

EVALUATION FORMATIVE

- DÉFINITION, JUSTIFICATIONS THÉORIQUES ET ILLUSTRATIONS -

M. Poumay, mai 2003

L'évaluation formative a fait l'objet de nombreux ouvrages, en Belgique et à l'étranger¹. Cependant, les centaines de pages produites sur le sujet n'offrent pas de définition à la fois complète et suffisamment féconde pour réagir sur le terrain, lors des accompagnements de projets, à des enseignants demandant des précisions. Les auteurs sont pourtant unanimes sur quelques éléments essentiels, que nous proposons ci-dessous au travers de la **définition** de G. De Landsheere (79).

Cette définition est complétée et illustrée par D. Leclercq dans son ouvrage « Edumétrie et Docimologie », dont nous joignons en annexe les pages en question.

En appui à ces définitions, nous trouvons fécond d'organiser une présentation de différents concepts et exemples liés, de façon à outiller et largement illustrer ces premiers apports. Nous les structurons en une **grille** mobilisable lors des accompagnements de projets, grille que nous compléterons ensemble lors de la co-formation par de nouveaux exemples issus des cours que vous avez accompagnés. Les catégories de la grille ainsi que les exemples de cours « non à distance » sont très largement empruntés au syllabus « Edumétrie et Docimologie » de D. Leclercq. Les modifications que vous constaterez vont dans le sens d'un allègement, mais surtout d'une plus grande fécondité pour l'EAD. Toutes ces modifications, qu'elles soient des ajouts, des suppressions de critères ou des précisions terminologiques, ont été approuvées par D. Leclercq.

Enfin, pour répondre à des questions posées par l'équipe LabSET à l'occasion du prétest du 4 février et pour en anticiper d'autres, nous proposons une série de **FAQ** suivie de références bibliographiques.

1. Définition

L'expression « évaluation formative » a été utilisée par Scriven pour désigner les procédures visant à améliorer le **curriculum**. En 1971, Bloom et ses collaborateurs ont popularisé l'expression en la faisant figurer dans le titre de leur ouvrage « Handbook of Formative and Sommative Evaluation » et en y appliquant l'expression à l'amélioration des **apprentissages**.

Gilbert de Landsheere, dans son Dictionnaire de l'évaluation et de la recherche en éducation (1979), a défini cette expression comme d'ailleurs les principaux termes liés à l'évaluation.

« *Evaluation intervenant, en principe, au terme de chaque tâche d'apprentissage et ayant pour objet d'informer élève et maître du degré de maîtrise atteint et éventuellement, de découvrir où et en quoi un élève éprouve des difficultés d'apprentissage, en vue de lui faire découvrir des stratégies qui lui permettent de progresser... L'expression « évaluation formative » due à Cronbach et à Scriven marque bien que l'évaluation fait ...partie intégrante du processus éducatif normal, les « erreurs » étant à considérer comme des moments dans ...l'apprentissage, et non comme des faiblesses répréhensibles ou des manifestations pathologiques. ...L'évaluation formative permet aussi de déterminer si un élève possède les prérequis.* » (G. DE LANDSHEERE, 1979, p. 113).

Les enseignants ont alors, à la suite de Bloom, pris l'habitude d'opposer évaluations formative et sommative. Estimant que cette distinction était beaucoup trop peu subtile pour résumer le grand nombre de dimensions qui caractérisent les évaluations, D. Leclercq, dans le premier chapitre de son

¹ Citons notamment Allal (79), Bloom (71), Cardinet (87), Perrenoud (98), Scriven (67) et Tardif (98).

cours d'Edumétrie et Docimologie, a défini une dizaine de facettes et une trentaine de dimensions souvent bipolaires caractérisant l'évaluation. Cette « rose des vents » de l'évaluation propose une classification, sorte de taxonomie, qui vient compléter la définition de Landsheere en lui apportant des nuances fécondes pour les praticiens de l'enseignement. De telles avancées, outillant les professionnels de la formation, sont ce qui vaut son nom à la « Technologie de l'Education », qui instrumente conceptuellement les sciences de l'éducation.

La « Rose des Vents » de D. Leclercq a servi de base au tableau ci-après, comprenant quant à lui moins de 20 dimensions.

Précisions préalables

Nous insérons ici deux précisions qui nous semblent de nature à faciliter la lecture du tableau ci-après. D'autres précisions, sont structurées en FAQ, en fin de document. Cette section FAQ sera complétée en fonction des questions posées lors de la coformation.

1. Contrairement à une représentation souvent véhiculée, le fait qu'une évaluation soit formative n'implique absolument pas l'absence de notation. Une évaluation peut très bien être à la fois formative et sanctionnante. La définition de Landsheere ne fait d'ailleurs aucun lien entre l'aspect formatif et d'éventuelles notes. Certains dispositifs sont formatifs et sanctionnants, d'autres sont formatifs et non sanctionnants. Ces deux points ne sont donc pas liés, comme le précise la seconde page du tableau.
2. Aucune évaluation ne peut jamais être considérée a priori comme formative. Une évaluation à visée formative n'est pas forcément à impact formatif.

Nous rapprochons cette distinction de celle proposée par Shanon et Weaver (1949) à propos de la distinction entre données et information, où l'information est ce qui réduit l'incertitude. Dans cette acception, ce qui est une donnée pour un étudiant peut être une information pour un autre et inversement.

En évaluation formative, nous pourrions de la même façon distinguer :

1. L'évaluation **documentée** (qui fournit à l'étudiant des données, par exemple ses sous-scores par matière)
2. L'évaluation **informative** (au sens de Shanon et Weaver, c'est-à-dire qui fournit à l'étudiant des données qui réduisent son incertitude quant au résultat de ses apprentissages)
3. L'évaluation **formative** (qui permet à l'étudiant de se mettre en mouvement pour pallier ses faiblesses et exploiter ses forces)

Dans cette acception, une évaluation formative est forcément aussi informative et documentée, le niveau supérieur incluant les niveaux précédents.

Dans le tableau qui suit, le terme Evaluation Formative (EF) est pris au sens large, le plus commun, c'est à dire au sens d'évaluation à visée formative. Ceci ne veut pas dire qu'elle sera formative pour un étudiant, par exemple, qui ne prend pas connaissance de ses notes pourtant très diagnostiques, ou à qui cela n'apprend rien car il se savait déjà faible dans tel point de la matière (pour lui, cette évaluation était juste documentée mais sans portée informative), ou encore qui, malgré la prise de connaissance de nouvelles données sur son apprentissage, ne souhaite en rien se mettre en mouvement, entrer en phase d'apprentissage (pour lui, cette évaluation était juste informative, sans suite).

2. Liens entre l'évaluation formative et d'autres caractéristiques de l'évaluation - illustrations

Caract. de l'évaluation (en général)		Lien avec l'évaluation FORMATIVE et illustrations (en rouge, les cas mentionnés) Notons que l'abréviation EF est utilisée pour <i>Evaluation Formative</i> , FB pour <i>Feed-back</i> .
Dimensions	Pôles extrêmes	
Pourquoi (fonctions)		
Référence	Critérielle	<p>Dans le cours d'Edumétrie et Docimologie (D. Leclercq & F. Demily), les étudiants connaissent à l'avance les critères par rapport auxquels leurs performances seront jugées et la pondération de chacun de ces critères. L'évaluation de chaque étudiant est donc réalisée en rapport avec ces critères et non selon les résultats de leurs pairs.</p> <p>C'est ce type d'évaluation qui est le plus adapté à l'EF qui peut se baser ainsi sur des repères-matière précis.</p>
	Normative	<p>Une évaluation normative peut se faire à partir du classement du score (FB : « Vous êtes premier, dixième, quarantième, etc. ») ou à partir de la moyenne (et de l'écart-type) de la cohorte (FB : « Vous êtes un écart-type au-dessus de la moyenne de votre cohorte »).</p> <p>Ainsi, Dans les Feed-Backs (FB) du PBL à Maastricht (« Progress Tests »), pour chacun des n domaines, les étudiants sont situés (par écart-type) par rapport aux scores des étudiants de leur cohorte. Une évaluation normative peut se faire à partir du classement (premier, dixième, quarantième, etc.) du score ou à partir de la moyenne (et de l'écart-type) de la cohorte. C'est ce dernier cas qui s'applique dans ces « Progress Tests » de médecine à Maastricht. Dans le livre PPUQ à la page 200, on trouve un tableau des résultats d'UN étudiant, avec 29 lignes correspondant aux 29 matières du test. A la ligne 1, on lit « Anatomie », Nbre de Q = 13 ; Correctes = 54%++. Le ++ signifie que ce résultat est très (au moins UN écart-type) au-dessus de la moyenne de la cohorte (39% avec un écart-type de 14%).</p> <p>Une telle évaluation normative est surtout pratiquée en sport : remonter au classement ATP, faire partie du top 10, etc.</p> <p>Une évaluation normative est très peu formative. Elle ne l'est que si elle donne aux étudiants des moyens de s'améliorer par la suite, puisqu'ils auront pris conscience de leur classement (motivation au changement : « Qu'est-ce qu'elles ont donc fait mieux que moi, ces 6 personnes qui sont classées avant moi ? »).</p>
Résultat	Sommatif	Dans les FB du PBL à Maastricht (« Progress Tests »), pour chacun des domaines, on ne donne à l'étudiant qu'un seul score (cfr PPUQ p 200), qui ne l'informe pas des lacunes intra-domaines (ex. quelles questions de cardiologie n'ont pas été réussies ?). C'est

