L'HISTOIRE DE L'EVALUATION SCOLAIRE DES ORIGINES A DEMAIN

Jean CARDINET

RECHERCHES 91.101 - Mai 1991

L'HISTOIRE DE L'EVALUATION SCOLAIRE DES ORIGINES A DEMAIN

Jean CARDINET

CARDINET, Jean. - L'histoire de l'évaluation scolaire des origines à demain. - Neuchâtel : Institut romand de recherches et de documentation pédagogiques, 1991. - 19 p. ; 30 cm. - (Recherches ; 91.101)

évaluation critère d'évaluation instrument d'évaluation sélection historique docimologie psychosociologie interaction relation maître-élève

La reproduction, totale ou partielle, des publications de l'IRDP est en principe autorisée à condition que leur(s) auteur(s) en ai(en)t été informé(s) au préalable et que les références soient mentionnées.

L'HISTOIRE DE L'EVALUATION SCOLAIRE

Jean CARDINET

DES ORIGINES A DEMAIN

Résumé

Le système des notes était cohérent au 19ème siècle, quand la bourgeoisie était en compétition avec les anciennes classes dirigeantes. La sélection scolaire n'a plus de sens aujourd'hui, où la société a besoin de développer au maximum les compétences de tous. D'ailleurs, aucun moyen ne permet de prédire le succès avec assez de précision pour justifier des décisions à long terme. Même si l'on veut vérifier après coup les apprentissages réalisés par un élève, il est impossible de contrôler l'ensemble des compétences que certifie un diplôme. L'évaluation devrait être avant tout formative, c'est-à-dire qu'elle devrait fournir à l'étudiant les informations utiles pour conduire son apprentissage. Mais suivre ses progrès ne suffit pas. C'est toute l'histoire des interactions du maître et de ses élèves qui détermine ce qu'ils apprennent. L'évaluation est moins un problème de mesure que de dialogue.

DIE GESCHICHTE DER SCHULISCHEN EVALUATION VON DEN ANFAENGEN BIS MORGEN

Jean CARDINET

Zusammenfassung

Das notensystem war im 19. Jahrhundert sinnvoll, als das Bürgertum mit den alten herrschenden Klassen rivalisierte. Die schulische Selektion hat heute keinen Sinn mehr, braucht doch die Gesellschaft die optimale Entwicklung der Fähigkeiten aller. Uebrigens gibt es kein Mittel, das es erlaubt, den Erfolg mit genügender Präzision vorauszusagen, um langfristige Entscheidungen zu rechtfertigen. Selbst wenn man gewillt ist, die von einem Schüler erworbenen Fertigkeiten nachträglich zu überprüfen, ist es unmöglich, die durch ein Diplom attestierten Fähigkeiten vollumfänglich zu kontrollieren. Die Evaluation sollte vor allem formativ sein, das heisst sie sollte dem Studierenden die zu seiner Ausbildung erforderlichen Informationen geben. Es genügt jedoch nicht, seine Fortschritte zu verfolgen. Vielmehr bestimmt der Verlauf der Wechselwirkung Lehrer/Schüler, was dieser lernt. Die Evaluation ist weniger ein Problem der Messung als des Dialogs.

L'HISTOIRE DE L'EVALUATION SCOLAIRE DES ORIGINES A DEMAIN

Jean CARDINET

En guise d'introduction, je voudrais saluer l'esprit européen de ceux qui ont organisé cette journée, et le dynamisme du Conservatoire de Strasbourg. C'est le première fois que je vois une administration française faire appel à des Belges le matin, à des Suisses l'après-midi, pour préparer une réflexion aussi importante. Je trouve que c'est vraiment la marque d'une ouverture d'esprit dont nous sommes, nous étrangers, profondément heureux. C'est pour cette raison que Jean-François Perret et moi sommes venus aujourd'hui. Nous souhaiterions en Suisse Romande pouvoir dialoguer par-dessus la frontière avec les régions avoisinantes. Nous venons donc en voisins, d'autant plus que nous vivons des événements qui nous font sentir la fraternité des peuples européens. Nous avons la chance de vivre des moments historiques et c'est dans cet esprit de retrouvailles que nous sommes venus aujourd'hui.

Je placerai aussi sous le signe de l'esprit européen la conférence que je veux faire, dans le sens que je ne voudrais pas vous présenter des techniques, mais plutôt des idées générales. Je voudrais développer devant vous quelques réflexions un peu historiques, un peu philosophiques, sur l'évolution des idées en matière d'évaluation. Je chercherai à situer les unes par rapport aux autres les différentes conceptions de l'évaluation qui ont eu cours depuis cinquante ans. Je voudrais mettre en évidence une progression, une direction et un sens (et je prends le mot "sens" dans les deux acceptions du terme). Voilà pourquoi je m'en tiendrai à des idées générales: je pense qu'au début de la réflexion que vous voulez conduire, il est important pour vous de faire le tour des possibilités et de repérer les choix qui vous sont ouverts. Vous serez mieux à même ainsi de discuter valablement, en ayant une sorte de base pour vous déterminer.

^{*} Communication présentée au Colloque "Objectifs et Evaluations dans l'Enseignement" qui s'est tenu au C.N.R. de Strasbourg, les 2 et 3 février 1990

Troisième point d'introduction: j'ai hésité quant au titre à donner à mon exposé. Je l'avais d'abord intitulé: "Les contradictions de l'évaluation scolaire", parce que je voulais montrer l'évolution des idées, la progression dialectique (par contradiction et synthèse), des idées sur le sujet. Puis il m'a semblé qu'il valait mieux parler de "L'histoire de l'évaluation", parce que je décrivais, en même temps qu'un développement logique, une progression relativement historique, où l'on pouvait situer diverses étapes chronologiques de la réflexion sur le sujet.

