

Auto-évaluation, évaluation par les pairs et évaluation par les formateurs dans un programme de formation-action

par Alain ROGER et Robert WEISZ
Professeurs à l'IAE d'Aix-en-Provence

Résumé :

L'implication de la personne elle-même ou de ses collègues de travail dans la démarche d'évaluation de la performance peut contribuer à améliorer la qualité de l'évaluation, à réduire l'anxiété ou à renforcer la satisfaction et la motivation des personnes concernées, mais la plupart des études font ressortir des biais de surestimation difficiles à éviter. Pour analyser ces phénomènes, nous avons étudié la façon dont des étudiants, leurs collègues et leurs enseignants évaluaient leur performance dans une série d'exercices. Les résultats montrent que les collègues arrivent difficilement à jouer totalement leur rôle d'évaluateurs : ils donnent bien un feedback pertinent et utile, mais ont plus de mal à évaluer sous forme de note la performance de leurs collègues. Une comparaison de la performance évaluée par les groupes d'observateurs avec une auto-évaluation faite par le formé ou avec une évaluation par des enseignants observateurs extérieurs montre une nette tendance à la surestimation, surtout dans l'évaluation faite par les pairs. Il est important de bien apprécier les avantages et les limites de ces approches avant de les mettre en application dans les établissements d'enseignement comme dans les entreprises.

L'une des questions essentielles concernant l'évaluation de la performance des individus est celle des acteurs impliqués dans cette évaluation. Le supérieur hiérarchique dans les entreprises, l'enseignant dans les organismes de formation ont généralement pour rôle de faire cette évaluation. Beaucoup fondent leur identité en grande partie sur l'affirmation de leur statut. Il leur est alors difficile de déléguer et de faire confiance tout en restant responsables des résultats. Toutefois, certains considèrent qu'il est important d'impliquer la personne elle-même dans sa propre évaluation. Les collègues de travail sont également parfois impliqués dans la démarche. Ces approches peuvent contribuer à améliorer la qualité de l'évaluation de la performance, à réduire l'anxiété ou à renforcer la satisfaction et la motivation des personnes concernées.

Dans les entreprises, des systèmes de ce type sont parfois présentés comme des sortes de remèdes miracles. Pourtant, l'auto-appréciation ou l'appréciation par les pairs sont encore peu utilisées. Thomas et Bretz (1994) constatent que, dans 70 grandes entreprises américaines, moins de 5% de l'information utilisée réellement pour évaluer la performance finale du

personnel vient des personnes elles-mêmes ou de leurs collègues. «Malgré les arguments théoriques et pratiques sur l'utilité de l'appréciation par les pairs et de l'auto-appréciation, et malgré les affirmations des journaux selon lesquelles cette forme d'appréciation connaîtrait un "boom", on constate qu'elle est peu appliquée et qu'elle n'a que peu d'impact sur l'évaluation des performances dans ces grandes organisations industrielles»(p. 31).

L'exemple de la mise en place d'une procédure d'auto-évaluation dans un programme de formation nous permet d'illustrer certains biais qui limitent le recours à ce type de système dans les entreprises. Après une brève revue de la littérature sur l'intérêt et les limites du recours à différents évaluateurs, nous précisons comment l'évaluation des individus est abordée dans le domaine de la formation, puis nous présenterons le contexte et les résultats de l'étude.

1. Intérêt et limites du recours à différents évaluateurs

1.1 Le supérieur hiérarchique

Le supérieur hiérarchique est généralement considéré, dans les entreprises, comme l'évaluateur naturel de la performance de ses subordonnés. Il s'agit d'une des composantes principales de son rôle. Pourtant, même si les cadres considèrent que leur entreprise a un bon programme d'évaluation, beaucoup sont insatisfaits de l'évaluation que leur supérieur hiérarchique porte sur leur performance (Hammer, 1977) : ils la considèrent comme trop subjective et ils n'ont pas toujours confiance dans le jugement porté. Dans une étude précédente, nous avons montré que l'évaluation d'enseignants par des inspecteurs pédagogiques était très liée à l'évaluateur (certains privilégiant l'autorité, d'autres les connaissances techniques ou la prévision), mais aussi à l'évalué (selon leur âge et leur sexe) et au contexte de l'évaluation.

Ilgen et Favero (1985) soulignent que les interactions qui suivent l'appréciation et les conséquences qui peuvent en résulter conduisent souvent les évaluateurs à modifier plus ou moins consciemment leur jugement. L'évaluation traduit aussi parfois le désir des supérieurs d'améliorer la performance autant que leur souci de juger la personne qu'ils évaluent. Des sentiments de menace ou de culpabilité peuvent biaiser leurs évaluations : ils peuvent inconsciemment sous-évaluer un subordonné très compétent, mais source d'anxiété et de jalousie, ou surévaluer un autre pour éviter de lui communiquer une évaluation négative. Parfois, ces «erreurs» sont même conscientes et délibérées avec l'objectif, par exemple, d'améliorer l'image du service (Campbell et Lee, 1988).

De Vries et coll. (1981) font ressortir que le supérieur hiérarchique n'est pas toujours dans la meilleure position pour apprécier la performance d'un salarié : il a parfois peu de contacts avec lui, connaît mal les aspects spécifiques de son poste, et il peut être biaisé par ses propres préoccupations et par les conséquences que peut avoir l'appréciation. Trois raisons au moins font que le chef est mal placé pour évaluer les membres de son unité (ou le formateur pour évaluer ses étudiants) :

- la valeur de ses subordonnés dépend de son action (et de son style d'animation)
- son évaluation peut être arbitrairement influencée par ses propres intérêts, en jeu dans la situation
- dans l'entretien, son rôle hiérarchique peut faire qu'il apparaît particulièrement intimidant aux intéressés (Lemaître, 1983).

