que faire se peut les cadres de référence des enseignants de manière à ce qu'ils jugent les élèves sur une base commune. Cela est réalisable en fournissant aux enseignants un étalon de mesure commun à partir duquel ils pourraient situer le niveau de leurs élèves par rapport à l'ensemble de la population nationale correspondante. De ce point de vue, les épreuves nationales de masse (6^e, 2^{nde}) sont un bon support. Il en va de même de certains critères IPES. Mais il faudrait s'inquiéter de l'usage qui est fait par les enseignants des résultats de leur établissement. En sont-ils réellement informés, et se servent-ils de cette information pour moduler leurs critères de jugement ? A cet égard, une sensibilisation à ces effets par la formation continue serait profitable. De même, le rôle des conseillers d'orientation-psychologues, non pris en compte jusque là dans ce rapport, est précieux pour apporter des informations sur les élèves grâce aux bilans standardisés qu'ils réalisent (Aubret, 2006).

Pour aller plus loin, faudrait-il que chaque année, ou tout au moins à chaque palier d'orientation, tous les élèves passent des épreuves standardisées de manière à adosser la décision d'orientation sur un étalon commun ? Il est tout à fait illusoire de penser que les scores obtenus à des épreuves standardisées serviront de critère unique d'orientation. D'une part, cela n'est pas souhaitable parce que les épreuves standardisées elles-mêmes ne sont pas parfaites. D'autre part, cela n'est guère réalisable parce que l'orientation met en jeu un pari sur l'avenir dont on voit mal comment il ne pourrait pas relever de la prise en compte d'un ensemble d'éléments dont la synthèse peut difficilement évacuer le jugement humain. Gageons cependant que, subordonné au jugement enseignant, conçu comme une aide à la formation de ce jugement, cela pourrait constituer une réponse intéressante aux variations de cadres de référence des enseignants. Outre que cela devrait avoir pour fonction de limiter les effets de contexte, cela devrait aussi permettre, ponctuellement, de corriger le jugement d'un enseignant, erroné ou biaisé pour des élèves particuliers. Le conseil de classe devrait pouvoir s'emparer des résultats des élèves aux épreuves standardisées pour moduler, pondérer, les notes obtenues par les élèves. Une éventuelle « correction » des notes devrait être réalisée, à notre sens, surtout lorsque les scores aux épreuves standardisées se trouvent à l'évidence supérieurs aux notes obtenues en classe. D'une part parce qu'il s'agit surtout de corriger les inégalités qui se font au détriment des individus. D'autre part parce que les épreuves standardisées, si elles permettent d'attester que la réussite exprime que l'élève possède bien les habiletés qui permettent d'atteindre ledit niveau de réussite (sous réserve de non-tricherie

bien sûr), n'ont pas la qualité inverse, celle d'attester que la non-réussite révèle l'absence des habiletés nécessaires à la réussite. Nous l'avons vu dans le rapport, le rappel de souvenirs autobiographiques d'échec, la menace du stéréotype ou d'autres facteurs contingents (maladie, préoccupations passagères créant un déficit attentionnel, etc.) nuisent à la performance réalisée lors d'une épreuve. Il est donc davantage fondé de corriger des notes « vers le haut » que « vers le bas » sur la base des scores obtenus à une épreuve standardisée. Il reste que tout ne s'évalue pas facilement sur la base d'épreuves standardisées : l'évaluation de l'expression écrite par exemple, ne laisse pas de poser problème, surtout dans sa partie expressivité et créativité. De même, comment évaluer les qualités artistiques des élèves ?

La standardisation à l'aune des procédures de sélection et d'autosélection

La standardisation ne pose pas que des problèmes techniques dans ses qualités métrologiques ou dans sa capacité à corriger d'éventuels biais du jugement enseignant. Raisonner ainsi supposerait que l'on adopte le postulat que chacun est disposé à s'ajuster à une « objectivité » qui serait fondée sur un étalon commun de mesure de performances. Or, un tel postulat ne manquera pas de se heurter d'emblée à la logique même des processus engagés dans le jugement et l'orientation des élèves.