1. Première esquisse historique

Au début des temps modernes, le maître d'école était un travailleur indépendant. Il savait lire et il commercialisait son savoir en établissant un contrat avec les parents qui lui envoyaient des élèves. On peut retrouver aujourd'hui des annonces parues dans le Feuille d'Avis de Neuchâtel au XVIIIème siècle, où des maîtres d'école offraient leurs services en précisant leurs conditions. C'était vraiment une activité économique, et qui était soumise à la concurrence. Enseigner à compter était payé aussi et faisait l'objet d'un contrat supplémentaire.

L'apprentissage de la lecture était décomposé en étapes. première phase, selon le contrat, était d'apprendre les lettres. deuxième phase impliquait d'apprendre à lire des mots, et ensuite Ces étapes étaient contrôlées régulièrement. des phrases. existe encore des gravures représentant cette activité d'évaluation. On voit des élèves qui se présentent l'un après l'autre devant le maître, qui leur demande de montrer ce qu'ils savent Le résultat était important pour le maître comme pour faire. l'élève. Pour ce dernier, le risque était d'être battu des verges. Pour le maître, l'enjeu était de perdre l'élève si les parents n'étaient pas satisfaits, de perdre donc ses clients et son gain Dans les écoles du peuple, l'évaluation se faisait économique. ainsi tout naturellement "par objectifs",

Je voudrais opposer maintenant cette situation à celle des Collèges des Jésuites, où a prévalu très vite dans l'évaluation scolaire une logique toute différente. La stratégie des Jésuites, pour christianiser le monde, était de commencer par christianiser les dirigeants, ceux qui avaient le pouvoir d'influencer ensuite la société. C'est dans ce but qu'ils s'étaient approchés de l'Empereur de Chine. C'est aussi dans l'espoir d'avoir une influence sur les

pays européens qu'ils se chargeaient de former les élites. Pour mener à bien cette tâche, il était logique qu'ils sélectionnent au départ les meilleurs élèves et qu'ils concentrent leurs soins sur les plus capables, ceux qui naturellement monteraient le plus haut dans l'échelle sociale, et auraient ainsi le plus d'influence. La Constitution des Jésuites montre que ces derniers s'efforçaient de classer leurs élèves en six groupes, depuis les meilleurs jusqu'à ceux qui devaient être rejetés. L'idée d'une sélection effectuée annuellement était ainsi prévue dans le règlement. L'Allemagne et la Suisse ont conservé cette échelle de notes allant de 1 à 6. On classait les élèves par ordre de valeur globale, avec l'idée de les comparer les uns aux autres pour sélectionner les meilleurs.

A la fin du siècle dernier, quand a été créée dans toute l'Europe l'école obligatoire, on avait donc le choix entre ces deux systèmes: le système des Jésuites et celui des petites écoles du peuple, centré sur les objectifs pédagogiques et non sur le classement des élèves. Sans doute sous l'influence de la France, c'est le système compétitif qui a été choisi. Je pense que si ce choix a été fait, c'est que la bourgeoisie française, au début de la troisième République, finissait de conquérir le pouvoir. Il s'agissait pour elle de disputer la place aux héritiers des grandes familles qui tenaient encore souvent le haut du pavé. L'école pouvait devenir le moyen de vraiment accéder aux positions dirigeantes de la société, si l'on créait une société méritocratique, un peu comme dans la Chine classique, où c'était par des examens successifs que l'on montait dans la hiérarchie sociale.

Nous vivons maintenant à la fin du XXème siècle, et nous préparons le troisième millénaire. La situation dans toute la société a évolué. Il ne s'agit plus maintenant de former un petit nombre de cadres compétents pour encadrer, diriger de grandes masses d'ouvriers ou de gens peu spécialisés, mais disciplinés. Il faut maintenant, dans une économie libérale, diffuser au contraire les compétences. Il faut qu'un maximum d'individus puissent prendre des initiatives. Toute la situation économique de l'Europe en dépend, par rapport à l'Extrême Orient en particulier. Ce n'est pas par hasard que ce soit au retour d'un voyage au Japon que le Ministre Chevènement ait décidé de conduire 80% des élèves français au baccalauréat.

Il sera évidemment nécessaire, pour y parvenir, d'améliorer la qualité de l'enseignement. Cela demandera des changements dans l'organisation de la scolarité, dans le type de pédagogie utilisé, mais aussi des changements dans la conception de l'évaluation.

On peut, très grossièrement, bien sûr, opposer deux conceptions de la pédagogie: celle de l'enseignement et celle de l'apprentissage. Dans la première, centrée sur l'activité d'instruction, le maître intervient devant un groupe d'élèves. Soumis à cet enseignement identique pour tous, certains d'entre eux apprennent d'avantage, d'autres moins. Les résultats se distribuent selon une courbe de Gauss. L'évaluation consiste alors à placer l'élève dans cette distribution en disant, s'il est dans les meilleurs, que c'est un bon élève, ou en affirmant à l'inverse, s'il est en retard, que c'est un mauvais élève.

