

L'INSOUTENABLE LÉGÈRETÉ DE L'ÉVALUATION OU LES AVATARS DE L'ÉVALUATION CRITÉRIÉE

Annie BARTHÉLÉMY,
Université de Savoie
I.U.F.M. de Chambéry

La mode est aux valises pédagogiques et s'il en existait sur l'évaluation, nul doute que bon nombre d'enseignants s'attendraient à y trouver des grilles d'évaluation, tant paraît rassurant ce genre d'instrument aux évaluateurs soucieux d'améliorer leurs pratiques. Or cette recherche de sécurité dans un instrument d'évaluation est illusoire. Employer une grille peut ne rien changer sur le fond si le choix de procédure ne vient pas concrétiser une conception de l'évaluation dans son objet, son but et ses destinataires ; en revanche, si la grille est perçue comme un moyen de préciser pour l'enseignant comme pour les élèves le cadre de référence de l'évaluation – ce à partir de quoi on juge, apprécie le degré de maîtrise de compétences en voie d'acquisition –, la définition des critères et leur utilisation conduisent à relativiser l'instrument. La volonté de disposer d'un référent pertinent, communicable, maniable mais révisable empêche de céder au fétichisme de l'outil.

L'analyse des pratiques d'évaluation critériée qui va suivre entend montrer leur intérêt et pointer les dérives et les leurres qui guettent l'évaluateur qui s'en remet à la fragile garantie qu'offre tout instrument d'évaluation aussi pertinent fût-il. Dans un premier temps, nous indiquerons les attentes auxquelles elles peuvent répondre, puis nous examinerons à la lumière des obstacles rencontrés, comment en faire un usage réaliste et lucide.

UNE ÉVALUATION PLUS EXACTE, PLUS TRANSPARENTE, PLUS OUVERTE, PLUS DYNAMIQUE

L'évaluation critériée consiste à apprécier un comportement ou une performance par rapport à un comportement-cible ou une performance-cible, caractéristiques de

celui qui a atteint les objectifs de l'épreuve d'évaluation. Dans le vocabulaire technique de l'évaluation, on oppose ce moyen d'évaluation à l'évaluation normative qui procède par comparaison de comportements ou de performances : par exemple, une note prend sens par rapport à une distribution de notes. L'élève qui reçoit une note perçoit plus aisément sa valeur par rapport aux notes inférieures ou supérieures qu'il ne saisit les repères en fonction desquels cette note a été attribuée. Lors d'une évaluation critériée, l'élève doit pouvoir rapporter des aspects de sa démarche ou de sa production à des repères qui indiquent soit les activités à mobiliser soit les conditions de réussite. Comme le soulignent O. et J. Veslin, une liste de critères maladroitement formulés, mais posés sur le papier, et utilisés comme outil de communication entre l'enseignant et ses élèves est de loin préférable à des critères impeccablement formulés... mais qui resteraient dans la tête de l'enseignant¹.

L'explicitation des critères n'a pas été la première préoccupation des recherches en évaluation. Les travaux de docimologie qui débutent en France en 1920 ont d'abord porté sur l'amélioration des procédures de notation, il s'agissait de les rendre plus fiables. Tant qu'il y aura des notes, les acquis en la matière méritent d'être rappelés pour tempérer les excès de zèle quantitatifs : mise en évidence des écarts d'un correcteur à l'autre, inventaire des sources de distorsions, remèdes pour réduire ces distorsions avant la correction (harmonisation des attentes des correcteurs par la production de barèmes de notation) ou après (multicorrection, harmonisation a posteriori des moyennes et écart-type).

