L'évaluation mutuelle interrogée par les principes d'efficacité et d'équité

Françoise Campanale

▶ **To cite this version:**

Françoise Campanale. L'évaluation mutuelle interrogée par les principes d'efficacité et d'équité. 16ème colloque international de l'ADMEE-Europe. Université (ULG). Evaluation : entre efficacité et équité, ADMEE-Europe & Université de Liège (ULG), Sep 2003, Liège, Belgique. hal-01175521

HAL Id: hal-01175521
https://hal.science/hal-01175521

Submitted on 9 Sep 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

L'évaluation mutuelle interrogée par les principes d'efficacité et d'équité

Françoise Campanale
IUFM de Grenoble, Laboratoire des Sciences de l'éducation, UMPF-Grenoble2 -France

La situation d'évaluation mutuelle (Allal et Michel, 1993 ; Allal, 1999 ; 2000) entre élèves s'inscrit dans une démarche d'évaluation formative conçue comme évaluation-régulations (Bonniol & Vial, 1997 ; Vial, 2001), autoévaluations interactives des acteurs du processus pédagogique (Campanale, 1997). Allal (2000) la définit ainsi : « Deux ou plusieurs apprenants (pairs ayant un même statut dans la situation de formation) évaluent leurs productions et/ou leurs procédures respectives ou conjointes, en se servant éventuellement d'un référentiel externe » (p. 88). Elle vise à provoquer chez l'élève une autoévaluation de son apprentissage, à travers les informations qui lui sont données en retour par des pairs sur sa production, à travers la mise en question de ses démarches et représentations sous l'effet à la fois d'une utilisation de critères et de verbalisations. Autrement dit, cette situation vise à provoquer une réflexion métacognitive, dont on sait qu'elle est associée à des autorégulations efficaces (Bouffard-Bouchard, 1990 ; Noël, Romainville & Wolfs, 1995 ; Allal & Michel, 1993). Sa mise en œuvre suppose l'organisation d'un dispositif, dont peuvent varier les modalités de mutualisation et les outils, qui sollicitent différemment les élèves.

Nous avons analysé 4 dispositifs d'évaluation mutuelle, mis en œuvre dans des classes d'écoles primaires par des professeurs d'école stagiaires qui en ont étudié les effets sur les apprentissages des élèves.

Les données, tirées de leurs 4 mémoires professionnels[1], comportent :

- les descriptions des dispositifs,
- les analyses faites par les enseignants stagiaires des différents états des productions des élèves,
- les analyses des fiches d'évaluation mutuelle remplies par les élèves,
- des retranscriptions des propos que des échantillons d'élèves ont échangé en remplissant ces fiches
- les entretiens d'explicitations (Vermersch, 1994) conduits a posteriori par les enseignants avec chacun des élèves des échantillons, pour savoir s'ils avaient conscience des régulations qu'ils avaient effectuées, pour savoir en quoi l'évaluation mutuelle les avait aidés.

[1] Il s'agit des mémoires professionnels réalisés respectivement par Anne-Sophie Vaufredaz et Magali Vivès, Sophie Bour et Timothy Murphy, Chloé Lecomte et Magali Chulia, Isabelle Jourdet et Myriam Pagneux, professeurs d'école stagiaires à l'IUFM de Grenoble en 2002-2003 et que je remercie pour l'intérêt de leur travail.

Les questions, auxquelles cette communication se propose de répondre, sont les suivantes : Quel dispositif semble le plus pertinent pour provoquer de l'autoévaluation ? Quelle est l'efficacité d'un dispositif d'évaluation mutuelle, du point de vue des performances, de celui des apprentissages ? Comment se pose la question de l'équité dans un tel dispositif ?

1. Quel dispositif semble pertinent pour l'évaluation mutuelle ?

Avec les mêmes sources théoriques, comment les enseignants stagiaires mettent-ils concrètement en œuvre l'évaluation mutuelle ?