		<p>déjà plus diagnostique qu'un score global pour les 250 questions du test, mais beaucoup moins que ce que serait un feed-back sur les solutions choisies à CHAQUE question.</p> <p>Ce type d'évaluation est peu approprié à l'EF. Le fait, pour l'étudiant, de connaître la somme de ses points, est souvent la partie la moins potentiellement formative des données qui lui sont communiquées. Ceci dit, une évaluation sommative peut être AUSSI formative, si d'autres données que le score final sont communiquées aux étudiants. Donc, une évaluation peut être sommative ET formative, par exemple si elle est aussi diagnostique, mais ce n'est pas l'aspect sommatif qui est formatif dans les données communiquées aux étudiants.</p>
	Diagnostique	<p>Ce type d'évaluation est nécessaire pour l'EF, tout comme en médecine on ne peut prescrire et « soigner » que si l'on a diagnostiqué le problème. Nous distinguons deux objets de diagnostic :</p> <ul style="list-style-type: none"> - <u>la maîtrise du contenu</u> : dans le test de chimie de Claude Houssier : 40 questions se rapportent à chacun des chapitres, avec FB différenciés selon les distracteurs, non communication de la réponse correcte de façon à permettre plusieurs essais et renvois vers les endroits exacts du syllabus liés à chacune des questions/erreurs. Donner les réponses correctes à chacune des questions n'est pas suffisant pour assurer l'aspect diagnostique d'une évaluation. Encore faut-il que les étudiants aient conscience du pourquoi de leurs erreurs ou de leurs réponses correctes et qu'on leur donne les moyens de pallier les manques détectés (ex. renvoi vers les pages précises du cours qui se réfèrent à ces points faibles). - <u>les processus mentaux</u> : dans l'opération RESSAC - FAPSE - ULg 2000, les étudiants recevaient des FB non seulement quant à leurs performances dans 4 domaines (cours) différents, mais aussi, pour chacun de ces cours, un sous-score de compréhension et un sous-score de mémorisation. Ils pouvaient ainsi voir si dans l'un ou l'autre de ces deux processus, ils étaient forts ou faibles par rapport à la moyenne de leur cohorte (normatif). L'intérêt de ce diagnostic est qu'il est révélateur de deux façons d'étudier : mémoriser et comprendre en profondeur, deux modalités sur lesquelles l'étudiant a pris ! Voilà donc un diagnostic normatif !!!! Décrit dans Leclercq, Dupont et al. RESSAC, chap 8 de Leclercq <i>Diagnostic Cognitif et Métacognitif au seuil de l'université</i> 2003. Il aurait été insuffisamment diagnostique de s'en tenir au score total pour chacun des cours car cela n'aurait pas permis de distinguer les processus en œuvre. Bien que diagnostique, cette opération RESSAC était insuffisamment formative, ce qu'elle aurait été si elle avait par exemple invité à consulter une liste de FAQ ou encore à contacter Guidance-Etudes pour tout problème transversal de compréhension ou de mémorisation.
Visée	Sanctionnante	<p>Ce type d'évaluation est en principe l'opposé même de l'EF. Les exemples ci-après montrent cependant qu'un même dispositif peut délivrer les 2 types d'évaluation.</p> <ul style="list-style-type: none"> - Dans le cours de physique de Y. Lion, une banque de questions a été développée de façon à permettre à la fois l'évaluation sanctionnante et, à d'autres moments, des auto-évaluations formatives. Vu l'investissement en temps nécessaire à la réalisation

		<p>de certains tests sanctionnants, il est effectivement parfois intéressant d'augmenter le nombre de questions produites pour une telle banque de questions et d'en utiliser une partie avec une visée formative.</p> <ul style="list-style-type: none"> - Dans le cours de Déontologie (D. Leclercq et al.), chaque étudiant reçoit à la fois une note sur 20 (sanctionnante) et un ensemble de remarques (formatives). Une évaluation peut donc être à la fois sanctionnante ET formative. C'est aussi le cas pour toute défense de mémoire de licence qui sanctionne et forme.
Formative	NA	
Quoi (objets)		
Centrage sur	Processus	<p>C'est souvent un défi plus grand de centrer l'EF sur les processus que sur les matières. Plusieurs types de recueil de données et de FB sur le processus peuvent être envisagés :</p> <ul style="list-style-type: none"> - Dans le cours de A. Corhay (comptabilité des entreprises), les étudiants réalisent des exercices, puis comparent eux-mêmes leurs productions avec des corrigés qui mettent l'accent sur le processus, sur le raisonnement qui mène à la solution correcte. Les corrigés peuvent être assez longs, puisqu'ils doivent donner à l'étudiant l'occasion de comprendre chacune des étapes du développement. - Dans les tests de Halleux (1969), les étudiants répondent à des QCM dont les solutions présentent des démarches (transformation d'équations mathématiques) aboutissant toutes à la même réponse finale correcte,... mais parfois par compensation de plusieurs erreurs. Chaque distracteur est donc une démarche et l'étudiant est donc jugé sur la démarche choisie et non sur le résultat du calcul. - Dans le test de R. Burton (1999), les résultats intermédiaires permettent de détecter les modalités ou styles (intuitive ou algorithmique) de résolution pratiquées par les étudiants sur des équations du 1^{er} degré à une inconnue. - Les portfolios permettent à la personne elle-même, mais aussi aux enseignants, de réfléchir à son parcours de formation et de porter un œil critique sur son processus d'apprentissage. Ces portfolios sont à la fois formatifs et tout à fait centrés sur le processus. Ils ne sont cependant formatifs que si les enseignants (ou les pairs) fournissent un feed-back sur ce travail réflexif et des conseils pour s'améliorer. Exemple : Volet 3 de l'évaluation des étudiants de FORMASUP. (Bilan réflexif) - Les traces informatiques : étant donné les possibilités de traces, par exemple le tableau de bord WebCT, on imagine la série de FB formatifs du style « Vous avez très peu consulté le glossaire, mais beaucoup utilisé les hyperliens,... ». Exemple : PMTIC, physique (Y. Lion),...
Produits	Dans le cours de F. Pichault (Thalys), les groupes d'étudiants sont jugés sur les rapports de missions renvoyées par e-mail au	

		<p>professeur, qui leur attribue une note sans tenir compte de la manière dont les étudiants sont arrivés à ce rapport de mission. Qu'ils y aient consacré plusieurs réunions ou qu'ils ne se soient jamais rencontrés, qu'ils aient éprouvé de grandes difficultés ou qu'ils aient détricoté plusieurs fois leur travail (le processus d'atteinte du résultat) n'est pas pris en compte dans la note.</p> <p>Notons que le produit n'est pas toujours aussi complexe que dans l'exemple ci-dessus. Il peut être la réponse correcte (RC) ou incorrecte (RI) à des questions isolées, QCM, QROC, QROM.</p>
Dimension	Uni-dimensionnelle	<p>UNE SEULE dimension du travail est prise en compte pour l'évaluation. C'est le plus souvent le nombre de bonnes réponses. C'est parfois une seule autre qualité (ex. originalité de la réponse, délai de réponse, logique de raisonnement,...) de ce travail. C'est le cas de la plupart des examens de candidature qui se font sous forme de QCM (en grands amphithéâtres, sur feuilles permettant la lecture optique de marques - LOM) et pour lesquels le calcul du score se fait uniquement sur le nombre de réponses exactes. Ces examens peuvent être sanctionnateurs ou formatifs. Ils sont formatifs lorsqu'ils renseignent l'étudiant sur ses erreurs (sous-scores par chapitre par exemple) ou que celui-ci peut être informé des réponses correctes et incorrectes (si possible commentées) et les comparer aux siennes.</p>
	Pluri-dimensionnelle	<ul style="list-style-type: none"> - L'utilisation des degrés de certitude assure une pluridimensionnalité car sont alors mesurés à la fois l'exactitude des réponses (cognition) de l'étudiant et le réalisme (métacognition) de cet étudiant. D. Leclercq les utilise largement dans ses cours (en Edumétrie et Docimologie à distance, dans les autres cours en présentiel). A l'ULg, plusieurs professeurs utilisent ce système. Il a été appliqué dans MOHICAN 1999-2001. Il est utilisé pour tester les connaissances de pilotes et de mécaniciens d'avions, celles des intervenants en médecine d'urgence et celles de patients diabétiques (Paris-Padoue). - On peut aussi combiner l'exactitude avec la durée (délai de réponse), ou encore avec l'originalité, la logique de raisonnement (processus) ou d'autres critères. Par exemple, dans les TPs du cours d'Edumétrie et Docimologie (D. Leclercq et F. Demily), la note finale est une combinaison des appréciations données pour l'exactitude des réponses données, l'animation d'un forum, l'originalité du cas créé, le respect des deadlines, etc. <p>Notons que parfois, certaines dimensions interfèrent. C'est le cas de la vitesse de réponse. Dans un test en temps limité, certaines questions (de la fin du test en particulier) peuvent ne pas même avoir été atteintes par l'étudiant, qui n'a pas eu le temps d'arriver aussi loin dans son test. C'est évidemment un désavantage pour l'EF.</p>
Qui (sujets)		
Cible	Individuelle	Cas majoritaire : on évalue UN apprenant sur ses connaissances, ses compétences, ses conduites, etc.
	De groupe	François Pichault, dans son cours sur les théories d'organisation (Thalys), donne aux étudiants une note de groupe suite à leurs missions. Chaque groupe reçoit donc plusieurs notes, une par mission, mais chacune des notes est collective. L'évaluation est

