

2007-2008

CHAPITRE 6

Dieudonné Leclercq & Marianne Poumay

Avec la collaboration de Séverine Delcomminette et Marilyn Quoilin

COMMENT SAVOIR CE QUE L'ON SAIT ?

LA METACOGNITION



Enjeu :

Au siècle de la « connaissance », l'idéal serait que toute personne donc chacun de nos étudiants soit un apprenant autonome, efficace et enthousiaste. Le rôle du professeur est de les y aider. Comment ? De quatre façons. (1) En ayant lui-même une position claire sur ce qu'est la connaissance. Ainsi, les doutes, les certitudes, les représentations erronées sont-elles de la connaissance ? (2) En aidant les étudiants à améliorer leur capacité de détecter QUAND ils ont besoin d'apprendre QUOI et JUSQU'OU. (3) En créant des situations où les étudiants pratiquent ces capacités : juger sa performance, analyser les causes qui en expliquent la qualité, modifier éventuellement sa stratégie en conséquence. (4) En utilisant un système de notation qui reflète la définition de la connaissance qui a été adoptée (voir point 1).



A. Comment opérationnaliser le concept de métacognition ? _____	205
B. Pourquoi la métacognition est-elle si peu connue des enseignants ? _____	208
C. Comment distinguer les situations métacognitives entre elles ? _____	211
D. Quelles sont les modalités possibles de collecte de jugements de vos étudiants ? _____	216
E. Comment recueillir, mesurer et expliquer la première opération ? _____	223
F. L'étudiant peut-il expliquer ses jugements sur sa connaissance ? _____	230
G. La métacognition sert-elle l'action ? _____	236
H. Y a-t-il des styles cognitifs et métacognitifs ? _____	243
I. Utiliser les Degrés de Certitude dans l'évaluation des performances des étudiants : pourquoi et comment ? _____	245
J. Quelles qualités peut et devrait avoir une approche métacognitive ? _____	255
K. Conclusions et perspectives _____	259
Bibliographie _____	260

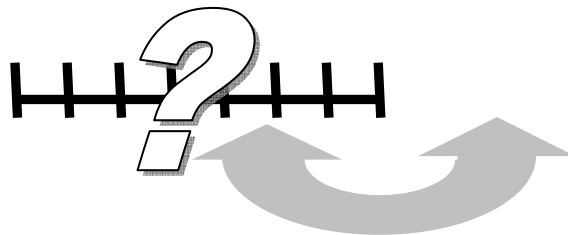
Le présent document peut prendre place dans plusieurs schémas théoriques.

Il peut être pris en considération dans le I de ASCID. Rappelons que l'acronyme ASCID (déployé dans les versions successives de l'ouvrage « Psychologie éducationnelle » de D. Leclercq) signifie : Affectivité (Affectivity), Savoir-faire (Skills), Cognition (Cognition), Image de soi (I ou Self-cognition), et Décision (Decision).

Disons de suite que la Métacognition n'est qu'une partie de l'image de soi, la partie relative au domaine cognitif, la seule partie traitée dans le présent document. Nous en ferons cependant un (gros) chapitre en soi, étant donné, à nos yeux, l'importance de la métacognition dans l'apprentissage humain, qui contraste avec le peu d'utilisation que l'on en fait actuellement en éducation.

Ce même document peut être utilisé dans un autre schéma théorique, des (8) Événements d'Apprentissage en ce sens que la Métacognition est une des façons d'apprendre, probablement la moins connue et la moins entraînée de huit. Ces Événements d'Apprentissage / Enseignement ont été décrits par Leclercq et Poumay (2005) dans leur « Learning Events Model » et qui a fait l'objet de plusieurs versions de l'ouvrage de D. Leclercq « Méthodes de Formation et Théories de l'Apprentissage » (le dernier en date étant de 2007).

L'acronyme ASCID a essentiellement une fonction mnémotechnique : aider les praticiens à penser aux grands domaines de l'activité mentale humaine. Ceux-ci ont d'ailleurs fait l'objet de taxonomies, ou classifications, elles-mêmes check-lists, permettant de détailler chacun de ces domaines. Les 5 lettres ASCID ont la même signification en anglais et en français, bien qu'elles ne soient pas, pour la lettre « I », la première lettre du même mot. En anglais, c'est le « I », bref le « je ». En français, cela peut tout aussi bien évoquer « ipsatif » (tourné vers soi), « identité » ou encore « image de soi ». Dans le présent chapitre, nous n'aborderons que le concept de métacognition qui ne se superpose pas totalement avec les notions évoquées ci-avant.