Dans une enquête auprès de 60 cadres dans sept grandes entreprises américaines, Longenecker et Goff (1992) font ressortir que les deux premières causes de l'inefficacité de l'appréciation des performances sont le manque d'informations de l'évaluateur sur la performance réelle du subordonné et l'attitude négative ou défensive de l'évalué. L'auto-évaluation et l'évaluation par les pairs peuvent, sinon éliminer, du moins limiter ces causes d'inefficacité.

1.2L'auto-évaluation

L'auto-appréciation est un complément utile à l'évaluation faite par le supérieur, en particulier si elle est orientée vers le développement plutôt que vers la sanction. Elle peut aider les personnes à améliorer leur performance future au travail. L'auto-appréciation est surtout utile lorsqu'elle apporte des informations que d'autres méthodes n'auraient pas permis de rassembler. Les recherches antérieures montrent que, par rapport à l'évaluation du supérieur, l'auto-appréciation présente deux biais importants : une plus grande indulgence et un éventail de notation plus fermé. Selon Campbell et Lee (1988), les différences entre l'auto-appréciation et celle du supérieur peuvent être expliquées par trois causes générales :

- les contraintes «informationnelles» liées au niveau de connaissance de ce que requiert le poste

- les contraintes «cognitives» liées à la façon de traiter les informations

- les contraintes «affectives» liées aux mécanismes de défense mis en oeuvre dans les situations d'évaluation.

Roberts (1994) souligne que la participation de l'intéressé au processus d'appréciation de sa performance améliore sa satisfaction, son acceptation et sa confiance dans le système. L'auto-appréciation devrait permettre de réduire les comportements de défense, d'améliorer la performance et d'augmenter l'accord entre l'évaluateur et l'évalué sur cette performance. Lawrie (1989) précise que, «si les salariés évaluent leur propre performance, il y a plus de chances pour 1-qu'ils croient en cette évaluation, et 2-qu'ils adoptent une attitude responsable et proactive envers leur travail» (p. 22). Toutefois, les conditions ne sont pas toujours réunies dans les entreprises pour que cette procédure fonctionne dans de bonnes conditions : les auto-appréciations sont particulièrement adaptées aux salariés qui ont un travail assez isolé ou qui possèdent une aptitude particulière, car ces salariés ont plus d'informations que n'importe qui d'autre sur leur propre comportement (Latham et Wexley, 1981). Sire (1990) souligne que «la principale difficulté de mise en oeuvre réside finalement dans la gestion des contradictions entre responsabilisation et contrôle hiérarchique d'une part, responsabilisation et prise de risque d'autre part». Sur la base notamment de l'audit d'un système de rémunération mis en place pour les cadres et ETAM d'une unité de la Lyonnaise des Eaux, il fait ressortir trois conditions principales d'efficacité pour une procédure d'auto-évaluation : 1-une entreprise participative et un système de communication adéquat, fondé sur la confiance; 2-une appropriation du système par les salariés, liée à des actions de sensibilisation et de suivi; et 3-un niveau de rétribution extrinsèque adéquat, considérée comme motivante par les intéressés. Pour lui, l'auto-évaluation n'est possible que si l'implication et la fidélité à l'organisation est forte : la délégation de pouvoirs qu'elle induit implique d'ailleurs qu'elle est aujourd'hui quasiment réservée dans les entreprises aux cadres dirigeants. Sa généralisation suppose en effet une remise en cause de la structure hiérarchique traditionnelle, avec une autonomie réelle de décision, et une absence de contrôle systématique (Sire, 1993).

Les recherches donnent des résultats contrastés sur la différence (positive ou négative) entre les appréciations faites par les personnes elles-mêmes et celles de leurs supérieurs. Pour Lemaître (1983), l'intéressé lui-même est plus sévère que ne le serait son chef, sauf dans certains cas rares (de surestimation systématique). Par contre, Hamner (1977) rapporte les résultats de plusieurs études montrant que la plupart des managers (90% à la General Electric) à qui l'on demande une auto-évaluation tendent à évaluer leur propre performance au dessus de la moyenne. Mc Farlane Shore et Thornton (1986) trouvent que les ouvriers d'assemblage d'une usine d'électronique américaine ont tendance à évaluer leur performance plus haut que ne le font leurs responsables hiérarchiques, et ce quel que soit le sexe de l'évalué et de son supérieur. Cette tendance serait due à l'estime de soi ou à l'estime des autres que recherche la personne à travers l'évaluation qui est faite. Inversement, l'évalué pourrait avoir tendance à se dévaloriser en se présentant de façon plus modeste pour assurer une auto-protection.

London et Stumpf (1982) expliquent que l'auto-appréciation améliore généralement la compréhension et la conscience de soi, mais que, généralement, «les gens ont tendance à s'évaluer eux-mêmes plus haut qu'ils ne sont évalués par les autres» (p. 34), même s'ils sont plus capables que les autres d'analyser les composantes de leur propre performance. Meyer et al. (1965) considèrent que «le fait que presque chacun pense qu'il a une performance supérieure à la moyenne est probablement à l'origine de la plupart de nos problèmes concernant les plans de rémunération au mérite... La majorité des gens pense qu'ils sont l'objet d'une discrimination car, bien sûr, l'encadrement ne reconnaît pas leur vraie valeur»(p. 13).

