

L'analyse individuelle de concordance des jugements auto- et allo-évaluatifs portés sur une performance complexe : un outil diagnostique au service d'un *feed-back* formatif

Véronique Jans

Volume 27, numéro 2, 2004

Auto-évaluation en situation de formation

URI : <https://id.erudit.org/iderudit/1087975ar>

DOI : <https://doi.org/10.7202/1087975ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (imprimé)

2368-2000 (numérique)

[Découvrir la revue](#)

Citer cet article

Jans, V. (2004). L'analyse individuelle de concordance des jugements auto- et allo-évaluatifs portés sur une performance complexe : un outil diagnostique au service d'un *feed-back* formatif. *Mesure et évaluation en éducation*, 27(2), 9-31.
<https://doi.org/10.7202/1087975ar>

Résumé de l'article

Dans le souci d'affiner l'évaluation des performances complexes des étudiants (universitaires dans notre cas), nos travaux de recherche s'intéressent, d'une part, à la problématique de l'auto-évaluation (à partir de grilles de critères) et, d'autre part, à la confrontation instrumentée (graphique et avec calcul d'indices) des jugements auto- et allo-évaluatifs émis à propos d'une performance complexe. L'analyse de concordance des jugements proposée dans cet article repose sur une telle confrontation, puisqu'elle compare les évaluations (émises par l'étudiant et par l'enseignant) à partir d'une représentation graphique des écarts, d'indices numériques et des justifications écrites par l'étudiant et l'enseignant sur la grille de critères. Une telle analyse individuelle favorise la préparation de *feed-back* personnalisés à l'intention des étudiants. Nous l'avons utilisée sur le terrain comme base à une discussion formative entre l'enseignant et chaque étudiant.

L'analyse individuelle de concordance des jugements auto- et allo-évaluatifs portés sur une performance complexe: un outil diagnostique au service d'un *feed-back* formatif

Véronique Jans

Université de Liège

MOTS CLÉS: Auto-évaluation, confrontation des jugements, *feed-back*, outil diagnostique, performance complexe

*Dans le souci d'affiner l'évaluation des performances complexes des étudiants (universitaires dans notre cas), nos travaux de recherche s'intéressent, d'une part, à la problématique de l'auto-évaluation (à partir de grilles de critères) et, d'autre part, à la confrontation instrumentée (graphique et avec calcul d'indices) des jugements auto- et allo-évaluatifs émis à propos d'une performance complexe. L'analyse de concordance des jugements proposée dans cet article repose sur une telle confrontation, puisqu'elle compare les évaluations (émises par l'étudiant et par l'enseignant) à partir d'une représentation graphique des écarts, d'indices numériques et des justifications écrites par l'étudiant et l'enseignant sur la grille de critères. Une telle analyse individuelle favorise la préparation de *feed-back* personnalisés à l'intention des étudiants. Nous l'avons utilisée sur le terrain comme base à une discussion formative entre l'enseignant et chaque étudiant.*

KEY WORDS: Self-assessment, confrontation of judgments, feedback, diagnostic tool, complex performances

In order to evaluate with accuracy complex cognitive performances of students (graduate students in this case), our research focuses, on the one hand, on self-assessment (based on criteria grids) and, on the other hand, on instrumented confrontation (graph with index calculations) of self and external assessments concerning a complex performance. This article presents an analysis of the degree of agreement found in the confrontation of these two judgements. It compares the assessment judgements (expressed by the student and by the teacher) based on a graphical representation of discrepancies, on numerical indicators and on justifications written by the student and by the teacher on the criteria grid. This type of individual analysis can be used to prepare personalized feedback for the benefit of the students. It was used in a university course as a basis for a formative discussion between the teacher and each student.

PALAVRAS-CHAVE: Auto-avaliação, confrontação dos juízos, *feed-back*, utensílio diagnóstico, *performance* complexa

Com o objectivo de melhorar a avaliação das performances complexas dos alunos (universitários no nosso caso), os nossos trabalhos de investigação interessam-se, por um lado, com a problemática da auto-avaliação (a partir de grelhas de critérios) e, por outro lado, com a confrontação instrumentada (gráfico e cálculo de índices) dos juízos auto e allo-avaliativos emitidos a propósito de uma performance complexa. A análise da concordância dos juízos, proposta neste artigo, assenta nessa confrontação, uma vez que compara as avaliações (emitidas pelo aluno e pelo professor) a partir de uma representação gráfica dos desvios, dos índices numéricos e das justificações escritas pelo aluno e pelo professor na grelha de critérios. Esta análise individual favorece a preparação de feed-back personalizado à disposição dos alunos. Utilizámo-lo no terreno como suporte a uma discussão formativa entre o professor e cada aluno

Contexte : les enjeux actuels de la pédagogie universitaire

À juste titre, l'étudiant universitaire d'aujourd'hui revendique une formation de qualité ainsi qu'un « droit de regard » sur ses apprentissages. La révolution copernicienne, qui a déjà touché d'autres ordres d'enseignement et certains secteurs de l'enseignement supérieur (en formation médicale par exemple), est en marche pour tout l'enseignement supérieur.

La recherche en ingénierie de la formation universitaire, dans laquelle notre travail s'inscrit, est confrontée à un double enjeu. Si un de ses rôles consiste à élaborer et expérimenter des concepts, des outils et des modèles favorisant l'analyse des phénomènes éducatifs (enjeu de recherche), il lui appartient également de proposer des concepts, des outils et des modèles qui soient adaptables aux situations réelles vécues par les étudiants universitaires (enjeu de formation).

Beaucoup de travaux en pédagogie universitaire sont basés sur l'apprentissage « spontané » des étudiants universitaires en tentant de répondre à des questions du type : « Comment les étudiants apprennent-ils ? » En nous plaçant dans une option de formation, nos questions prennent inévitablement une tournure plus engagée, du type « Comment aider les étudiants à apprendre tel contenu ou à développer telle compétence ? Comment organiser les situations de formation favorisant ces apprentissages ? » Notre recherche est ici davantage basée sur un apprentissage soutenu, organisé, guidé.

Un intérêt d'étude : l'auto-évaluation en éducation

Nous avons choisi d'approfondir un domaine particulier de la recherche en éducation, celui de l'auto-évaluation, considérée comme *une activité qui engage l'étudiant à réfléchir à ses connaissances, à ses performances et à ses capacités d'apprentissage, dans l'optique de les développer et de les réguler si nécessaire*. Il existe à nos yeux un intérêt particulier à étudier l'auto-évaluation pour elle-même. Dans tout processus d'apprentissage, l'étudiant s'auto-évalue régulièrement, sans nécessairement s'en rendre compte. Il importe, pour la formation de l'étudiant comme pour la recherche en pédagogie universitaire, de traiter cette activité de façon plus consciente et plus systématique. L'intérêt de cette problématique de recherche nous semble indéniable puisqu'elle rencontre un des objectifs prioritaires de la formation universitaire : former des apprenants professionnels, conscients de la qualité de leurs performances, dans le but de les gérer et de planifier leurs apprentissages (Boud, 1995 ; Frenay, Noël, Parmentier & Romainville, 1998 ; Leclercq, 1998 ; Nyhan, 1991). Dans cette optique, la capacité de «se connaître en tant qu'apprenant» ou «se connaître en tant qu'auto-évaluateur» constitue une compétence de niveau stratégique importante à développer dans le cadre d'un cursus universitaire. Pour être formative, l'auto-évaluation doit, selon nous, dépasser la seule introspection personnelle. Nos travaux de recherche (Jans, 2000) s'intéressent, notamment, à la question méthodologique suivante : Comment aider l'étudiant à mieux se connaître, à pratiquer une auto-évaluation de ses performances cognitives plus «précise», plus «conforme à la réalité» ?