1.1 Présentation des contextes

Dans les 4 cas, il s'agit de séquences de 8 séances, visant une production écrite ou orale en français pour des élèves de cycle 3, n'ayant aucune expérience préalable de l'évaluation mutuelle. Les séquences, finalisées par une valorisation externe (recueil de textes / envoi aux correspondants), sont structurées de façon similaire : présentation du projet, analyse d'exemples et dégagement de critères de réalisation, premier jet individuel total ou partiel, première correction individuelle par les élèves après exercice de remédiation sur des erreurs constatées pour l'ensemble des élèves, évaluations mutuelles après le 2ème jet, voire le 3ème jet et réécriture, éventuellement autoévaluation au sens strict à la fin.

Par contre, apprentissages et contextes de classe diffèrent :

Cas 1 : Décrire un monstre imaginé - classe de CM1 (26 élèves de 9/10ans)

Cas 2 : Rédiger une nouvelle version d'un conte traditionnel (le petit poucet) – classe de CM2 (29 élèves de 10/11ans)

Cas 3 : Rédiger une règle d'un jeu de cartes (une réussite) à partir d'une explication orale et d'une appropriation de cette règle – classe de CM2 (22 élèves) en zone d'éducation prioritaire

Cas 4 : Présenter oralement une courte nouvelle – classe de CM1 (21 élèves), en zone d'éducation prioritaire et au climat très conflictuel.

1.2 Caractéristiques du dispositif adopté

Le but du dispositif est de provoquer de l'autoévaluation et de l'autorégulation dont on saisit les manifestations à travers ce que se disent les élèves au cours de l'évaluation mutuelle et à travers les améliorations qu'ils effectuent au niveau de leur production. Plusieurs dispositifs sont essayés par les enseignantes, qui finalement s'accordent sur un type de mutualisation et un type d'outil jugés plus pertinents.

Pour la mutualisation, ne sont pas retenues les modalités suivantes :

- l'évaluation réciproque simultanée mais séparée (chaque élève évalue à l'aide d'une grille de critères la production d'un pair qui en même temps

mais séparément évalue celle de son évaluateur, puis chacun remet la grille remplie à l'évalué), car les élèves disent ensuite ne pas bien comprendre ce que leur pair a écrit sur la grille ;

- l'évaluation réciproque en triade (3 élèves ensemble évaluent successivement la production de chacun des 3), car les échanges enregistrés montrent que l'élève évalué adopte une position passive face au dialogue des 2 évaluateurs, et, de plus, cela prend environ 30 minutes, les élèves se lassent.

La modalité de mutualisation finalement adoptée dans les 4 cas comme gérable par les élèves et produisant un échange constructif est :

- L'évaluation mutuelle réciproque commune en dyades (2 élèves évaluent ensemble successivement la production de l'un et de l'autre à l'aide d'un outil qu'ils remplissent ensemble).

L'outil adopté comme support de l'évaluation mutuelle est un questionnaire interactif, relatif aux critères (*cf.* Allal & Michel, 1993), qui sollicite davantage l'analyse et la réflexion qu'une grille de contrôle (critères sous forme d'énoncés affirmatifs à confirmer ou infirmer). Il comprend des questions fermées et des questions ouvertes pour les critères difficiles à prendre en compte et sollicite des propositions d'améliorations (*cf.* exemple en annexe). Les élèves sont alors en situation non seulement de se prononcer sur le texte de l'autre mais aussi de s'interroger sur leur propre texte, de s'expliquer, de réclamer le point de vue de l'autre. Cela favorise la prise de distance, l'interrogation de ce qui a été écrit et de ce qui pourrait l'être, et ainsi des régulations à effectuer.

Dans l'exemple qui suit, enregistré lors d'une évaluation mutuelle pour la description d'un monstre, on voit apparaître un discours de l'évalué, qui est un questionnement tout autant pour soi que pour l'autre et l'amorce de régulations.

Exemple d'échanges entre Fanny et Benoit (élèves moyens), à propos des questions : « Est-ce que l'ensemble est organisé ? Quelle organisation a été choisie ? »

Fanny évalue Benoit

F : C'est plutôt du particulier au général, non ?

B : En fait, j'sais pas trop.

F : Donc, c'est à améliorer. Voilà. Ca change tout le temps ! Il faut mettre « plus organisé ». Après, le texte comporte t-il des paragraphes ?