1.3 L'évaluation par les pairs

Les collègues de travail sont souvent en meilleure position que le supérieur pour voir l'employé tel qu'il est (Bazinet, 1980). Les appréciations des pairs sont fondées sur les interactions journalières qu'ils ont avec l'évalué et utilisent des jugements multiples. Elles sont généralement plus fiables et plus valides que les évaluations des supérieurs, souvent entachées de subjectivité et de biais (Latham et Wexley, 1981). Ces systèmes sont pourtant encore assez rarement appliqués en entreprise, mais quelques exemples témoignent de leur intérêt : l'entreprise de constructions électriques Peintamelec a mis au point un système dans lequel chaque salarié (anonymement) note et place ses collègues de travail dans une grille. Une commission composée pour trois quart de membres de la direction, tranche ensuite à partir des avis émis (Gestion sociale, 1987). Amado (1982) relate le cas d'une entreprise multinationale dans laquelle des pairs choisis par la personne évaluée participent à la procédure d'évaluation. Ils constatent que «la surévaluation que craignaient les expérimentateurs ne s'est pas produite, ce qui semble signifier que la liberté laissée à l'évalué de choisir ses appréciateurs ne conduit pas ceux-ci nécessairement à se vivre comme "avocats de la défense"»(p. 14).

Pourtant, Fox et al. (1994) rapportent des études selon lesquelles les évaluations des collègues seraient approximativement à un quart d'écart-type au dessus de celles des supérieurs hiérarchiques, et les auto-évaluations seraient à un demi écart-type au dessus de celles des supérieurs hiérarchiques. Meyer (1980) trouve aussi que l'évaluation de 92 ouvriers faite par leurs collègues dans une entreprise américaine est inférieure à leur auto-évaluation.

Besson (1988) considère qu'une appréciation réciproque est plutôt adaptée à des entreprises où le style de direction tend vers une coopération et une participation maximales des collaborateurs. «Pour l'instant, cependant, ajoute-t-il, la réciprocité formelle de l'appréciation n'est guère envisagée. Elle apparaît comme idéale, pouvant diminuer bien des craintes et des peurs mais sans doute trop perturbante puisqu'elle remettrait en cause le pouvoir du chef -et les conceptions que chacun s'en fait- dans l'entreprise»(p. 133). L'auto-appréciation, l'appréciation par les pairs ou par les subordonnés sont probablement plus appropriées dans le contexte d'un style de management favorisant une forte implication des salariés (Mohrman et al., 1989). Cependant, on observe que les pairs prennent surtout en compte dans l'évaluation de la performance le type de relation qu'ils ont avec la personne évaluée et qu'ils se laissent souvent influencer par sa popularité (Lemaître, 1983).

2. L'évaluation des individus en formation

Dans le contexte d'un programme de perfectionnement pour cadres, Hofstede (1975) montre que l'évaluation des collègues de formation permet de prédire le succès de carrière des cadres plus efficacement que l'évaluation des formateurs. Le plus grand nombre d'appréciateurs peut expliquer ce résultat, mais il est probable que le formateur prend en

compte plus que les collègues des critères tels que l'implication dans la formation, l'âge, ou le niveau hiérarchique du participant. Il est possible également que l'autorité hiérarchique, même bienveillante, du formateur biaise sa perception et influence le comportement des participants. Les collègues, qui ne sont pas dans une relation d'autorité entre eux, peuvent mieux observer et se comporter de façon plus naturelle.

Barbier (1985) insiste sur les effets que l'évaluation peut engendrer sur les individus en formation. Ce n'est pas une activité neutre: le jugement émis va participer à la formation du rapport à soi car il implique une comparaison entre les acteurs. La pratique de l'évaluation discrimine, distingue, hiérarchise. Elle peut donc contribuer à stimuler ou à inhiber la volonté des individus d'apprendre et d'évoluer.

Les approches modernes de formation-action, d'évaluation formative ou d'andragogie font ressortir l'intérêt que peut avoir, dans un contexte adapté, la responsabilisation de l'évalué, le développement de son autonomie et son implication dans le processus d'évaluation lui-même. Le rôle du formateur ou de l'enseignant est alors transformé: de juge tout-puissant, il devient un animateur et un conseiller. L'évaluation devient une auto-évaluation, considérée comme faisant partie intégrante du processus d'apprentissage. Les approches de formation-action supposent un apprentissage de la part du formé qui doit prendre suffisamment de recul et acquérir des capacités d'observation et de réflexion avant de pouvoir jouer pleinement le rôle qui lui est attribué dans une auto-évaluation de groupe (Mc Laughlin et Thorpe, 1993). La formation-action est considérée comme «un processus d'apprentissage avec et par les autres» (Mumford, 1992). Elle permet aux participants, placés à tour de rôle en position d'évaluateurs, d'arriver à une vision «meta», à un niveau supérieur d'observation et de réflexion. L'objectif n'est plus seulement d'apprendre, mais aussi d'apprendre à apprendre.