L'auto-évaluation de performances cognitives complexes

Notre recherche s'intéresse plus particulièrement à l'auto-évaluation des performances cognitives des étudiants (et donc peu aux performances sensori-motrices ou au vécu affectif). Romainville (1992) définit la performance comme «l'actualisation observable, sur une tâche donnée, des modifications des connaissances de l'étudiant. [...] Ces performances finales ou intermédiaires aboutissent souvent à des productions : une réponse écrite à l'examen (performance finale), un plan, un résumé (performances intermédiaires)» (p. 24). Dans notre étude, nous considérons comme *performance cognitive* une production de l'étudiant reflétant, sur une tâche donnée, son état de connaissance ou de compétence. Le terme «reflétant» sous-entend que l'on n'obtiendra pas une «copie conforme» de l'état de connaissance ou compétence de l'étudiant.

Les performances dépendent notamment de l'interprétation que font les individus de la situation d'évaluation pour « deviner » l'intention de l'interrogateur. « Alors qu'on croyait saisir une compétence personnelle [...], on saisit souvent un construit social, dont les manifestations fluctuent selon le contexte social ! » (Hadji, 1997, p. 35). Connaissance et compétence sont donc toujours inférées.

Dans nos travaux de recherche, nous avons distingué deux types de performances par rapport à l'évaluation que l'on peut en faire : d'une part, la performance *isolable*¹, qui est une performance non imbriquée dans d'autres et « strictement contrainte par la forme de la question », pour reprendre l'expression de Beckers (1997, p. 97) ; d'autre part, la performance *complexe*, qui relève d'une combinaison de critères de jugements. Ce type de performance « cherche spécifiquement à mesurer la manière dont l'étudiant conjugue et articule des connaissances, des raisonnements, des savoir-faire pour produire un tout intégré et qui lui est personnel » (Beckers, 1997, p. 97). On parle aussi dans ce cas de « performance globale » (Leclercq, 1995).

Scallon (1988, cité dans Hadji, 1997) propose quatre caractéristiques pour définir une tâche complexe : (1) la pluralité des « habiletés » requises, (2) l'autonomie de réponse laissée à l'individu, (3) la multidimensionnalité de la tâche et (4) la variabilité du produit issu de la tâche. Les aspects à prendre en compte pour l'évaluation d'une performance complexe sont multiples. Une performance complexe est généralement évaluée au moyen d'une note « subjective », intégrant une série de critères, ce qu'illustre la caractéristique de « multidimensionnalité » de Scallon.

Thèse de départ

Le contenu présenté dans le cadre de cet article s'inscrit dans un programme de recherche plus vaste, dont la thèse de départ peut se formuler de la façon suivante (Jans, 2000)² :

Il est pertinent (c'est-à-dire « axiologiquement désirable, scientifiquement souhaitable et contextuellement réalisable »³) de concevoir et d'organiser en milieu universitaire une évaluation des performances des étudiants qui soit à la fois plus diagnostique (plus précise) et plus démocratique (plus dialectique, plus négociée).

Pour rendre l'évaluation des performances plus diagnostique, nous proposons de recourir à l'auto-évaluation et à une « confrontation instrumentée » des jugements portés sur ces performances ; pour la rendre plus démocratique,

nous proposons de recourir à une « confrontation dialectique » des jugements. Par confrontation dialectique de points de vue, nous entendons une mise en présence (des points de vue) pour les comparer. Dans le cadre de notre recherche, nous avons expérimenté (en nous inspirant de la théorie des conflits sociocognitifs bénéfiques à l'apprentissage) des séances de confrontation dialectique des jugements, c'est-à-dire des rencontres organisées entre l'étudiant et l'enseignant, ou avec ses pairs, pour confronter leurs points de vue à propos de la performance de l'étudiant et de l'auto/allo-évaluation⁴ de celle-ci, en vue de comprendre leurs éventuelles différences (d'appréciation de la performance cognitive complexe produite par l'étudiant) et s'en enrichir mutuellement. La visée de nos séances de confrontation dialectique des jugements est réflexive plutôt que prescriptive. Plus qu'à amener l'étudiant ou l'enseignant à « entendre raison », notre démarche cherche à « faire prendre conscience » à l'étudiant de ses connaissances et lacunes, de son réalisme, des progrès pouvant être réalisés, des décisions à prendre pour la suite de ses apprentissages, ainsi qu'à faire prendre conscience à l'enseignant de l'impact de son enseignement, de son degré de sévérité, du niveau de compréhension de ses étudiants, etc. Un précédent article (Jans, 2002) est consacré à l'analyse d'une expérience de confrontation dialectique des jugements évaluatifs entre un professeur et chacun de ses étudiants. Mentionnons simplement à ce propos que pour viser le bon déroulement (objectif de démocratie) ainsi que l'efficacité (objectif de *feed-back* formatif pour l'étudiant et pour le professeur) de ces séances, nous faisons deux recommandations :

- Ces séances doivent respecter certaines règles de fonctionnement clairement définies et partagées par les interlocuteurs (relevant de la psychologie sociale et clinique, de la méthodologie d'enseignement, notamment).
- Ces séances demandent à être instrumentées adéquatement à deux niveaux : (1) l'auto/allo-évaluation de la performance considérée doit avoir été faite à l'aide d'un outil *ad hoc*⁵ et (2) la confrontation dialectique des jugements gagne à prendre appui sur un premier traitement comparatif des jugements émis. C'est ce que nous appelons la « confrontation instrumentée des jugements ».

L'analyse individuelle de concordance des jugements auto/allo-évaluatifs portés sur une performance complexe

Nous exposerons dans cette section les concepts et les outils relatifs à l'analyse individuelle de concordance⁶. Ce type d'analyse repose largement sur une confrontation instrumentée des jugements auto/allo-évaluatifs. Les exemples illustratifs ainsi que les quelques résultats de l'analyse individuelle proposés sont issus d'une expérience de terrain menée dans le cadre du cours de «Technologie de l'éducation» de première licence en Sciences de l'éducation à l'Université de Liège.

Contexte d'expérimentation

Dans le cadre de ce cours, les 18 étudiants ont été invités à auto-évaluer leur rapport écrit de travaux pratiques⁷. L'activité d'auto-évaluation a été réalisée à partir de la même grille de critères que celle utilisée par le professeur pour évaluer les travaux. Cette activité comportait deux types de tâches : l'attribution d'une note chiffrée et la rédaction d'une justification pour chaque critère.