B : En haut, c'est plutôt la tête, après, en fait… heu, c'est du haut en bas, je dirais, moi.

Benoit évalue Fanny

B : Ouais. Est-ce que l'ensemble est organisé ? Alors, il a trois têtes (…)

F : Oui, c'est bien. Par contre, est-ce que c'est du général au particulier ou du particulier au général ?

B : Je dirais plutôt du général au particulier. Le texte comporte t-il des paragraphes ?
F : Est-ce que j'ai bien mis ensemble ce qui allait ensemble ?
B : Ben… pas trop, parce qu'après, t'as mis la tête, après t'as dit comment il vivait, après, t'as redit des petits machins…
F : Ben, oui. J'ai bien mis ensemble, heu… ah oui !
B : Ca, t'aurais dû le mettre plutôt au début.
F : D'accord. Paragraphes, tu mets « non ». J'aurais peut-être dû le mettre dans le début et ça serait allé.

Dans un entretien a posteriori avec l'enseignant, Fanny dira : «*En fait, c'était Benoît qui m'a dit qu'il y avait trop de « il », qu'il fallait mieux mettre ensemble ce qui allait ensemble* ». Or, Benoît, comme en témoigne l'extrait précédent, n'a pas fait des préconisations aussi précises. Mais, cela signifie que c'est bien au cours de l'interaction avec lui que Fanny a enclenché de l'autorégulation, d'où la pertinence de ce dispositif d'évaluation mutuelle commune où les échanges sont sollicités par le questionnaire interactif.

2. En quoi ce dispositif est-il efficace ?

Pour juger de l'efficacité, on met généralement en rapport les résultats et les objectifs visés (Nacuzon Sall et De Ketele, 1997). L'atteinte de objectifs s'observe au terme de la séquence à travers les performances des élèves, mais on espère aussi des apprentissages à plus long terme.

2.1 Les élèves améliorent-ils significativement leurs performances ?

Dans le cas 1 (description d'un monstre), les productions des 26 élèves, lors du 1er jet et lors du 3ème et dernier jet, ont été évaluées en regard de 4 critères généraux : présence de caractéristiques physiques et morales, organisation cohérente visible à travers des paragraphes, correction syntaxique et orthographique (langue), variété du vocabulaire. Entre ces deux écrits ont été organisées deux situations d'évaluation mutuelle (1er jet, évaluation mutuelle portant sur la présence de caractéristiques physiques et morales, réécriture, évaluation mutuelle sur les 4 critères cités et réécriture). La comparaison des performances de chaque élève entre ces 2 écrits nous montre l'évolution.
A la fin, 15 productions sur 26 (57,7%) satisfont tous les critères, ce qui n'était le cas d'aucune lors du premier jet. Une analyse plus détaillée de la progression critère par critère donne les résultats suivants :

- 57% des élèves ont amélioré leur texte pour les caractéristiques physiques et morales (4 élèves sur 8 en difficulté au 1er jet ne progressent pas),
- 56% ont amélioré l'organisation (6/17 en difficulté au 1er jet ne progressent pas)
- 100% améliorent la correction orthographique et syntaxique (résultat dû aussi à un étayage de l'enseignant),

- 47,8% seulement la variété des anaphores.
- 2 élèves sont en échec (1 ne finit pas son texte, l'autre ne progresse en rien).

On peut considérer le dispositif d'évaluation mutuelle comme efficace car on approche des 60% pour une progression sur les 2 gros critères spécifiques à cette tâche, qui n'est pas facile pour des élèves de CM1, au premier trimestre. On constate sans étonnement une progression moindre en ce qui concerne la variété du vocabulaire. Mais, pour vraiment donner du sens à cette mesure, il faudrait disposer d'une mesure comparative sans évaluation mutuelle, ce que nous n'avons pas.

2.2 Quelle efficacité par rapport à l'apprentissage ?

Au-delà des rectifications et remaniements effectués, les élèves ont-ils modifié leurs représentations, construit et mis en mémoire une procédure de référence pertinente ? Sauraient-ils réutiliser ces critères pour une tâche similaire ultérieure ?