La recherche d'une notation plus juste, il faudrait mieux dire moins arbitraire, passe au second plan lorsque, dans les années soixante-dix, la Pédagogie par Objectifs se diffuse en formation continue puis en formation initiale : la priorité est alors donnée au contrôle des résultats de la formation au regard des objectifs visés. Pour déterminer la congruence entre performances et objectifs, deux techniques sont proposées :

- l'opérationnalisation des objectifs qui consiste à les définir par la manière dont ils sont évalués. Mager, qui a popularisé cette technique², réclame un calibrage strict des performances attendues et des circonstances de l'évaluation ainsi que la précision du niveau d'exigence avec des critères de réussite clairement énoncés. On perçoit l'influence de ce procédé dans certains référentiels, utilisés pour le contrôle continu dans l'enseignement technique, présentés sous forme de tableaux à trois colonnes où sont précisés premièrement ce dont disposera le candidat lors de l'épreuve d'évaluation, deuxièmement les capacités dont il devra faire preuve et troisièmement pour chacune des capacités, les critères qui décideront de la maîtrise ou de la non-maîtrise du niveau requis.
- les taxonomies d'objectifs dont la plus connue est celle de Bloom³, il s'agit d'une classification hiérarchique des objectifs relevant du domaine intellectuel en six ni-

1. Veslin O. et J., *Corriger des copies. Evaluer pour former*, Hachette, 1987, p. 87.

2. Mager R. F., *Comment définir des objectifs pédagogiques*, Bordas, 1977 (première éd. fr. 1972).

3. Bloom B. et Coll., *Taxonomie des objectifs pédagogiques*, tome 1, Presses de l'Université de Québec, 1975.

veaux allant de la mémorisation de connaissances à l'évaluation en passant par la compréhension, l'application, l'analyse et la synthèse. A l'origine simple pense-bête pour examinateurs de l'enseignement américain en vue de les inciter à varier leurs épreuves d'examen, cette taxonomie qui connut un immense succès vaut surtout par la démarche car l'outil proposé manque de fondements psychologiques et didactiques.

Ces techniques d'opérationnalisation et de hiérarchisation des objectifs ont fourni des points d'appui à l'évaluation critériée. Bien entendu, on n'a pas attendu la Pédagogie par Objectifs pour formuler des critères d'évaluation mais bon nombre d'enseignants et de formateurs ont mesuré la distance existant entre les rubriques d'un barème de notation et des critères d'exigence qui permettent de situer un élève par rapport à des objectifs spécifiés.

Beaucoup d'espoirs ont été mis dans la définition rigoureuse des objectifs et des critères de réussite qui semblait répondre aux attentes de renouvellement de l'évaluation.

– attente d'une évaluation plus rigoureuse et plus exacte

L'évaluation critériée méthodiquement pratiquée oblige à lire la production de l'élève avec un questionnement précis ; soit, à titre d'exemple, la fiche d'évaluation présentée par A. Béguin pour l'évaluation du commentaire littéraire⁴.

Les critères sont présentés par ordre de complexité croissante, ils spécifient la compétence : « produire un commentaire de texte » pour un élève de seconde. Le jugement sur la copie se fait à partir de tout ou partie de ces douze points de vue explicites et objectifs. On peut en effet penser que la correspondance entre les critères d'évaluation et les indices relevés sur la copie sera plus assurée.

– attente d'une évaluation plus transparente qui facilite la communication entre enseignants et élèves

La fiche jointe à chaque devoir corrigé crée un espace de dialogue. Les élèves savent par rapport à quelle norme ils sont jugés ; une confrontation avec l'enseignant peut avoir lieu grâce à la pratique de la coévaluation. Les critères ne sont pas en eux-mêmes porteurs de sens, celui-ci passe par l'interprétation qui en est donnée lors des échanges autour de leur formulation et de leur application. La communication ne s'établit pas par les critères mais à leur propos ; ce qui compte c'est l'élaboration patiente des critères à partir d'une analyse des exigences de la tâche grâce à la lecture conjointe des Instructions Officielles et l'analyse d'extraits de commentaires produits par les élèves en vue de faire apparaître les problèmes d'écriture et les solutions utilisées. Trois

4. Béguin A., « Aide au travail en seconde : autour du commentaire de texte », *Innovations*, n° 16, 1990, CRDP Lille.

des méthodes identifiées par O. et J. Veslin sont ici mobilisées : partir des erreurs et des réussites de élèves, explorer les actions requises pour réaliser un commentaire précis et leurs incidences sur l'écrit à produire ou plus abstraitement faire l'inventaire des composantes de ce genre d'écrit scolaire.