		<p>formative car les étudiants reçoivent des commentaires sur leur rapport, ce qui leur permet de progresser au fur et à mesure des missions qui leur sont virtuellement confiées.</p> <p>Pour assurer un minimum d'équité dans l'évaluation et refléter les performances individuelles, les professeurs ont souvent recours à des méthodes mixtes, alliant les évaluations de groupe et individuelle.</p> <ul style="list-style-type: none"> - L'évaluation des étudiants du cours de philosophie de Thomas de Pratere (UCL) prend en compte quatre notes, qui sont combinées pour donner le score de chacun. Parmi ces notes, deux sont individuelles (pour un total de 60% de la note finale) et deux collectives (pour un total de 40% de la note finale). La participation continue aux forums est prise en compte au niveau individuel et au niveau du groupe, de même pour la qualité des rapports : richesse, cohérence, originalité, clarté, précision et organisation du rapport final déterminent une note pour chacun des groupes et une note tenant compte de l'apport personnel de chacun. Cet apport personnel est évalué selon la façon dont chacun a assumé l'un des 8 rôles qui lui avait été confié au départ : explorateur, secouriste, reporter, politique, logicien, traducteur, éditeur ou académicien (groupes de 8 étudiants, un étudiant par rôle et par groupe). - Dans le cours de PMD (D. Leclercq), les groupes de 4 étudiants devaient remettre un travail écrit coté globalement (la même cote pour les 4). Ils recevaient une cote commune, nuancée par la suite au cours d'une <u>interview individuelle en groupe</u> où les 4 étudiants présents étaient interrogés oralement tour à tour et au hasard sur divers aspects de leur production commune. Dans ce cas, l'évaluation alliait une phase groupale et une phase individuelle, la note finale étant une combinaison des deux notes. <p>Dans le cours de MFPA (D. Leclercq), les étudiants s'attribuent des <u>répartitions de tâches inconnues du professeur</u> au moment où il juge la performance du groupe. Pour chaque rubrique, le score groupal du professeur est ensuite combiné avec les répartitions déclarées par les membres de ce groupe. Pour certains étudiants, cette évaluation se transforme en évaluation individuelle, lorsqu'ils ont seuls pris en charge une rubrique. Pour d'autres, qui ont décidé de partager la responsabilité d'une rubrique avec des pairs, l'évaluation est groupale car ils obtiennent alors tous le même score pour cette rubrique.</p> <p>L'EF délivrée à un condisciple peut bénéficier à un étudiant par apprentissage vicariant, par observation (Bandura). A l'inverse, une qualité ou un défaut d'un travail peut n'être attribuable qu'à un seul membre du groupe, d'où les précautions décrites ci-avant.</p>
Destinataires	Privée	<p>Il arrive souvent qu'il soit préférable de garder confidentielles (privées) certaines données destinées à l'EF. Les rendre publiques risque d'entraîner un mécanisme de déni (ces évaluations sont non valides !), de défense de l'image sociale du moi, pour éviter une blessure narcissique.</p> <p>EvalEns est l'opération d'évaluation des enseignements de l'ULg par recueil des avis des étudiants. Cette opération annuelle est menée par le SMART (2700 cours en 2002). La facette privée de cette évaluation est le résultat détaillé de tous les scores moyens obtenus à toutes les (18) questions par un professeur particulier. Celui-ci est le seul (avec le Doyen) à recevoir ces</p>

	<p>données, ainsi que tous les commentaires rédigés. Il reçoit son feedback pour chacune des questions sous la forme d'une distribution (de -3 à +3) des moyennes des enseignants (ici : 9), la case grise étant celle où lui-même est situé.</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td>L'exposé est bien structuré</td><td>-3</td><td>-2</td><td>-1</td><td>+1</td><td>+2</td><td>+3</td></tr> <tr> <td></td><td>2</td><td>1</td><td>1</td><td>2</td><td>2</td><td></td></tr> </table> <p>Le professeur peut voir qu'il partage la moyenne +2 avec un collègue et que deux autres collègues ont un avis moyen plus favorable que lui (+3).</p> <p>En évaluation formative, les notes obtenues par chaque étudiant seront souvent connues de lui seul, de façon à respecter chacun et à favoriser la multiplication des essais : l'étudiant se testera d'autant plus volontiers s'il sait que ses notes resteront connues de lui seul et n'auront aucun impact sur ses résultats au cours.</p> <p>Par exemple, dans le cours de Philippe Lefèuvre (ORL), après avoir étudié le cas clinique proposé en ligne, chaque étudiant envoie son diagnostic au professeur, qui renvoie un FB personnalisé, visible par cet étudiant uniquement puisqu'il est acheminé par mail. Cette procédure est d'autant plus formative que les FB sont personnalisés.</p>	L'exposé est bien structuré	-3	-2	-1	+1	+2	+3		2	1	1	2	2	
L'exposé est bien structuré	-3	-2	-1	+1	+2	+3									
	2	1	1	2	2										
Publique	<p>La même opération EvalEns a une facette publique : les résultats d'une section et d'une faculté sont mis à disposition des étudiants, mais sans que l'on puisse identifier de quel professeur il s'agit. Ce qui est privé (confidentiel) pour chaque enseignant est public pour la faculté qui, elle aussi, pourrait en ressentir une blessure narcissique. C'est pourtant l'options choisie par l'ULg. Tous voient donc la même distribution, mais sans case grise.</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td>L'exposé est bien structuré</td><td>-3</td><td>-2</td><td>-1</td><td>+1</td><td>+2</td><td>+3</td></tr> <tr> <td></td><td>2</td><td>1</td><td>1</td><td>2</td><td>2</td><td></td></tr> </table> <p>La FAPSE-ULg a utilisé ces feedbacks pour améliorer (au cours des années successives) ses performances sur certains points, par exemple l'existence et la qualité des feedbacks aux étudiants suite à des évaluations.</p> <p><i>"L'EF est une sorte de dialogue particulier entre l'éducateur et son élève"</i> (de Landsheere, 1979, 115).</p> <p>Dans la pratique, nombreux sont les enseignants qui publient les résultats des étudiants en les affichant, en les rendant ainsi disponibles au regard des autres étudiants, des parents. Certains parents viennent vérifier aux valves si leur enfant ne leur a pas menti ; en conséquence, pas mal d'étudiants viennent subir un examen sans y être préparés, juste pour avoir une note et ne pas</p>	L'exposé est bien structuré	-3	-2	-1	+1	+2	+3		2	1	1	2	2	
L'exposé est bien structuré	-3	-2	-1	+1	+2	+3									
	2	1	1	2	2										

		<p>subir le reproche parental "tu n'as même pas essayé, tu ne t'y es pas préparé".</p> <p>Cet affichage a d'autres effets, notamment cette blessure narcissique racontée dans le cas (inspiré d'un fait réel) traité dans le cours de Déontologie (D. Leclercq et al.) "La demoiselle en larmes", où une étudiante est fort touchée par le fait que ses pairs ont tous pu voir aux valves qu'elle avait lamentablement raté son examen. A l'université de Durham (UK), les notes sont carrément affichées sur le mur extérieur de l'université, donc sur le trottoir... Ces affichages publics déplaisent fortement à une frange des étudiants, principalement ceux qui échouent. Leurs inconvenients peuvent être contournés par l'affichage des matricules plutôt que des noms des étudiants. D. Leclercq pratique cette formule pour des cohortes d'étudiants en 1ère candi, 2e candi, Agrégation et CAPAES.</p> <p>Par contre, dans le cours de PAV, D. Leclercq corrige publiquement certaines activités, affichant sur le site du cours les réponses des uns et des autres de façon à donner des idées et des points de comparaison à chacun. C'est très formatif pour l'étudiant qui prend la peine d'analyser ses propres erreurs... mais aussi celles des autres. Nous sommes ici dans un plus petit groupe, où les étudiants se connaissent, et face à des activités à visée très formatrice.</p>
Opérateurs	Allo-opératrice par expert(s)	<p>En EAD, il est fécond de distinguer l'auto-évaluation et l'allo-évaluation (par expert ou prof) en fonction du degré de dépendance/d'autonomie de l'étudiant par rapport au professeur. Si l'étudiant dépend du professeur ou d'un autre expert pour recevoir son FB, nous parlerons d'allo-évaluation. Sinon, nous parlerons d'auto-évaluation.</p> <p>Dans l'allo-évaluation, l'enseignant (voire un autre expert) juge (de façon subjective) ou jauge (mesure de façon objective) les performances de ses étudiants. Souvent, cette allo évaluation n'entraîne pas l'étudiant à mener ces opérations de façon autonome. En effet, l'EF menée par autrui constitue ce que Jérôme Bruner appelle un étayage (scaffolding), ou des bêquilles, qui devront disparaître. C'est un processus normal d'autonomisation progressive, Vygotsky (1932) définissant la Zone Proximale de Développement comme « ce que l'apprenant peut faire aujourd'hui avec l'aide d'autrui, et demain seul ». Bien entendu, remplacer l'enseignant par un algorithme qui, lui aussi, fait le travail d'évaluation en lieu et place de l'étudiant, ne favorise pas non plus son autonomie.</p> <p>Des outils peuvent aussi supplanter une partie du travail mental. Ainsi, en Réanimation Cardio Pulmonaire (RCP), on utilise un mannequin grande taille appelé « Ressusci Anne », relié à un ordinateur qui mesure divers indices difficiles à déceler à l'œil nu. Par exemple, l'endroit où l'apprenant a appuyé et avec quelle force (sur les côtes flottantes, si bien qu'on les casserait) ; le rythme des pressions (combien par minute ?) ; le volume insufflé dans les poumons lors de chaque « bouche à bouche ». La difficulté est qu'il va falloir s'habituer à s'auto-évaluer SANS ce système d'évaluation « assistée ».</p>