Dans la pédagogie future, celle de l'apprentissage, il faudra se centrer sur les comportements de l'élève, les objectifs pédagogiques précisant ce qu'il doit savoir faire. Il faudra, pour que tous les élèves atteignent ces objectifs, que l'enseignement On devra, pour certains élèves, employer telle soit différencié. pédagogie, pour d'autres telle autre pédagogie. Il se pourrait même que le coût et l'effort consentis soient plus grands pour certains que pour d'autres. Il est bien clair que, dans les hôpitaux, on dépense d'avantage pour les personnes âgées que pour les jeunes. De la même façon, il faudrait que sortent de l'école des gens qui soient bien formés, même si le coût devait en être plus élevé pour les élèves défavorisés au départ. L'évaluation qui correspond à cette pédagogie ne nécessite plus de classer les élèves, mais consiste simplement à voir si chaque individu a atteint ou non l'objectif.

Cette présentation tout à fait schématique avait pour but de montrer que ce qui peut paraître futuriste n'est en réalité qu'un retr à une situation antérieure où le maître prenait en charge l'apprentissage de l'élève et était lui-même financièrement intéressé à sa réussite. Le bilan d'apprentissage fondait alors l'évaluation à la fois du maître et de l'élève.

Les nouvelles conceptions de l'évaluation correspondent donc simplement aux nouvelles exigences de la société, qui a besoin maintenant de faire réussir tous les élèves. Ainsi, il n'existe pas de bon ou de mauvais système d'évaluation, mais des systèmes plus ou moins cohérents avec le cadre social dans lequel ils se situent. A l'heure actuelle, conserver le système des notes, en s'en tenant à sa logique purement comparative, n'est plus cohérent avec les nouveaux objectifs de l'école. La compétition n'est pas mauvaise en soi, mais elle a pour résultat de restreindre les possibilités de formation, d'empêcher les gens d'accéder à la culture, ou à la compétence. Elle entre en contradiction avec les besoins de notre société.

J'en arrive maintenant à voir de façon moins schématique et plus respectueuse des faits l'évolution des idées au cours du XXème siècle. Pour annoncer un peu où nous allons, je donnerai mon plan et une vue d'ensemble très globale.

On peut dire que, dans cette évolution, il y a eu trois phases. Dans la première, on a pris conscience des problèmes des notes et du dysfonctionnement du système d'évaluation actuel. Dans la deuxième, on a essayé d'améliorer les choses, par une série de tentatives que je vous détaillerai. Mais ces tentatives, aussi intéressantes soient-elles, ont révélé assez vite leurs limites. Dans la troisième phase, que je considère contemporaine, la conception psycho-sociale de l'évaluation représente comme l'aboutissement de ces déceptions successives. On y accepte le fait que l'évaluation n'est pas objective, et ne peut pas l'être, mais on essaye d'en tirer les conséquences sur le plan pédagogique et social.

2. Les contradictions des notes

2. 1. La docimologie

Le point de départ de la recherche sur l'évaluation scolaire fut une étude critique des notes mises au baccalauréat, étude qui a été faite par Laugier et Weinberg vers le début des années 30, et publiée en 1938. Henri Piéron y avait collaboré, mais n'a publié son ouvrage sur la docimologie qu'en 1963. Ces études révélaient des différences d'appréciation incroyables entre correcteurs. vu d'abord, en étudiant les jurys de bachot, que pour une même épreuve de mathématique, la moyenne des notes données par un jury était de 5,8; celle d'un autre jury était de 9,1 (sur 20). L'écart croissait encore pour les épreuves orales, où l'on trouvait 5 points sur 20 de différence entre un jury et un autre. En conséquence, le pourcentage d'admissibles était pour un jury de 31 %, pour un autre de 53 %. Toute une série d'autres résultats faisaient dire à Piéron: "La réussite au baccalauréat dépend plus de l'examinateur que de l'examiné". On a cherché à améliorer les choses en tenant compte de la note du carnet scolaire. mais était-ce vraiment une

On peut en douter si l'on considère les résultats amélioration? d'une étude faite à l'INOP (Institut National d'Orientation Professionnelle) à Paris en 1969. Les auteurs Reuchlin et Bacher ont utilisé des tests très bien faits, d'une part de français, d'autre part de mathématique, et les ont fait passer dans 400 classes françaises. Ils ont vu que, pour le français (où l'échelle avait 80 points), les moyennes de classes variaient de 37 points. Pour la mathématique, où l'échelle n'avait que 40 points, ils trouvèrent 31 points de différence entre les moyennes des classes extrêmes. Certaines classes se trouvaient ainsi près de zéro et d'autres près du maximum. Quand on regardait les notes, pourtant, avaient la même distribution pratiquement dans toutes les classes. Cela provient de ce que, dans chaque classe, le maître met des bonnes notes et des mauvaises notes sans savoir où se situe sa classe par rapport aux autres.

2.2. L'approche systémique

Devant ces constats, pour tenter tout de même d'y remédier, on a cherché à s'inspirer de l'analyse des systèmes.

2.2.1. L'évaluation comme moyen de régulation

C'était l'époque où l'on commençait à lancer des fusées. Atteindre Mars, par exemple, nécessite d'avoir prévu une trajectoire avant le lancement, en tenant compte des diverses accélérations à l'oeuvre. Ensuite, lorsqu'on essaye d'envoyer la fusée sur cette trajectoire, on doit comparer les déplacements effectifs à ce qui était attendu et corriger immédiatement les poussées des moteurs. Au bout de quelques minutes, on doit faire le point pour préparer des corrections de trajectoire plus importantes. On intervient donc à trois moments différents pour guider la fusée: avant le lancement, pour calculer le déplacement optimum, pendant, pour asservir la fusée sur sa trajectoire, et après, pour la corriger.