– attente d'une évaluation moins répétitive, plus diversifiée

Dans l'exemple cité, la liste des critères indique quel type d'écrit est attendu dans un « commentaire composé », cette liste n'est ni intangible ni univoque ; « repérer des techniques d'écriture » recouvre des aspects différents s'il s'agit d'un poème d'Apollinaire ou un extrait des Caractères de La Bruyère. Dans un autre registre, disposer de critères pour évaluer l'oral en fonction de l'objectif, l'objet et le destinataire de la communication c'est ouvrir sur des appréciations plus modulées que celles qui s'appuient sur la fréquence des prises de paroles ou des qualités formelles : articulation, intonation...

– attente d'une évaluation plus dynamique qui soit une aide dans l'apprentissage

L'évaluation critériée permet de concrétiser la fonction formative de l'évaluation. Historiquement, effectivement, la notion d'évaluation formative telle qu'elle est définie chez Bloom et reprise par De Landsheere⁵ apparaît liée à la centration sur les objectifs et les critères. Bloom suggère de vérifier, à chaque étape de l'apprentissage, par un test quels sont les objectifs atteints ; en cas de résultat négatif, il convient de renvoyer l'élève à une ressource pédagogique : logiciel, fichier, etc. afin d'assurer la maîtrise des objectifs à un maximum d'élèves. Les améliorations escomptées ne se produisent pas automatiquement parce que la méthode préconisée engage une conception de la formation d'inspiration néobehavioriste. Dans cette optique, les caractéristiques de la performance visible, du comportement observable sont les garants de la maîtrise d'une ou de plusieurs capacités alors que celles-ci ne sont que le sommet d'un iceberg difficilement contrôlable⁶.

Par la suite, cette première approche de l'évaluation formative – correspondant à la pédagogie de la maîtrise – est remise en question non dans sa visée mais dans ses modalités. Les rôles respectifs des instruments d'évaluation et des actions de remédiation sont relativisés. L. Allal, s'appuyant sur le modèle constructiviste de l'apprentissage, fait remarquer que l'adaptation de l'enseignement aux démarches d'apprentissage des élèves ne s'opère pas seulement rétroactivement mais s'engage dès la conception et la conduite des situations d'apprentissage⁷. L'évaluation n'est qu'un moment, qu'une forme de régulation

5. Evaluation intervenant, en principe, au terme de chaque tâche d'apprentissage et ayant pour objet d'informer élève et maître du degré de maîtrise atteint et, éventuellement, de découvrir en quoi un élève éprouve des difficultés d'apprentissage en vue de lui faire découvrir des stratégies qui lui permettent de progresser. *Dictionnaire de l'évaluation et la recherche en éducation*.

6. Cardinet J., *Evaluation scolaire et pratique*, Bruxelles, De Boeck, 1985, p. 218.

7. Allal L., Bain D., Perrenoud P., *Evaluation formative et didactique du français*, Delachaux et Niestlé, 1993.

parmi d'autres (interventions du professeur et des autres élèves durant l'apprentissage en classe, ajustement de l'activité à la tâche proposée...). A l'inverse, J.J. Bonniol et G. Nunziati accordent une place centrale à l'évaluation en considérant la formation à l'auto-évaluation un levier pour faire évoluer les représentations que les élèves se font des tâches scolaires et développer leurs capacités à anticiper, à planifier, à contrôler leur activité. Cette proposition qui fait de l'évaluation une démarche didactique est appelée évaluation formatrice, le vocable traduit l'importance donnée à l'implication et à la prise de conscience des élèves⁸.