Nous l'avons vu, les différences de notation et d'orientation entre établissements (et même entre enseignants) ne relèvent pas que d'un défaut d'objectivité. Elles relèvent aussi de politiques (ou de conceptions du métier au niveau des enseignants) mises en place au niveau de l'établissement. Certains établissements (et certains enseignants) se veulent – et s'affichent – sélectifs. D'autres adoptent davantage une optique de promotion. A cet égard, l'information standardisée sur le niveau des élèves risque de ne produire que des effets bien insignifiants s'il ne s'accompagne de quelque injonction à les prendre en compte dans la décision d'orientation. Dans le fil de ce raisonnement, le rôle du chef d'établissement apparaît ambigu dans la procédure d'orientation. Rappelons que le chef d'établissement confirme ou non la décision d'orientation prise en conseil de classe, et l'arrête réglementairement. S'il peut être celui qui essaie de limiter d'éventuels écarts qu'il serait amené à percevoir entre les jugements des enseignants et les performances, voire entre les différents degrés de sévérité dans la notation des enseignants, il est aussi le principal artisan de la politique de l'établissement. Or, comme nous avons tenté de le mettre en évidence, celle-ci ne s'accorde pas toujours d'un principe de standardisation car elle obéit parfois à une logique opposée, de singularisation, voire de distinction. Se vouloir sélectif (en vue de l'intégration dans les classes préparatoires par exemple) c'est aussi se vouloir plus sélectif que les autres. On rejoint ainsi la marge d'autonomie dont bénéficie tout établissement dans la définition de ses missions.

Du point de vue des familles, l'un des problèmes est que, bien souvent, elles ne savent pas quel est le degré de sélectivité de l'établissement. Qui plus est, en raison de la carte scolaire, elles n'ont pas le choix de l'établissement d'affectation. L'opacité et la contrainte du système rendent la combinaison bien inéquitable¹².

¹² Cette réflexion ne doit pas être prise comme un argument pour ou contre la carte scolaire. Nous insistons ici sur l'inéquité que crée la conjonction de plusieurs caractéristiques (politiques différentes, opacité, inégalité des

N'oublions pas non plus que les familles formulent des choix et s'autosélectionnent. De ce point de vue, un regard objectif donné sur le niveau de l'enfant pourrait être salutaire. Bien que différents éléments entrent en compte dans le choix d'orientation (coûts supportés, risques encourus, bénéfices escomptés pour se limiter aux éléments d'une vision utilitariste de l'orientation), il reste que la valeur scolaire intervient dans l'arbitrage entre ces éléments, en affectant leur pondération, notamment pour les risques encourus, qui apparaissent moindres quand la valeur scolaire est élevée. Notons que, même dans une interprétation différente de l'autosélection, la valeur scolaire intervient à deux niveaux, comme « goût pour... » (j'ai toutes chances de préférer les matières qui me réussissent, qui sont par conséquent celles où ma probabilité de réussir est élevée), et comme propension à investir dans l'éducation (j'investis d'autant plus que ma réussite actuelle rend l'accomplissement futur possible – par l'Ecole). La valeur scolaire, telle que produite par l'Ecole, est par conséquent une donnée fondamentale des choix d'orientation. Tout montre même qu'elle est d'autant plus fondamentale que les familles sont plus modestes car se sont elles les plus promptes à se limiter dans les choix vers les filières d'excellence.

Du point de vue de l'autosélection, le problème est sans doute moins celui des choix trop ambitieux que celui des choix trop peu ambitieux. D'une part, parce que les procédures d'orientation corrigent plus le premier élément que le second. D'autre part, parce que la surévaluation est surtout le fait des catégories aisées, qui ont les moyens (notamment économiques) d'une seconde chance. Ce constat vaut surtout pour le supérieur attendu que, dans le secondaire, l'échec dans une filière peut avoir des conséquences (presque) aussi dramatiques pour les élèves de milieu aisé que pour ceux de milieu modeste (redoublement, réorientation subie, etc.).