On peut travailler de façon exactement parallèle pour guider un élève vers un objectif pédagogique quelconque. On postule une certaine énergie, qui est la motivation de l'élève à apprendre, et l'existence d'activités didactiques, qui servent de moteur à l'apprentissage. Avant l'étude, d'après les caractéristiques de l'élève, on choisit la didactique la plus appropriée. On peut donc ensuite suivre ses progrès et analyser ses difficultés. Sur cette base, on peut introduire une pédagogie correctrice et modifier les activités d'apprentissage. A la fin du trimestre, ou à certaines échéances, on peut regarder où il se situe et, sur cette base, décider de continuer ou de changer de stratégie, (par une réorientation scolaire).

Dans un cas comme dans l'autre, l'évaluation n'est au fond que l'instrument destiné à apporter les informations nécessaires pour corriger le fonctionnement du système. L'évaluation fournit toujours l'un des trois types d'information suivants: données préalables, permettant de prévoir le type d'activité souhaitable, observations simultanées, recueillies par exemple pendant la classe, permettant de corriger tout de suite les incompréhensions de l'élève, bilans ultérieurs, permettant d'intervenir après coup pour corriger.

On peut donc distinguer trois fonctions principales de l'évaluation, que j'appelerai les fonctions de sélection, de régulation et de certification. La première fonde la sélection, parce qu'il s'agit de choisir dans quelle filière on va mettre l'élève, d'après ce qui paraît le plus utile pour lui. La seconde permet la régulation, quand on essaye d'aider l'élève à apprendre. La troisième fonde la certification, l'examen visant à établir un bilan de ce que l'élève a appris. Autrement dit, l'évaluation vise à améliorer soit les conditions d'apprentissage, soit le processus de l'apprentissage, soit le résultat de l'apprentissage.

2.3. Le conflit entre les trois finalités de l'évaluation

Malheureusement, ces distinctions logiques n'ont pas encore été admises par le système scolaire. Dans l'école telle que nous la connaissons, ces trois fonctions sont prises en charge par un système d'information unique, celui des notes. Il n'est pas besoin d'être mathématicien ni logicien pour voir qu'une seule coordonnée ne suffit pas à situer un point dans un espace à trois dimensions.

Il en résulte, dans la réalité scolaire, une série de contradictions, que j'illustrerai par trois exemples réels, bien que caricaturaux. Pour la première finalité, celle de la sélection, je citerai le cas d'un professeur de français, qui mettait une note à ses élèves au bout d'un mois de classe et disait: "J'ai mis mes notes pour l'année; si je devais en changer, c'est que je me serais trompé la première fois". Sa logique était celle du système des Jésuites: la note indique le niveau de compétence globale de l'élève, niveau que l'on suppose stable.

Pour la deuxième finalité, celle du soutien de l'apprentissage, je présenterai un professeur d'allemand qui donnait chaque semaine six exercices à faire et donnait ensuite à ses élèves une note proportionnelle au nombre d'exercices qu'ils avaient traités correctement. La note reflétait donc la qualité du travail effectué chaque semaine.

Pour la troisième conception de la note, celle du bilan, je donnerai l'exemple d'un instituteur qui faisait faire un premier travail, le corrigeait, puis le rendait aux élèves. Ceux-ci voyaient les erreurs qu'ils avaient commises et ils avaient le droit de recommencer leur travail, jusqu'à ce qu'il soit irréprochable. Lorsque ce niveau était atteint, les élèves recevaient la note maximale, puisque le bilan était devenu totalement positif.

Ces trois conceptions, bien qu'ici poussées à l'extrême, sont un peu présentes chez tous les enseignants. Ceux-ci voudraient, à la fois, indiquer le niveau de l'élève, suivre son apprentissage journalier, et faire la différence entre ce qui est acquis et ce qui ne l'est pas. Or les trois exemples montrent bien que ces trois conceptions sont radicalement contradictoires. On ne peut pas exprimer ces trois informations avec un seul chiffre.

2.4. La nécessité d'instruments diversifiés.

La seule conclusion que l'on puisse tirer de cela, c'est qu'il nous faut créer trois types d'instruments différents, si nous voulons obtenir des informations pour prévoir, réguler et certifier.

C'est bien la conclusion qu'a tirée, vers 1980, le Ministère de l'Education de Québec, et c'est la raison pour laquelle il a mis sur pied tout un système d'épreuves répondant à chacune de ces finalités, épreuves prédictives, formatives et sommatives, collectives et individuelles, pour suivre le fonctionnement des élèves, des classes et des régions, etc.

3. Les tentatives de rationalisation

Essayons de voir maintenant ce que donne cette approche, ce qu'on a réussi à faire, en allant au bout de cette logique systémique.

3.1 La sélection rationalisée

Ce qu'il faut bien comprendre à propos de la fonction sélective, c'est que sa logique première n'est pas de voir si l'élève a bien travaillé ou non. On remarque souvent une confusion chez les enseignants en fin d'année, quand ils décident de la promotion d'un Beaucoup d'entre eux croient qu'ils sont en train de juger l'élève: s'il a bien travaillé, ils veulent le promouvoir; s'il n'a pas bien travaillé, ils demandent qu'il redouble. C'est faux. Le Conseil de classe doit examiner s'il est plus favorable pour l'élève de passer à de nouveaux apprentissages, ou bien de reprendre les notions qu'il n'a pas bien assimilées. C'est donc selon une perspective d'avenir, avec un regard tourné vers le futur, en se demandant ce qui se passera si l'on prend telle ou telle décision, qu'il faut décider. Si un élève n'a pas bien étudié le programme de l'année, mais si l'on sait qu'il est brillant, et qu'il sera tout à fait capable de se débrouiller l'année d'après, il n'y a pas de raison de le faire redoubler sa classe, par exemple.