Leurs représentations sur la tâche et les moyens de la réaliser n'ont pas été recueillies au début ni à la fin. La durée des expériences pour le mémoire professionnel ne permettait pas de donner une nouvelle tâche aux élèves. Toutefois, on peut inférer d'un apprentissage à plus long terme s'il y a eu réflexion métacognitive au cours de cette séquence. L'analyse des entretiens évaluatifs rétro et proactif conduits, dans le 1er cas (décrire un monstre), avec un échantillon de 5 élèves (fort et moyens) nous permet d'amorcer une réponse.

2 élèves (niveau fort et moyen) sur 5 tiennent un discours clair sur ce qu'ils ont amélioré. L'élève de niveau fort développe peu (il sait faire) et s'étend davantage sur la situation d'évaluation mutuelle qu'il a trouvé très utile. Par contre Fanny (niveau moyen) témoigne d'une prise de conscience particulièrement significative : « *…avant, par exemple, je mettais toujours « il », je m'en étais pas rendu compte, puis aussi, je m'en étais pas rendu compte qu'il fallait mettre du général au particulier ou du particulier au général, qu'il fallait regrouper. Quand je lisais des portraits dans les livres, j'avais jamais vu ça. En fait, ça m'a appris des choses.* » On peut dire que pour Fanny la situation a provoqué un remaniement de ses représentations, un apprentissage en ce qui concerne l'organisation d'un texte.

Les trois autres élèves de niveau moyen sont évasifs (« *j'ai changé des choses, j'en ai enlevé et j'ai rajouté des choses* ») et ne semblent pas avoir vraiment construit une référence consciente. L'un d'eux déclare que la prochaine fois, même s'il s'agit de décrire un monstre, ce sera un autre monstre, donc un autre texte avec d'autres améliorations à faire. On touche là la question du rapport au savoir et de la différence entre les élèves centrés sur la tâche et ceux qui dépassent la situation pour véritablement apprendre (Beautier, Charlot, Rochex, 2000).

Le résultat est mince et ne concerne qu'un petit échantillon d'élèves non habitués à ce type d'évaluation et de réflexion. En reproduisant ce type de dispositif (évaluation mutuelle suivie d'un entretien avec l'enseignant), pourrait-on obtenir davantage de réflexion métacognitive de la part des élèves moyens ? Une étude longitudinale serait à faire.

3. Et l'équité dans ce dispositif ?

Nous reprendrons la distinction de Nacuzon Sall et De Ketele (1997) entre équité horizontale (« traitement égal pour des individus égaux », p. 131) et équité verticale (traitement inégal pour des individus inégaux, qui viserait à compenser des inégalités dont les individus ne sont pas responsables).

3.1 Equité horizontale ?

Les élèves, pairs par le même statut, sont-ils dans ce dispositif traités pareillement ? Cela signifierait que chacun bénéficie de la même qualité de feed back évaluatif. Par qualité, nous entendons à la fois la pertinence des appréciations et la précision des propositions d'amélioration.
Dans les 4 cas, la comparaison entre l'évaluation des élèves et celle de l'enseignant montre qu'un tiers des élèves évalue de façon experte (ils se sont appropriés, du moins pour cette fois, les critères), les autres ne remarquent pas toutes les erreurs, 3 ou 4 élèves font des remarques non pertinentes. Deux tiers des élèves font des propositions précises (Exemples : « dire qu'il faut étaler les cartes en ligne sur la droite » dans le cas de la règle du jeu ; « Il faut ajouter un objet magique qui va les sauver » dans le cas du conte), alors qu'un tiers s'en tient à des conseils généraux du type « il faut mieux écrire ». Les évaluations faites par les élèves sont donc de qualité inégale.

Cette inégalité se répercute-t-elle sur les améliorations effectuées ?
La mise en relation de la qualité de l'évaluation mutuelle et des améliorations effectuées pour un échantillon hétérogène de 7 élèves (cas 3 concernant la règle du jeu) montre que :

- la qualité de l'évaluation mutuelle est déterminante chez des élèves de niveau très moyen en production écrite, pour provoquer des améliorations de l'écrit ;
- elle n'est pas déterminante pour un élève qui dispose d'un niveau plutôt élevé en production écrite, et dont les capacités lui permettent de pallier les lacunes de l'évaluation mutuelle[2] ;
- elle est sans effet pour les élèves démotivés qui ne cherchent pas à améliorer leur production.

