



L'ÉVALUATION DES APPRENTISSAGES

TRENTE ANS DE RECHERCHES ET DE DÉBATS

CHARLES HADJI

Professeur de sciences de l'éducation
à l'IUFM de Grenoble,
auteur de *L'Évaluation, règles du jeu*,
Paris, ESF, 1995.

L'évaluation est devenue l'une des questions majeures de la formation et de l'enseignement. On parle d'évaluation formative quand elle permet à l'apprenant d'améliorer ses démarches d'apprentissage et de renforcer ses connaissances et ses compétences. Charles Hadji explique pourquoi l'évaluation formative est un idéal, un guide pour l'action pédagogique.

DANS LE MONDE DE L'ÉDUCATION et de la formation, le concept d'évaluation a fait l'objet, depuis une trentaine d'années, d'un engouement qui est allé en s'accroissant. On prône même, aujourd'hui, une véritable « culture de l'évaluation », c'est-à-dire « l'émergence d'un état d'esprit, d'habitudes, de réflexes même, grâce auxquels sont appréciés régulièrement l'état courant et les actions conduites pour, en retour, infléchir ces dernières si nécessaire ». (1)

Quels sont donc les pouvoirs de l'évaluation, pour ce qui concerne la transmission des savoirs et des compétences ? Que sommes-nous en droit d'attendre d'elle, aujourd'hui ?

La première difficulté est de pouvoir caractériser les pratiques d'évaluation : un même terme convient-il pour désigner des activités diverses, telles que noter, comparer, s'interroger sur le sens, et portant sur des objets aussi différents que des élèves, des

connaissances, des actions ou des systèmes ? Prendre l'évaluation pour objet, c'est tout autant poser la question du sens des pratiques abordées (approche compréhensive), que tenter de décrire et d'analyser ces pratiques (approche descriptive) et, enfin, tenter de contribuer à l'amélioration de ces pratiques (approche prescriptive). Ces trois grandes orientations peuvent offrir une entrée commode pour tenter un bilan des acquis de la recherche des dernières décennies.

Les formes de l'activité évaluative

L'évaluation peut être implicite : par exemple, l'autoélimination d'un grand nombre d'étudiants qui ne se présentent même pas à l'examen. L'évaluation implicite, précédant les multiples décisions que prend tout sujet humain, est ainsi un « *acte social universel* », mais qui est demeuré, en fonction de sa dimension cachée, relativement peu étudié. Seule l'évaluation instituée constitue « *un acte délibéré et socialement organisé* » (2) qui s'exprime essentiellement selon deux directions.

La première s'oriente plutôt vers le quantitatif (estimer la valeur d'un bijou). La seconde privilégie un point de vue qualitatif : apprécier la valeur d'une action, c'est dire si elle peut être jugée bonne ou non. Cette dichotomie met également en jeu le domaine de référence de l'acte évaluatif. Quand l'évaluation porte sur un comportement ou sur une production, elle est tentée de les appréhender en tant que performances. On pourra se contenter de mesurer la performance, ce qui peut être facile lorsqu'il s'agit de performances physiques ou de réalisations matérielles, mais déjà beaucoup plus difficile pour une production intellectuelle. C'est pourquoi on peut apprécier ce qui aura été tout d'abord mesuré, ou estimé, en le situant par rapport à un barème de performances jouant le rôle de « norme ». Les réponses, les performances d'un sujet sont jugées par comparaison à celles d'un groupe de référence. C'est la distribution des résultats de la population de référence, à laquelle appartient l'apprenant, qui donne sens à la performance individuelle. Mais si ce sens est recherché, non pas dans l'affectation d'un rang, mais dans la comparaison avec un critère de contenu, on parlera d'évaluation à « référence critériée ». On situe alors le sujet par rapport à une « performance cible », qui est de l'ordre de l'objectif à atteindre. L'évaluation a ici pour but de fournir des informations sur l'écart à la cible (comportement visé) et le degré de maîtrise des acquisitions attendues.