Quoi qu'il en soit, qu'elle soit relativisée ou majorée, l'évaluation n'apparaît plus comme une simple méthodologie, elle cherche son enracinement théorique dans trois orientations bien distinguées par L. Allal : la didactique, la psychologie cognitive, la psychologie sociale⁹. L'élaboration des critères d'évaluation s'articule plus explicitement à une vision de la discipline enseignée ou une conception de l'apprentissage ou une manière d'engager la négociation sur le sens des apprentissages dans le cadre de la classe. L'enseignant effectue bien souvent le lien entre ces trois perspectives, de manière pragmatique.

En conclusion, l'évaluation critériée ne transforme en profondeur la démarche de l'évaluateur que lorsqu'elle est liée à d'autres modifications ; la pratiquer revient à objectiver le cadre de référence qui sert à apprécier les productions et les conduites des élèves. Mais un modèle précis n'est pas forcément pertinent ni légitime : il peut traduire des options didactiques cohérentes comme codifier des routines pédagogiques, voir par exemple le nombre de critères formels que comportent certaines grilles d'évaluation des écrits des élèves.

Par ailleurs, même si le modèle renvoie à des options didactiques fondées – et il peut y avoir en ce domaine des modèles alternatifs – il reste à s'assurer que l'épreuve d'évaluation et les situations d'apprentissage ne sont pas en porte à faux avec les priorités affichées dans la grille. M. Constant et A. Noyère, ont pointé dans l'évaluation nationale de 1989 à l'entrée en sixième une contradiction entre l'objectif déclaré d'une épreuve – reconnaître un type de *texte* – et la nature de celle-ci invitant à repérer les caractéristiques d'un écrit¹⁰. D. Toupry a montré, dans une analyse rigoureuse des codages attribués à une même série de textes, que différents correcteurs ne donnaient pas la même importance ni la même signification aux critères imposés pour la correction des récits produits par les élèves soumis à l'évaluation nationale 1991 à l'entrée du CE2¹¹.

8. Nunziati G., « Pour construire un dispositif d'évaluation formative », *Cahiers Pédagogiques*, n° 280, janvier 1990.

9. Op. cit., p. 14.

10. Constant M., Noyère A., « Comment vivent les Eskimos ? Ce que peut cacher une épreuve d'évaluation de la compréhension d'un texte », *Spirales* n° 4, avril 1990, Ecole Normale de Lille.

11. Toupry D., *L'évaluation de l'écrit à l'entrée du CE2*, Mémoire de maîtrise, U.F.R. des Sciences de l'Éducation, Université de Lille III, 1992.

Il faut donc relativiser la portée d'une évaluation critériée standard qui ignorerait les conditions d'apprentissage et d'évaluation. Au delà, peut-être faut-il modérer les espoirs en une évaluation rigoureuse, transparente, élargie et dynamique.

LA VALEUR RELATIVE DES CRITÈRES

Avoir une confiance mesurée dans les critères d'évaluation, c'est avoir à l'esprit que ceux-ci ne prennent sens qu'à l'usage. S'ils ne sont pas portés par des enseignants et des élèves qui ont pris le temps de stabiliser provisoirement leur signification dans le contexte des apprentissages en classe, ils ne sont que des étiquettes sous lesquelles vont perdurer les biais des évaluations impressionnistes, unilatérales, stigmatisantes. L'habit ne fait pas le moine. Mais il est aussi des biais qui naissent des bonnes intentions, l'expérience permet d'avancer quelques propositions pour éviter les désillusions.

Première proposition : ne pas détailler inconsidérément la liste des critères

Une dérive fréquente de l'évaluation critériée est l'excès d'analyse. Soit à titre d'exemple, une fiche-contrat rédigée en Sciences Physiques pour des élèves de sixième qui, sur le thème : volume de solides et de liquides, énonçait une douzaine de capacités parmi lesquelles : nommer des unités de volume et écrire leur symbole, convertir les volumes avec un tableau d'unités, déterminer le volume d'une goutte de liquide, vérifier que le volume d'un liquide ne change pas quand on le transvase ... Il n'est pas difficile de mesurer les conséquences de l'accumulation d'objectifs parcellaires et cloisonnés si tous les points du programme étaient ainsi démultipliés. L'évaluation devient pointilliste et pesante. Juger ainsi séparément de la maîtrise de capacités fait perdre de vue la cohérence de la formation.