Pour l'orientation dans l'enseignement supérieur, le baccalauréat peut jouer le rôle d'étalon commun du fait qu'il s'agit d'une épreuve commune aux élèves d'une même filière. Certes, il ne s'agit pas d'une évaluation standardisée proprement dit (en particulier dans son étalonnage psychométrique et dans son mode de correction) mais le baccalauréat permet de situer le résultat obtenu par rapport à une population large (à l'inverse de la note à un contrôle, contrainte à ne rester valide que dans le cercle étroit de la classe). Le résultat au baccalauréat peut alors jouer comme un indicateur, voire un révélateur de la valeur scolaire, dont on sait qu'elle joue de manière considérable sur l'évaluation des risques encourus lors du choix de la filière.

Le problème, c'est que certaines demandes (nous pensons ici en particulier aux demandes d'inscription en classe préparatoire) s'effectuent avant l'obtention des résultats au baccalauréat. Si cela peut toujours produire un effet « descendant » tel que renoncer finalement à un choix risqué et se rabattre sur une filière moins sélective, restée ouverte à l'inscription, il n'y a plus de possibilité d'effet « ascendant » : décider, finalement, de s'inscrire en classe préparatoire où, au vu d'excellents résultats obtenus au baccalauréat par exemple, on envisagerait raisonnablement courir des risques qui, de ce fait, apparaîtraient moindres.

opportunités à échapper à la carte scolaire, contraintes d'affectation, ségrégation sociale spatiale, etc.). Une modification d'une ou plusieurs caractéristiques pourrait rendre la synthèse bien différente.

Qu'est-ce qu'une orientation réussie ?

Faut-il se contenter de dire qu'une orientation réussie est une orientation qui satisfait le choix des élèves (sous réserve que le niveau scolaire rende ledit choix possible bien sûr) ? La réponse à cette question n'est, en fait, pas évidente. Il faut rappeler à cet égard que la politique suivie actuellement en France, qui invite surtout le conseil de classe à suivre le choix des familles, semble entériner, au moins implicitement, ce postulat. Dans notre pays, l'orientation scolaire est conçue, tout au moins au niveau secondaire, comme une réponse aux demandes familiales, les conseils de classe jouant, pour l'essentiel, de manière réactive (Duru-Bellat, 2002) : « Les conseils se contentent en fait de contester les choix qui leur apparaissent irréalistes vu le niveau scolaire de l'élève, sans chercher à "tirer vers le haut" les choix prudents de certains jeunes. De par ce mode de fonctionnement, conformes aux textes qui les invitent, autant que faire se peut, à suivre les choix des familles, les conseils de classe entérinent des demandes socialement différenciées » (p. 79).

On ne peut évidemment pas aller contre le gré des individus (sous réserve, nous le répétons, que le niveau scolaire permette d'agréer le vœu formulé). Il est en revanche fondamental que l'institution scolaire sache « tirer vers le haut » les demandes d'orientation des plus modestes, celles qui font manifestement l'objet d'une autosélection forte. Cet élément doit s'entendre non seulement d'un point de vue individuel qui viserait la maximisation de l'utilité personnelle, mais aussi d'un point de vue collectif, afin qu'une nation ne gaspille pas sa « réserve de talents », c'est-à-dire en assurant par là qu'elle distribue bien les postes les plus haut placés aux individus les plus talentueux, et qu'elle ne les distribue pas seulement à ceux qui ont bénéficié des conditions, notamment économiques, qui leur permettent de satisfaire aux conditions pour les briguer.

Cela nous conduit à traiter la question du droit aux parents et aux élèves de décider *in fine* de leur orientation. Il faut alors remarquer que, si une telle perspective se mettait en place, cela conduirait à substituer entièrement l'autosélection à la sélection. La question est, par conséquent, que perd-on ou que gagne-t-on à enlever la sélection par l'Ecole ?