Evaluation sommative et évaluation formative

Les pratiques d'évaluation sommative sont celles par lesquelles on s'efforce de faire un bilan des acquisitions d'un élève, ou

d'une action de formation, à la fin d'un apprentissage ou d'une formation. On fait la somme de ces acquisitions. Souvent une telle évaluation est certificative : le bilan effectué permet de délivrer (ou non) un diplôme, tel que le baccalauréat. Ce sont ces pratiques d'examen qui ont été l'objet premier de la docimologie, précisément définie par Henri Piéron comme « *la science des examens* », à laquelle il donne pour objectif général, en 1963 (3), de contribuer au perfectionnement des méthodes destinées à assurer les « classements humains ». Car la scientificité des analyses docimologiques est vue comme un garant de l'efficacité future des pratiques. Se centrant sur l'examen, la docimologie est amenée, en fonction même du résultat de ses recherches, à remettre en cause les formes et les modalités de l'examen traditionnel. Car ce que la « science des examens » révèle en premier lieu, c'est que l'examen n'est pas une science !

Partie avec l'ambition d'assurer la mesure, la docimologie montre abondamment que l'activité de l'examineur n'est en rien comparable à celle du physicien mesurant un phénomène naturel. La notation des copies, par exemple, est toujours aléatoire. Des examinateurs différents n'évaluent pas de la même façon les mêmes objets, n'utilisent pas de manière identique les échelles de notes, et produisent des jugements qui n'ont pas de stabilité dans le temps. Des travaux ultérieurs viendront préciser certains mécanismes à l'origine des errements constatés (voir encadré p. 68) (4). Trois principales causes de biais ont été étudiées : la possession *a priori* d'informations sur le producteur de la copie, l'ordre des copies, la dynamique du recueil de l'information dans la correction d'une copie. Plutôt que de succomber au mythe de la note « vraie », ne vaut-il pas mieux se centrer sur ce à quoi sert, (ou devrait servir !) l'évaluation, quand elle s'intègre au processus de formation ?

Ainsi se justifie la notion d'évaluation formative, proposée en 1967 par Michael Scriven (5) à propos de l'évaluation des *curricula*, puis étendue en 1971 par Benjamin S. Bloom (6) à l'évaluation des élèves et des étudiants. On désigne comme formative une évaluation qui joue un rôle de régulation des apprentissages. Qu'il s'agisse de programmes d'études, ou d'apprentissages scolaires, l'objectif essentiel est de les améliorer. L'évaluation contribue, par les informations qu'elle apporte, à une bonne régulation du processus concerné. S'il s'agit d'un processus d'apprentissage, le but est de guider l'élève, de lui permettre de reconnaître, de comprendre et de corriger lui-même ses erreurs (fonction correctrice), de l'informer sur les étapes franchies ou non, tout en informant l'enseignant sur les effets réels de son action pédagogique (fonction régulatrice). L'évaluation décrite comme formative présente ainsi trois caractères principaux. Elle s'inscrit dans la durée d'une action pédagogique (intégration au processus). Elle se met délibérément au service de la fin poursuivie

dans le contexte de sa mise en œuvre (subordination à l'action éducative et de formation). Elle se propose moins de dire la réalité telle qu'elle est que de l'aider à devenir ce qu'elle pourrait être (visée transformatrice). De l'évaluation sommative à l'évaluation formative, il y a toute la distance qui sépare un paradigme de la saisie objective (attribuer la note vraie) d'un paradigme de l'aide au développement (accompagner le mouvement de progression en l'éclairant).

Mais l'ambition fondatrice de l'évaluation formative – qui est de se centrer sur l'élève pour mieux le connaître, afin de mieux l'aider à apprendre – ne s'est pas vraiment traduite dans les faits. L'évaluation formative correspond à des pratiques qui sont davantage prônées, et rêvées, que réelles. Le modèle de l'évaluation formative est un modèle idéal. Et sans doute parle-t-on d'autant plus de cette forme de pratique que l'on déplore que tant de formateurs ou d'enseignants se refusent à l'adopter...

En effet, si l'essentiel est d'aider l'élève à repérer, analyser et comprendre ses erreurs pour ne plus les commettre, l'évaluateur doit disposer d'un modèle théorique rendant possible cette analyse et, plus généralement, d'un modèle du fonctionnement cognitif de l'élève. Les progrès de l'évaluation sont de ce point de vue suspendus aux progrès de la science cognitive. Il devient nécessaire de maîtriser des savoirs nombreux et complexes pour bien évaluer. Le discours sur l'évaluation formative n'est précisément qu'un discours pédagogique. Son intérêt est sans doute moins de décrire ce qui doit être absolument réalisé, que d'offrir une perspective pour penser la pratique (7). En troisième lieu, l'objectif de l'évaluation formative est de progresser vers la transparence, qu'il s'agisse des comportements analysés aussi bien que des procédures utilisées.