	<p>L'allo-évaluation a tout son sens lorsque la complexité de la tâche permet à l'enseignant de vraiment faire la différence avec une machine, ce qui est très fréquent. L'étudiant est alors dépendant de son enseignant pour recevoir un FB. Il s'agit souvent de tâches telles des corrections de dissertations, des avis critiques sur des processus (résolution de problèmes complexes, analyses de cas) ou encore des débats et argumentations.</p>
Auto-opératrice (auto-évaluation)	<p>Un des buts centraux de l'EF est de rendre l'apprenant autonome, en particulier de favoriser l'internalisation de la confirmation : qu'il puisse lui-même détecter et juger de l'exactitude / inexactitude de ses réponses, pour qu'il passe aussi à la phase formative de façon flexible et autonome. Nunziati (90), cité par Jans (2000, 28), trouve crucial d'aider l'étudiant à se construire « un bon système interne de pilotage ». Pour cet auteur, la régulation opérée par l'enseignant n'est qu'une « roue de secours » qui ne devrait intervenir que « lorsque les mécanismes d'autorégulation de l'élève sont grippés ».</p> <p>Donnant les rennes à l'étudiant, le cours Méthodes en ligne (Guidance-Etudes), l'amène à se questionner sur ses propres pratiques. Il choisit lui-même le champs de ses investigations et se promène dans le cours en fonction de ce qu'il ressent comme étant ses faiblesses, ses lacunes.</p> <p>En EAD, il est possible de fournir automatiquement à l'étudiant quantité de données le renseignant quant à ses consultations, sa trajectoire, ses multiples résultats, le nombre de ses interactions, etc (cfr tableau de bord de François Georges, utilisé dans le cours de physique de Y. Lion). Ces données peuvent être formatives (en fonction de la façon dont l'étudiant les exploite), mais il faut prendre garde à ce que l'étudiant puisse aussi réguler ses apprentissages en se passant peu à peu de ces aides. En effet, autant il est simple de donner à l'étudiant un FB automatique sur des problèmes simples (performances isolables), autant il est impossible de tout régler de la sorte. Il est donc très intéressant pour l'étudiant de s'équiper des compétences nécessaires à son autonomisation dans ses propres évaluations, pour s'affranchir au maximum des dépendances extérieures.</p> <p>Notons qu'en cas de correction automatisée d'un QCM, par exemple, nous parlerons d'auto-évaluation même si, évidemment, l'enseignant a conçu l'algorithme par lequel l'étudiant reçoit ses FB. Quel qu'ait été l'investissement de l'enseignant dans la construction de l'outil, nous considérons qu'il s'agit d'auto-évaluation puisque l'étudiant est autonome dans l'accès à son FB.</p> <p>Parmi les actes d'évaluation habituellement exercés par le professeur, l'étudiant a parfois l'occasion d'en prendre lui-même certains en charge : choisir l'objectif à évaluer, choisir le moment, choisir l'instrument, corriger sa copie (avec des critères inventés par lui ou fournis par l'enseignant), comparer lui-même ses performances à ce qui était attendu, émettre un diagnostic, procéder à la phase formative. Prendre en charge ces tâches demande évidemment un entraînement, un accompagnement, une progressivité.</p> <p>Dans les tests de Chimie (Cl Houssier), l'étudiant reçoit les résultats de son test avec FB diagnostiques, souvent même après</p>

	<p>chaque question. En Comptabilité des entreprises (A. Corhay), il compare ses réponses à celles d'un corrigé (ex. il reçoit en ligne des corrigés d'écritures comptables complexes, à comparer avec ses propres productions).</p> <p>Le recours aux degrés de certitude (ex. Edumétrie et Docimologie de D. Leclercq) est une façon systématique et transversale d'habituer les étudiants à se questionner sur leur propre certitude. On vise par là plus de réalisme, donc une capacité plus importante à prendre en charge certaines phases de sa propre formation. L'étudiant réaliste sait quand il peut arrêter de creuser une matière car il sait quand il la maîtrise, il sait quand il doit retourner au cours car il se rend compte de ce qu'il n'a pas compris, etc. Cette compétence du réalisme dans d'auto-estimation de ses performances sont cruciales pour un bon « auto-apprenant ».</p> <p>Les liens entre auto-évaluation (solo) et la réussite à l'examen mené de façon allo-évaluative ont été montrés dans l'expérimentation du logiciel Crânes (Poncin et Gilles, 2000).</p>
Allo-opératrice par pairs	<p>L'évaluation n'est susceptible d'être menée par les pairs QUE lorsqu'elle est formative car on ne peut leur demander de prendre la responsabilité de sanctionner leurs condisciples.</p> <p>Dans les TP d'Edumétrie et Docimologie en ligne, chaque étudiant est responsable de l'évaluation des cas créés par deux de ses pairs. Il est d'ailleurs notamment jugé sur la pertinence des FB qu'il leur renvoie. Cette évaluation est très formative pour les pairs, qui reçoivent le FB de leur pair mais qui sont aussi guidés au travers de discussions en forums (animés eux-aussi par les étudiants) pour tout point qu'ils n'auraient pas bien compris.</p> <p>On pratique souvent ce genre d'évaluation par les pairs lors d'une présentation orale devant les pairs, ou dans des contextes de supervision (lors d'un stage clinique) ou de <i>mentoring</i> (par un étudiant plus âgé). Ce type d'évaluation mérite d'être plus développé en formation. En effet, Falchikov (1995, 175) cite des expériences montrant :</p> <ul style="list-style-type: none"> - que l'évaluation par les pairs constituait le meilleur prédicteur de la performance médicale d'interne (en médecine). - que l'évaluation par les pairs est bien corrélée avec celle des experts sur les comportements interpersonnels, ce que n'est pas l'auto-évaluation.
Comment ? (méthodes, conditions)	
Périoricité	<p>Ponctuelle</p> <p>« <i>Evaluation effectuée à un moment donné pour répondre à une question à propos d'un individu ou d'un programme (passage d'une classe, attribution d'une bourse,...). Ce type d'évaluation ne semble se justifier pleinement que dans les cas où une évaluation continue est impossible.</i> » (DE LANDSHEERE, 1979, p.114)</p> <p>Ce type d'évaluation n'est pas l'idéal pour l'EF, qui est ainsi privé de la vérification de l'efficacité des remédiations.</p>

	<p>Il arrive qu'une performance soit évaluée une seule fois dans la vie d'une personne. C'est le cas le plus souvent, du mémoire de fin d'études universitaires, car les étudiants n'ont jamais eu auparavant l'occasion de mener à bien un travail d'une telle ampleur et présentant de telles exigences et la plupart d'entre eux ne produiront pas de doctorat. On s'efforce, cependant, de les y préparer par des travaux similaires « en réduction ». C'est une EF s'il y a possibilité de transfert à d'autres situations : on ne refera plus un mémoire, mais on fera des rapports, etc.</p> <p>Il existe des situations où le formé lui-même souhaite être libéré d'une évaluation continue, par exemple dans l'enseignement supérieur. Il n'est donc pas rare de voir des institutions offrir le choix à leurs étudiants : SOIT une évaluation « étalée » (dans l'enseignement universitaire par des « partiels dispensatoires »), SOIT une évaluation « concentrée » (examens uniquement durant la session).</p> <p>En EAD, on trouve rarement de telles situations puisque l'auto-évaluation est fort répandue. Les étudiants se testent grâce à de nombreux QCM durant l'année. L'évaluation ponctuelle, dangereuse et risquée pour les étudiants, est donc généralement évitée.</p>
Continu	<p>« Collecte systématique de scores ou d'appréciations...aboutissant à une note finale. Peuvent intervenir dans l'élaboration de cette note : les travaux faits en classe, les travaux à domicile, les résultats d'interrogations écrites, de tests, ainsi que le jugement subjectif de l'enseignant. L'évaluation continue est...un processus cumulatif, suivant le développement de l'élève et réfléchissant les changements qui interviennent dans ses réactions au cours. » (DE LANDSHEERE, 1979, p. 112).</p> <p>Sauf si elle porte sur des contenus différents à chaque fois, l'EF continue permet de ne pas laisser s'ancrer des erreurs, de ne pas les laisser s'étendre aux nouvelles matières.</p> <p>Contrairement à l'évaluation « répétée », ce ne sont pas les mêmes épreuves qui sont utilisées en évaluation continue (ou « diluée » ou encore « répartie ou fractionnée » dans le temps). Les épreuves sont variées, soit parce qu'elles portent sur des matières différentes, soit parce que les questions posées varient, sur une même matière.</p> <p>L'évaluation est continue dans le cours « Formasup », où les participants sont amenés à rentrer des travaux, des valorisations et des rapports réflexifs tout au long de l'année, éléments qui sont tous pris en compte dans l'évaluation finale (soit de façon conditionnelle, soit au travers d'une note). Dans ce cas, l'évaluation est aussi clairement formative, puisque tout élément fait l'objet de commentaires personnalisés de l'équipe du LabSET.</p> <p>De la même façon mais en proposant des tâches plus similaires tout au long de l'année les cours de Chimie (Claude Houssier), de Biostatistiques (A. Leroy) et de Comptabilité des entreprises (A. Corhay), par exemple, appliquent aussi un principe d'évaluation continue.</p> <p>La pédagogie de la maîtrise (Bloom, 1969 et 1986) propose aussi ce principe d'évaluation continue, mais en lui ajoutant le</p>