C'est l'occasion d'expliquer maintenant les trois symboles graphiques représentant les trois composantes de ce logo.

La métacognition n'est pas un des facteurs ASCID mais plutôt une combinaison de plusieurs de ces facteurs.

Dans sa composante « jugement », représentée par l'échelle d'évaluation, elle relève de l'Image de soi et de l'Affectivité sur un terrain particulier : la cognition (ce que je sais est-il satisfaisant ? idem sur mes méthodes d'apprentissage ou d'étude).

Dans sa composante « analyse », représentée par le point d'interrogation, (pourquoi je suis sûr de ce que j'avance, comment j'apprends), elle relève de la Cognition.

Dans sa composante « régulation », représentée par une flèche, (que dois-je changer ? quelles dispositions dois-je prendre ?) elle relève de la Décision.

A. Comment opérationnaliser le concept de métacognition ?

Notre définition de la métacognition

De nombreux chercheurs (Brown, 1994, Noël, 1991 ; Noël, Romainville et Wolfs, 1995) reconnaissent que la plupart des définitions ne sont pas opérationnelles (elles ne peuvent servir l'action). Nous essayons ici de les pallier par une définition plus pratique et plus précise, plus susceptible de guider l'action.

Nous (Leclercq et Poumay, 2007) définissons la **métacognition** comme : « un ensemble

- d'**opérations** (Jugement – Analyse – Régulation)
- sur des **objets** (ses processus ou ses productions)
- à certains **moments** (PRE – PER – POST performance)
- dans certaines **situations** (d'Apprentissage ou d'Evaluation)
- et observables via des **performances** (comportements et/ou conduites) ».

D. Leclercq



M. Poumay



Les exemples qui suivent sont loin d'être exhaustifs. Ils portent sur les différentes composantes (cellules du tableau de Leclercq et Poumay) de la métacognition. Certains de ces exemples illustrent certaines difficultés de classement, ce qui n'enlève rien à la fécondité des catégories.

La question essentielle que nous invitons le lecteur à se poser est

« Sur quoi pouvons-nous agir en tant que formateur ? ».

PRE	Exemple 1 (plutôt centré sur le <u>processus</u>) : L'exposé en anglais qui aura lieu demain
Jugement	Je pense que je serai incapable de prendre de prendre des notes demain
Analyse	Parce que souvent je ne comprends une phrase qu'à la fin... Au moment où l'autre commence et j'aurai besoin de toute ma concentration.
Régulation	J'emporterai un enregistreur sonore à la conférence et je la réécouterai au calme chez moi.

PRE	Exemple 2 (plutôt centré sur le <u>produit</u> ou résultat) : Réussirai-je mon année ?
Jugement	Etudiant de 1° bac, aujourd'hui en novembre, je m'attribue une chance sur deux de réussir en première session (juin) et 3 chances sur 4 de réussir soit en première soit en deuxième session (août).
Analyse	Parce que j'ai toujours bien réussi dans le secondaire mais je ne me rends pas bien compte de la difficulté de compréhension et de mémorisation de cours universitaires.
Régulation	Je grappille le plus possible de points aux examens partiels (produit). Je ne réserve pas de longues vacances d'été à l'étranger (processus).

PER	Exemple 3 (plutôt centré sur le <u>processus</u>) : Pendant l'examen par QCM
Jugement	La moitié du temps est passée et je n'ai pas encore lu la moitié des questions.
Analyse	Parce que j'ai passé trop de temps sur UNE question, au détriment des autres.
Régulation	Je vais lire toutes les questions qui restent, répondre à celles pour lesquelles je suis assez sûr et je reviendrai sur les autres plus tard.