A l'origine de telles dérives, un asservissement à la technique d'opérationnalisation des objectifs.

L'antidote existe : hiérarchiser les objectifs et en donner une présentation structurée. Le défaut de la fiche précédente, c'est la juxtaposition de capacités apparaissant détachées les unes des autres et strictement délimitées à un contenu disciplinaire alors que certaines ont manifestement des relations entre elles et avec des capacités mobilisées dans d'autres parties du programme. Un référentiel utile, c'est au contraire un référentiel qui rompt avec une lecture trop linéaire du programme. Les grilles d'évaluation trop lourdes sont peu maniables ; à vouloir multiplier les points de repères, on n'est pas plus efficace, mieux vaut en sélectionner quelques-uns à la croisée des exigences de la tâche dont l'enseignant est le garant et des cheminements des élèves. La valeur des critères tient, comme le souligne C. Hadji, à leur valeur comme outil de développement et il ajoute « la difficulté réside davantage dans le choix d'une cible appropriée que dans la mesure de l'écart à la cible »¹².

12. Hadji C., *L'évaluation, règles du jeu. Des intentions aux outils*, Paris, ESF, 1989, p. 113.

Deuxième proposition : mettre en perspective les critères

L'application de critères peut donner le sentiment d'une plus grande objectivité de l'évaluation alors que la grille ne fait qu'objectiver les valeurs de référence de l'évaluateur. Cette illusion guette surtout ceux qui se fient, sans appropriation critique, à un référentiel tout élaboré qui tient lieu de « prêt à porter ». Pour éviter ce que certains ont pu appeler la dérive objectiviste, un travail en amont et en aval de la grille s'impose. En amont, il s'agit de savoir le lire avant de s'en servir pour apprécier les productions des élèves. En aval, il s'agit de reconnaître l'indispensable travail d'interprétation.

En amont, un modèle en impose d'autant plus qu'échappe à celui qui s'y réfère la logique de construction. Les meilleurs outils peuvent produire des effets pervers. Au début du contrôle continu en EPS au baccalauréat, certains ont cru bon d'évaluer les connaissances sur les techniques des activités physiques et sportives par des interrogations écrites, faute d'analyse du sens de cette rubrique du référentiel !

En aval, la possession d'une grille ne suffit pas à harmoniser les attentes des évaluateurs. Une recherche menée à l'INRAP de Dijon a montré que les membres des jurys, munis pour l'évaluation de rapports de stage d'une grille renvoyant à sept capacités commentées dans une liste indicative de critères, n'abandonnaient pas pour autant leurs habitudes anciennes. Chaque correcteur avait tendance à vérifier ce qui dans le rapport pouvait être relié aux connaissances de sa discipline¹³. On pourrait en tirer argument pour mettre en cause l'instrument et chercher une grille qui canaliserait mieux les évaluateurs. Mais le perfectionnement de l'instrument laisse entier la question de son maniement. Les travaux de docimologie ont depuis belle lurette prouvé qu'un barème n'harmonise les référents des divers évaluateurs que s'il est le produit de leur concertation. Dans le cas cité, remplacer une grille par une autre ne changerait sans doute rien si les membres du jury ne se l'appropriaient pas en discutant, à partir de deux ou trois rapports, des liens entre épreuve, capacités attendues, critères et indices prélevés dans l'écrit soumis à leur appréciation.

Troisième proposition : harmoniser la procédure et les choix didactiques

Faute d'une telle articulation, les grilles produites sont très formelles ou très contraignantes. Si la généralité de certains critères – vocabulaire diversifié – n'offre que peu de prises à l'enseignant comme à l'élève, à l'inverse d'autres très fonctionnelles conduisent à un calibrage étroit des productions attendues – cf les évaluations du conte réduites au schématisme des fonctions, selon la version scolaire de Propp. Dans les deux cas, la grille enferme l'évaluateur au lieu d'être l'expression provisoire de choix didactiques assumés lucidement.