Le sens des activités d'évaluation

Mais un excès de transparence ne risque-t-il pas d'être dommageable pour le sujet ainsi mis à nu ? Faut-il vouloir tout analyser, et tout dévoiler ? Enfin, de cette manière, ne va-t-on pas perdre son temps ? Si l'évaluation est au service de la pédagogie, au lieu de passer son temps à évaluer, ne faudrait-il pas mieux agir ? C'est la question du sens même de l'acte d'évaluation qui se trouve posée. Cette question a été au cœur des recherches « compréhensives ».

L'évaluation est un acte qui se situe dans le cadre d'un rapport social de communication. On peut alors décrire les différents usages sociaux de cette activité. Mais la question a d'emblée une dimension théorique, ou analytique. En effet, pour évaluer il faut soit posséder, soit construire à cet effet un référent, constitué pour l'essentiel par un ensemble de critères qui spécifient ce que l'on estime être légitimement en droit d'attendre de l'apprenant ; et lire ce réel à

Des notes baladeuses

Noter un devoir est un art difficile qui peut laisser une place importante à la subjectivité. Il y a presque vingt ans, deux chercheurs ont mené une série d'expériences qui mettent en évidence un certain nombre d'effets caractéristiques. Ces travaux, qui font toujours référence, doivent néanmoins être aujourd'hui relativisés, précisément parce que l'attention portée sur ce problème est plus importante qu'autrefois.

Il y a, d'une part, un effet d'ordre. Si l'on fait corriger des copies dans un certain ordre par un enseignant et dans l'ordre inverse par un autre enseignant, on constate que les évaluateurs ont tendance à surévaluer les copies corrigées en premier et à sous-évaluer les copies corrigées en dernier.

Il y a également un effet d'assimilation, qui consiste à rapprocher une note d'une note antérieure.

Ainsi, on demande à des enseignants d'évaluer six copies de version anglaise, de niveau à peu près équivalent. Mais, sur chacune figure une note, censée avoir été obtenue par le même élève quelque temps auparavant. Si cette note antérieure est élevée, la note est en moyenne supérieure de 2 points à celle attribuée si la copie antérieure est faible (moyennes de 11,86 et 9,84).

De même, face à une série de copies identiques (8 rédactions), les enseignants évaluent différemment (moyennes de 11,16 et 9,65) selon que les devoirs sont censés provenir d'une classe de 6^e forte ou faible.

Même effet encore selon le lycée de provenance. Des copies identiques sont surévaluées si elles sont censées provenir du lycée Janson-de-Sailly (16^e arrondissement de Paris) et sous-évaluées si elles sont censées provenir d'un lycée de banlieue.

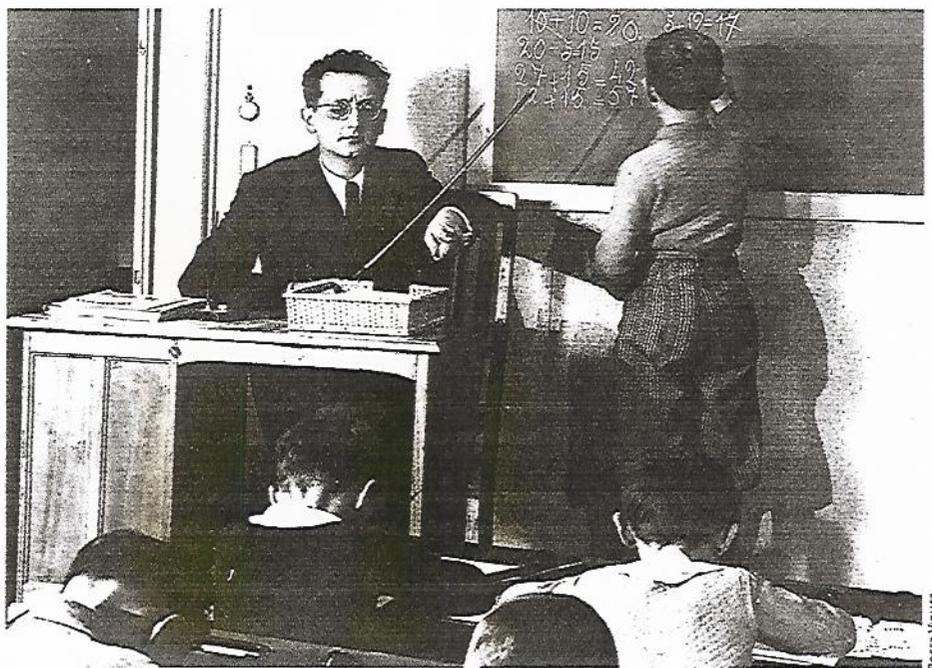
Enfin, l'effet de contraste est un processus inverse. Une copie est généralement surévaluée quand elle est corrigée après une copie faible et sous-évaluée quand elle vient après une copie forte.