Nous l'avons vu, la sélection à l'orientation vise surtout à corriger des demandes jugées irréalistes au vu du niveau des élèves. On sait que le maintien de la demande de filières prestigieuses malgré des résultats peu probants est surtout le fait des familles aisées. Plus on laisse de choix aux familles, plus on entérine par conséquent les biais sociaux. Nous sommes ainsi dans un dilemme car la liberté de choix, souhaitable en elle-même, conduit à des inégalités sociales dans l'orientation scolaire, non désirables dans un système méritocratique et qui, de plus, risquent de conduire à une exploitation non optimale de la « réserve de talents » d'une nation. Le hiatus entre liberté et égalité, pointé il y a fort longtemps par Tocqueville, est ici bien apparent. A cet égard, ce qui se passe au lycée est éclairant. Dans la lutte pour le placement aux bonnes filières, quand la course s'accélère et devient décisive car de la note et de la série du bac se profileront ensuite des possibilités et des impossibilités fortes quant au cursus envisagé, autrement dit, lorsque la sélection scolaire se fait forte, les inégalités sociales d'orientation sont plus faibles.

Puisque la course à la recherche de la maximisation de l'utilité semble biaisée dans ses fondements, faut-il du coup se résigner à laisser s'exprimer les goûts, les vocations ? Autrement dit, puisque tous ne veulent pas maximiser leur utilité économique, ne vaut-il pas mieux les rendre heureux ! C'est oublier que les goûts, les vocations, sont aussi faits des possibilités ou impossibilités perçues à réaliser un destin. Il n'est pas sûr que les élèves en

classe préparatoire déclarent avoir plus de goût pour ce type d'études que les élèves relégués à des filières d'apprentissage ne peuvent en exprimer pour les leurs (j'ai choisi coiffure parce que ça me plaît, j'ai toujours voulu faire mécanique auto, etc.).

Il y a souvent un caractère paradoxal dans la construction des projets. On risque fort d'écouter ses goûts de manière d'autant plus intense qu'on a en effet renoncé (dû renoncer) aux filières d'excellence et que, faute de bénéfices importants à escompter de la filière choisie, il reste à se faire plaisir, voire à prolonger autant que possible le plaisir de faire des études (on rejoint par là le modèle d'éligibilité que Lévy-Garboua avait proposé il y a déjà longtemps pour expliquer l'allongement du temps des études à l'université).

Les individus ont sans doute, dans la plupart des cas, de « bonnes raisons » de se diriger là où ils se dirigent parce que tout contribue, en quelque sorte, à les y pousser : intérêts « objectifs » tels que la capacité à trouver un emploi après les études, multiples injonctions et rappels à l'ordre (ce n'est pas pour toi ! Qu'iras-tu y faire ? C'est pour les filles ! Etc.), notes obtenues, décisions de conseils de classe, etc. Le cas des filles est ici trop évident pour qu'on ne le cite pas : plus promptes à aller vers les filières littéraires (féminines !), elles sont aussi, généralement, pénalisées dans leurs notes par rapport aux garçons du fait « de la précision scientifique et de la bonne compréhension des principes » attribuée à ces derniers, à copie pourtant... identique (Duru-Bellat, 1990) ! N'ont-elles pas de bonnes raisons de fuir les filières scientifiques où la « précision scientifique » et la « bonne compréhension des principes » sont requises ? On peut se demander cependant si le rôle de l'institution scolaire ne consiste qu'à entériner, voire à légitimer des destinées sociales hautement différenciées, fondées sur des principes générés en dehors d'elle.

Addendum

Cet *addendum* résulte de questions posées en dehors de la commande initiale. Les réponses ne sont pas présentées dans le corps du rapport parce qu'elles ne traitent pas directement l'objet initial de la commande et parce qu'elles reflètent plus une position personnelle de l'auteur qu'un avis longuement élaboré sur la base d'informations scientifiques.