PER	Exemple 4 (plutôt centré sur le <u>produit</u> ou résultat à un niveau très <u>micro</u>) : L'examen par réponses rédigées
Jugement	Ma certitude quant à « Pas une seule faute d'orthographe » est de 50%.
Analyse	Parce que j'hésite sur le terme «un problème <i>parallèle</i> » ou « un problème <i>parallèle</i> ».
Régulation	J'écris « un problème <i>voisin</i> », avec certitude 80% sur « pas une seule faute d'orthographe » (une erreur, peu probable, pouvant venir d'un autre endroit de mon texte.

PER	Exemple 5 (plutôt centré sur le <u>produit</u> ou résultat à un niveau très <u>micro</u>) : Apprentissage durant la période de préparation des examens
Jugement	Je ne parviens pas à assimiler la matière efficacement.
Analyse	Parce que je manque de vue d'ensemble, ce qui me contraint à mémoriser des contenus épars, sans liens entre eux.
Régulation	Je vais donner une structure personnelle à cette matière.

NB : La catégorisation entre processus et produit n'est pas aussi claire que la distinction entre PRE, PER et POST. Par ailleurs, une analyse de produit(s) peut déboucher sur une analyse, remise en cause et régulation des processus. Ce que montre l'exemple 5.

POST	Exemple 6 (plutôt centré sur les <u>processus</u>) Le pré-mémoire ou Rapport préparatoire au Travail de fin d'études (TFE) : réflexions APRES critiques faites par le superviseur
Jugement	Ma méthode de traitement des données (corrélations entre variables en colonnes de mon tableau de données) était inadéquate, comme mon superviseur m'en a convaincu(e).
Analyse	Parce que ce ne sont pas les différences inter-individus qu'il fallait étudier, mais les différences intra-individuelles.
Régulation	Je vais calculer les corrélations entre les lignes du tableau.

POST	Exemple 7 (plutôt centré sur les <u>produits</u>) Le pré-mémoire (suite de l'exemple 5) réflexions APRES critiques faites par le superviseur
Jugement	Mes graphiques sont sans intérêt (étant donné la méthode de calcul signalée ci-dessus).
Analyse	Parce que traduisant des données sans intérêt (cf ci-dessus).
Régulation	Je n'utiliserai pas ces graphiques dans mon mémoire.

Les deux exemples (6 et 7) illustrent la relation étroite entre processus et produit. Parfois la distinction est difficile à établir. Ainsi, une démonstration mathématique telle que la développe un étudiant au tableau est un produit révélateur de son processus de raisonnement.

POST	<p>Exemple 8 (centré sur le <u>produit</u> au départ mais débouchant sur un <u>processus</u>) Lors de la communication (et discussion) collective des réponses correctes et du débat à leur propos juste APRES (dans les minutes qui suivent) un examen par QCM.</p> <p>Dans ses cours à des grands groupes à l'université de Liège, D. Leclercq organise une communication des réponses correctes immédiatement discutées une à une par les participants à l'examen qui restent parfois une heure supplémentaire pour cet « exercice » (de la démocratie). Il arrive fréquemment que le professeur accepte séance tenante de considérer comme valable des réponses qu'il n'avait pas anticipées.</p>
Jugement	Je crois que ma réponse « Absurdité » à la question 17 est « défendable » alors que le professeur juge que la réponse correcte est « solution 4 ».
Analyse	Parce que j'avais une autre interprétation (que le professeur) de l'énoncé de cette question, une interprétation de romaniste (que je suis), ignorée par le professeur (qui n'est pas romaniste).
Régulation	Lors de la communication orale des réponses correctes immédiatement après l'examen, je suis intervenu(e) pour exposer mon point de vue, mais le professeur semble ne pas l'avoir compris, contrairement à d'autres objections (qu'il a admises) venant d'autres personnes sur d'autres questions. Je vais donc lui envoyer un courriel reformulant mes explications alternatives ¹ .

L'exemple ci-dessus illustre qu'il arrive souvent que l'analyse précède le jugement. Ainsi, la conviction de certitude de cet étudiant vient de l'analyse qu'il fait de ses propres processus mentaux.