La grille de critères utilisée porte sur les compétences entraînées lors des travaux pratiques et plus particulièrement sur la capacité de produire le rapport de recherche demandé. Le tableau 1 illustre la cohérence entre les onze objectifs des travaux pratiques (TP), annoncés dès le départ et par écrit aux étudiants, et les critères d'évaluation de ces TP.

Définition

Concordance signifie «conformité de deux ou plusieurs choses entre elles ; correspondance, accord» (Le Petit Larousse, 1999, p. 245). L'analyse de concordance des jugements auto/allo-estimatifs portés sur une performance complexe vise à rendre compte du degré d'adéquation entre les estimations de deux personnes : l'auto-estimateur (l'étudiant dans notre cas) et l'allo-estimateur (le professeur ou un pair dans nos études). Cette analyse permet d'étudier le degré de concordance des jugements par le calcul et la représentation graphique des écarts de points observés entre les estimations de l'étudiant et celles du professeur. Il s'agit donc d'une «confrontation» (*mise en présence en vue de comparer*) de jugements. Cette confrontation est «instrumentée» (*calcul d'indices et représentation graphique*). Il s'agit essentiellement d'un acte «technique» qui ne fait pas (ou plus) intervenir l'avis subjectif d'une personne. Ce type de confrontation est «aveugle» ou «autoritaire» (contrairement à la confrontation dialectique ; dans le cas de la confrontation instrumentée, c'est l'ordinateur qui établit la comparaison).

Tableau 1
Parallélisme entre objectifs et critères d'évaluation des travaux pratiques

<i>Objectifs des TP</i>	<i>Critères d'(auto-)évaluation du rapport</i>	<i>Pondération (sur 100)</i>
<i>À la fin de ces TP, les étudiants auront progressé (objectivement et subjectivement) dans leurs capacités suivantes</i>		
1. Être ponctuel tout au long d'un travail (respect des échéances, présence aux rendez-vous)	1. Ponctualité tout au long des TP	5
2. Décrire une situation vécue d'apprentissage et ses propres réactions sur le déroulement de la situation	2. Description du déroulement des TP	5
3. Lire et résumer un article en anglais	3. Présentation de l'article lui-même	10
4. Exploiter les idées d'un article et faire des liens avec d'autres situations, d'autres notions, d'autres théories	4. Liens avec le travail (et ses prolongements)	10
5. Traiter des données avec EXCEL (les encoder, les présenter et dresser des graphiques)	5. Réalisation des graphiques (en fonction des consignes)	10
6. Être créatif, produire des idées, aller «plus loin que ce qui est demandé», etc.	6. Originalité dans la présentation des données: ajouts, modifications, etc.	5
7. Interpréter des graphiques	7. Rigueur de l'interprétation des données	10
8. Formuler des hypothèses personnelles pour expliquer des observations	8. Formulation d'hypothèses explicatives	10
9. Tirer les conclusions d'un travail, préciser ses lacunes et proposer des perspectives	9. Conclusions - Perspectives – Lacunes	10
10. Rédiger un rapport écrit structuré, syntaxiquement et sémantiquement rigoureux	10. Intelligibilité - clarté – structuration	5
	11. Précision, rigueur du vocabulaire	5
	12. Rigueur syntaxique	5
11. Soigner la mise en page d'un rapport écrit (titre accrocheur, lisibilité, orthographe, titre de bas de page, esthétique, bibliographie normalisée, etc.)	13. Présence d'un titre accrocheur	1
	14. Présence d'un titre en bas de page	1
	15. Présence d'une bibliographie normalisée	3
	16. Orthographe	3
	17. Qualité de la mise en page (lisibilité, esthétique, etc.)	2

Représentation graphique du « nuage de concordance »

Un « nuage de concordance » est un graphique présentant en abscisse les auto-estimations de l'étudiant et en ordonnée les estimations du professeur. Cette représentation graphique favorise la visualisation des écarts évaluatifs. Le nuage de concordance peut être construit à partir des données chiffrées absolues (notes accordées par l'étudiant et notes accordées par l'enseignant) si chaque critère d'évaluation a le même poids. Autrement, il faut travailler sur base des données relatives (notes traduites en valeur relative, comme dans la figure 1), mais l'interprétation devra tenir compte des différences de pondération.

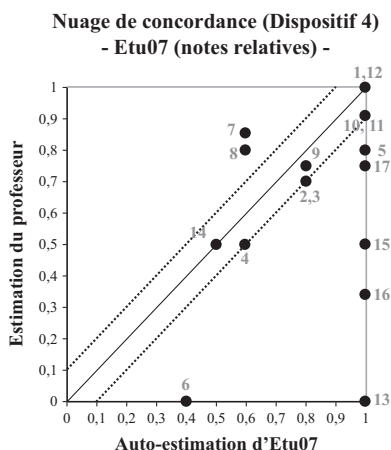


Figure 1. *Exemple d'un nuage de concordance pour l'Etu07*

Sur ce graphique illustrant l'évaluation de la performance de l'étudiant 7 (Etu07), les 17 critères d'évaluation sont numérotés de 1 à 17.

- Les points situés *sur* la diagonale représentent des cas de concordance parfaite (pour deux critères).
- Ceux situés *entre les deux lignes pointillées* (écart de 0,1 par rapport à la concordance parfaite) témoignent d'une bonne concordance ou d'une adéquation acceptable (pour quatre critères). Ces deux lignes délimitent ce que nous appelons la « zone de bonne concordance ».
- Les points situés *au-dessus* de la zone de concordance représentent des cas de sous-estimation de la part de l'étudiant (trois critères).
- Ceux situés *en dessous* de la zone de concordance représentent des cas de surestimation de la part de l'étudiant (huit critères).

Observations générales

Une analyse préalable des nuages de concordance par critère (Jans, 2000) a mis en évidence deux difficultés :

- Lorsque les critères permettent très peu de nuance (cas des critères «légers» à 1, 2, 3 points), une différence d'un seul point (ou de 1/2 point) rend les avis très discordants. C'est le cas pour quatre des cinq critères «poids légers» du graphique ci-dessus (13, 15, 16 et 17).
- Lorsque l'échelle est plus nuancée (5 ou 10 points), le phénomène de tendance centrale (évitement des extrêmes) aussi bien chez l'étudiant que chez l'enseignant risque de «gonfler» la concordance. C'est le cas pour les critères 2, 3, 4, 7, 8 et 9 du graphique ci-dessus.

Observations à partir du nuage de concordance d'ETU07

Le nuage de concordance d'Etu07 montre une bonne concordance entre ses estimations et celles du professeur pour neuf des 17 critères : un critère «poids léger» (c'est-à-dire évalué sur un, deux ou trois points) ; cinq critères «poids moyens» (c'est-à-dire évalués sur cinq points) et trois critères «poids lourd» (c'est-à-dire évalués sur dix points).

Comme pour la majorité des étudiants, le critère 6 concernant l'originalité «dans la présentation des données» (c'est-à-dire ce qui «dépasse» positivement les consignes données pour «améliorer les graphiques et tableaux») n'a sans doute pas été compris comme tel par Etu07. Comme nous l'avons mis en évidence lors de l'analyse des séances de confrontation dialectique des jugements qui ont suivi, le professeur a profité de ces séances pour justifier sa position. Il n'a cependant pas révisé ses cotes à la suite du constat d'un biais «assez systématique».