– La sélection à l'université. Disons en préambule que tout ce qui engendre des risques limités et des coûts faibles est susceptible de favoriser les candidatures des moins aisés. En ce sens, une sélection initiale, qui serait non coûteuse (avec des frais de dossier très modérés), et sans autre risque que le refus de la candidature, qui permettrait donc de se rabattre sur un second choix, pourrait être attractive pour les plus modestes économiquement. Cependant, chacun sait que, dans la réalité, les frais de dossiers sont parfois élevés et, même lorsqu'ils sont modestes, peuvent se révéler au bout du compte fort coûteux en cas de candidatures multiples. Une fois la sélection initiale passée, les coûts seraient toutefois maintenus constants. Les risques d'échouer peuvent être alors envisagés de diverses manières : du point de vue global, l'effectif plus restreint entraînerait une probabilité globale de succès sans doute plus élevée. Au niveau individuel cette fois, la sélection risque fort d'élever le niveau général et ainsi, comparativement aux autres étudiants, de rendre la réussite finale moins probable, à niveau scolaire donné.

– Le *numerus clausus* jouerait un effet quelque peu différent. Il n'engendrerait sans doute pas de frais d'inscriptions supplémentaires et, dans l'optique où la « sélection » ne s'effectuerait que par l'effet liste d'attente (premier arrivé premier servi), ne conduirait pas à une hausse du niveau général des étudiants. La probabilité individuelle de succès devrait donc être, de ce fait, plus élevée par un simple effet de réduction des effectifs. Cela dit, au niveau collectif, cela priverait la filière d'un certain nombre de « bons » étudiants sous le seul prétexte qu'ils seraient arrivés trop tard et qu'ils seraient placés trop loin sur la liste de candidature. Ce qui n'est pas très efficace pour ladite filière. De plus, cette pratique suppose, de la part des usagers, un bon niveau d'information pour se trouver au bon endroit au bon moment. Or, l'information n'est pas forcément d'accès égal pour tous. Le marché idéal composé d'individus parfaitement informés n'existe que dans les modèles des économistes et ne vaut que parce qu'il est une épure de la réalité, qui permet de faire « comme si ».

– La course au diplôme peut être vue selon au moins deux points de vue. D'une part, elle est rentable puisque malgré la dépréciation globale des diplômes qui s'observe en valeur absolue (on ne peut prétendre avoir le même poste maintenant avec un bac qu'il y a trente ans), ceux-ci restent, relativement parlant, très discriminants. Ils sont notamment protecteurs du chômage. Chaque diplôme couvre, dans la hiérarchie des emplois, une « zone » plus étroite maintenant qu'auparavant, mais cette zone est plus étanche qu'elle ne l'était naguère. Autrement dit, l'ascension sociale est plus difficile si l'on ne dispose pas des titres scolaires requis au « franchissement » des diverses zones. A cet égard, la course au diplôme est parfaitement raisonnable au sens d'une maximisation de l'utilité. D'autre part, la course au diplôme est aussi parfois un moyen de prolonger le temps des études car il s'agit d'un temps protégé, et la multiplication des « petits boulots » étudiants peut être vue comme un moyen de vivre cette période sans avoir à trop se priver financièrement et sans se sentir une charge trop lourde pour les parents. Le temps consacré à ces petits boulots constitue en effet un coût d'opportunité par rapport au travail qu'exige l'obtention du diplôme (et limite donc les chances de succès). Inversement, les petits boulots permettent de limiter le coût d'opportunité,

financier cette fois, que représente le fait de faire des études plutôt que d'être en emploi, avec les gains que l'on pourrait espérer compte tenu du niveau d'études déjà atteint. On comprendra cependant d'emblée que ce coût d'opportunité sera d'autant plus faible que les perspectives de rémunération immédiate sont peu avantageuses et que la probabilité d'être au chômage augmente. On comprend ainsi aisément que, à l'heure actuelle, les étudiants ont « intérêt » à prolonger leurs études. Si cette prolongation des études se traduit par une accumulation de capital humain qui est susceptible d'accroître la productivité des individus, alors on peut penser que la société en tirera aussi des bénéfices. En revanche, dans le cas où cette prolongation des études ne se traduit pas par une accumulation de capital humain, le diplôme jouant surtout un rôle de signal pour l'employeur potentiel, alors la quasi-gratuité des études supérieures en France pose problème. Dans une telle hypothèse, la société supporte un coût dont seuls les individus tireraient bénéfice à titre personnel. Qui plus est, la société financerait surtout les études des moins nécessaires puisque les individus aisés sont surreprésentés dans l'enseignement supérieur.