[2] Dans les autres cas, on constate aussi qu'une évaluation non pertinente est susceptible d'être rectifiée par les élèves évalués eux-mêmes.

Autrement dit, le dispositif est inéquitable pour les élèves de niveau moyen. Et, pour eux, *a fortiori* pour ceux de niveau faible, variations de dyades et interventions de l'enseignant seraient nécessaires (Allal et Michel 1993).
Cette dernière remarque appelle la question de l'équité verticale, car on sait bien que Reste à savoir apprécier la responsabilité personnelle des élèves dans cette inégalité de capacités. Sur ce point épineux, nous renvoyons à la réflexion de Meuret (2000), qui soulève le problème du droit de chacun d'utiliser ses capacités à sa façon et selon ses motivations.

3.2 Qu'en est-il de l'équité verticale ?

Les pairs ne sont égaux que par le statut et non par les capacités. Non équitable dans le traitement égal des élèves de même niveau, du fait de la différence de qualité de l'évaluation fournie par des pairs, le dispositif serait-il équitable au sens de compensation des capacités initiales inégales ? Le dispositif permet-il aux élèves en difficulté de progresser plus que les élèves forts et d'approcher leurs performances ?
Dans le cas 3 (celui de la règle du jeu), les progrès d'un échantillon hétérogène de 7 élèves (3 de niveau faible, 2 de niveau moyen, 2 de niveau fort) ont été analysés et figurent ci-dessous.

Elèves		**Nombre de critères satisfaits**	
		1er jet	**Dernier jet**
2 élèves forts ensemble		6/7	7/7
2 élèves moyens	Anissa	5/7	6/7
	Yoann	4/7	7/7
3 élèves faibles	Gregory	3/7	5/7
	Enis	3/7	4/7
	Mathilda	1/7	6/7

Les 2 élèves de niveau moyen réalisent des performances finales quasiment équivalentes à celles des élèves de niveau fort. Par contre, dans le cas des élèves en difficulté, on note une disparité entre les résultats de Mathilda et ceux de Enis, qui ne progresse que sur un critère. Ainsi pour 4 élèves de niveau moyen et faible, on peut considérer que le dispositif est compensatoire. Il ne l'est pas pour un élève de niveau faible. Or, il se trouve que du fait du nombre impair de l'échantillon, Enis (niveau faible) s'est trouvé avec 2 élèves forts, situation qui a été jugée comme non pertinente précédemment. Par contre, les 2 autres dyades ont été organisées de façon plutôt homogène (2 élèves de niveau moyen ensemble et 2 élèves de niveau faible ensemble). Il serait intéressant d'étudier les effets de la composition des dyades en fonction de leur niveau initial, dans la situation d'évaluation mutuelle. Quel degré d'hétérogénéité est profitable et à qui ? Retrouverait-on ce qui a été observé pour les dyades ayant à résoudre à deux un problème ?

Conclusion

Cette analyse a permis de dégager un dispositif pour l'évaluation mutuelle qui paraît pertinent en regard de la réalisabilité et de l'autorégulation qu'il veut susciter. Toutefois, dans la classe conflictuelle, il n'a pu être mis en place que tout à fait à la fin de la séquence. La régulation des comportements des élèves a pris le pas sur la régulation des apprentissages et incité les enseignantes à utiliser d'abord des grilles de critères suscitant plus du contrôle par les pairs que de l'évaluation interactive ; ce qui ne contribuait pas forcément à atténuer les conflits. Dans les autres cas, on peut considérer le dispositif comme relativement efficace au niveau de l'amélioration des performances, mais pour en juger vraiment une analyse comparative serait nécessaire. Il permet toutefois de rendre les élèves de niveau fort et certains de niveau moyen plus autonomes et ainsi de dégager du temps pour des interventions de l'enseignant auprès des autres élèves. Il s'avère décevant pour un apprentissage à plus long terme. On peut toutefois le considérer comme prometteur s'il est repris pour d'autres séquences et que les élèves se rôdent à sa pratique. Pour provoquer, au-delà de l'amélioration des performances, une réflexion métacognitive des élèves, il faudrait le compléter par des entretiens enseignant-élève visant à faire verbaliser les prises de conscience.