L'apprenant est considéré comme un adulte responsable, «acteur et partenaire» (De Baets, 1991). Belet (1993) montre comment les modèles d'éducation adaptés aux adultes (andragogie) tels qu'ils ont été développés aux Etats-Unis, notamment par Kowles, peuvent être adaptés en France. Ils reposent sur la motivation intrinsèque des apprenants travaillant en petits groupes sur des projets, et supposent que l'accent est mis sur le processus d'apprentissage plus que sur le contenu. Celui qui apprend évalue l'apprentissage lui-même ou avec ses collègues, et l'enseignant ne joue qu'un rôle d'aide et de conseil. Nadeau (1988) insiste sur cette démarche d'auto-évaluation, précisant même qu'il est nécessaire «de donner à l'adulte en situation d'apprentissage la possibilité d'élaborer les critères qui mesurent ses progrès». Quel que soit le programme, la plus grande efficacité est en général obtenue par des méthodes qui favorisent une implication importante des apprenants. Nous avons souligné par ailleurs (Weisz, 1988) l'intérêt d'une participation ou d'une négociation des acteurs dans les processus qui les concernent: ils ont alors toutes les chances d'être mieux acceptés et plus efficaces.

Pour étudier les réactions individuelles face à ces processus d'évaluation, nous avons suivi la mise en place d'une procédure d'évaluation formative dans un programme universitaire de troisième cycle.

3. Le contexte de l'étude

Dans un établissement d'enseignement supérieur de gestion, plusieurs expériences ont été faites visant à impliquer les étudiants dans l'évaluation de leurs collègues. Une première, plus limitée, concerne les étudiants d'un cours de Méthodes qualitatives de recherche sur le terrain en DEA; une seconde, plus ambitieuse, concerne les étudiants d'un cours de Développement Relationnel.

Dans le cours de méthodes qualitatives, les étudiants doivent présenter en fin de trimestre un exposé présentant les résultats d'une étude sur le terrain qu'ils ont réalisée individuellement ou en binôme. Pour impliquer leurs collègues dans l'évaluation de cet exposé, il leur

est demandé de remplir chacun une grille d'évaluation de la performance selon une grille de critères établie par l'enseignant, et reprenant, outre la qualité de la présentation, les principaux critères d'une bonne étude de terrain abordés en cours (rigueur du recueil des données, distinction entre les faits et les opinions, recul, ...). La notation de chaque critère est faite sur une échelle en cinq points (de «pas du tout d'accord» à «tout à fait d'accord»). L'évaluation se termine par une question ouverte demandant des commentaires et suggestions pour aider l'étudiant dans la rédaction du rapport écrit qu'il doit rendre par la suite. La note de l'exposé compte pour un quart de la note globale du cours (un quart étant réservé à la participation en cours, et la moitié au rapport écrit). De son côté, l'enseignant évalue également les étudiants sur la même grille, mais il prend largement en compte pour la note de l'exposé l'évaluation faite par les collègues. Deux groupes de DEA ont été évalués par cette procédure en 1994 et 1995 par deux enseignants différents.

L'évaluation de la performance des étudiants du cours de Développement Relationnel est fondée sur une approche participative fondée sur l'autonomie et l'implication de chacun. Dans cette formation, le contexte favorable à une intériorisation des normes par les étudiants permet de leur demander d'évaluer la performance de leurs pairs en jouant eux-mêmes le rôle d'évaluateurs. La procédure d'intégration en groupe devient leur procédure.

Pour préparer l'évaluation, un groupe d'étudiants volontaires met au point chaque année les modalités d'évaluation avec l'aide de l'enseignant responsable : organisation, grille de critères, type d'exercices, etc. Le déroulement est toujours fondé sur l'analyse par un petit groupe d'environ quatre collègues («observateurs») de la façon dont un étudiant (le «praticien») a réalisé un exercice tiré au sort parmi un ensemble d'exercices mis à sa disposition par l'enseignant responsable. Le groupe s'organise pour que chacun ait l'occasion de réaliser devant ses collègues trois exercices durant les trois jours de cette «intégration». Il s'agit donc d'une auto-évaluation de groupe, tournante, dans laquelle chacun est successivement évaluateur et évalué¹. A l'issue de chaque exercice, un feed-back est prévu, dont la durée est au moins égale à celle de l'exercice lui-même. Huit enseignants «observateurs» impliqués dans le programme assistent à certains exercices et peuvent, si besoin est, aider et conseiller les groupes, mais en aucun cas ils ne sont en position d'intervenants. La note définitive résulte de la grille d'appréciation remplie par le groupe de collègues observateurs. Cette grille comporte neuf critères (adaptation relationnelle, gestion du temps, connaissance des concepts et des outils, professionnalisme, déontologie, ...). L'évaluation est faite sur une échelle allant de A («très bon») à E («très perfectible»), avec possibilité de moduler chaque niveau par un (+) ou un (-). Comme toutes les notes étaient au moins égales à C- («moyen-»), nous avons utilisé dans notre analyse une notation de 1 (C-) à 9 (A+). En 1994, il était demandé également à l'étudiant lui-même d'évaluer sur la même grille sa propre performance, et aux enseignants d'évaluer les séances auxquelles ils avaient assisté. Nous disposions alors, pour un exercice donné, de l'auto-évaluation de sa performance par l'étudiant, de l'évaluation de ses collègues, et de celle de l'enseignant. Seule la note donnée par les collègues était prise en compte dans le contrôle des connaissances.