Par exemple, une copie, corrigée après une copie mauvaise, a reçu une note qui lui a valu d'être classée 3^e.

Le même devoir, corrigé après une bonne copie, s'est retrouvé en 14^e position.

J.L.

Georges Noizet et Jean-Paul Caverni, *Psychologie de l'évaluation scolaire*, Puf, 1978.



ROGER VIOLET

partir de cette grille en recherchant des signes (les indicateurs) susceptibles de témoigner de la satisfaction de l'attente, et dont l'ensemble constituera le référent de l'évaluation. Par exemple, on attend d'un élève de 6^e qu'il sache tirer des informations ponctuelles d'un écrit (référent), et on constate qu'il a exécuté de façon satisfaisante un exercice l'invitant à retrouver dans le texte des indicateurs de lieux et de temps, et à dresser une liste de personnages (référé). L'évaluation est bien ainsi une opération de lecture orientée de la réalité, qui suppose une série d'impératifs.

L'enseignant ou le formateur ont à préciser leur projet en termes de compétences attendues, dont la légitimité s'appréciera d'un triple point de vue : scientifique (l'attente est-elle conforme à ce que la psychologie cognitive du développement nous apprend d'un élève de cet âge ?); institutionnel (respect des textes réglementaires et institutions officielles ?) et idéologique (est-ce bien cela qu'il faut socialement souhaiter ?).

Le processus évaluatif, c'est-à-dire l'ensemble des activités et démarches décrites ci-dessus, pourra être mis au service :

- d'un acte de certification comme valider des acquis ou accorder un diplôme (l'évaluation, dite alors certificative, est le plus souvent terminale, bien qu'elle puisse reposer sur un contrôle continu; elle a pour fonction de vérifier, afin de pouvoir l'attester, la « présence » des compétences visées);

- d'un acte d'orientation qui tend à harmoniser les particularités d'un individu et les exigences d'une filière de formation (à cette fin, l'évaluation s'efforce d'être diagnostique, c'est-à-dire de faire un bilan des difficultés et lacunes d'un côté, des acquis et potentialités de l'autre);

- d'un acte de régulation qui permet un ajustement ou une réorientation de l'action

pédagogique elle-même, ce qui n'est vraiment possible que dans la mesure où l'évaluation est devenue formative. On comprend, en fonction de la pluralité de ses usages sociaux possibles, que l'activité évaluative puisse osciller entre le contrôle (simple vérification de la présence de ce qui était attendu) et l'interprétation (effort fait pour rendre intelligible la réalité concrète examinée).