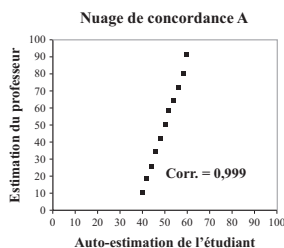
Quatre des cinq critères «poids léger» témoignent d'une surestimation d'Etu07 qui s'est attribué le maximum, le professeur étant plus sévère que lui. Parmi les critères «poids lourd», les critères 5, 7, 8 relatifs à la réalisation des graphiques et au traitement des résultats présentent des écarts plus importants (mais au maximum 0,15) par rapport à la zone de bonne concordance. Comme plusieurs étudiants l'ont souligné lors des séances de confrontation dialectique des jugements, ils sont peu familiarisés avec les critères d'excellence dans ce domaine.

Quantification de la concordance par des indices d'écart

Imprécision du coefficient de corrélation

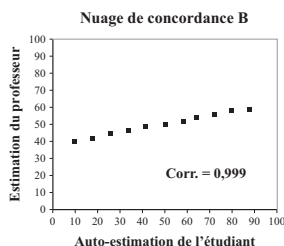
Le coefficient de corrélation est l'indice habituellement utilisé pour comparer deux séries de données. Il mesure la relation existant entre les deux séries de données (mises à l'échelle pour être indépendantes de l'unité de mesure). Nous nous sommes posé la question de savoir si la corrélation était l'indice le plus approprié pour juger de la concordance des auto/allo-estimations. Il apparaît que non.

La corrélation est «une certaine forme de concordance», mais elle ne tient pas compte de la dispersion des cotes des deux évaluateurs (qui pourrait varier grandement, comme l'illustrent les nuages de concordance des figures 2A et 2B), ni des moyennes des deux évaluateurs (qui pourraient également différer grandement, comme l'illustre le nuage de concordance de la figure 2C). Ces trois nuages présentent pourtant un indice de corrélation extrêmement élevé (0,999).



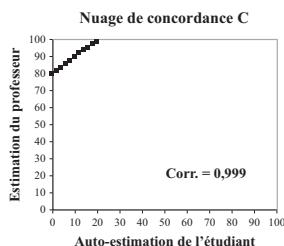
Dans ce graphique :

- La moyenne-étudiant = +/-50
- La moyenne-prof. = +/-50
- L'étudiant est moins discriminant que le professeur.



Dans ce graphique :

- La moyenne-étudiant = +/-50
- La moyenne-prof. = +/-50
- L'étudiant est plus discriminant que le professeur.



Dans ce graphique :

- La moyenne-étudiant = +/-10
- La moyenne-prof. = +/-90
- Les dispersions des cotes de l'étudiant du professeur sont semblables.

Figures 2A, 2B, 2C. **Observation de l'indice de corrélation sur trois nuages de concordance**

Les indices d'écart selon les critères

L'analyse des deux séries de données disponibles autorise le calcul d'autres indices basés sur les écarts (ou différences) entre les auto-estimations de l'étudiant et les allo-estimations du professeur.

La colonne de droite du tableau 2 présente les écarts par critère. Zéro signifie une *absence d'écart*. Un écart négatif reflète une *sous-estimation* de l'étudiant par rapport au professeur. Un écart positif reflète une *surestimation* de l'étudiant par rapport au professeur.

La moyenne des écarts *relatifs* (comprise entre -1 et +1) reflète la tendance «globale» de l'étudiant à se surestimer ou se sous-estimer par rapport au professeur. Attention cependant: cet indice masque les compensations éventuelles entre les sur- et les sous-estimations. La moyenne des écarts pris en *valeur absolue* (comprise entre 0 et 1) reflète l'ampleur de la discordance d'estimation entre l'étudiant et le professeur pour l'ensemble des critères.

Tableau 2
Indices d'écart entre les estimations (val. relatives) d'Etu07 et de l'enseignant

<i>Critères</i>	<i>Etu07</i>	<i>Prof</i>	<i>Écart</i>
1	1	1	0
2	0,8	0,7	0,1
3	0,8	0,7	0,1
4	0,6	0,5	0,1
5	1	0,8	0,2
6	0,4	0	0,4
7	0,6	0,85	0,25
8	0,6	0,8	0,2
9	0,8	0,75	0,05
10	1	0,9	0,1
11	1	0,9	0,1
12	1	1	0
13	1	0	1
14	0,5	0,5	0
15	1	0,5	0,5
16	1	0,33	0,67
17	1	0,75	0,25
Moy. des écarts relatifs			0,18
Moy. des écarts absolus			0,24

Observations à partir du calcul des écarts de points d'ETU07

On observe : trois cas de concordance parfaite (critères 1, 12 et 14); deux cas de sous-estimation de 25 % et 20 % pour des critères jugés sur dix points (critères 7 et 8); douze cas de surestimation, présentant des écarts relatifs compris entre 0,05 et 1 (pour un critère léger, ce qui est «normal»).

La moyenne élevée des écarts relatifs (0,18 alors que l'indice moyen pour l'ensemble des étudiants = 0,012; $\sigma = 0,11$) s'explique par le grand nombre de cas de surestimation par rapport aux sous-estimations. Etu07 présente une moyenne des écarts pris en valeur absolue parmi les plus élevées (0,24 alors que l'indice moyen pour l'ensemble des étudiants = 0,20; $\sigma = 0,05$), ce qui signifie qu'il y a, en moyenne, 24 % de différence entre les estimations d'Etu07 et celles du professeur, ce qui représente une discordance globale assez importante.

La confrontation des justifications écrites

Il est possible et souhaitable de compléter l'analyse de concordance des jugements estimatifs par une analyse des données qualitatives disponibles : dans notre cas, nous pouvons confronter les justifications écrites sur la grille de critères par l'étudiant et par le professeur. Pour chaque critère, l'étudiant comme le professeur ont eu l'occasion de fournir une justification écrite à leur note chiffrée. Ces justifications devraient pouvoir entrer en ligne de compte dans l'évaluation finale de la performance de l'étudiant. Il pourrait en effet s'avérer, à la lecture des justifications données par l'étudiant, que le professeur ait évalué «trop sévèrement» la performance de l'étudiant.

Le traitement des justifications écrites doit se faire, à notre avis, en grande partie de façon qualitative. Il convient d'apprécier :

- la présence ou l'absence de justifications (L'étudiant a-t-il «profité» de l'occasion qui lui était donnée de se justifier?);
- le contenu de la justification (Celle-ci se rapporte-t-elle bien au critère d'évaluation considéré? L'étudiant parle-t-il de façon générale ou précise? Donne-t-il des exemples? Met-il davantage en évidence les côtés positifs ou les côtés négatifs de son travail? etc.);
- l'adéquation entre les justifications de l'étudiant et celles du professeur (Y a-t-il des liens entre les justifications données? Les avis sont-ils partagés ou en opposition? etc.).