Références bibliographiques

- Alvidrez, J., & Weinstein, R. S. (1999). Early teacher perceptions and later student academic achievement. *Journal of Educational Psychology, 91*(4), 731-746.
- Aubret, J. (2006). Les conseillers d'orientation. In J. Beillerot, & N. Mosconi (Eds), *Traité des sciences et des pratiques de l'éducation* (pp. 305-315). Paris :Dunod.
- Babad, E. Y., Inbar, J., & Rosenthal, R. (1982). Pygmalion, Galatea and the Golem : investigations of biased and unbiased teachers. *Journal of Educational Psychology, 74*(4), 459-474.
- Balzer, W. K., & Sulsky, L. M. (1992). Halo and performance appraisal research: A critical examination. *Journal of Applied Psychology, 77*, 975-985.
- Beauvois J.-L. (1976). Problématique des conduites sociales d'évaluation. *Connexions, 19*, 7-30.
- Beauvois, J.-L., (1984). *La psychologie quotidienne*. Paris : PUF.
- Beauvois, J.-L., & Dubois, N. (1988). The norm of internality in the explanation of psychological events. *European Journal of Social Psychology, 18*, 299-316.
- Bonniol, J. J., Caverni, J. P., & Noizet, G. (1972). Le statut scolaire des élèves comme déterminant de l'évaluation des devoirs qu'ils produisent. *Cahiers de Psychologie, 15*, 83-92.
- Boudon, R. (1973). *L'inégalité des chances*. Paris : Armand Colin.
- Bourdieu, P. (1974). Avenir de classe et causalité du probable. *Revue Française de Sociologie, 15*(1), 3-42.
- Bourdieu, P. (1979). *La distinction. Critique sociale du jugement*. Paris : Editions de Minuit.
- Bourdieu, P. (1989). *La noblesse d'état*. Paris : Editions de Minuit.
- Bourdieu, P. (2002). *Le bal des célibataires*. Paris : Seuil.
- Bourdieu, P., & de Saint Martin, M. (1975). Les catégories de l'entendement professoral. *Actes de la recherche en sciences sociales, 3*, 68-93.
- Brattesani, K. A., Weinstein, R. S., & Marshall H. H. (1984). Student perceptions of differential teacher treatment as moderators of teacher expectation effects. *Journal of Educational Psychology, 76*(2), 236-247.
- Bressoux, P. (1993). *Les effets des écoles et des classes sur l'apprentissage de la lecture*. Thèse de doctorat en Sciences de l'Éducation, non publiée, Université de Bourgogne, Dijon.
- Bressoux, P. (à paraître). *La modélisation statistique en sciences sociales*. Bruxelles : De Boeck.
- Bressoux, P., & Pansu, P. (1998). Norme d'internalité et activités évaluatives en milieu scolaire. *Revue Française de Pédagogie, 122*, 19-29.
- Bressoux, P., & Pansu, P. (2001a). Effet de contexte, valeur d'internalité et jugement scolaire. *L'Orientation Scolaire et Professionnelle, 30*(3), 353-371.
- Bressoux, P. & Pansu, P. (2003). *Quand les enseignants jugent leurs élèves*. Paris : PUF.
- Brookover, W. B., Schweitzer, J., Beady, C., Flood, P., & Wisenbaker, J. (1979). *School Social Systems and Student Achievement : Schools Can Make a Difference*. New York: Praeger.
- Brophy, J. E., & Good, T. L. (1974). *Teacher-student relationships : causes and consequences*. New York : Holt, Rinehart and Winston.
- Caille, J.-P. (2004). Le redoublement à l'école élémentaire et dans l'enseignement secondaire : évolution des redoublements et parcours scolaires des redoublants au cours des années 1990-2000. *Education et Formations, 69*, 79-88.
- Caverni, J. P., Fabre, J. M., & Noizet, G. (1975). Dépendances des évaluations scolaires par rapport à des évaluations antérieures. *Le Travail Humain, 38*, 213-222.
- Coladarci, T. (1986). Accuracy of teachers judgments of student responses to standardized test items. *Journal of Educational Psychology, 78*(2), 141-146.
- Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin, 90*, 218-244.