POST	<p>Exemple 9 (plutôt centré sur le <u>produit</u>) APRES connaissance des jugements du correcteur sur divers aspects d'une copie d'examen</p> <p>Dans les réponses construites (d'une longueur d'une page ou d'une demi-page par exemple), D. Leclercq et al. (2006) demandent aux étudiants d'accompagner leur réponse de degrés de certitude sur divers critères. Par exemple :</p> <ul style="list-style-type: none"> -pertinence et structure de la réponse -caractère suffisant de détail pour permettre à un non expert en la matière de comprendre -utilisation (et explicitation correcte) des termes techniques -pas une seule utilisation incorrecte d'un terme (technique ou général de la langue française) -pas une seule faute de syntaxe -pas une seule faute d'orthographe grammaticale -pas une seule faute d'orthographe d'usage <p>Bien que l'étudiant soit invité à donner un degré de certitude (de 0% à 100%) sur la qualité de sa réponse en regard de chacun de ces critères, le poids des différents critères n'est pas le même : le premier est plus important que le dernier, par exemple. Le professeur ne révèle pas forcément ce poids à l'avance, pour éviter les comportements « stratégiques » (négliger les critères à faible poids).</p>
Jugement	J'aurais dû accompagner certaines de mes réponses d'un degré de certitude moins élevé.
Analyse	Parce que mon indice de Confiance (certitude moyenne accompagnant mes réponses correctes) est de 70%, mais mon indice d'Imprudence (accompagnant mes réponses incorrectes) est de 65% ; je fais ainsi trop peu la différence entre ce que je sais correctement et ce qui est incorrect dans mes réponses. A l'analyse plus fine, il apparaît que c'est dans les critères portant sur les aspects syntaxiques, sémantique (usage pertinent de mots de vocabulaire), d'orthographe d'usage et d'orthographe grammatical que je me surestime.
Régulation	Je fournirai des degrés de certitude plus faibles sur ces critères et, en conséquence, les vérifierai plus longuement avant de remettre ma copie.

¹ A l'université de Maastricht, lors des Tests de Progression (Progress test) en faculté de médecine, les étudiants ont 5 jours ouvrables pour contester (par écrit) les réponses présentées comme correctes lors de leur affichage aux valves.

B. Pourquoi la métacognition est-elle si peu connue des enseignants ?

Une étonnante absence

Alors qu'au terme d'une méta-analyse² sur les facteurs affectant l'apprentissage Wang, Haertel et Walberg (1990) concluent « Parmi les variables cognitives et affectives, c'est la métacognition qui influence le plus l'apprentissage » (cité par Tardif (1992)), il faut bien constater, avec Bereiter & Scardamalia (1989), que la formation scolaire actuelle « supporte peu les apprenants dans le développement de connaissance sur la connaissance » (p. 380).

Nous pensons que cette situation est due à trois causes fondamentales.

La première est que la pratique généralisée de la métacognition dans l'apprentissage est un phénomène très récent dans l'histoire de l'humanité.

La deuxième raison est le manque de modèles porteurs de clarifications théoriques suffisantes relatives à l'apprentissage, à la connaissance et à la métacognition.

La troisième raison est le manque de méthodes, d'outils et d'indices qui soient dotés à la fois de validités théorique et conséquentielle et de qualités de simplicité dans l'utilisation quotidienne. Nous reprenons séparément ces trois causes.

B1. Cause 1 : Un événement d'apprentissage non généralisé

Contrairement à d'autres « Événements d'Apprentissage »³ tels que la « modélisation / observation » ou la « transmission / réception », etc., la métacognition est encore loin d'être un mécanisme généralisé (dans toutes les classes, pour tous les élèves, à tout moment) d'éducation. Et pourtant, bien des penseurs et bien des philosophes, anciens et modernes, non seulement ont eu recours à la métacognition, mais en ont proposé des analyses et des théories. Par exemple, Socrate (« Connais-toi toi-même » : γνῶθι σεαυτόν), Bernouilli qui a introduit la notion d'utilité (subjective) dans la Théorie des décisions, De Finetti qui a montré que seules les probabilités subjectives pouvaient donner un sens objectif à toute réponse et à toute méthode de scoring.