Nous retranscrivons dans le tableau 3 les notes chiffrées et les justifications figurant sur les feuilles d'auto-évaluation de la performance d'Etu07 et sur celles d'allo-évaluation par le professeur de cette même performance. Les données de ce tableau permettent plusieurs observations.

Trois critères reflètent une *concordance parfaite* (1, 12, 14). Il y a très peu de choses à dire à propos de ces trois critères. La justification écrite par Etu07 au critère 1 est sans équivoque et en accord avec la cote du professeur. Aucune justification n'est donnée pour le critère 12. Seul le critère 14 présente une justification écrite de la part des deux acteurs. Ceux-ci justifient d'une façon identique leur cote.

Six critères se trouvent dans la *zone de bonne concordance* (2, 3, 4, 9, 10, 11). Seuls les critères 2, 3 et 4 présentent une justification écrite de la part des deux acteurs. On retrouve de nombreux parallélismes dans les commentaires, ce qui témoigne d'un accord partagé quant au jugement de ces critères. Au critère 2, l'idée d'un manque d'exhaustivité dans la description du déroulement des TP se retrouve des deux côtés (« *bref* » ; « *manque* »). Au critère 3, on retrouve une marque équivalente d'appréciation positive des deux côtés (« *bonne synthèse* » ; « *bien expliqués* »). Un aspect négatif, le manque de précision est mentionné de façon équivalente par les deux acteurs (« *manque de précisions* » ; « *n'est pas donnée* » ; « *incompréhensibles si on n'a pas lu l'article en anglais* »). Au critère 4, les justifications traduisent des insatisfactions chez les deux acteurs ; celles du professeur sont plus diagnostiques, celles de l'étudiant plus floues.

Tableau 3
*Retranscription des grilles d'auto-évaluation d'Etu07
et d'allo-évaluation de l'enseignant*

CRITÈRES	Sur	Cote Etu07	Justifications écrites d'Etu07	Cote Prof	Justifications écrites du professeur
1 Ponctualité tout au long des TP	5	5	Je n'ai jamais eu de retard et j'ai toujours respecté le rendez-vous.	5	/
2 Description du déroul. des TP	5	4	Peut-être trop bref!	3,5	Synthétique, mais manque de précision à certains endroits (voir rapport lui-même).
3 Présentation de l'article	10	8	Bonne synthèse, avec un manque de précisions dans l'expérience elle-même.	7	La signification de T.P.R. n'est pas donnée. Ne pas traduire <u>gender</u> par genre, mais <u>sexe</u> , ni <u>challenge</u> par challenge, mais <u>défi</u> . Qq fautes d'orthographe. Certaines descriptions sont incompréhensibles si on n'a pas lu l'article en anglais. <u>Top down</u> et <u>bottom up</u> sont bien expliqués.

CRITÈRES	Cote Sur Etu07		Justifications écrites d'Etu07	Cote Prof	Justifications écrites du professeur
4 Liens avec le travail(et ses prolongements)	10	6	Plus difficile à réaliser. Pas toujours très clair ! Liaison restreinte, c'est le point le moins bien réalisé du TP.	5	Faible, <u>sauf</u> allusion à l' <u>éventuelle</u> approche top down de ETU60. Idée non <u>creusée</u> : «faire la même chose»! (p. 13) Déclaration à l'emporte pièce! (p. 13)
5 Réalisation des graphiques (en fonction des consignes)	1	10	Toutes les consignes ont été respectées et appliquées.	8	Quelques erreurs dans le tableau EXCEL (p. 3, 7 et 9). Quelques petites imperfections dans les graphiques (voir rapport).
6 Originalité dans la présentation des données (ajouts, modifications, créations, etc.)	5	2	Suppression de certaines données peu intéressantes (lien entre -2 et 0 sur les graphes). Pas d'ajouts spécifique-ment personnels.	0	/
7 Rigueur de l'interprétation	10	6	Le plus possible mais comme c'est la première fois, je dois encore avoir de grosses lacunes.	8,5	Très bonne analyse des données... un peu plus faible cependant à la question 2 (quantification manquante à certains endroits)
8 Formulation d'hypothèses explicatives	10	6	Elles ne sont pas toujours claires, manquent de précisions ou se répètent.	8	Bonnes tentatives, pertinentes. Bons liens entre les diverses sources de données.
9 Conclusions, perspectives, lacunes	10	8	Synthèse assez souple du travail avec une mise en perspective, les lacunes sont peut-être trop personnalisées.	7,5	/
10 Intelligibilité, clarté, structuration	5	5	/	4,5	Très agréable à lire, bien formulé!
11 Précision, rigueur du vocabulaire	5	5	/	4,5	/
12 Rigueur syntaxique	5	5	/	5	/
13 Présence d'un titre accrocheur	1	1	/	0	/
14 Présence d'un titre en bas de page	1	0,5	Haut de page (originalité)	0,5	Haut de page: OK.
15 Présence d'une bibliographie normalisée	3	3	Bibliographie basée sur un document de la Communauté française: «Réaliser vos TFE».	1,5	Plusieurs aspects (notés sur le texte lui-même) non conformes aux règles de l'A.P.A.
16 Orthographe	3	3	/	1	Couverture: expériences; p. 4: faiblir; p. 7: degrés utilisés (2x1); p. 14: notions comblées; p. 14: en commun.
17 Qualité de la mise en page	2	2	/	1,5	Table des matières: TB.
	100	79,5		71	

Deux critères reflètent une *sous-estimation de la part d'Etu07* par rapport au professeur (7, 8). Comme mis en évidence dans nos analyses par critère (Jans, 2000, pp. 273-285), les critères 7 et 8 (Rigueur de l'interprétation et Formulation d'hypothèses explicatives) n'ont pas suffisamment reflété auprès des étudiants le niveau d'exigence de l'enseignant par rapport à ces tâches de traitement des résultats (même si des exemples de traitement de résultats ont été proposés aux étudiants). Etu07 indique dans ses commentaires son manque d'expérience pour réaliser la tâche et par conséquent la difficulté pour lui de juger de la valeur de son travail sur ce point : *« mais comme c'est la première fois, je dois encore avoir de grosses lacunes »*. Le professeur a, quant à lui, tenu compte du fait que c'était une des premières fois que les étudiants étaient invités à réaliser ce genre d'exercice ; il s'est montré « indulgent » (selon ses dires).

Un critère « poids lourd » (5) et un critère « poids moyen » (6) reflètent une *surestimation de la part d'Etu07* par rapport au professeur. Au critère 5, Etu07 semble ne pas avoir été conscient des erreurs relevées par le professeur dans ses tableaux de données. C'est une « surprise » pour lui, ce qui explique qu'il se soit surestimé à ce critère-là. Au critère 6, Etu07 s'est accordé 2/5 pour la *« suppression... du lien entre -2 et 0 sur les graphes »*. Cette suppression faisant cependant partie des consignes, ce qui explique que le professeur ne l'ait pas considérée comme une « originalité ». Etu07 reconnaît cependant qu'il n'y a *« pas d'ajouts spécifiquement personnels »* dans la présentation de ses données, avis partagé par le professeur et reflété dans sa cote 0/5.