- Crahay, M. (1996). *Peut-on lutter contre l'échec scolaire ?* Bruxelles : De Boeck.
- Croizet, J.-C. (2001). *L'intelligence et ses déterminants : perspectives comportementales*. Habilitation à diriger des recherches, non publiée, Université Blaise Pascal, Clermont-Ferrand.
- Dépret, E., & Filisetti, L. (2001). Juger et estimer la valeur d'autrui : des biais de jugement aux compétences sociales. *L'orientation Scolaire et Professionnelle*, 30(3), 297-315.
- Dompnier, B., Pansu, P., & Bressoux, P. (2006). An integrative model of scholastic judgments: pupils' characteristics, class context, halo effect and internal attributions. *European Journal of Psychology of Education*, 21(2), 119-133.
- Dubois, N., & Beauvois, J.-L. (2001). Désirabilité et utilité : deux composantes de la valeur des personnes dans l'évaluation sociale. *L'Orientation Scolaire et Professionnelle*, 30(3), 391-405.
- Dubois, N., & Le Poutier F. (1991). Internalité et évaluation scolaire. In J.-L. Beauvois, R.-V. Joule, & J.-M. Monteil (eds), *Perspectives cognitives et conduites sociales* (vol. 3). Cousset, Fribourg (Suisse) : DelVal.
- Duru-Bellat, M. (1990). Une orientation orientée. *Le Monde l'Education*, juillet-août.
- Duru-Bellat, M. (2002). *Les inégalités sociales à l'école : genèse et mythes*. Paris : PUF.
- Duru-Bellat, M., Jarousse, J.-P., & Mingat, A. (1992). De l'orientation en fin de 5^e au fonctionnement du collège, vol. 3. *Les Cahiers de l'IREDU*, 45.
- Duru-Bellat, M., & Mingat, A. (1988). De l'orientation en fin de 5^e au fonctionnement du collège, Vol. 2. *Les Cahiers de l'IREDU*, 45.
- Dusek, J. D., & Joseph, G. (1983). The bases of teacher expectancies: a meta-analysis. *Journal of Educational Psychology*, 75(3), 327-346.
- Eccles, J., & Wigfield, A. (1985). Teacher expectations and student motivation. In J. Dusek (Ed.), *Teacher expectancies* (pp. 185-226). Hillsdale, NJ : Erlbaum.
- Engel, P. (1996). *Philosophie et psychologie*. Paris: Folio.
- Felouzis G. (1997). *L'efficacité des enseignants*. Paris : PUF.
- Filisetti, L. (2005). *Politesse et responsabilité sociale comme aspects importants de la compétence sociale à l'école : nouvelles perspectives concernant les causes et les conséquences des comportements socialement approuvés*. Thèse de doctorat en Psychologie Sociale, non publiée, Université Pierre Mendès France, Grenoble.
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. New York: McGraw-Hill.
- Good, T. L. (1987). Two decades of research on teacher expectations : findings and future directions. *Journal of Teacher Education*, 38(4), 32-47.
- Good, T. L., & Brophy, J. E. (2000). *Looking in classrooms* (8^e éd.) New York : Longman.
- Heider, F. (1958). *The psychology of interpersonal relations*. Hillsdale NJ : Lawrence Erlbaum Associates.
- Hoge, R. D., & Butcher, R. (1984). Analysis of teacher judgments of pupil achievement levels. *Journal of Educational Psychology*, 76(5), 777-781.
- Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement : a review of literature. *Review of educational research*, 59(3), 297-313.
- Jarlegan, A. (1999). *La fabrication des différences : sexe et mathématiques à l'école élémentaire*. Thèse de doctorat en Sciences de l'Education non publiée, Université de Bourgogne, Dijon.
- Jussim, L. (1989). Teacher expectations : self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology*, 57(3), 469-480.
- Jussim, L., & Eccles, J. S. (1992). Teacher expectations II : construction and reflection of student achievement. *Journal of Personality and Social Psychology*, 63(6), 947-961.
- Kahneman, D., & Tversky, A., (1982). The simulation heuristic. In D. Kahneman, P. Slovic & A. Tversky (eds), *Judgment under uncertainty : heuristics and biases*. New York: Cambridge University Press.
- Kruglanski A. W. (1990). Conditions for accuracy : general or specific ? In J.-P. Caverni, J.-M. Fabre & M. Gonzalez (eds), *Cognitive biases*. Amsterdam : North Holland.