Dans une première étude (Roger et Weisz, 1994), nous avons étudié les réactions des participants aux différentes formes d'évaluation. Les participants souhaitent une évaluation qui ne soit pas unilatérale, qui inclue un feed-back avec des commentaires rapides, et qui leur donne une occasion supplémentaire de se former. Ils recherchent une évaluation constructive

1. Les étudiants de ce programme (à temps plein sur un an) sont en majorité des étudiants inscrits en première formation, mais environ un tiers des répondants suivent la formation au titre de la formation continue. La moyenne d'âge est de 27 ans. Les femmes représentent plus des deux tiers du groupe. Tous ont une formation supérieure préalable de niveau bac + 4 et ont été sélectionnés sur dossier, test et entretien parmi un nombre de candidats important (moins d'un admis sur quatre). Pour l'évaluation, ils devaient choisir en 1994 un exercice dans chacun des trois pôles suivants : 1 - «Coaching» ou «Prise de décision», 2 - «Négociation», «Médiation» ou «Conduite de réunion», 3 - «Méta-miroir» ou «Stratégie d'objectif»

qui réponde mieux à leurs objectifs. A la question (posée avant les exercices) demandant si les procédures d'évaluation devraient plutôt permettre une évaluation par le formateur ou une auto-évaluation, ils répondent qu'il doit y avoir un équilibre entre les deux, tout en constatant qu'aujourd'hui, l'évaluation est plutôt centrée sur le formateur.

Par opposition à des évaluations traditionnelles fondées sur la sanction et la contrainte, l'auto-évaluation de groupe semble favoriser le développement professionnel et personnel des formés dans une ambiance dynamique et positive. Elle est perçue comme très proche des souhaits exprimés, notamment par l'implication qu'elle entraîne et par son caractère formateur. Par rapport à des évaluations traditionnelles, la procédure est perçue comme moins stressante et plus équitable. Ces résultats confirment largement l'impression générale que l'on pouvait avoir en observant les groupes pendant ces trois journées d'intégration, une forte implication des acteurs dans un climat à la fois de sérieux et de convivialité.

4. Les résultats

Le tableau 1 présente, pour chacun des critères d'évaluation et pour chacun des deux groupes d'étudiants de Méthodes qualitatives en DEA, la moyenne des évaluations faites par les collègues, les variations observées entre ces évaluations (moyenne des écarts à la moyenne du groupe), l'évaluation de l'enseignant, et la moyenne des écarts entre les évaluations de l'enseignant et celles des collègues (d'abord en compensant les écarts positifs et négatifs, puis en prenant les écarts en valeur absolue). Il confirme la tendance notée dans la littérature à une évaluation plus généreuse de la part des collègues que de l'enseignant (dont la position peut être assimilée à celle d'un supérieur hiérarchique). Cette tendance est très marquée pour l'enseignant du groupe 2 qui note nettement plus bas que les collègues de l'étudiant (écart moyen de plus de un point sur une échelle de cinq : 2,66 contre 3,68, $p < 0,001$). Par contre la différence entre les évaluations de l'enseignant du groupe 1 et celles des collègues n'est pas significative.

Tableau 1 : synthèse des évaluations du cours de Méthodes qualitatives au DEA

	Groupe 1					Groupe 2				
	Moyenne des collègues	Variations entre collègues	Enseignant	Moyenne des écarts Enseign. Collègues	Moyenne Ecarts (val. abs) Enseign. Collègues	Moyenne des collègues	Variations entre collègues	Enseignant	Moyenne des écarts Enseign. Collègues	Moyenne Ecarts (val. abs) Enseign. Collègues
rigueur	4.32	0.56	4.33	0.02	0.36	4.26	0.62	2.64	-1.55	1.59
choix des thèmes	4.22	0.60	4.56	0.33	0.51	3.92	0.74	2.73	-1.21	1.37
faits	4.01	0.72	3.56	-0.46	0.52	3.66	0.91	2.64	-0.99	1.25
recul	4.17	0.87	4.00	-0.15	0.24	3.67	0.85	3.18	-0.64	0.99
identification des facteurs importants	3.89	0.75	3.78	-0.09	0.42	3.59	0.87	3.00	-0.47	0.57
interprétations *	3.01	1.22	2.89	-0.12	0.34	3.01	1.15	2.36	-0.53	0.83
méthode inadaptée*	3.06	0.98	2.11	-0.94	1.04	2.93	1.18	1.73	-1.22	1.28
compréhension de la réalité	4.22	0.56	4.33	0.12	0.31	3.94	0.89	2.55	-1.28	1.55
clarté de la présentation	4.36	0.46	4.11	-0.25	0.42	4.14	0.62	3.09	-1.10	1.17
(moyenne)	3.92		3.74	-0.17	0.46	3.68		2.66	-1.00	1.18

* échelle inversée (une meilleure évaluation correspondait à une note plus faible)

° moyenne des écarts entre l'évaluation de chacun et la moyenne des évaluations des collègues

Une analyse plus fine montre que, sur certains critères, les écarts sont plus faibles : recul («le chercheur a pris du recul par rapport à l'organisation et n'a pas pris parti»), et interprétations («la présentation comportait beaucoup d'interprétations et de jugements de valeur - échelle inversée -); par contre, des différences plus nette apparaissent sur d'autres

critères : l'enseignant du groupe 2 est par exemple beaucoup plus exigeant que les collègues par rapport à celui du groupe 1 sur la rigueur et le choix des thèmes. La taille des groupes (13 personnes, 9 exposés dans le groupe 1; 16 personnes, 11 exposés dans le groupe 2) ne permet pas d'aller beaucoup plus loin dans l'analyse de ces résultats.

Dans la deuxième étude, qui porte sur 40 évaluations concernant 29 personnes, l'évaluation des enseignants est également inférieure, dans la plupart des cas, à l'évaluation portée par les pairs. Les différences varient beaucoup selon les enseignants : les uns sont très proches de l'évaluation des pairs (tableau 2 : enseignants C et G); d'autres en sont plus éloignés (E, F, et surtout A). L'écart moyen entre l'évaluation de l'enseignant et l'auto-évaluation varie aussi selon les évaluateurs, de +0,54 pour C (le seul à apprécier la performance au dessus du groupe de collègues) à -2,18 pour H. On remarque aussi que, globalement, l'auto-évaluation est moins éloignée de l'évaluation de l'enseignant que l'évaluation des pairs.