Quatre critères « poids léger » reflètent une *surestimation importante de la part de l'Etu07* (13, 15, 16, 17). Au critère 13, le professeur n'a pas considéré comme « accrocheur » le titre donné par Etu07 à son travail : *« Technologie de l'Éducation - Rapport sur les T.P. »* avec en sous-titre *« L'apprentissage d'une langue étrangère à l'aide d'un ordinateur, confrontation entre des expériences sur base d'un cd-rom et d'un article de langue anglaise basé sur un logiciel de type T.P.R. »*. Même si ce critère engendre inévitablement un jugement fortement teinté de subjectivité, on ne pourra cependant nier le manque d'attrait du titre fourni par Etu07. Or, le professeur avait fourni aux étudiants des exemples de « titres accrocheurs ». Mais, pour l'étudiant, s'attribuer zéro reviendrait à reconnaître qu'il n'avait pas fait le travail demandé. Certains étudiants ont souligné ce problème lors des séances de confrontation dialectique en faisant aveu de « surévaluation consciente ».

Au critère 15, un malentendu entre les deux acteurs explique la discordance des jugements, puisque le professeur avait demandé une bibliographie rédigée en fonction des normes de l'*American Psychological Association* (A.P.A.) et qu'Etu07 a cru bon d'en proposer une en fonction d'autres normes (présentées dans un document sur les Travaux de fin d'études (TFE) à l'usage des écoles normales de la Communauté française de Belgique). Nous avons constaté que bon nombre d'étudiants n'ont pas utilisé les références recommandées. À ce propos, le professeur avait ici recommandé à ses étudiants de lire une de ses brochures (*Comment saloper son mémoire*). Au critère 16 relatif à l'orthographe, Etu07 a considéré son rapport exempt de fautes. Le professeur en a relevé, quant à lui, six et a appliqué, comme principe de cotation, «retrait d'un point par trois fautes».

Enfin, le critère 17 relatif à la qualité de la mise en page présente peu de discordance entre les cotes des deux acteurs (2/2 et 1,5/2). Le fait que ce critère soit évalué sur seulement deux points provoque sur le nuage de concordance l'illusion optique d'un écart relativement important (en dehors de la zone de bonne concordance).

Mentionnons pour terminer que le professeur a relevé au travers de ses justifications écrites autant de points positifs que de points négatifs. Une telle attitude de la part d'un enseignant est importante s'il souhaite que son évaluation influence favorablement les décisions futures de l'étudiant. Les travaux de Viau (1995) sur la motivation ont illustré ce principe.

Intérêt de l'analyse individuelle de concordance

L'analyse individuelle de concordance, telle que nous l'avons présentée et illustrée ci-dessus, nous paraît intéressante dans un contexte de formation. Le fait de pouvoir aller en profondeur dans l'analyse des cas individuels, de mettre en parallèle et de recouper des données issues d'horizons divers, peut favoriser considérablement la préparation de *feed-back* personnalisés à l'intention des étudiants. Baser ces *feed-back* à la fois sur des données «objectives» et sur d'autres plus «cliniques» constitue, à notre avis, une démarche solide et riche en vue d'aider l'étudiant à poser un diagnostic précis sur ses apprentissages.

Les séances de confrontation dialectique des jugements que nous avons organisées à la suite de la production et de l'(auto)évaluation de performances cognitives variées reposent sur de tels *feed-back* (Jans, 2000). Ces deux types

de confrontation (instrumentée et dialectique), successifs et complémentaires, ont contribué à fournir un feed-back formatif à *l'étudiant* (accord ou non avec l'enseignant, degré de sévérité de l'enseignant, prise de conscience de certaines lacunes ignorées, etc.). Ils ont également été utiles pour fournir un feed-back formatif à *l'enseignant* (information sur les critères d'évaluation, information sur le degré de conscience de ses étudiants face à leur performance, prise de conscience de son degré de sévérité, prise de conscience de son (im)partialité, décision de revoir ses notes chiffrées, etc.).

À la suite à la séance de confrontation dialectique des jugements organisée entre le professeur et chaque étudiant, sur base à la fois du rapport de TP annoté par le professeur des grilles d'auto/allo-évaluation et de la confrontation instrumentée des jugements émis sur ces grilles, nous avons demandé aux étudiants de répondre par écrit à la question suivante: «Quelle(s) décision(s) as-tu prise(s) pour tenter d'améliorer tes auto-évaluations à l'avenir? À quoi vas-tu être plus attentif?» La réponse d'Etu09 illustre l'impact de la procédure sur sa réflexion: «À l'avenir, je demanderai les exigences d'évaluation (si elles ne sont pas données) pour pouvoir mieux comparer avec ce que j'ai fait. L'auto-évaluation est, cependant, une compétence difficile à acquérir et j'ai tendance à "gonfler" artificiellement mon opinion. C'est le deuxième point que je travaillerai dans le futur.»

Avis des étudiants sur les outils proposés

Les 18 étudiants impliqués dans notre recherche ont été invités à exprimer leur avis sur les concepts et outils vécus relatifs à l'auto/allo-évaluation des performances complexes et à la confrontation des jugements. Ces avis ont été livrés «avec recul» (plusieurs mois après le déroulement des activités) lors d'entretiens individuels. Nous présentons ci-dessous la synthèse à deux questions posées.

Avis sur la technique de la grille de critères

La synthèse des avis est présentée dans le tableau 4. Les réponses montrent que les étudiants ont très largement apprécié la technique d'auto-évaluation proposée. Ils soulignent l'intérêt de pouvoir fournir des justifications écrites des notes chiffrées, mais cette exigence est parfois perçue comme une contrainte.

Cette technique est perçue comme facilitant (1) une confrontation des points de vue après la performance, (2) une meilleure connaissance de soi et (3) une possibilité de régulation, trois objectifs que notre travail cherchait à atteindre à travers l'utilisation des techniques d'auto-évaluation proposées aux étudiants. Des étudiants relèvent aussi le rôle de «pense-bête⁸» joué par la grille d'(auto)évaluation.

Tableau 4
*Avis sur la technique de la grille de critères émis par les étudiants
lors d'entretiens individuels (N=18)*

<i>Concernant la technique d'auto-évaluation proposée (grille de critères + justifications), qu'en penses-tu ?</i>	
<i>Idées résumées</i>	<i>Illustration par un verbatim d'étudiant(e)</i>
Marque explicite de satisfaction (N=16)	«Pour moi, c'est ce qu'il faudrait mettre dans tout rapport.»
Rôle de «pense-bête» (N=2)	«Je l'ai utilisée en deux temps. J'ai d'abord fait un premier jet de mon rapport puis la grille d'auto-évaluation, et ça m'a permis de me rendre compte en devant remplir les points que je ferais bien de corriger telle ou telle lacune.»
Découverte des exigences du professeur (N=2)	«Ce qui est bien, tu découvres les exigences du professeur et tu regardes si elles sont tiennes.»
Information sur sa capacité d'auto-évaluation (N=3)	«Moi, ça m'intéressait de voir si je m'étais bien évaluée.»
Possibilité de confronter son auto-évaluation à un avis extérieur (N=6)	«Pour remarquer que ce qu'on croyait être une qualité, en fait c'était un défaut, et ce qu'on croyait être un défaut, c'était une qualité.»
Possibilité de donner des justifications (N=4)	«Les justifications, c'est bien, c'est donner du sens à ce qu'on fait, c'est génial. À l'école primaire même, on ne le fait presque pas, apprendre à se justifier.»
Contrainte des justifications (N=1)	«Il y a des fois où tu mets quelque chose pour mettre quelque chose. Parfois tu te justifies pour ne pas laisser une case blanche.»
Favorise la régulation (N=4)	«Peut-être que parfois on a une mauvaise opinion de soi-même, ça permet de changer son fusil d'épaule.»
Tâche inhabituelle et non aisée pour l'étudiant (N=4)	«On est tellement pas habitué à devoir dire ce qu'on pense de nous qu'à la limite on est un peu bloqué face à ça, surtout qu'on sait qu'après on est évalué un peu à partir de ce qu'on dit.»