- Madon, S., Jussim, L., & Eccles, J. (1997). In search of the powerful self-fulfilling prophecy. *Journal of Personality and Social Psychology*, 72(4), 791-809.
- Madon, S., Smith, A., Jussim, L., Russell, D. W., Eccles, J., Palumbo, P., & Walkiewicz, M. (2001). Am I as you see me or do you see me as I am ? Self-fulfilling prophecies and self-verification. *Personality and Social Psychology Bulletin*, 27(9), 1214-1224.
- Monteil, J.-M., & Huguet, P. (2002). *Réussir ou échouer à l'école: une question de contexte ?* Grenoble : PUG.
- Maurice, J.-J. (1996). *Modélisation du savoir-faire de l'enseignant expérimenté : adaptation aux contraintes, anticipation, négociation, pilotage de la classe par les tâches scolaires*. Thèse de doctorat en Sciences de l'Éducation, non publiée, Université Pierre Mendès France, Grenoble.
- Mueller, C. M., & Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of Personality and Social Psychology*, 75(1), 33-52.
- Noizet, G., & Caverni, J. P. (1978). *Psychologie de l'évaluation scolaire*. Paris : PUF.
- Noizet, G., & Caverni, J. P. (1983). Les procédures d'évaluation ont-elles leur part de responsabilité dans l'échec scolaire ? *Revue Française de pédagogie*, 62, 7-14.
- Perry, N. E., & Meisels, S. J. (1996). *How accurate are teacher judgments of students' academic performance?* National Center for Educational Statistics, Working Papers Series. US Department of Education, Office of Educational Research and Improvement.
- Piéron, H. (1969). *Examens et docimologie*. Paris : PUF.
- Prost, A. (1968). *Histoire de l'enseignement en France 1800-1967*. Paris : Armand Colin.
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom : teacher expectations and pupils' intellectual development*. New York : Holt, Rinehart and Winston.
- Smith, A. E., Jussim, L., & Eccles, J. (1999). Do self-fulfilling prophecies accumulate, dissipate or remain stable over time ? *Journal of Personality and Social Psychology*, 77(3), 548-565.
- Smith, J. J., Jussim, L., Eccles, J., VanNoy, M., Madon, S., & Palumbo, P. (1998). Self-fulfilling prophecies, perceptual biases, and accuracy at the individual and group levels. *Journal of Experimental Social Psychology*, 34(6), 530-561.
- Steele, C. M. (1997). A threat in the air : how stereotypes shape the intellectual identities and performance of women and African Americans. *American Psychologist*, 52, 613-629.
- Stevenson, H. W., Parker, T., Wilkinson, A., Hegion, A., & Fish, E. (1976). Predictive value of teachers' ratings of young children. *Journal of Educational Psychology*, 68, 507-517.
- Swann, W. B. (1984). Quest for accuracy in person perception: a matter of pragmatics. *Psychological Review*, 91(4), 457-477.
- Trouilloud, D., Sarrazin, P., Bressoux, P., & Bois, J. (2006). Relation between teachers' early expectations and students' later perceived competence in physical education classes: Autonomy-supportive climate as a moderator. *Journal of Educational Psychology*, 98(1), 75-86.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty : heuristics and biases. *Science*, 185, 1124-1131.
- Van Der Pligt, J. (1996). Judgement and decision making. In G. R. Semin & K. Fiedler (Eds.), *Applied social psychology* (pp. 30-64). London : Sage.
- Yserbyt, V., & Schadron, G. (1996). *Connaître et juger autrui*. Grenoble : PUG.