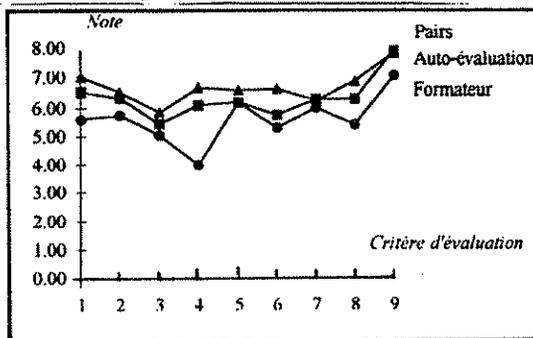
Tableau 2 : écarts entre les évaluations des différents enseignants, les évaluations des pairs et les auto-évaluations

Enseignant	Moyenne des écarts Enseignant-Pairs		Moyenne des écarts Enseignant-Auto-évaluation	
A	-2.25	(4)	-0.84	(4)
B		(1)	-1.79	(3)
C	-0.10	(3)	+0.54	(4)
D		(0)	-0.57	(3)
E	-1.75	(4)	-1.41	(4)
F	-1.33	(6)	-0.65	(8)
G	+0.09	(5)	-0.22	(7)
H		(1)	-2.18	(3)
<i>Moyenne</i>	<i>-1.26</i>	<i>(24)</i>	<i>-0.76</i>	<i>(36)</i>

(le nombre d'évaluations correspondant est indiqué entre parenthèses)

Le graphique 1 montre que les écarts varient beaucoup selon les critères d'évaluation : sur les neuf critères utilisés, certains recueillent des évaluations à peu près homogènes (5 -gestion du temps- et 7 -connaissance des concepts et des outils-). Sur d'autres, les avis divergent nettement (1 -diagnostic relationnel-, 8 -professionnalisme- et surtout 4 -gestion des processus relationnel -).

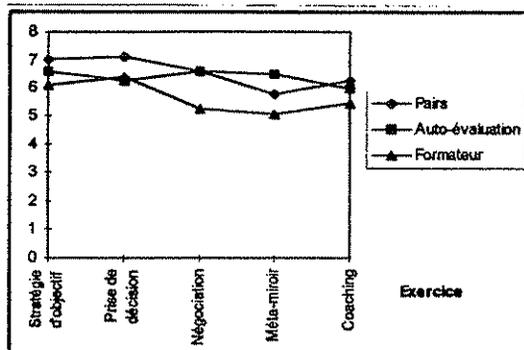
Graphique 1 : comparaison des trois formes d'évaluation pour chacun des neuf critères (critères : 1-diagnostic relationnel 2-adaptation relationnelle 3-adéquation méthode et outils à la gestion de la réalité 4-gestion des processus relationnels 5-gestion du temps 6-capacité à meta-communiquer 7-connaissance des concepts et des outils 8-professionnalisme 9-déontologie)



Les mêmes différences subsistent à peu près dans tous les exercices, sauf dans l'exercice «meta-miroir» où les collègues sont plus sévères que les individus eux-mêmes (graphique2).

Les études antérieures suggéraient que les évaluations pouvaient varier selon l'évaluateur, selon le contexte, mais aussi selon l'évalué lui-même. Le tableau3 montre que les écarts entre les auto-évaluations et celles de leurs pairs ou de leurs enseignants varient nettement selon

Graphique 2 : comparaison des trois formes d'évaluation selon le type d'exercice proposé



qu'ils sont ou non inscrits au programme en formation continue et selon leur sexe. Les étudiants en formation continue semblent s'évaluer au dessus du niveau où les évaluent les formateurs, mais aussi leurs pairs. Les hommes ont également cette tendance à s'évaluer à un niveau plus élevé.

Tableau 3 : écarts entre les évaluateurs selon l'origine et le sexe des évalués

	Global	Première formation	Formation continue	Hommes	Femmes
Formateur > auto-évaluation (nombre)		8	1	2	7
Formateur < auto-évaluation (nombre)		8	9	9	9
Moy. For.-Auto	-0.71	-0.11*	-1.76*	-1.80*	-0.02*
Pairs > auto-évaluation (nombre)		13	5	5	14
Pairs < auto-évaluation (nombre)		1	5	5	2
Moy. Pairs-Auto	0.38	-1.08*	-0.70*	-0.54*	+0.96*

* différence significative au seuil de $p < 0.01$

Ces résultats doivent cependant être modérés lorsque l'on s'aperçoit qu'une personne contribue pour une large part à l'écart observé : contrairement à la plupart des autres étudiants, un cadre (homme) en formation continue s'est évalué lui-même largement au dessus de l'évaluation de ses pairs ou des enseignants (trois enseignants différents dans trois exercices différents). En enlevant ce cadre de l'échantillon, les résultats ne changent pas de sens, mais les différences ne sont plus significatives, sauf pour l'écart entre l'auto-évaluation et celle des pairs pour les premières formations par rapport aux formations continues ($p < 0.05$). Ce cadre est pratiquement le seul à être insatisfait de la procédure qu'il considère comme une sanction : pour lui, les pairs ne sont pas compétents pour le juger et la note qu'ils donnent n'est pas équitable.