Néanmoins, l'usage scolaire qui en a été fait jusqu'ici n'a concerné qu'une proportion infime d'êtres humains en regard de tous les pratiquants potentiels. Pour la très grande majorité des hommes, leurs représentations personnelles relatives aux contenus aussi bien qu'aux mécanismes d'apprentissage restent largement **inconscients** et ne font l'objet ni de jugements ni d'analyses systématiques fondés sur des démarches ou des références stables. Nous disons « systématiques » parce qu'il arrive, au hasard des circonstances, que de tels processus interviennent, plus chez certaines personnes que chez d'autres.

C'est un défi pour l'éducation contemporaine de provoquer l'irruption massive de la métacognition parmi les Événements d'Apprentissage, comme elle a fait irruption, sous le terme de « réflexivité » chez les enseignants sous l'impulsion des travaux de Schön (*Le praticien réflexif*, 1983). Ce dernier distingue la capacité de réfléchir à la fois « dans (le feu de) l'action » (en situation d'urgence) et « sur l'action » (après coup).

² Par « méta-analyse, on entend le rassemblement, à l'aide d'une métrique commune (le calcul d'indices d'Ampleur de l'Effet) de résultats issus d'expériences ayant la même méthodologie expérimentale (comparaison de groupes expérimentaux et de groupes contrôle), mais pas les mêmes instruments, donc pas la même métrique.

Nous aurons l'occasion, plus en avant dans cet article de discuter de l'utilisation du préfixe « méta », issu du grec, dans « méta-analyse » et dans « méta-cognition ».

³ Voir Leclercq et Poumay (2005)

B2. Cause 2 : Un flou conceptuel sur la définition

A la lecture de la définition initiale de Flavell et de celle proposée par Gombert quinze ans plus tard, on est frappé par la filiation, voire la quasi équivalence, si ce n'est le renforcement par des termes des notions de conscience et d'intentionnalité⁴, déjà présentes mais moins explicitement chez Flavell :



Flavell (1976, p.232) : « *La métacognition fait référence à la connaissance qu'on a de ses propres processus cognitifs et de leurs produits ou de ce qui leur est relié, par exemple, les propriétés différentes des informations ou des données pertinentes pour leur apprentissage. La métacognition se rapporte, entre autres choses, au contrôle actif, à la régulation et à l'orchestration de ces processus en fonction des objets cognitifs et des données sur lesquels ils portent, habituellement pour servir un objectif ou un but concret.* »

Gombert (1990, p.27) : « *Métacognition : Domaine qui regroupe (1). les connaissances introspectives conscientes qu'un individu particulier a de ses propres états et processus cognitifs ;(2). les capacités que cet individu a de **délibérément** contrôler et planifier ses propres processus cognitifs en vue de la réalisation d'un but ou d'un objectif déterminé* ».

Nous avons déjà signalé que de nombreux chercheurs (Brown, 1994, Noël, 1991 ; Noël, Romainville et Wolfs, 1995) reconnaissent que ces définitions ne sont pas opérationnelles (elles ne peuvent servir l'action).

Notre définition (cf. section A), elle, détermine un tableau à 9 cellules contenant chacune deux aspects, donc 18 possibilités de classement.

<i>Opérations</i>	PRE (avant)	PER (pendant)	POST (après)
Jugement	Sur le processus	Sur le processus	Sur le processus
	Sur le produit	Sur le produit	Sur le produit
Analyse	Sur le processus	Sur le processus	Sur le processus
	Sur le produit	Sur le produit	Sur le produit
Régulation	Sur le processus	Sur le processus	Sur le processus
	Sur le produit	Sur le produit	Sur le produit

Nous explicitons chacun de ces concepts à la section B.

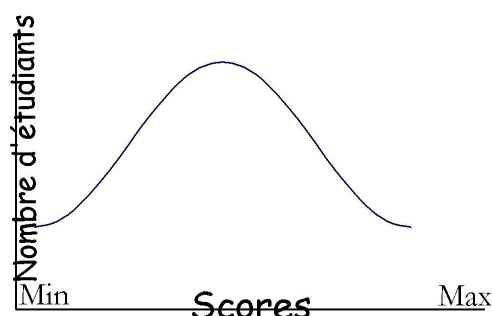
⁴ Nous avons mis en gras ces deux mots dans la définition de Gombert.