	<p>principe de progressivité, de difficulté croissante. Ce principe est appliqué dans des cours comme celui de Cenforgil (préparation à la recherche d'emploi) ou de Handelscorrespondentie (FOREM), où la réussite d'une tâche conditionne l'accès à la suivante, dans cet esprit de maîtrise progressive.</p>
Répétée	<p>L'évaluation répétée est la situation idéale pour l'EF. C'est d'ailleurs la situation de l'athlétisme ou de la natation : c'est toujours la même épreuve dans laquelle on mesure ses progrès. Les QCM d'une quarantaine de questions disponibles pour chacun des chapitres du cours de chimie (Cl Houssier) en sont un bel exemple : l'étudiant qui ne maîtrise pas la matière peut se confronter au test autant de fois qu'il le désire. Il trouvera les mêmes questions et constatera ses progrès. De même, et les langues étrangères sont un domaine dans lequel cette répétition est particulièrement nécessaire pour fixer l'imprégnation, les cours de Rudiments d'Espagnol (EAD-CFWB) et de British and American English (EAD-CFWB) permettent à l'étudiant de se confronter plusieurs fois à un même test, comprenant notamment des extraits sonores.</p> <p>Dans un Progress Test, des questionnaires parallèles (questions semblables mais pas identiques) constituent une sorte de répétition. Ainsi, les étudiants en médecine à Maastricht passeront 24 fois « le » Progress Test, une fois par trimestre.</p> <p>La répétition au sens strict n'existe que quand les mêmes questions sont passées à des semaines, des mois voire des années d'intervalle. Ce fut le cas dans l'opération PREDIC (Boxus 1971) où le même test a été appliqué 4 fois aux mêmes sujets à 3 mois d'intervalle chaque fois. Les courbes obtenues sont successivement en I, en Gauss et en J. (PPUQ, 1998, 64). Le passage du I à la distribution Gaussienne puis au J est l'évolution normale d'une courbe d'apprentissage.</p> <p><i>En pédagogie, les « retests » sont difficiles, car le simple fait de passer le test produit un apprentissage.</i> G. de LANDSHEERE (1979, p. 238). Notons que par « retest », de Landsheere entend « le test répété dans les mêmes conditions et à l'aide du même instrument ou d'une forme parallèle (équivalente) ». Tout comme de Landsheere, nous considérerons en EAD que des formes parallèles d'un même test constituent une évaluation répétée. En effet,</p> <ul style="list-style-type: none"> A - il est de toute façon bien difficile, au vu d'un QCM en ligne, de savoir s'il est exactement semblable à celui du mois précédent ou si certaines questions ont été conçues avec des variables qui prennent une valeur différente à chaque passage du test. B - l'important étant la démarche mentale de l'étudiant et l'aspect formatif du test, passer EXACTEMENT le même test ou en passer une forme parallèle importe peu. Au contraire, le fait de s'attaquer à une forme parallèle peut assurer l'étudiant qu'il n'a pas mémorisé un pattern mais qu'il a effectivement compris un principe. <p>En raison du point B ci-dessus, nous conseillerons d'ailleurs la création de formes parallèles en EF. Ce principe est appliqué dans le cours de Génétique Quantitative (J. Detilleux).</p>

Révisabilité des scores/notes	Définitive	<p>Ce principe est peu propice à l'EF, qu'il décourage.</p> <p>La plupart du temps, les notes sont pourtant sans appel. C'est le cas de la plupart des cours à distance également : Chimie, Comptabilité, ...</p>
	Améliorable	<p>Ce principe d'évaluation améliorables encourage l'EF.</p> <p>Sur un travail complexe (ex : dans le cours PMD de D. Leclercq), l'étudiant reçoit une première note ainsi que des critiques et des conseils sur son travail. Il a alors un mois pour re-soumettre son travail après l'avoir amélioré. Le score final est la moyenne des deux. Ce principe évite que l'étudiant remette une première version bâclée, car elle pèsera pour la moitié des points. En outre, la capacité d'un étudiant à tirer parti des critiques qui lui sont faites mérite en soi d'être prise en compte. La seule situation dramatique est celle où la première version est tellement mauvaise que sa note (faible) est « irrattrapable ».</p> <p>Une autre modalité est la <u>défense orale d'un écrit</u> ...en vue d'améliorer sa performance sur cette base (Cours de Déontologie, mémoire de fin d'études). Le professeur peut n'octroyer ce « droit à l'amélioration » que si la copie atteint un minimum fixé au préalable. Il est fréquent que les professeurs aient fixé leur note pour la version écrite avant cette défense.</p> <p>François Pichault agit de même dans son cours de QA de gestion des ressources humaines, où les étudiants doivent faire le point sur un thème (ex. le temps de travail, la communication,...) mais en plusieurs étapes : préparation, exposé, débat, « working paper ». Les étudiants reçoivent des FB et améliorent leur travail sur base de ces conseils à 3 étapes. Ils ont donc 3 possibilités d'amélioration de leur travail.</p>
Source	Objective	<p>L'objectivité n'existant pas, on la définit comme « le consensus entre experts ». En évaluation, pour y parvenir, on s'efforce souvent d'obtenir des échantillons de comportements dont la mesure puisse être la même, quels que soient les juges, à condition qu'ils marquent leur accord sur les instruments et les critères (idéalement prédéfinis) et les appliquent. C'est le cas des QCM. Des échelles d'évaluation descriptives (c-à-d où les échelons ou niveaux sont décrits) sont d'autant plus objectives que la description de chaque échelon comporte moins d'ambiguïté (ex. un échelon « plus de deux fois par semaine » est plus objectif que « souvent »).</p> <p>Souvent, l'EF se base sur des données objectives, mais leur interprétation comporte une composante subjective.</p> <p>Dans tous les QCM des cours en ligne accompagnés par le LabSET, les solutions possibles sont objectives.</p>
	Subjective	<p>L'évaluation à source subjective est aussi nommée estimation ou jugement.</p> <p>Le jugement (par définition subjectif) le mieux connu est la notation de dissertations par un professeur, avec tous les problèmes que l'on connaît de non concordance inter-juges et même de non stabilité intra-juge. C'est l'objet même de la docimologie critique,</p>

	<p>fondée par Piéron (1963).</p> <p>L'EF peut grandement bénéficier de la confrontation des subjectivités (conflits socio-cognitifs), comme c'est le cas dans le paradigme du débat.</p> <p>Les évaluations subjectives sont aussi fréquentes dans les cours en ligne, sous forme de :</p> <ul style="list-style-type: none"> - Questionnaires d'avis en fin de cours (FORMADIS, FORMASUP et pratiquement tous les cours développés avec le soutien du LabSET) - Autoestimations des performances au travers des indices de certitude (dans la plupart des cours de D. Leclercq, et notamment Edumétrie et Docimologie). - Portfolios, qui permettent à la personne elle-même, mais aussi aux enseignants, de réfléchir à son parcours de formation et de porter un œil critique sur son processus d'apprentissage. Ces portfolios sont à la fois formatifs et tout à fait centrés sur le processus. Ils ne sont cependant formatifs que si les enseignants (ou les pairs) fournissent un feed-back sur ce travail réflexif et des conseils pour s'améliorer. Exemple : Volet 3 de l'évaluation des étudiants de FORMASUP (bilan réflexif).
Procédure	<p>Standardisée</p> <p>En standardisation totale, tous les étudiants d'une classe reçoivent les mêmes questions, y répondent dans les mêmes conditions (la surveillance le garantit) et sont évalués selon les mêmes critères, avec la même procédure (les Américains parlent de « tests objectifs » quand la correction est automatisable). Le problème de la surveillance se posant avec acuité dans les examens en EAD, des « <i>Trusted Third Parties</i> », partenaires fiables, assurent cette surveillance, comme le font les ambassades des USA où se passent les épreuves EAD d'universités américaines.</p> <p>A l'université, les examens de candidatures (grands groupes) relèvent généralement de cette catégorie, ce qui évite de nombreux recours au conseil d'état pour les étudiants qui s'estimerait différemment défavorisés par les questions reçues.</p> <p>La standardisation peut aussi ne porter que sur l'une des composantes de l'évaluation : son contenu, ses conditions de passation ou encore sa correction.</p> <p>En EAD, les conditions d'administration (l'endroit où chacun passe le test, le temps passé sur le test, le nombre d'essais, etc) ne sont généralement pas standardisées mais les contenus des tests le sont par contre souvent. Par exemple, dans les cours de Chimie (Cl Houssier) et Physique (Y Lion), les tests sont disponibles pour tout qui veut s'y essayer, mais les étudiants peuvent tout autant les passer chez eux que sur le campus, à toute heure et sans restrictions de temps. La correction, par contre, est standardisée et automatique. Dans le cours d'échographie en reproduction bovine, (Ch. Hanzen), chaque participant prend des mesures (manuelles et échographiques) sur des bovins différents mais les questions de départ sont les mêmes pour tous, de</p>