5. Discussion et conclusion

La surestimation observée en entreprise par plusieurs auteurs (London et Stumpf, 1982, Campbell & Lee, 1988, Fox et coll., 1994) entre la performance d'une personne évaluée par ses pairs, ou son auto-évaluation, et l'évaluation de ses supérieurs hiérarchiques se retrouve ici dans la relation entre les étudiants et les enseignants. Les étudiants arrivent difficilement

à jouer totalement leur rôle d'évaluateurs sans être influencés par un souci d'indulgence, voire de complaisance, d'autant plus compréhensible qu'ils savent qu'ils seront à leur tour évalués par les mêmes personnes. Les participants, lors des évaluations de fin de programme, insistent aussi sur trois points principaux :

- ils accordent une grande place aux relations et ont du mal à ne pas privilégier les aspects affectifs dans leurs évaluations (comme le soulignent Hofstede, 1975, Lemaître, 1983 ou Campbell et Lee, 1988)

- la partie la plus formatrice pour eux est d'avoir à donner un feed-back et d'avoir à évaluer

- ils n'hésitent pas à dire ce qu'ils pensent dans le feed-back, mais ils ne se sentent pas légitimés pour noter leurs collègues.

Les jugements que les individus portent sur leurs propres actions dépendent beaucoup de l'image qu'ils se font d'eux-mêmes : «certains sont facilement satisfaits tandis que d'autres se considèrent en perpétuel échec; certains surévaluent leurs capacités alors que d'autres les sous-évaluent» (Sekiou et al., 1992, p. 348). Monteil (1992) note, en se référant aux travaux de Festinger sur la dissonance cognitive, que l'évaluation, dans un processus de comparaison sociale, risque d'entraîner un sentiment d'infériorité pour l'évalué. Le sujet ainsi «infériorisé» adopte alors une stratégie de comportement (démission, émulation, isolement ou marginalisation) pour préserver une image de soi positive. Contrairement aux études de Meyer (1980) et de Fox et al. (1994) qui trouvent que les auto-évaluations sont supérieures à celles des pairs, nos résultats font ressortir globalement une plus grande indulgence chez les collègues que dans l'auto-évaluation. Pour les cadres en formation continue, qui ont peut-être plus confiance en eux et sont plus soucieux de leur image, l'auto-évaluation dépasse par contre l'évaluation des pairs.

De Vries et al. (1981) font ressortir que chaque évaluateur valorise plus ou moins les différents critères de performance : par exemple, le supérieur attache plus d'importance aux connaissances et aux aspects techniques du travail; les pairs et les personnes elles-mêmes ont tendance à accorder plus de poids aux aptitudes interpersonnelles. Ainsi, il est difficile de dire qui fait l'évaluation la plus exacte : des évaluations faites avec des points de vue différents ont chacune leur validité. L'exemple des formations que nous avons étudiées montre bien que les critères d'évaluation sont valorisés différemment par les différents acteurs. Certains critères sont évalués de façon plus homogène, probablement parce qu'ils donnent moins lieu à interprétation (gestion du temps, connaissances des concepts et des outils). Sire (1993) insiste sur le fait que l'auto-évaluation repose sur la précision des critères utilisés, sur la qualité des objectifs définis et sur la qualité des processus de feed-back mis en place. L'étude réalisée par Fox et al. (1994) dans un centre de formation de policiers en Israël montre que certaines précautions, comme la définition d'un type d'échelle ou le choix de critères adaptés, peuvent réduire les biais de l'auto-évaluation.

L'utilisation de plusieurs sources d'information peut aussi améliorer la validité de la procédure. Un cabinet américain tente par exemple d'introduire en France «l'évaluation à 360°» : à partir des appréciations du salarié, mais aussi de son supérieur hiérarchique, de ses collègues et de ses collaborateurs, «on obtient ainsi, sans risques, une vision panoramique de ses comportements, de ses manières d'être et de faire» (Girard, 1996, p. 66). Constatant que les gens ont tendance à être trop durs ou trop indulgents envers eux-mêmes, Lawrie (1989) conseille de compléter l'auto-évaluation par une appréciation de pairs, dont l'avis peut être utile, et du supérieur hiérarchique qui peut apporter une aide et des conseils. Ce procédé est long, lourd, complexe, coûteux en temps (Lemaître, 1983), et il exige une grande maturité et de bonnes relations dans l'équipe (Bazinet, 1980), mais quand les conditions favorables sont réunies dans une entreprise, l'évaluation provenant de plusieurs sources est plus objective et plus complète.

Dans le cadre d'une formation, l'auto-évaluation de groupe n'est pas la panacée qui pourrait résoudre comme par miracle tous les problèmes d'évaluation. Outre les conditions

de mise en oeuvre d'une telle démarche que nous avons déjà mentionnées, il faut en ajouter deux qui limitent beaucoup sa généralisation. La première est le temps nécessaire pour organiser, puis pour mettre en place le processus. Même si il fait partie intégrante de l'enseignement et dispense l'enseignant de la correction de copies, il faut qu'il dispose du temps et des locaux nécessaires, ce qui peut s'avérer difficile, notamment lorsque la taille du groupe est importante. La deuxième limite est l'acceptation par l'enseignant d'une certaine perte de pouvoir. Beaucoup d'enseignants fondent leur identité en grande partie sur la différenciation avec leur auditoire et sur l'affirmation de leur statut. Il leur est alors difficile de déléguer et de faire confiance tout en restant responsables des résultats. La démarche appliquée ici dans un contexte d'éducation en milieu universitaire correspond pour l'essentiel à celles qui sont menées dans les organisations soucieuses de promouvoir le changement en profondeur, et qui ont le souci, à partir d'une vision partagée, de responsabiliser les gens et d'accélérer leur apprentissage.