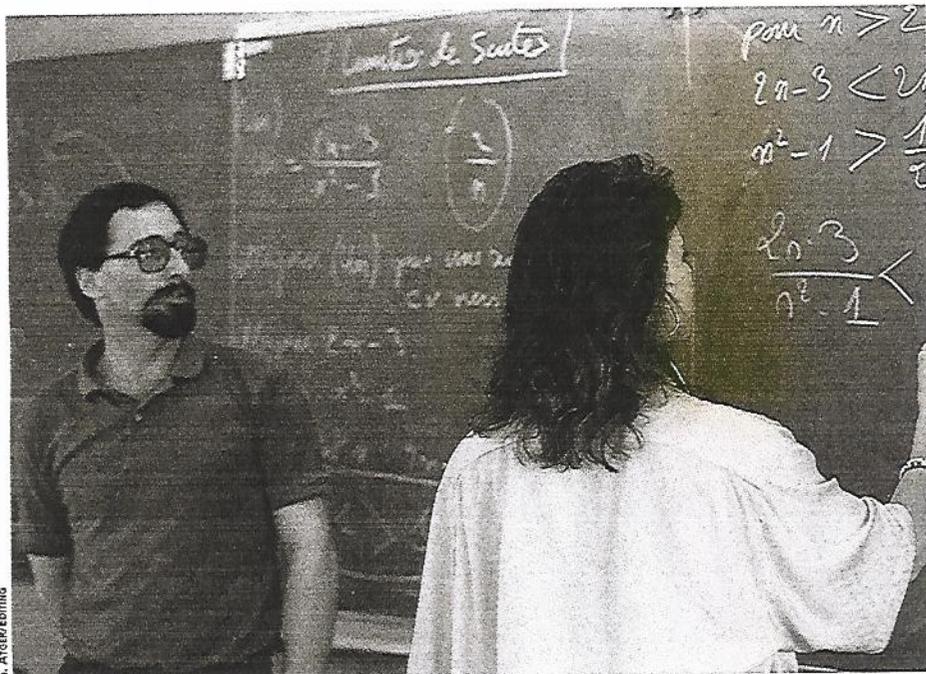
Les conditions d'une évaluation efficace

L'évaluation efficace est celle qui atteint son but. Bien évaluer, c'est en premier lieu comprendre ce que l'on fait, pourquoi on le fait. La difficulté est de savoir de quel point de vue il faut se placer pour trancher de la pertinence d'une pratique.

Voilà sans doute un acquis fondamental des trente dernières années : il n'y a pas d'évaluation correcte en soi - dans l'absolu. Mais il y a des évaluations pertinentes, en fonction d'une intention donnée et pour un usage social précis.

C'est sans doute pourquoi le discours sur l'évaluation formative est devenu dominant dans le champ de la pédagogie. L'évaluation formative est l'évaluation que l'on pourra dire « pédagogiquement correcte » (comme on dirait « politiquement correcte »), dans la mesure où les enseignants et les formateurs ont pour tâche de contribuer au développement positif des élèves et des formés en facilitant leurs apprentissages. C'est cet objectif pédagogique qui donne son sens à l'évaluation formative comme objectif à atteindre, laquelle s'intègre au processus pédagogique qu'elle éclaire.

Et l'évaluation efficace est celle qui éclaire bien, d'où la nécessité d'explicitier les attentes en termes de compétences, d'analyser et d'interpréter les erreurs, d'identifier les



Beaucoup de choses ont changé dans la classe. La règle du maître n'est plus de mise, mais l'évaluation de la compréhension de l'élève par la venue au tableau demeure une pratique répandue.

caractéristiques des apprenants, de faire un diagnostic précis des acquis et des lacunes, des forces et des faiblesses de chacun.

Les débats contemporains

Nous ne prendrons que les trois grandes questions « vives » autour desquelles se sont cristallisées les interrogations, et qui correspondent, en fait, à trois grandes alternatives.

1. Mesurer ou interpréter ?

À l'époque de la docimologie naissante, et encore pleine d'espoir, le dilemme paraissait être : mesurer correctement ou disparaître. Faut-il dire aujourd'hui : interpréter correctement ou disparaître ? En ce sens, avec l'évaluation formative, c'est bien ce second dilemme qui est d'actualité. Toutefois, beaucoup de formateurs et enseignants, et encore plus de parents, ont du mal à se détourner de l'idéal de la mesure « objective ». Si bien qu'aujourd'hui, les évaluateurs paraissent écartelés entre un fait et un impératif. Le fait, c'est l'absence de fiabilité des notes, que de nouvelles recherches viennent constamment rappeler au bon souvenir de ceux qui voudraient l'oublier. Et l'impératif, c'est d'apprécier le plus objectivement possible le degré d'atteinte des objectifs pédagogiques ou sociaux assignés aux apprenants ou aux formés. Il est toujours fréquent que des évaluations reposent ainsi sur des relevés quantitatifs (pourcentages, notes, moyennes, etc.). Or, « avant d'interpréter ces résultats, il est toujours nécessaire d'examiner s'ils méritent effectivement d'être pris en considération » (8). Cependant, pour évaluer, il faut bien à la fois observer, le plus rigoureusement possible, et interpréter, le plus pertinemment possible.