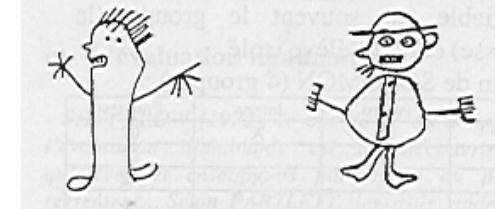
	<p>même que la façon dont seront traités les résultats de ces mesures et leurs interprétations.</p> <p>Exception à ces pratiques : dans le cours d'Edumétrie et Docimologie (D. Leclercq), l'examen final s'est déroulé en 2003 à CAFEIM avec une standardisation totale : les étudiants ne choisissaient pas l'ordre des questions, ils répondaient linéairement et en temps limité (bien qu'aucun étudiant n'ait déclaré avoir eu trop peu de temps pour passer le test).</p> <p>Comme dans le cas du cours d'Edumétrie et Docimologie, l'EF peut exiger une procédure standardisée puisque c'est à une telle procédure qu'il faut préparer les étudiants, mais c'est assez rare. Généralement, l'EF n'exige pas de standardisation. Il est même préférable d'offrir aux étudiants une flexibilité maximale plutôt qu'une standardisation qui ne tient pas compte de leurs différences interindividuelles.</p>
Adaptative	<p>La composante adaptative peut être entre les mains de l'étudiant ou entre celles de l'enseignant, voire même du système.</p> <p>Entre les mains de l'étudiant : la procédure est adaptative lorsqu'elle permet par exemple à chaque étudiant de choisir sa question (ex. le thème d'une dissertation ou le chapitre sur lequel il sera questionné), le thème du projet sur lequel il sera évalué (voir aussi évaluation « négociée »), ou, plus simplement, l'ordre dans lequel il souhaite voir traiter les questions.</p> <p>Dans cette acception, une procédure non standardisée est automatiquement adaptative. Cette flexibilité est souvent fortement en cohérence avec l'EF.</p> <p>En EAD, de nombreux dispositifs prévoient des activités optionnelles, que le participant choisit en fonction de ses préférences ou de ce qu'il pense être ses lacunes. Ainsi, en Chimie (CI Houssier), il choisira de ne pas passer le test sur le Chapitre 2 s'il est certain de le maîtriser au point de n'avoir pas besoin de vérification. Autre exemple : dans le cours FORMASUP, les participants choisissent eux-mêmes les thèmes des 6 ECTS de valorisations souhaitées. Chacun sera donc évalué sur un travail qui traite d'un thème différent, mais aussi d'un type différent (ex. l'un sur son compte-rendu d'une conférence sur l'enseignement à distance, l'autre sur son résumé d'un ouvrage traitant d'évaluation formative, un troisième sur sa propre évaluation d'une expérience qu'il mène avec ses étudiants). Dans le cours de Simulation du lancement d'un nouveau produit (HEMES Ste Marie), les groupes d'étudiants choisissent le thème de leur projet, qui servira de base à leur évaluation.</p> <p>Entre les mains de l'enseignant : L'examen oral est souvent adaptatif, évoluant dans le temps, puisque l'interrogateur modifie ses questions subséquentes (leur nature et leur nombre) en fonction des réponses de l'évalué aux précédentes et s'arrête quand il dispose de suffisamment d'informations pour « trancher ». C'est le principe du testing séquentiel de WALD (1943). L'EF exige souvent un tel questionnement adaptatif, non standardisé. C'est aussi la fameuse « méthode clinique » de Piaget, qui ne se prive pas de poser des sous-questions en fonction des réponses de l'apprenant et qui pense que c'est la seule façon d'y « voir clair », de comprendre.</p>

		<p>Entre les mains » du système : le principe du testing séquentiel de WALD peut être automatisé, par ordinateur. D. Leclercq a expérimenté cette formule dans son cours de Technologie donné à l'université d'Aoste. Le principe adaptatif entre les mains du système est poussé à son paroxysme avec les « Intelligent Tutoring Systems », qui prennent des décisions pour l'utilisateur sans lui laisser percevoir ni la nature de ces décisions (ex. le système ne posera plus aucune question sur la matière du Chap 2), ni les raisons pour lesquelles elles ont été prises (ex. l'étudiant ne maîtrise pas les prérequis, qui se trouvent dans le Chap 1). Le système analyse les erreurs de l'étudiant et lui envoie des questions de niveau plus facile ou plus difficile, en fonction de ces erreurs, sans commentaires.</p> <p>Au point de vue pédagogique, nous préférons généralement les systèmes qui éventuellement orientent l'étudiant en lui proposant des trajectoires contrastées mais jamais en prenant les décisions à sa place, au risque d'hypothéquer le développement des compétences métacognitives de l'étudiant : sa capacité à prendre de bonnes décisions pour son propre apprentissage, analysant réaliste-ment ses forces et ses faiblesses.</p>
Implication	Interne ou participante	<p>« L'anthropologue Malinovski parlait d'<i>observation participante</i> qui « amène le chercheur à vivre la vie des groupes qu'il étudie, à partager le plus possible leurs activités pour mieux comprendre leur vision du monde » (de Landsheere, 1979,192). L'évaluateur est « participant » quand il a été impliqué dans la formation de l'évalué.</p> <p>L'EF se produit majoritairement de cette façon, ce qui est normal puisque l'EF fait partie du processus NORMAL d'apprentissage et d'ENSEIGNEMENT.</p>
	Externe	<p>Les évaluateurs n'ont pas été impliqués dans la formation des évalués. En France, c'est le cas du Baccalauréat et des Concours d'entrée aux grandes écoles. Aux USA, c'est le cas pour le SAT (Sholastic Aptitude Test, charnière secondaire - université) et pour les Medical Boards Exams comme le FLEX (permettant de pratiquer la médecine dans un état autre que celui du diplôme). Internationalement, le TOEFL (Test of English as a Foreign Language) n'est administré que par des organismes indépendants des étudiants. En Belgique, les examens cantonaux (primaire), l'examen d'entrée aux ingénieurs et bientôt l'examen de sélection des vétérinaires, les examens du SELOR (Sélection et Orientation de la Fonction Publique), sont autant d'exemples d'évaluations externes.</p> <p>L'EF est souvent absente de ce type d'opération, mais ce n'est pas incompatible. Tout dépend de la comparabilité permise entre les réponses et les réponses attendues. Or souvent on entretient un maximum d'obscurité, de secret sur les procédures et les détails des notes pour éviter les recours légaux (Conseil d'Etat). Ces évaluations externes ont pourtant souvent comme caractéristique une grande « validité de construct » (leurs questions ont été étudiées pour couvrir la matière, calibrées pour correspondre exactement aux compétences attendues, etc.), leurs résultats sont souvent très diagnostiques,... mais malheureusement rarement communiqués avec le degré de détail qui permettrait aux étudiants d'en faire un réel outil de</p>

		formation,
Contrat	Imposé	<p>Système le plus fréquent : l'évaluation « vient d'en haut » et il n'y a pas de place pour une discussion entre l'évaluateur et l'évalué. En formation d'adultes, ce type d'évaluation est de plus en plus rarement pratiqué car il ne tire aucun parti de la richesse des interactions. L'EF y est moins souvent rencontrée, bien qu'il n'y ait pas d'incompatibilité. Le danger est que les apprenants n'y perçoivent pas la visée formative éventuellement pourtant présente.</p> <p>Dans les cours de candidature, les contrats sont généralement imposés aux étudiants. Des situations mixtes se rencontrent souvent en EAD, où l'évaluation finale est imposée mais une grande flexibilité s'installe quant à l'EF continue. Ainsi, en Chimie (Cl Houssier), en Compabilité des Entreprises (A. Corhay) ou encore en Biostatistiques (P. Leroy), les étudiants ne se voient imposer que l'évaluation finale, les tests auto-formatifs étant facultatifs, ce qui est une forme primaire de négociation.</p>
	Négocié	<p>Dans le processus d'Audit des universités, on procède en deux temps. L'université rédige son propre rapport interne précisant ses objectifs, ses contraintes, et les critères selon lesquels elle souhaite être évaluée. C'est sur base de ce rapport d'auto-évaluation que les experts extérieurs évalueront.</p> <p>Sur base de ce type de négociations, l'EF devrait, en principe, avoir ici son maximum d'efficacité puisque l'apprenant est demandeur et que l'évaluation va donc se produire « <i>just in time</i> ».</p> <p>Dans le cours FORMASUP, les 6 ECTS de valorisations externes font l'objet de négociations entre chaque participant et l'équipe d'encadrants. Le contrat est négocié avec chacun.</p> <p>Dans une expérience, Jans et Leclercq (Jans 2000) ont demandé à leurs étudiants d'auto-évaluer leur rapport écrit (ipso-évaluation) sur 8 aspects. Les deux encadrants ont ensuite évalué les travaux (allo-évaluation). Un dialogue transactionnel s'est alors engagé entre les deux évaluateurs et chaque étudiant, en se concentrant sur les notes divergentes. Ces dialogues ont été filmés et analysés par V. Jans. Tantôt l'étudiant "acceptait" les arguments (et la note) des encadrants, tantôt c'était l'inverse ; le professeur était donc parfois amené à modifier sa note.</p> <p>STAKE désigne par l'expression « évaluation répondante » une démarche focalisée plus sur les processus que sur les résultats. Pour DE LANDSHEERE (1979), « <i>une l'évaluation est dite répondante [notamment] si elle porte plus sur les actions réelles que sur les intentions initiales...</i> » (p.115). On retrouve dans l'évaluation répondante la flexibilité de la négociation. L'EF peut (et parfois doit) être répondante elle aussi, c'est-à-dire porter sur des effets inattendus d'une intervention de formation plutôt que de se focaliser uniquement sur ce qui était planifié au départ.</p> <p>Par exemple, en EAD, il est fréquent de constater des difficultés dans la maîtrise des outils technologiques. Ces freins handicapent certains étudiants. Il est alors conseillé d'adapter le dispositif et de prévoir des mises à niveau technologiques pour</p>