Prendre des décisions dans une perspective d'avenir suppose que les maîtres soient capables de savoir ce qui va se passer pour l'élève, si on lui fait répéter sa classe, ou bien si on le promeut. Or cette condition n'est pas souvent satisfaite dans les classes. même l'inverse qui est le plus souvent vérifié. Des milliers de recherches effectuées dans le monde entier sont là pour le prouver. Je présenterai des résultats trouvés à Lausanne en Suisse, mais i'aurais pu en choisir beaucoup d'autres. Il s'agit d'études qui ont été effectuées sur des milliers d'élèves. On a cherché à déterminer la corrélation entre les résultats obtenus à dix ans, au moment de l'entrée à l'école secondaire, et les notes reçues année après année, jusqu'à six années après. On a essayé différentes mesures, différents tests, différentes façons de combiner les épreuves, etc., mais la tendance générale est toujours la même: plus l'échéance est lointaine, plus la prédiction est mauvaise.

Pour faire comprendre ce que représente la validité prédictive obtenue, on peut estimer sur une base statistique les proportions de décisions erronées qui en résultent. A l'examen d'entrée à l'école secondaire, qui existe malheureusement encore souvent en Suisse, on accepte généralement le tiers supérieur des élèves. Admettons qu'il y ait dans la population aussi un tiers d'élèves capables de réussir. Avec la validité à long terme obtenue en pratique, il est facile de calculer qu'il y a plus d'élèves qui échouent, parmi ceux qui sont admis, que d'élèves qui réussissent. Par contre, parmi ceux qui seraient capables de réussir, il y a plus d'élèves qu'on n'admet pas que d'élèves qu'on admet. C'est la preuve d'un gâchis social terrible, à une époque où l'on aurait pourtant besoin de toutes les compétences.

La conclusion de cette discussion de la fonction prédictive, c'est qu'il n'est pas souhaitable de vouloir, dès le début des études, évaluer les capacités futures et sélectionner sur cette base. Ceci ne veut pas dire qu'on ne sache rien des élèves. Les enseignants peuvent certainement citer nombre de cas où leur prédiction s'est avérée juste. Notre capacité de prédiction est effectivement bien supérieure au hasard. Mais elle n'est pas suffisante pour rendre la sélection économiquement et socialement rentable, compte-tenu de ses erreurs inévitables.

Le modèle de l'examen d'entrée à l'Ecole Polytechnique n'est donc pas le modèle qu'il faut prendre pour rationaliser l'évaluation scolaire. Il faut plutôt donner aux élèves la chance de se former, et en fonction de leurs résultats, décider de leur certification. C'est après l'apprentissage que l'on doit conduire l'évaluation, et non pas avant.

3.2 La certification rationalisée

Alain Lamée a bien clarifié, plus haut, ce qui faisait une bonne évaluation certificative, en parlant des trois cohérences nécessaires entre les objectifs, la formation et l'examen. C'est effectivement cette cohérence qui constitue le premier critère de qualité pour une évaluation certificative.

J'en ajouterai cependant un deuxième, dont il n'a pas parlé, mais avec lequel je sais qu'il serait parfaitement d'accord, c'est la portée suffisamment générale de l'examen. A quoi servirait une évaluation très liée à l'apprentissage, mais qui ne porterait que sur des exercices scolaires, sans rapport avec autre chose? Qui s'intéresserait, par exemple, à un examen de mécanicien automobile, même parfaitement fidèle, s'il ne prenait en compte qu'un seul type de panne? Il faut que le mécanicien soit capable de réparer toutes les pannes possibles, et pas seulement celles qu'on lui a présentées à l'examen. Il faut que les compétences sociales qui sont certifiées aient un vaste champ d'application, qu'elles ne soient pas uniquement scolaires.

Or cela n'est pas aussi simple qu'on le pense. Prenons un objectif qui paraît parfaitement clair: savoir calculer l'aire d'un triangle. En fait, si l'on demande aux élèves de nous donner la formule pour ce calcul, on obtient 85% de bonnes réponses. Mais si on leur donne seulement les dimensions du triangle, en leur laissant faire le calcul, il n'y a plus que 70% de bonnes réponses. Ce qui est très intéressant, c'est que, si on leur donne le même triangle, avec une équerre et une régle graduée, en leur demandant de calculer son aire, la réussite descend de 70 à 46 %. En prenant un triangle dont la base n'est pas horizontale, ou dont la hauteur ne coupe pas la base, il n'y a plus que 20 % de réussite. Enfin, le fait de présenter un triangle la pointe en bas est pour beaucoup d'élèves une source de difficulté telle qu'ils ne le reconnaissent plus comme un triangle et ne sont donc pas capables de lui appliquer les formules d'aire. Ainsi, selon la façon dont on pose le problème, on peut faire passer le taux de réussite à l'objectif "savoir calculer l'aire du triangle" de 85% à peut-être 10%.

Quand peut-on dire alors que l'objectif est atteint de façon générale? Et s'il n'est pas atteint de façon générale, quelle est l'utilité d'une évaluation par objectif?

Il faut réfléchir à cette source d'erreur dans la certification et essayer d'en tirer les conséquences. On voit que le maître est pris entre Charybde et Scylla. Ou bien il veut une mesure précise, et pour cela il faut qu'il opérationalise très étroitement son objectif: il va obtenir quelque chose de répétable, mais valable uniquement pour, disons, le calcul de l'aire d'un type de triangle. Un tel objectif n'aurait aucun intérêt pratique. Ou bien il veut une formulation générale de l'objectif, mais comme le résultat de l'élève dépendra néanmoins de l'opérationalisation choisie, le certificat n'aura plus aucune valeur.