La critique qu'a faite D. Brassart d'une des épreuves de l'évaluation Français 6ème (1989) révèle le rôle que peut jouer la didactique¹⁴. L'épreuve analysée

13. *Actes du colloque de Dijon : Evaluer l'évaluation*, Dijon, INRAP, 1986.

14. Brassart D. G., « Lire les Eskimos, Retour sur l'épreuve 21 de l'Evaluation Français 6ème », *Cahiers Pédagogiques*, n° 283, avril 1990.

testait la compréhension par les élèves d'un texte informatif par une liste de questions censées exiger une prise d'indices plus ou moins aisée selon que la réponse demandait ou non à dépasser le cadre de la phrase, obligeait ou non à des retours en arrière. Le taux élevé d'échecs aux questions les plus complexes a suscité des propos alarmistes sur le niveau en lecture des jeunes collégiens. D. Brassart incitait à plus de prudence : l'objectif affiché de l'épreuve se révèle réducteur par rapport aux compétences dès lors que le texte vient heurter des représentations du lecteur, au-delà de la prise d'indices c'est la capacité à lire pour apprendre des connaissances nouvelles qui est en jeu.

Si un objectif peut en cacher un autre, et une épreuve servir d'écran plutôt que de révélateur, il est aussi des critères faussement transparents. L'apport des didactiques, quand il fait l'objet d'une appropriation dans les démarches d'apprentissage, conduit l'évaluateur à être plus circonspect dans ses choix et ses jugements.

Quatrième proposition : les critères ne sont pas la panacée de l'apprentissage

La confusion entre modalité et fonction de l'évaluation produit l'illusion que l'énoncé de critères suffit pour que l'évaluation soit une aide à l'apprentissage. On attend, par exemple, de l'évaluation critériée qu'elle donne des indications sur les voies de la remédiation alors que la plupart des grilles sont conçues pour évaluer les acquis : elles permettent de localiser réussites et échecs mais ne les rendent pas nécessairement compréhensibles. Comment trouver alors les leviers qui feront progresser l'élève sans disposer d'hypothèses sur la source de ses difficultés ?

Ainsi les critères d'évaluation se révèlent notoirement insuffisants pour caractériser les productions d'élèves les plus éloignées des attentes de l'enseignant. F. Darras et J.P. Ledoux offrent un témoignage parlant de leur perplexité face aux résultats d'une élève à l'évaluation CE2 (1989)¹⁵. Sur l'épreuve de production d'histoire avec mots imposés, l'application du critère de lisibilité se révèle franchement délicate :

- s'il réfère au tracé des lettres, aucun problème, la production ne satisfait pas au critère ;
- s'il réfère à la maîtrise du code écrit, de graves problèmes de correspondance grapho-phonétique le rendent illisible, cf. beaucoup transcrit botou ;
- s'il réfère à la compétence textuelle testée par l'épreuve, paradoxalement, la production peut être jugée satisfaisante puisqu'un patient décryptage du texte fait apparaître une histoire avec un scénario construit.

Dans cet exemple, les codes prévus dans la grille d'évaluation se révèlent en porte-à-faux. On pourra rétorquer que l'épreuve a néanmoins permis de mettre le doigt sur les difficultés graphophonétiques mais, sur ce point, des résultats très hétérogènes aux différentes épreuves qui testent cette dimension contrai-

15. Darras F., Ledoux J.-P., « Il était une fois un cha et un chen... ou l'histoire d'une évaluation », *Spirales*, n° 4, avril 1990, Ecole Normale de Lille.

gnent les évaluateurs à formuler quelques hypothèses provisoires et partielles pour observer de plus près cette élève en situation de lecture-écriture.