Bibliographie

- Amado G., Entretien d'évaluation ou multi-appréciation, *Personnel*, n° 238, janv. 1982
- Barbier J.M., *L'évaluation en formation*, PUF, 1985
- Besson B., *L'appréciation du personnel*, Chotard et ass., 1988
- Bazinet A., *L'évaluation du rendement*, Agence d'Arc, 1980
- Belet D., Gallic pedagogics and anglo andragogics, *EFMD Quarterly Review*, 1993
- Campbell D.J. & Lee, C., Self-appraisal in performance evaluation : development versus evaluation, *Academy of Management Review*, Vol. 13, n° 2, 1988, p. 302-314
- De Baets P., La formation : performante; l'apprenant responsable, *Gestion 2000*, 6-1991
- De Landsheere G., EPS interroge Gilbert De Landsheere, *E.P.S.*, n° 222, mars-avril 1990
- De Vries D.L., Morrison, A.M., Schullman, S.L., Gerlach, M.L., *Performance appraisal on the line*, Wiley, 1981
- Fox S., Caspy T. & Reisler A., Variables affecting leniency, halo and validity of self-appraisal, *Journal of Occupational and Organizational Psychology*, Vol. 67, 1994, p. 45-56
- Gestion Sociale*, Quand les salariés notent leurs collègues, n° 142, 13 juillet 1987
- Girard B., Evaluation des performances: méthodes américaines pour entreprises françaises, *Liaisons Sociales Mensuel*, avril 1996, p. 66
- Hammer W. C., How to ruin motivation with pay, dans Hackman J.R., Lawler E.E. III & Porter L. W., *Perspectives on behavior in organizations*, Mc Graw Hill, 1977
- Hofstede G., Predicting managers' career success in an international setting: the validity of ratings by training staff versus training peers, *Management International Review*, Vol. 15, 1/1975, p. 43-50
- Ilggen D.R. & Favero J.L., Limits in generalization from psychological research to performance appraisal processes, *Academy of Management Review*, Vol. 10, n° 2, p. 311-321, 1985
- Latham G.P. & Wexley K.N., *Increasing productivity through performance appraisal*, Addison-Wesley, 1981
- Lawrie J.W., Your performance : appraise it yourself !, *Personnel*, Jan. 1989, p. 21-23
- Lemaître P., *Appréciation du personnel et entretien de bilan*, Editions d'Organisation, 1983
- London M. & Stumpf S.A., *Managing careers*, Addison Wesley, 1982
- Longenecker C.O. & Goof S.J., Performance appraisal effectiveness : a matter of perspective, *SAM Advanced Management Journal*, Spring 1992, p. 17-23
- Mallet J., Dynamique des systèmes ouverts, politique qualité et politique formation dans l'entreprise, *Journal de la formation continue et de l'EAO*, n° 258, juin 1991
- McFarlane Shore L. & Thornton G.C., III., Effects of gender on self-and supervisory ratings, *Academy of Management Journal*, Vol. 29, n° 1, p. 115-129
- Mc Laughlin H. & Thorpe R., Action Learning -a paradigm in emergence : the problem facing a challenge to traditional management education and development, *British Journal of Management*, Vol. 4, 19-27, 1993
- Meyer H.H., Self-appraisal of job performance, *Personnel Psychology*, Vol. 33, n° 2, sum 1980, p. 291-295
- Meyer H.H., Kay, E. & French, J.R.P., Split roles in performance appraisals, *Harvard Business Review*, Jan. -Feb. 1965

Mohrman A.M. jr., Resnick-west S. & Lawler E.E.,III, Designing performance appraisal systems, Jossey-Bass, 1989

Monteil B., Eduquer et former, EO, 1992

Mumford A., Apprendre par l'action, Personnel, n° 333, juin 1992

Nadeau J.-R., L'andragogie existe-t-elle ?, Personnel, n° 293, février 1988

Roberts G.E., Maximizing performance appraisal system acceptance : perspectives from municipal government personnel administrators, Public Personnel Management, Vol. 23, n° 4, winter 1994, p. 525-549

Roger A. et Roger J., L'évaluation des enseignants : les composantes de l'appréciation et leurs déterminants, Cahiers de Recherche n° WP 312, IAE d'Aix-en-Provence, 1985

Roger A. et Weisz R., L'auto-évaluation de groupe comme processus d'apprentissage, Cahiers de Recherche n° WP 438, IAE d'Aix-en-Provence, 1994

Sekiou L., Blondin L., Fabi B., Chevalier F. & Besseyre des Horts C.H., Gestion des Ressources Humaines, Editions 4L, 1992

Sire B., Auto-fixation des objectifs et auto-évaluation des performances dans les systèmes de rémunérations individualisées : fondements théoriques et conditions d'application, Perspectives de l'entreprise et recherche en GRH, les défis des années 1990, 1° Congrès de l'AGRH, Bordeaux, 1990

Sire B., Gestion stratégique des rémunérations, Editions Liaisons, 1993

Thomas S.L. & Bretz R.D. Jr., Research and practice in performance appraisal : evaluating employee performance in America's largest companies, SAM Advanced Management Journal, spring 1994, p. 28-34

Weisz R., Efficacité du travail : fausses pistes et vrais enjeux, Personnel, n° 293, février 1988