La question de l'équité amène à se préoccuper de la composition des dyades, qui n'a pas encore été bien étudiée dans ce type de dispositif. Toutefois, il a une dimension compensatoire des inégalités de capacités initiales en permettant à des élèves de niveau faible et moyen de rejoindre les performances des élèves de niveau fort.

Références bibliographiques

Allal, L. & Michel, Y. (1993). Autoévaluation et évaluation mutuelle en situation de production écrite. In L. Allal, D. Bain, P. Perrenoud. *Evaluation formative et didactique du français* (pp. 239-264). Delachaux et Niestlé.

Allal, L (1999). Impliquer l'apprenant dans le processus d'évaluation : promesses et pièges de l'autoévaluation. In C. DEPOVER & B. NOËL, *L'évaluation des compétences et des processus cognitifs* (pp. 35-56). De Boeck.

Allal, L. (2000). Acquisition et évaluation des compétences en situation scolaire (pp. 77-94). In J. Dolz &E. Ollagnier (éds), *L'énigme de la compétence en éducation*. De Boeck.

Bonniol, J.J. & Vial, M. (1997). *Les modèles de l'évaluation – Textes fondateurs avec commentaires*. De Boeck.

Bouffard-Bouchard, T. (1990). Capacité Cognitive, Sentiment d'Auto-Efficacité et Autorégulation. *European Journal of Psychology of Education*, Vol. V, n° 3, 355-364.

Bautier, E., Charlot, B., Rochex, J.Y. (2000). Entre apprentissage et métier d'élève : le rapport au savoir. In A. Van Zanten (sous la dir.), *L'école, l'état des savoirs* (pp. 179-188). La découverte.

Campanale, F. (1997). Autoévaluation et transformations de pratiques pédagogiques. *Mesure et Evaluation en Education*, Vol. 20, n°1, 1-24.

Meuret, D. (2000). L'équité en éducation selon les théories de la justice. In A. Alcouffe, B. Fourcade, J.M. Plassard, G. Tahar (coord.), *Efficacité versus équité en économie sociale :* tome 1, XX^{èmes} Journées de l'A.E.S. Paris : L'Harmattan.

Nacuzon Sall, H. & De Ketele, J.M. (1997). L'évaluation du rendements des systèmes éducatifs : apports des concepts d'efficacité, d'efficience et d'équité. *Mesure et Evaluation en Education*, Vol. 19 n° 3, 119-142.

Noël, B., Romainville, M. et Wolfs, J. L. (1995). La métacognition : facettes et pertinence du concept en éducation. *Revue Française de Pédagogie*, 112 (3), 47-56.

Vermersch, P. (1994). *L'entretien d'explicitation en formation initiale et en formation continue*. ESF.

Vial, M. (2001). *Se former pour évaluer.* De Boeck-Université.

Annexe

GRILLE D'EVALUATION MUTUELLE N°2 (conçue par A.S. Vaufredaz et M. Vivès)

Date : **Nom :** **J'évalue le monstre de :**

	Oui	**Non**	**Que proposes-tu pour améliorer ?**
Caractéristiques physiques, morales et habitudes de vie du monstre Y en a-t-il assez pour que tu puisses t'imaginer le monstre ?			
Organisation générale Est-ce que l'ensemble est organisé ?			
Quelle organisation a été choisie ?			
Du général au particulier?			
Du particulier au général ?			
Le texte comporte-t-il des **paragraphes** ? Les caractéristiques de même nature sont dans un même paragraphe : physique, habitudes de vie…			
Langue **Les phrases** sont complètes : Elles commencent par une majuscule et se terminent par un point.			
les verbes conjugués sont au présent ou au passé composé			
Variété Les **sujets** sont variés			
Les **verbes** sont variés			
Les **adjectifs** sont variés			
Il y a des **comparaisons** (des dents comme des couteaux)			