*Concernant la technique d'auto-évaluation proposée
(grille de critères + justifications), qu'en penses-tu ?*

<i>Idées résumées</i>	<i>Illustration par un verbatim d'étudiant(e)</i>
	«C'est difficile à faire l'auto-évaluation, parce qu'on fait le travail [...] mais après, on se dit : l'évaluation, c'est pas nous.»
Subjectivité dans l'attribution des points (N=1)	«C'est fort relatif par rapport à nos ambitions et nos exigences personnelles. [...] C'est subjectif dans la mise des points.»
Rôle controversé de la prise en compte de l'auto-évaluation dans le score final (N=3)	«C'est une bonne idée, sauf qu'on est évalué après dessus. Mais si on n'en tient pas compte, peut-être qu'on ne ferait pas ça sérieusement.»

Malgré l'enthousiasme des étudiants, il apparaît que la tâche d'auto-évaluation d'une performance complexe à partir d'une grille de critères n'a pas été perçue comme aisée par tous, d'une part parce qu'elle est inhabituelle, d'autre part parce qu'il est difficile de s'attribuer des points. «L'activité d'auto-évaluation doit-elle peser dans la note "sanctionnante" ?» Les avis des étudiants sur la question sont partagés et continuent à alimenter le débat présent dans la documentation actuelle (Jans, 2000).

Avis sur l'outil «Nuage de concordance»

Le tableau 5 présente la synthèse des avis. La majorité des étudiants et des étudiantes apprécient la clarté du «nuage de concordance», principalement de par son côté graphique, ce qui permet notamment de mieux visualiser les écarts entre les auto et les allo-estimations.

Les étudiants apprécient également la distinction des critères de poids différent (poids léger, moyen ou lourd), rendue apparente (dans nos expériences de terrain) par l'utilisation de couleurs. Ils estiment que cet outil suscite réflexion et discussion et favorise la régulation, deux des objectifs que nous lui assignions. Selon certains étudiants, une explication préalable plus approfondie serait nécessaire pour bien comprendre le graphique. Enfin, quelques étudiants se montrent réticents face à cet outil graphique, soit parce qu'ils n'y voient pas d'utilité pour eux-mêmes, soit par le côté subjectif des auto/allo-estimations de points.

Tableau 5

***Avis sur l'outil «nuage de concordance» émis par les étudiants lors
d'entretiens individuels (N=18)***

<i>Est-ce que ce genre d'outil te semble intéressant, utile, ... ? Et pourquoi ?</i>	
<i>Idées résumées</i>	<i>Illustration par un verbatim d'étudiant(e)</i>
Marque explicite d'accord, de satisfaction (N=9)	«Ce qui peut être bien apprécié, c'est combien on arrive à être rigoureux ici alors qu'on part de quelque chose tellement à la tête du client. [...] Montrer [aux étudiants] qu'on part de là pour réfléchir, c'est super, c'est bien.»
Marque de désaccord (N=2)	«Ça a été intéressant pour vous. [...] Pour moi, non, franchement. Moi, je me contente de ça [calcul des écarts de points].»
Nécessite une explication (N=3)	«Il est nécessaire de donner des explications car toute seule il me faudrait longtemps pour saisir.»
Permet de constater les différences de cotation prof/étudiant (N=3)	«Ça permet simplement, à mon sens, de constater les différences de cotation, d'appréciation d'un critère entre l'étudiant et le professeur qui cote.»
Clarté du <i>feed-back</i> graphique (N=9)	«Je trouve que les graphiques sont toujours mieux que des lignes descriptives, des tableaux, parce qu'ils ont quelque chose en plus, et pas à la place.»
Suscite la réflexion et la discussion (N=4)	«Ça permet à l'étudiant de voir : voilà, je ne suis pas du tout réaliste et je reprends mon critère, je lis, et j'essaie de comprendre pourquoi il y a un tel écart, et si possible j'essaie d'en discuter avec le professeur.»
Nécessité de la discussion (N=1)	«Un système comme ça est intéressant si on continue et si on argumente.»
Favorise la régulation (N=1)	«C'est plutôt un résumé qui permet de vite aller voir les points sur lesquels on doit travailler.»
Appréciation de la distinction des critères en fonction de leur poids (N=3)	«En voyant ça, on peut directement dire : tel critère, c'est un lourd, attention, je me suis évaluée comme ça, pourquoi ? Et puis, c'est vrai que les critères légers, c'est peut-être pas là-dessus qu'il faut passer son temps.»
Subjectivité (N=2)	«C'est tellement subjectif. Bon, moi, je me mets 70 %, et puis voilà, c'est mon avis, et ça dépend tu vois, 70 % pour moi à l'Université, c'est très bon.»

Conclusions

L'étudiant universitaire a une certaine conscience de la qualité de ses performances, au moment où elles sont soumises à l'(allo)évaluation des encadrants. Dans nos expériences relatives aux performances complexes, ce jugement s'est exprimé par une note chiffrée et un commentaire qualitatif (sur une grille de critères). On peut espérer qu'ultérieurement ce jugement soit de plus en plus précoce, bien avant l'évaluation, et engage l'apprenant dans des procédures «préventives». Une telle démarche n'a de sens que si l'étudiant est conscient de ses forces et faiblesses, et s'il a le sentiment que le jugement d'autorité (allojugement) est pertinent, sans quoi il se conformera à la norme en attendant que celle-ci cesse d'être appliquée (changement de professeur par exemple). D'où l'importance de tout un travail préalable de formation de l'étudiant à s'auto-évaluer de façon précise et conforme à la réalité, à réfléchir sur ses performances, à discuter de ses performances avec d'autres acteurs, à prendre des décisions de régulation proactive ou rétroactive.

Dans cette perspective, tant les encadrants que les étudiants doivent s'armer de concepts et d'outils pour analyser les situations. Notre recherche s'inscrit dans un contexte d'innovation pédagogique: nous avons cherché à élaborer des instruments pour approcher les phénomènes d'apprentissage, d'évaluation et d'auto-évaluation à l'université; notre tâche a également consisté à concevoir des «scénarios» d'utilisation de ces instruments en pédagogie universitaire.