Alors comment font les enseignants? M. Guislain l'a esquissé dans sa conclusion. Ils n'ont pas découvert la solution parfaite, mais ils font des choix: ils décident de proposer un certain nombre d'exercices et ils pilotent la classe d'après les erreurs qu'ils observent. Ils mènent les élèves à la réussite pour ce type particulier d'exercices.

Ceci signifie que l'évaluation qu'ils peuvent faire est toujours limitée à un certain savoir scolaire et n'a pas une portée sociale générale. La valeur donnée aux diplômes n'est donc pas justifiée par les études métriques précises que l'on peut en effectuer. Ce qu'un enseignant peut réaliser convenablement, c'est simplement certifier ce qu'il a vu, en disant: "mes élèves sont capables de faire précisément ceci..."

Ce ne serait pas aussi limité qu'on le pense. Pour l'illustrer, je citerai une parole de Churchill, qui avait été un mauvais élève et disait plus tard: "Je me demande pourquoi les examinateurs ont toujours insisté pour m'interroger sur ce que je ne savais pas, au lieu de me demander ce que je savais".

Cet humour anglais est plein de bon sens. Il faudrait rendre compte des savoirs particuliers des élèves plutôt que de vouloir certifier la maîtrise d'objectifs trop généraux. Les anglais et les écossais sont en train de modifier leur évaluation certificative pour essayer d'aller dans ce sens. Ils décernent maintenant, à la fin des études secondaires, non pas un bachot, mais un profil, qui dit ce que l'élève est capable de faire en langue maternelle, en mathématique, etc., et même dans des savoirs non scolaires: "il a conduit un groupe de jeunes à faire felle activité, il a monté telle exposition ou tel concert, etc.". Avec ce profil, l'élève peut se présenter devant des employeurs et faire valoir certaines compétences, quel que soit son niveau général. Il n'y a plus d'échec à l'examen final.

Il est intéressant de noter, en guise de transition, que cette transformation de l'évaluation certificative la rapproche de l'évaluation formative. L'élève est conduit vers certains savoirs qui ne sont pas nécessairement les mêmes dans toutes les classes: l'essentiel est que l'élève progresse; le maître rendra compte plus tard de ce qu'il sait faire.

3.3. La régulation rationalisée

Alain Lammé a très bien défini l'évaluation formative et il me suffira d'illustrer cette démarche à l'aide d'un exemple: pour moi, l'évaluation formative type, c'est la correction d'un exercice. Un élève a traité un problème de mathématique; le maître vérifie sa démarche et relève, s'il s'est trompé, quel type d'erreur il a fait; il lui explique alors pourquoi c'était faux, et ce qu'il faudrait faire à la place. Je pense que c'est exactement la même chose en musique. Remarquons bien ce qui caractérise cette évaluation formative: on ne dit pas quelque chose sur l'élève, mais on dit quelque chose à l'élève. Le retour d'information lui apprend quelque chose qu'il ne connaissait pas sur la discipline.

Il faut donc rechercher toutes les occasions possibles d'assurer ce feedback, ce retour d'information à l'élève. Le contrôle peut être assuré par des corrigés-types, des exercices autocorrectifs, des corrections mutuelles ou en groupes, des correspondants, un ordinateur, etc.

Tous les facteurs qui influencent l'apprentissage méritent d'être pris en compte dans ce retour d'information, qui doit rendre l'élève conscient de la démarche qu'il utilise, et de la façon dont il peut l'améliorer. Sa méthode est-elle correcte? Son attitude est-elle favorable? etc.

Le problème, si l'on cherche à tenir compte de tous ces facteurs individuels, comme il semble qu'on devrait le faire, c'est que l'on est bien vite nové dans une masse d'informations que l'on ne peut plus exploiter. On se trouve devant une impossibilité et qualitative et quantitative. Qualitative, dans le sens qu'on ne sait pas quelle signification donner à chaque information. Vous observez une baisse de motivation chez tel élève; c'est très important, mais quelle en est la cause? Même si vous la connaissiez, il n'est pas certain que vous puissiez intervenir, s'il s'agit d'un problème familial, par exemple. La science didactique n'est pas suffisamment élaborée pour qu'on sache exploiter toutes les informations qu'on reçoit. D'autre part, le problème est également quantitatif: même avec un ordinateur, le maître n'en finirait pas de collecter toutes les données qui seraient en principe pertinentes à la gestion de l'apprentissage. Quand aurait-il le temps de gérer tout cela?

Nous avons parcouru ensemble les étapes successives de la réflexion sur l'évaluation scolaire. Nous étions partis du constat des problèmes que posait l'utilisation des notes. En réfléchissant à leur rôle, nous avons vu qu'elles avaient trois fonctions et qu'il était donc logique de créer des instruments pour chacune de ces fonctions.

On a essayé d'améliorer la fonction sélective, mais on a vu qu'il n'était pas possible de prévoir de façon suffisamment précise, et qu'il fallait plutôt certifier après-coup les apprentissages réalisés.

On a essayé d'améliorer la certification, mais on a vu qu'il n'était pas possible de porter des jugements suffisamment généraux et qu'il fallait plutôt conduire la formation aussi loin qu'on le pouvait, et dire ensuite ce que les élèves avaient appris de particulier.

Mais on voit, en cherchant à guider cet apprentissage, que l'évaluation formative pose, elle aussi, des problèmes insurmontables. Est-ce qu'il faut abandonner nos efforts d'amélioration et de rationalisation de la pédagogie?

Je crois qu'on peut sortir de cette situation à première vue bloquée, à condition de changer de modèle scientifique, c'est-àdire de modifier notre angle d'attaque.