		les étudiants chez qui des manques ont été diagnostiqués. On évite ainsi de laisser s'installer une situation de découragement. Ce fut le cas dans le cours de PAV de D. Leclercq, où les étudiants ont bénéficié de cours visant des maîtrises logicielles. Les intentions initiales ne prévoyaient pas ces remises à niveau mais l'enseignant a adapté son dispositif.
Repères/ échelle	Mobiles	<p>Il s'agit du système le plus fréquent : les questions de 4e année sont plus compliquées que celles de 3e année, les partiels de janvier ne couvrent pas la même matière que les examens de juin, ou encore le test portant sur le chapitre 5 ne reprend pas les questions des chapitres antérieurs.</p> <p>Cette évaluation ne permet pas d'avoir une vue permanente de la maîtrise des prérequis, ce qui peut handicaper l'EF. Par contre, côté avantages, elle est plus facile à organiser car comporte moins d'items et ceux-ci portent sur la seule "fenêtre d'incertitude", excluant les Q trop faciles ou trop difficiles pour le niveau de l'apprenant. Elle peut donc s'organiser très fréquemment et revêtir des formes diverses, propices à l'EF.</p>
	Fixes	<p>Les tests à repères fixes permettent d'avoir une vue d'ensemble des compétences, et tout spécialement de la maîtrise ou non des prérequis, ce qui est précieux dans un processus d'EF. Ils sont cependant plus complexes à organiser, principalement parce qu'ils nécessitent une coordination verticale des enseignements et des évaluations, donc un accord entre de nombreux enseignants de même que des décisions institutionnelles. Ils nécessitent aussi un très grand nombre de questions.</p> <p>Un exemple : les « Tests de progrès » pratiqués tous les 3 mois à la faculté de médecine de Maastricht (LECLERCQ et VAN DERVLUTEN, 1998, 200) sont obligatoires. Ils comportent 250 QCM Vrai-Faux, portent sur toute la médecine, et sont imposés à tous les étudiants en médecine (de la 1ère à la 6e année). Le contenu est donc le même pour tous. Cependant, les seuils de réussite diffèrent selon l'année de l'étudiant : on attend de l'étudiant de 1ère année qu'il réussisse 10% de ces questions, de ceux de 2e année qu'ils réussissent 20%, etc.</p>
Visibilité des critères	Annoncés	<p>Les critères sont précisés dès le départ. Par exemple, dans les QCM, la consigne précise :</p> <ul style="list-style-type: none"> - la nature de la <u>question</u> (une seule solution est correcte, etc.) ; - comment <u>répondre</u> (par exemple en accompagnant sa réponse d'un degré de certitude). - les <u>conséquences</u> en cas de Réponse Correcte (RC) ou Incorrecte (RI), en appliquant un barème de tarifs annoncé (TC = Tarif en cas de réponse correcte ; TO = en cas d'omission ; TI = en cas de Réponse Incorrecte). <p>Pour les performances complexes, on annonce quels critères seront appliqués (mais pas forcément les pondérations, pour ne pas encourager les réponses trop « stratégiques » d'étudiants qui laisseraient carrément tomber un volet trop peu pondéré).</p> <p>L'EF bénéficie souvent de l'annonce des critères car ainsi l'apprenant sait clairement ce qui attendu de lui ; par contre, cela ne permet pas de juger de la spontanéité du comportement.</p>

		<p>Les QCM de Chimie (Cl Houssier) ou Physique (Y. Lion) bénéficient de ce système très transparent quant aux critères d'évaluation. Sur des performances complexes cette fois, les cours de D. Leclercq (Edumétrie et Docimologie, PAV, Déontologie) sont eux aussi très précis quant aux critères d'évaluation, communiqués aux étudiants en début d'année. Cette communication de plus en plus transparente va dans le sens de la définition précise de contrats pédagogiques pour chaque cours, contrats qui sont dorénavant disponibles pour chaque cours de l'ULg en début d'année. Dans le curriculum FORMASUP, les participants seront évalués début juin sur base des critères communiqués, de façon à avoir encore l'occasion d'améliorer leurs productions avant la défense de leur mémoire, qui se fera fin juin sur base de ces mêmes critères. L'EF trouve ici un renforcement dans la transparence des critères.</p>
	Cachés	<p>La consigne est « ouverte » laissant une grande liberté d'expression à celui qui répond. C'est le cas du « test du bonhomme » où l'on demande à des enfants d'âges divers (et au même enfant à des âges différents) de dessiner un bonhomme « le plus complet possible ». On applique, ensuite, une « grille » permettant de situer le degré d'évolution dans la représentation mentale du schéma corporel (présence du tronc, du cou, de la paume, du talon, etc.).</p> <p>A part pour des expériences ponctuelles (comme ci-dessus) ou dans la procédure des SGI (voir ci-dessous), la dissimulation des critères d'évaluation est souvent peu propice à l'EF : soit l'étudiant ne peut se préparer correctement et a tendance à se révolter contre le système et à considérer ses notes injustifiées, ce qui ne l'incite guère à se former, soit il comprend l'intérêt a posteriori mais regrette alors de n'avoir pas été mis au courant en amont.</p> <p>L'EF implique pourtant parfois que certains critères soient cachés, ou, en tous cas, non rappelés. Le système des Solutions Générales Implicites (solutions « aucune », « manque », « toutes » et « absurdité », non rappelées à chaque QCM) est une procédure d'EF permanente... à condition bien sûr d'en analyser les résultats (cfr Edumétrie et Docimologie, de D. Leclercq).</p>
Proximité professionnelle	écologique	<p>Nous parlerons d'évaluation écologique lorsque cette évaluation est proche de la future situation professionnelle des participants. Cette proximité peut se cristalliser sur les contenus ou sur les conditions de passation. L'idéal est même de conjuguer ces deux critères, pour augmenter les chances de transfert, c'est-à-dire les chances qu'aura un étudiant d'être aussi (ou aussi peu) performant à cette évaluation qu'il le sera lorsqu'il rencontrera ce même défi dans sa vie professionnelle.</p> <p>Les ECOS, par exemple, Evaluations Cliniques Objectives et Structurées, pratiquées en pédiatrie (J-P Bourguignon) et en médecine générale (D. Giet), tentent de reproduire le plus fidèlement possible la situation d'un examen médical traditionnel, pour évaluer les étudiants médecins le plus objectivement possible, mais dans toute la complexité de leur environnement professionnel c'est-à-dire avec des patients simulés et des cas cliniques réels. De même, dans le cours de PAV, D. Leclercq et F. Georges</p>



	<p>proposent aux étudiants de répondre à des lettres fictives leur posant de réels problèmes, très proches de ceux qu'ils pourraient rencontrer dans leur profession de pédagogues.</p> <p>Dans ces deux exemples, il est fort probable que l'étudiant qui obtient un bon score au test reproduira ces mêmes comportements performants dans sa réalité professionnelle, ce que l'on ne peut pas déduire d'une évaluation non contextualisée.</p> <p>L'évaluation est aussi plus probablement écologique lorsqu'elle exploite des supports et méthodes divers, une grande variété de pratiques correspondant à la diversité des situations professionnelles rencontrées. C'est le cas du curriculum FORMASUP, qui propose et évalue à la fois des réflexions et analyses critiques, des participations à des événements professionnels réels et un projet complexe. Au sein du projet, l'enseignant tient son vrai rôle de professionnel, il est évalué dans son propre contexte de travail. Il s'agit ici du niveau le plus élevé d'une évaluation écologique, très proche de la consultance.</p>
Non (ou peu) contextualisée	<p>Nous parlerons d'évaluation non contextualisée lorsqu'elle met en œuvre des conditions de passation et des objectifs qu'il est difficile de rapprocher d'un contexte professionnel. La plupart des évaluations sont dans ce cas et les pédagogues préconisent une évolution des pratiques allant vers des évaluations comportant une plus grande « validité écologique ». Par exemple, varier les types d'évaluations au sein d'un même cours et proposer une combinaison de performances simples et complexes est l'un des gages d'écologie dans l'évaluation. Par contre, ne proposer que des QCM, même conçues de façon à évaluer des compétences d'analyse ou de vigilance cognitive des étudiants, est peu écologique car peu de situations professionnelles ne comportent aucun débat, aucune nécessité d'argumenter ou de produire une réflexion personnelle.</p>

Représentation synoptique des facettes et dimensions de l'évaluation formative

	Facettes	Dimensions
Pourquoi ? (fonctions)	Référence	Critérielle / Normative
	Résultat	Sommatif / Diagnostique
	Visée	Sanctionnante / Formative
Quoi ? (objets)	Centrage sur	Processus / Produits
	Dimension	Uni-dimensionnelle / Pluri-dimensionnelle
Qui ? (sujets)	Cible	Individuelle / De groupe
	Destinataires	Privée / Publique
	Opérateurs	Allo-opératrice par expert(s) / Auto-opératrice (auto-évaluation) / Allo-opératrice par pairs
Comment ? (méthodes, conditions)	Périodicité	Ponctuelle / Continue / Répétée
	Révisabilité des notes	Définitive / Améliorable
	Source	Objective / Subjective
	Procédure	Standardisée / Adaptative
	Implication	Interne (ou participante) / Externe
	Contrat	Imposé / Négocié
	Repères/ échelle	Mobiles / Fixes
	Visibilité des critères	Annoncés / Cachés
	Proximité professionnelle	Écologique / Non (ou peu) contextualisée

3. FAQ

La question X est-elle sommative ou formative ?

Une question, en soi, n'est pas indicative de son appartenance au groupe sommatif ou formatif. En effet, seul l'usage fait de la question nous permet de la classer dans l'une ou l'autre de ces catégories. Une même question pourra donc, dans deux dispositifs différents, servir une évaluation à visée sommative ou formative.

L'évaluation formative n'est-elle pas intrusive ?

On peut imaginer une gamme de procédures, de la moins intrusive à la plus intrusive.

A un extrême (le moins intrusif), on peut juste donner des feedbacks. C'est du **monitoring**, comme les appareils au pied du lit du patient qui renseignent en permanence sur la température, le rythme cardiaque, etc. Un superbe exemple est le **tableau de bord** proposé par François Georges. L'apprenant peut ou non le consulter, peut ou non s'en servir pour « rectifier » ses démarches.

A un autre extrême, le plus intrusif, l'étudiant DOIT corriger ses erreurs au fur et à mesure qu'elles apparaissent. C'est le cas du **PACELBRO** (Programme Auto-Correctif à Embranchements sous forme de Livre Brouillé à Réponses Ouvertes). Il branche (automatiquement) sur des remédiations à chaque erreur et tente donc de les extirper pour qu'elles ne se produisent plus dans la suite du test. Leclercq a montré (1970) que les statistiques d'erreurs diminuaient en effet au fur et à mesure que l'on avance dans le test (CICP, ch.7 p.17). C'est le cas aussi de beaucoup de didacticiels et jeux pour enfants (du petit poisson Marine Malice à Rayman en 3D, en passant par le pirate Max qui cherche son trésor), dans lesquels on ne peut avancer si on n'a pas résolu la question précédente.