2. Juger ou informer ?

Selon Jean Cardinet, à un paradigme décisionnel (mesurer, juger, décider), a succédé un paradigme informationnel (apporter de l'information utile à l'apprenant pour faciliter ses apprentissages). Le premier paradigme s'est incarné dans les figures de l'évaluation comparative (pratiques de classement fondées sur des techniques psychométriques), puis critériée (régulation de la pédagogie par un retour fonctionnel d'information). Le second, dans les figures de l'évaluation correctrice (on fournit un modèle approprié), puis enfin conscientisante (on fournit des repères permettant à l'apprenant de s'auto-évaluer correctement). À cette dernière figure correspond ce qui a été désigné comme « évaluation formatrice », laquelle fait pratiquement disparaître ce qui serait un jugement formulé par un évaluateur externe, au profit d'une participation au mouvement d'autonomisation des apprenants. C'est pourquoi il y aurait lieu désormais d'« évaluer sans juger » (9), une évaluation descriptive étant la seule qui soit compatible avec une relation d'aide : « L'évaluation formative souhaitable est un retour d'information multidirectionnel, s'adressant à l'élève, au lieu de porter sur l'élève ». Non plus juger pour décider mais informer, pour aider.

En premier lieu, il s'agit moins pour l'évaluateur de prendre l'élève comme objet, que de l'aider à devenir sujet. En second lieu, et dans le même sens, le jugement de l'évaluateur a tout à gagner à s'éloigner de ce qui ressemblerait au verdict rendu par un tribunal. L'évaluateur ne juge pas une personne, mais juge (ou l'aide à juger) de la façon dont elle s'est rapprochée (ou non) d'un but.

3. Évaluer ou agir ?

L'évaluation implique toujours, d'une certaine façon, un arrêt (même bref) dans le mouvement (de réalisation d'un projet, de progression vers un but). Ne vaudrait-il pas mieux alors évaluer moins et agir plus ? Le glissement progressif de l'évaluation vers la pédagogie durant les trente dernières années, du souci de saisir à celui d'aider, traduit bien cette relative désaffection pour une évaluation perçue comme une inutile perte de temps au profit d'une activité didactique se mettant plus rationnellement et efficacement au service des apprentissages.

Mais le choix est-il vraiment entre évaluation et pédagogie ? Il convient en effet de se méfier de l'activisme évaluatif, et de refuser une frénésie évaluative qui ne serait plus qu'un alibi pour l'inaction et l'immobilisme (plus on évalue, moins les élèves progressent...).

Mais une évaluation qui présenterait le triple caractère d'être « compréhensive » (capable d'interpréter la situation « mesurée »), « conscientisante » (fournissant des repères éclairants à l'élève au lieu de le tancer), et « formatrice » (préoccupée de donner les outils de la réussite) ne serait-elle pas le plus bel auxiliaire et le premier moyen d'une pédagogie enfin efficace ? ■

NOTES

- (1) Claude Thélot, *L'Évaluation du système éducatif*, Nathan, 1993.
- (2) Jean-Marie Barbier, *L'Évaluation en formation*, Puf, 1985.
- (3) Henri Piéron, *Examens et Docimologie*, Puf, 1963.
- (4) Georges Noizet et Jean-Paul Caverni, *Psychologie de l'évaluation scolaire*, Puf, 1978.
- (5) Michael Scriven, « The Methodology of Evaluation », *Perspectives of Curriculum Evaluation*, Rand Mc Nally, 1967.
- (6) Benjamin S. Bloom et al., *Handbook on Formative and Summative Evaluation of Student Learning*, Mc Graw-Hill Book Co., 1971.
- (7) Philippe Meirieu, *La Pédagogie entre le dire et le faire*, ESF, 1995.
- (8) Jean Cardinet, « Pour évaluer vos mesures », *Bulletin de l'ADME-Europe*, 1995-2.
- (9) Jean Cardinet, « Évaluer sans juger », *Revue française de pédagogie*, n° 88, INRP.

A LIRE...

- L. Allal et al., *L'Évaluation formative dans ses enseignements différenciés*, Peter Lang, 1990.
- M. Barlow, *L'Évaluation scolaire : décoder son langage*, Chronique sociale, 1992.
- J.M. de Ketele (ed), *L'Évaluation : approche descriptive ou prescriptive*, De Boeck, 1986.
- Ch. Hadji, *L'Évaluation, règles du jeu*, ESF, 1995.
- Ch. Hadji, *L'Évaluation des actions éducatives*, Puf, 1992.