Il me semble que ce qui caractérise toutes les tentatives qui viennent d'être présentées, c'est qu'elles considèrent toujours la pédagogie comme une technologie. Elles cherchent, de l'extérieur, à influencer l'élève, en changeant l'enseignement, l'évaluation, etc. Toutes ces tentatives ne prennent pas en compte le vécu de l'élève, ses motivations, sa vision subjective des choses. A mon avis, c'est de ce côté qu'il faut chercher maintenant, parce que les philosophes nous font bien remarquer qu'apprendre est une action purement individuelle. On ne peut pas apprendre pour son élève. On peut seulement lui préparer un milieu favorable.

Ce nouveau modèle, qui combine l'aspect subjectif et l'aspect objectif, c'est celui de la psychologie sociale. Nous allons voir, dans la dernière partie, qu'il semble ouvrir des perspectives pour l'évaluation, et même pour une autre conception de l'éducation.

4. L'approche psycho-sociale et ses conséquences

4.1. Travaux expérimentaux

Un courant de pensée s'appuie toujours sur des précurseurs, mais je crois légitime de dire que les conceptions psycho-sociales de l'éducation qui nous intéressent découlent de la recherche initiale de Anne-Nelly Perret-Clermont. Il vaut la peine de la décrire un peu en détail.

Anne-Nelly Perret-Clermont a cherché à dépasser les travaux de Piaget, en montrant comment se faisait effectivement le passage d'un stade logique au suivant, sous quelles influences sociales et dans quelles conditions. Pour cela, elle a repris le problème, bien étudié par lui, de la conservation des liquides. Si l'on donne du sirop à de jeunes enfants dans un verre étroit, ou bien dans un verre large, ils se repèrent uniquement au niveau du liquide, et ils choisissent le verre étroit où le sirop monte plus Après cinq ans, ils acquièrent la conservation et tiennent compte aussi du volume de liquide. Mais ils passent, entre les deux, par une phase d'incertitude. Anne-Nelly Perret-Clermont a constitué des groupes d'enfants qui en étaient à ce stade intermédiaire, leur a donné une tâche à faire en commun et leur a dit ensuite: "Vous avez bien travaillé; vous avez tous droit à du sirop; partagez vous bien cette bouteille; il ne serait pas juste que certains en aient moins que les autres". Comme les verres étaient évidemment de formes différentes, les enfants se sont mis à Motivés par le besoin de justice distributive, ils ont discuter. argumenté dans chaque groupe pendant 20 minutes en moyenne. Au bout de ce temps, tous les enfants étaient devenus conservants. c'est-à-dire qu'ils avaient appris à tenir compte du volume. questions posées à deux reprises ensuite ont permis de vérifier que l'apprentissage était bien stabilisé, alors que dans des groupes de contrôle, où l'on avait seulement expliqué la bonne réponse aux enfants, l'argumentation était restée sans effet.

Si un changement aussi fondamental peut être obtenu en si peu de temps, peut-on encore parler de stade logique, comme le faisait Piaget? Une autre expérience est venue récemment démontrer que cette notion était à abandonner, parce qu'on pouvait prouver que des élèves qui avaient atteint un stade dans une certaine situation sociale n'étaient pas au même niveau dans une autre.

La recherche a été faite au Brésil, où beaucoup d'enfants vendent ce qu'ils peuvent dans la rue, notamment des noix de coco. La psychologue qui menait cette étude leur en achetait, mais toujours un nombre fractionnaire, par exemple trois et demie. Les enfants calculaient généralement le prix correctement. Le lendemain, elle se présentait à leur école et leur posait par écrit le problème qu'ils avaient résolu correctement la veille. Presque toujours, les enfants échouaient.

On voit bien là que le savoir ne dépend pas seulement de l'élève. Comme nous l'avons vu tout à l'heure à propos du triangle, le savoir est lié à la fois à l'élève et au mode de questionnement. Il n'y a pas de niveau de connaissance absolu. La performance est une fonction à plusieurs variables. Cela n'a pas de sens de dire que tel élève maîtrise ou ne maîtrise pas telle notion. C'est l'interaction du maître et de l'élève qui aboutit à la manifestation de telle ou telle performance.

4.2. Conséquences théoriques

Une telle affirmation remet en cause fondamentalement toutes les conceptions de l'évaluation proposées jusqu'ici, puisqu'elle rejette l'idée qu'il existe une valeur vraie de la compétence. Elle oblige aussi à modifier notre conception des objectifs pédagogiques, qui ne sont plus donnés une fois pour toutes au départ, mais qui évoluent selon les interactions du maître et de l'élève, au cours de la vie de classe. Selon les événements qui ont marqué l'histoire de la classe, le maître arrive à tel ou tel résultat et finalement adapte son ambition à cette réalité extérieure à lui. C'est pourquoi les maîtres répugnent à s'échanger leurs épreuves et se trouvent très mal à l'aise lorsqu'ils doivent faire passer des épreuves communes à plusieurs classes. Les épreuves, comme les objectifs et comme la progression didactique, doivent varier d'une classe à l'autre.

L'évaluation, par conséquent, doit être adaptée, autant qu'il est possible, à chaque situation. Reprenons l'exemple de la correction des exercices. Ce devrait être un exemple de communication entre deux personnes: le maître qui essaye de comprendre l'erreur que fait l'élève, et l'élève qui essaye d'assimiler les critères du maître et de comprendre ce qu'il doit

rectifier. L'apprentissage progresse dans la mesure où l'élève arrive à savoir ce que le professeur attend de lui, et c'est à partir du dialogue évaluatif qu'il peut le comprendre.