Une des particularités de notre démarche réside dans la volonté d'entreprendre ces démarches *avec* les principaux bénéficiaires. Avant de les tester à grande échelle, il nous semblait plus pertinent et plus riche de les expérimenter avec un petit groupe d'étudiants, régulièrement interrogés sur la pertinence des outils et dispositifs. Dans ce sens, notre démarche est essentiellement qualitative. Il s'agit d'une procédure lourde, longue, mais bénéfique: elle vise à faire fonctionner la démocratie à l'université, à travers la prise en compte institutionnelle des sujets (non considérés comme objets). De plus, chez de futurs enseignants, vivre ces concepts et ces situations nous paraît crucial, par principe d'isomorphisme avec ce qu'ils feront vivre plus tard à leurs propres élèves.

L'analyse individuelle de concordance proposée dans cet article pour illustrer nos propos concerne la performance d'un seul étudiant (Etu07). Nous avons par ailleurs appliqué ce type d'analyse aux performances des 18 étudiants engagés dans la recherche. Les nuages de concordance obtenus nous ont

permis de rendre un *feed-back* personnalisé à ces étudiants et d'organiser des séances de confrontation dialectique des jugements avec le professeur sur base de la comparaison des jugements évaluatifs (Jans, 2002). Nous avons, dans ce cas, utilisé la confrontation instrumentée comme outil diagnostique au service de la confrontation dialectique. En croisant ainsi les regards d'étudiants avec ceux de personnes extérieures, nous faisons l'hypothèse que la capacité de réflexion métacognitive des étudiants mériterait d'être étudiée pour elle-même, en tant que compétence importante à développer à l'université pour former des adultes à porter un regard conscient et réaliste sur eux-mêmes.

NOTES

1. Dans le cadre de cet article, nous ne présentons pas nos travaux relatifs à l'auto-évaluation des performances isolables, mais renvoyons le lecteur intéressé à des écrits précédents (Jans, 1997; Jans, 1998; Jans & Leclercq, 1999).
2. Plusieurs publications (Jans, 1997, 1998, 2002, Jans & Leclercq, 1999), ainsi que le présent article, sont basées sur notre thèse de doctorat en sciences de l'éducation, réalisée à l'Université de Liège, sous la direction de Dieudonné Leclercq.
3. Selon Mouvet, dans Jans (2000, p. 80).
4. Par *allo-évaluation* d'une performance, nous entendons l'évaluation portée sur la performance d'*autrui* (allo), par opposition à l'auto-évaluation qui est l'évaluation portée sur sa propre performance. En éducation, le cas le plus fréquent d'allo-évaluation est celui de l'enseignant évaluant l'étudiant, mais c'est aussi, par exemple, le cas des pairs fournissant un commentaire sur la performance d'un de leurs condisciples.
5. Nous avons travaillé avec une échelle descriptive d'évaluation, élaborée conformément à la méthode proposée par De Bal, de Landsheere & Paquay-Beckers (1976).
6. Dans notre recherche, nous proposons également, mais à d'autres fins, une analyse de concordance par critère ainsi qu'une analyse par strates de concordance (Jans, 2000).
7. L'objet de ce rapport consistait en une analyse de données d'auto-apprentissage de l'anglais par CD-Rom (Jans, 1998).
8. Tardif (1998) souligne l'intérêt de l'existence d'une grille de critères *avant* l'activité d'auto-évaluation, c'est-à-dire avant et pendant la réalisation de la performance.

RÉFÉRENCES

- Beckers, J. (1997). *Cours de didactique générale et gestion des apprentissages en contexte scolaire destiné aux étudiants présentant l'agrégation, Notes de cours 1996-1997*. Université de Liège, Faculté de psychologie et des sciences de l'éducation.
- Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.
- De Bal, R., de Landsheere, G., & Paquay-Beckers, J. (1976). *Construire des échelles d'évaluation descriptives*. Bruxelles: Ministère de l'Éducation nationale et de la Culture française.
- Frenay, M., Noël, B., Parmentier, P., & Romainville, M. (1998). *L'étudiant-apprenant. Grilles de lecture pour l'enseignant universitaire*. Bruxelles: De Boeck.

- Hadji, Ch. (1997). *L'évaluation démystifiée*. Paris: ESF.
- Jans, V. (1997). Observer les progrès en anglais dus à l'utilisation d'un hypermédia par des étudiants universitaires: analyse spectrale de l'évolution des performances. In E. Boxus, V. Jans, J.-L. Gilles & D. Leclercq (éds), *Stratégies et médias pédagogiques pour l'apprentissage et l'évaluation dans l'enseignement supérieur* (pp. 201-219). Actes du 15^e colloque de l'Association internationale de pédagogie universitaire (AIPU), Université de Liège.
- Jans, V. (1998). L'auto-évaluation de performances simples et complexes par des étudiants universitaires: description et résultats d'une expérience. In C. Depover & B. Noël (éds), *Approches plurielles de l'évaluation des compétences et des processus cognitifs* (pp. 61-75). Actes du 12^e colloque de l'Association pour le développement des méthodologies d'évaluation en éducation (ADMEE), Mons, Université de Mons-Hainaut et Facultés universitaires catholiques de Mons, 9-11 septembre 1998.
- Jans, V. (2000). *Confrontations instrumentées et dialectiques des jugements auto- et alloévaluatifs. Contributions conceptuelle et méthodologique à l'étude de l'auto-évaluation réflexive chez des étudiants universitaires*. Thèse de doctorat en sciences de l'éducation, Université de Liège.
- Jans, V. (2002). *Pourquoi et comment dialoguer entre étudiant et enseignant sur base de leurs auto- et alloévaluations écrites? Une expérience de confrontation dialectique des jugements évaluatifs portés sur une performance complexe à l'université*. Communication présentée au colloque de l'Association internationale de pédagogie universitaire, Louvain-la-Neuve, 29-31 mai 2002.
- Jans, V., & Leclercq, D. (1999). Mesurer l'effet de l'apprentissage à l'aide de l'analyse spectrale des performances. In C. Depover & B. Noël (éds), *L'évaluation des compétences et des processus cognitifs. Modèles, pratiques et contextes* (pp. 303-317). Bruxelles: De Boeck.
- Leclercq, D. (1995). *L'évaluation des performances complexes*. Document interne au Service de technologie de l'éducation, Université de Liège.
- Leclercq, D. (1998). *Pour une pédagogie universitaire de qualité*. Sprimont: Mardaga.
- Nyhan, B. (1991). *Promouvoir l'aptitude à l'auto-formation. Perspectives européennes sur la formation et le changement technologique*. Bruxelles: Presses Universitaires Européennes.
- Romainville, M. (1992). *Savoir parler de ses manières d'apprendre. Contribution à l'étude de la relation entre la métacognition et la performance chez l'étudiant universitaire de première année*. Thèse de doctorat en psychopédagogie, Louvain-la-Neuve, Université catholique de Louvain.
- Tardif, J. (1998). *Intégrer les nouvelles technologies de l'information. Quel cadre pédagogique?* Paris: ESF.
- Viau, R. (1995). La motivation à apprendre à l'université: l'état des recherches et une démarche d'intervention. In *Enseignement supérieur: stratégies d'enseignement appropriées* (pp. 291-296). Actes du colloque de l'Association internationale de pédagogie universitaire, Hull.