Cependant la distinction entre critères de réalisation et critères de réussite, établie par J.-J. Bonniol et G. Nunziati, permet d'introduire dans l'élaboration de la grille une logique d'apprentissage. Les critères de réussite fixent les normes en fonction desquelles un résultat est jugé satisfaisant (ex : « pas d'oublis importants » pour un résumé) ; les critères de réalisation indiquent les opérations à mobiliser pour réussir (ex : « fixer des proportions qui respectent l'organisation du texte »).

La formulation de critères de réalisation suppose un travail d'élaboration avec les élèves : dans l'exemple cité plus haut du commentaire composé, la discussion avec les élèves à propos de leurs premiers jets en réponse à une consigne d'écriture demandant à présenter de courts extraits de textes. Le critère de réussite « le texte est présenté dans l'introduction » se trouve alors assorti des critères de réalisation suivants :

- repérer les références du texte et toutes les indications dans le titre et les notes permettant de le situer ;
- trouver un ordre pertinent pour restituer ces indications.

Il est évident que de tels critères ne parlent qu'à ceux qui en ont dégagé l'utilité à partir de leurs propres démarches ; ils ne sont pas communiqués aux élèves par l'enseignant mais définis en interaction, ils vont donc apparaître selon un ordre de complexité croissant pour l'élève qui cherche à se les approprier et non en fonction du poids qu'ils auront lorsque la compétence sera installée.

Quelque soit l'intérêt de ces démarches d'évaluation formatrice, elles ne peuvent à elles seules assurer la régulation des apprentissages ; les pratiques innovantes depuis une dizaine d'année ont évolué, on attend moins de l'évaluation et davantage des situations pédagogiques. L'illusion que ça ira mieux le jour où des critères plus pointus fourniront diagnostics plus fouillés est moins répandue ; le traitement de l'erreur, le recours à la métacognition à l'occasion des activités et échanges qui ont lieu en classe manifeste que souvent ce qui fait progresser c'est une remarque ou une relance au bon moment et l'enseignant n'a pas le monopole de ce type d'intervention.

Faire réagir les élèves sur plusieurs solutions trouvées, mobiliser un entretien d'explicitation pour débloquer une difficulté, c'est engager une évaluation qui ne passe pas par une interprétation à référence critériée. Examiner un échantillon de textes d'élèves pour décider d'une nouvelle consigne d'écriture, c'est pratiquer une évaluation ciblée qui ne passe pas par le crible d'une grille.

Au terme de cette analyse, revenons sur les attentes relatives à l'évaluation critériée :

1) attente d'une évaluation plus rigoureuse.

Oui, à condition de chercher moins la précision que la pertinence ;

2) attente d'une évaluation plus transparente.

La réponse est moins affirmative car il y a toujours une zone d'opacité dans le rapport évaluateur-évalué vu la complexité des enjeux institutionnels et affectifs ; il est cependant possible de s'entendre provisoirement sur des repères utiles à un moment donné pour un groupe d'élèves ;

3) attente d'une évaluation plus diversifiée.

Oui, si l'évaluation n'est pas plaquée sur des pratiques pédagogiques qui ne sont pas en cohérence avec elle. Vouloir élargir l'évaluation sans modifier les pratiques pédagogiques, c'est accroître la pression évaluatrice en la dissimulant sous une apparence « rationnelle » ;

4) attente d'une évaluation qui soit une aide à l'apprentissage.

En la matière, ce n'est pas la procédure qui décide du sens et de la visée de l'évaluation : toutes les formes d'évaluation critériée ne s'inscrivent pas dans une logique d'évaluation formative et cette dernière passe aussi par d'autres voies que l'appropriation des critères d'évaluation.

En définitive, le recours à l'évaluation critériée traduit un souci de ne pas évaluer à la légère ; c'est un moyen précieux de dialogue pédagogique à condition de préserver l'insoutenable légèreté de toute évaluation fût-elle méthodiquement construite, car des critères d'évaluation non relativisés peuvent devenir la version modernisée de verdicts scolaires sans appel.