L'intervention corrective, qui était tout le problème de la pédagogie classique, n'est plus envisagée de la même façon. On ne savait pas comment, de l'extérieur, influencer l'élève. Mais si on s'aperçoit qu'en réalité les objectifs et leurs critères sont progressivement transmis par ce dialogue, la pédagogie corrective n'est plus l'affaire du maître, mais de l'élève. C'est ainsi que la psychologie sociale conduit à repenser toute la pédagogie, dans l'optique de l'autoévaluation.

Je vais terminer par des exemples un peu plus concrets, pour mieux expliquer ce que j'entends par "communiquer aux élèves les objectifs et les critères pour qu'ils prennent eux-mêmes en charge leur adaptation".

4.3. Exemples pratiques

Le premier exemple est celui de Mme Georgette Nunziati qui enseignait à Marseille, dans des milieux défavorisés. Pour faire apprendre à rédiger une dissertation, elle a commencé par montrer à ses élèves des dissertations des années précédentes, déjà corrigées par elle. Ils pouvaient voir en marge les erreurs qu'avaient faites les élèves précédents. Ils pouvaient alors cataloguer les types d'erreurs et commencer à se donner quelques points de repères. Sur cette base, ils cherchaient les critères permettant de repérer ce qui faisait une bonne dissertation (il faut qu'elle ait trois parties, une introduction et une conclusion, des transitions, etc.). Cela ne veut pas dire qu'ils savaient le faire, mais qu'ils se représentaient correctement le produit à obtenir.

Le plus dur restait encore à faire. Alain Lammé a bien expliqué la différence entre les objectifs relatifs au produit et les objectifs relatifs aux processus. Il a montré que l'essentiel était de faire comprendre ces derniers. Jean-Jacques Bonniol, Professeur à l'Université d'Aix-Marseille a souligné, le premier, que l'essentiel du travail des maîtres se situait là: analyser avec les élèves les étapes par lesquelles ils devaient passer pour finalement être capables d'obtenir un produit satisfaisant. Ceci implique d'analyser tout le processus de travail et de pouvoir décrire, pour chaque étape, ce qui fait qu'un élève procède correctement ou non. L'enseignement devient une discussion avec

la classe pour voir comment chacun s'y prend et le résultat qui en découle. Une bonne partie du trimestre se passe à analyser les démarches d'apprentissage.

Une fois que ceci est au clair pour tout le monde, il est possible d'établir des contrats individuels, pour soutenir les élèves dans leur progression. Un collègue de Mme. Nunziati, M. Michel-Paul Vial, met par écrit ce qui est demandé et par quelles étapes les élèves peuvent le faire. C'est ce qu'il appelle la carte d'étude. Il les aide ensuite à établir une fiche individuelle, qui sert de contrat. Chaque élève analyse ce qu'il sait déjà faire (qu'il ne vaut pas la peine de recommencer), ce qui dépasse ses possibilités (qu'il ne faut pas aborder tout de suite), mais aussi ce qui lui paraît accessible (où l'on peut choisir un objectif à atteindre). élève se fixe donc sa prochaine étape, avec l'accord de son maître. A l'échéance suivante, l'élève rend compte de ce qu'il a fait, l'évalue et propose un nouveau contrat. Ainsi l'élève est motivé et capable de voir par lui-même, avant de rendre son travail à son professeur, ce qui ne va pas et ce qu'il lui faudra faire pour progresser.

On arrive là à une conception nouvelle à la fois de la pédagogie et de l'évaluation, et c'est normal que les deux soient liées puisque l'évaluation formative est le circuit d'information qui permet de réguler la didactique.

Ainsi tous les travaux que l'on a poursuivis sur l'évaluation depuis vingt ans n'ont pas été inutiles. Ils ont conduit, c'est vrai, à explorer un certain nombre de cul-de-sacs. Mais en même temps, par approximations successives, ils ont permis de s'approcher du but et d'arriver à des solutions plus positives.

5. Conclusion

En guise de conclusion, je me résumerai simplement. Nous avons vu l'évolution du rôle social de l'école et comment l'évaluation a été cohérente, en général, avec ces finalités successives. Mais à l'heure actuelle, le système des notes n'est plus cohérent avec les besoins de notre société. C'est pourquoi il serait utile de le modifier pour avoir une évaluation qui soit vraiment utile à l'apprentissage.

Pour clarifier la direction à prendre, nous avons analysé les fonctions de l'évaluation, afin de sortir de cette ambiguïté des

notes qui veulent tout mesurer à la fois et qui confondent tout. Pour se dégager de ces confusions, on peut, comme les Québécois, essayer de préparer des épreuves de toutes sortes. C'est très bien, mais cela ne résout pas tous les problèmes. Même si un bon instrument permet à un enseignant de situer sa classe dans l'ensemble des autres classes, ou l'aide à diagnostiquer les difficultés de certains de ses élèves, il ne peut lever les impossibilités majeures de l'évaluation: on ne peut pas prédire à long terme; on ne peut pas certifier de façon suffisamment générale pour que les diplômes aient la valeur que la société leur attribue; on ne peut pas vraiment aider les élèves, les guider, comme les contrôleurs de vol qui suivent les avions avec un radar. Il faut abandonner ce mythe de l'évaluateur à son tableau de bord.

En réalité, la vie de la classe est donnée par la suite des interactions maître-élèves et l'évaluation dépend de la qualité de leur communication. Le maître-mot est finalement celui de: "Dialogue" et c'est pourquoi il convient de s'arrêter sur lui: l'évaluation est moins un problème de mesure que de dialogue.