Au-delà du monitoring : le « tutoring ». Nous n'avons pas classé l'exemple du **PACELBRO** dans le **TUTORING**, encore plus intrusif, parce que dans **PACELBRO**, l'étudiant est conscient de ce qui se passe et qu'il a la liberté de suivre ou non le chemin conseillé. Dans certains didacticiels intelligents (**Intelligemment Assistés par Ordinateur**), c'est un système enfoui (non visible par l'apprenant) qui prend les décisions, souvent à l'insu de l'apprenant. Ces systèmes sont appelés intelligents pour deux raisons. Tout d'abord parce qu'ils font appel à des langages d'intelligence artificielle (IA) comme LISP. Ensuite parce qu'il sont basés souvent sur 4 modèles : un modèle de l'apprenant (ses préacquis, son style d'apprentissage, etc.), un de l'expert (capables de résoudre les problèmes et donc d'y fournir les réponses correctes) , un du tuteur (ses stratégies si l'étudiant se trompe deux fois de la même façon, comment il récompense et distribue ses feedbacks, sa stratégie de progressivité, etc.) et un de l'interface (sera-t-elle essentiellement verbale : des mots ? Ou permettra-t-elle des actions ? de quelle sorte ? sur des images ? de quelles sorte ? et quand ? etc.). Le « **tutoring** », assisté par intelligence artificielle et prenant des décisions à l'insu de l'étudiant, est peut-être formatif mais n'informe pas l'étudiant sur sa façon de fonctionner et de gérer ses erreurs. Le principe est formatif quant à la matière, mais pas au niveau métacognitif. Il est peu utilisé dans les universités, en raison des coûts de développement très importants.

Formatif... micro ou macro ?

Une évaluation sommative des apprentissages sert souvent la régulation d'un dispositif plutôt que des apprentissages qui ont fait l'objet de l'évaluation. Le professeur qui a fait passer un examen va utiliser les résultats de sa cohorte d'étudiants pour ajuster son cours et l'examen qui en découle les

années ultérieures, mais l'examen n'aura en rien régulé l'apprentissage des étudiants testés cette première année. L'examen a été formatif pour le cours, qu'il a permis de réguler.

Autre exemple : lorsqu'un professeur attribue un 8/20 (ou 12/20) à un étudiant, il participe à un processus non pas d'apprentissage, mais bien de sélection par lequel le système se régule (les étudiants passent - ou non - de classe). Par contre, fournir aussi à l'étudiant une série d'indications sur ses points forts et ses points faibles est une contribution possible à son apprentissage. Encore faut-il que l'étudiant l'exploite, mais nous sommes ici (et d'ailleurs souvent - ex dans l'évaluation des mémoires) dans une situation alliant les aspects formatifs et sommatifs de même que les aspects micro et macro. Le test sert à l'évaluation formative de l'étudiant mais rend aussi possible la régulation du dispositif de formation et du système scolaire en général.

Portée au niveau macro, l'évaluation formative est donc nommée régulation.

La régulation est le propre du vivant, donc de tout projet. Parmi les étapes classiques d'un projet (BOMAER), il s'agit des étapes ER, étroitement liées aux étapes BOMA. L'évaluation formative est donc, dans le domaine pédagogique, l'application d'un principe beaucoup plus large, universel, qui consiste à prélever des données et à les réinjecter dans le processus pour augmenter les chances d'atteindre les objectifs assignés, voire de modifier les objectifs eux-mêmes ou d'autres composantes du processus. C'est ce qu'on appelle la cybernétique (de GOUBERNAI, en grec, qui signifie piloter, conceptualisé par Norbert Wiener en 1947). On trouve des processus régulés à tous les niveaux de la vie. Un être unicellulaire comme la paramécie a des capteurs (captant les propriétés chimiques de son environnement) qui déclenchent des réactions adaptatives, par exemple l'enkystement (elle se met en position de ne pas avoir besoin d'eau, comme en hibernation, en attendant que les conditions redeviennent meilleures, ce que font aussi beaucoup de plantes). A un niveau plus élevé, les organes (ex l'estomac), les individus, les groupes sont aussi des ensembles régulés. Plus de détails sont donnés sur ces exemples dans D. Leclercq, Edumétrie et Docimologie, introduction.

4. Bibliographie

- Allal, L. (1979), Stratégies d'évaluation formative : conceptions psychopédagogiques et modalités d'application, in L. Allal, J. Cardinet et P. Perrenoud (Eds), *L'évaluation formative dans un enseignement différencié*, Berne : Lang, pp. 130-156.
- Bandura, A. (1977), *A social learning theory*, Englewood Cliffs, New Jersey: Prentice Hall. Traduit en français par J.A. Rondal (1979), *L'apprentissage social*, Bruxelles : Mardaga.
- Bloom, B.S. et coll. (1969, original 1956). *Taxonomie des objectifs pédagogiques*, Tome I, Domaine cognitif, Montréal : Education nouvelle
- Bloom, B. (1979). *Caractéristiques individuelles et apprentissage scolaire*. Bruxelles : Labor.
- Bloom, B.S. et al. (1971), *Handbook on Formative and Summative Evaluation of Student Learning*, Mc Graw-Hill Book.
- Boxus, E. (1971). Une méthode de prédiction du rendement en lecture en première année. *Education - Tribune Libre*. 127, 63-78.
- Bruner, J.S. (1991). « ... car la culture donne forme à l'esprit », Paris : Eshel.
- Bruner, J.S. (1986). *Savoir-faire et savoir dire*, Paris : PUF.
- Burton, R., *Approche psychologique de la résolution des équations du premier degré à une inconnue et de leur rôle dans la résolution de problèmes en physique*, mémoire de licence en Sciences de l'Education, Université de Liège, 1999.
- Cardinet, J. (1992), *Evaluation scolaire et mesure*, 2^e Edition, Bruxelles : De Boeck-Wesmael.
- De Landsheere, G. (1979), Dictionnaire de l'évaluation et de la recherche en éducation. Paris : PUF.
- Demily, F. (1999), *Une tentative de réponse aux défis des grands groupes : Quiz Training, un outil au service de l'évaluation formative et de la recherche*, mémoire de licence en Sciences de l'Education, Université de Liège.
- Falchikov, N. (1995), Peer Feedback Marking : Developing Peer Assessment, *Innovations in Education and Training International (IETI)*, 32, 2, 175-187.
- Halleux-Hendrick, J. (1969), Construction des questions à choix multiple : une seule solution correcte ?, in *Revue Belge de Psychologie et de Pédagogie*, Tome XXXI, n° 127, 113-125.
- Halleux-Hendrick, J. (1969a), L'attitude de l'élève de l'enseignement secondaire face aux questions à choix multiple, in *Revue Belge de Psychologie et de Pédagogie*, Bruxelles, Tome XXXI, n° 127, 87 et ss.
- Jans, V. (2000). *Confrontations instrumentées et dialectiques des jugements auto et allo-évaluatifs. Contributions conceptuelle et méthodologique à l'étude de l'autoévaluation réflexive chez des étudiants universitaires*, thèse de doctorat, ULg, STE, décembre.
- Leclercq, D. & Van der Vleuten, C. (1998), PBL - Problem Based Learning ou APP - Apprentissage Par Problèmes, in D. Leclercq (Ed.), *Pour une pédagogie universitaire de qualité*, Sprimont : Mardaga, pp. 187-205.
- Leclercq, D. (Ed.) (1998), *Pour une pédagogie universitaire de qualité*. Liège : Mardaga.
- Leclercq, D. (1998b), *Edumétrie et Docimologie*, document interne, Service de Technologie de l'Education, Université de Liège.
- Leclercq, D. (1995), *Conception d'Interventions et Construction de Produits de Formation*, document interne, Service de Technologie de l'Education, Université de Liège.

- Leclercq, D. (2003), *Diagnostic cognitif et métacognitif au seuil de l'université. Le projet MOHICAN mené par les 9 universités de la CFWB*, Liège : Editions de l'Université de Liège.
- Nunziati, G. (1990), Pour construire un dispositif d'évaluation formatrice, *Cahiers pédagogiques*, 280, pp 47-64.
- Perrenoud, Ph. (1998), *L'évaluation des élèves. De la fabrication de l'excellence à la régulation des apprentissages. Entre deux logique*, Paris, Bruxelles : De Boeck et Larcier
- Piaget, J. (1974), *Réussir et comprendre*. Paris : PUF.
- Shannon, C.E. & Weaver, W. (1949), *The Mathematical Theory of Communication*, University of Illinois Press.
- Scallon, G. (1988), *L'évaluation formative des apprentissages : l'instrumentation*. Québec : Presses de l'Université Laval.
- Scriven, M. (1967), *The Methodology of Evaluation, Perspectives of Curriculum Evaluation*, Rand Mc Nally
- Stake, R. (1994), Case Studies, in N.K. Denzin & Y.S. Lincoln (Eds), *Handbook of Qualitative Research*, Thousand Oaks : SAGE Publications, pp. 236-247.
- Stufflebeam, D., (1980), *L'évaluation au service de la décision*. Paris : NMP.
- Tardif, j. (1998), *Intégrer les Nouvelles Technologies de l'Information. Quel cadre pédagogique ?*, Paris : ESF.
- Van der Vleuten, C. & Wynen, W. (1990), *Problem-based learning : perspective from the Maastricht experience*, Amsterdam : Thesis.
- Vygostky, L. (1931). Les bases épistémologiques de la psychologie, titre attribué par Bronckart et Schneuwly (1985,25-38) à l'article « Histoire du développement des fonctions psychiques supérieures », in J.P. Bronckart et B. Schneuwly (Eds), 1985, *Vygotsky aujourd'hui*, Neuchâtel : Delachaux et Niestlé.
- Wald, A. (1943). *Sequential Analysis of Statistical Data*. New York : Columbia University Press.
- Wiener, N. (1947), *Cybernetics or Control and Communication in the Animal and the Machine*, New York, Wiley.