B3. Cause 3 : Manque d'instrumentation

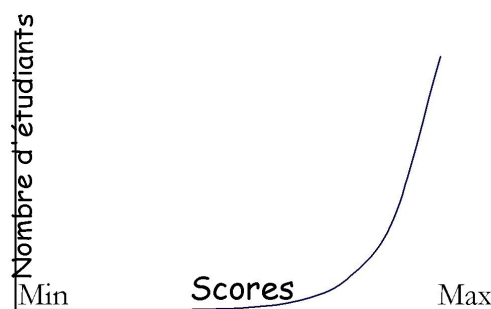
La psychologie scientifique, via les approches réflexologiques⁵, associationnistes⁶, constructivistes⁷, cognitivistes⁸, instrumentalistes⁹, a, depuis un siècle, permis une meilleure compréhension des mécanismes mentaux. Leur évaluation a, elle aussi, fait l'objet de travaux, en psychologie comme en pédagogie. Cependant, depuis la création en 1974 par Carver du terme « Edumétrie » (mesure en éducation) dans son article « *Two dimensions of tests : Psychometric and edumetric* », on ne peut que constater que c'est la psychométrie (mesure en psychologie) qui a le plus (encore) développé les concepts et instruments typiques de sa métrique.

Le présent chapitre renforcera et développera des pistes typiquement édumétriques fondées sur

- la connaissance partielle et son auto-évaluation via des degrés de certitude,
- le calcul de scores individuels et diagnostiques de réalisme (ou conformité avec la réalité), de résolution (distinction entre réponses correctes et réponses incorrectes), de Confiance, de Prudence (voir ci-après, section H6).
- le souci de la validité conséquentielle des mesures (leur impact sur les comportements ultérieurs des enseignants et des apprenants),
- la représentation des résultats à des épreuves au moyen d'une distribution spectrale (voir ci-après, section H4) tendant le plus possible vers la « courbe en J » et non la « courbe de Gauss », tant pour les réponses incorrectes que pour les réponses correctes. Dans les graphiques ci-dessous, en ordonnée figurent les nombres de scores observés (les scores eux-mêmes étant en abscisse)



Courbe de Gauss



Courbe en J

Nous pensons qu'il est utile de faire connaître de nouvelles formes de représentation des résultats de mesure de la connaissance partielle. Ce qui montrera entre autre qu'il est possible de mesurer et d'entraîner ces composantes de la métacognition.

⁵ Celles qui sont inspirées des travaux de Pavlov sur le conditionnement répondant.

⁶ Celles qui sont inspirées des travaux de Skinner sur le conditionnement opérant.

⁷ Celles qui sont inspirées des travaux de Piaget sur l'auto-structuration des connaissances et intériorisation de l'action.

⁸ Celles qui sont basées sur les mécanismes mentaux débouchant sur la conceptualisation, les réseaux conceptuels, etc.

⁹ Celles qui sont basées sur les travaux de Vygotsky et l'appropriation individuelle des acquis collectifs.

C. Comment distinguer les situations métacognitives entre elles ?

Les caractéristiques de la métacognition

Revenons graduellement sur le tableau déterminé par la définition de Leclercq et Poumay.

C1. Trois types d'opérations portant sur les performances

Opérations	PRE (avant)	PER (pendant)	POST (après)
Jugement	Sur le processus	Sur le processus	Sur le processus
	Sur le produit	Sur le produit	Sur le produit
Analyse	Sur le processus	Sur le processus	Sur le processus
	Sur le produit	Sur le produit	Sur le produit
Régulation	Sur le processus	Sur le processus	Sur le processus
	Sur le produit	Sur le produit	Sur le produit

a) Les **jugements** sont des évaluations par l'apprenant lui-même du degré de qualité (ou du degré de certitude dans l'exactitude) d'une réponse ou du degré de satisfaction qu'il attribue à sa performance, ou à sa compétence (par exemple par anticipation de la performance) si la performance n'a pas encore eu lieu. Ex : « *J'estime mon score le plus probable au test de demain à 60% de réussite* ».

b) Les **analyses** sont les justifications et explications (les pourquoi), c.-à-d. les élaborations, les explications que l'apprenant donne à la qualité de sa performance, ainsi que les « attributions causales » de ce qui lui arrive (voir plus loin). Ex : « *J'ai tendance à lire trop superficiellement la question et à tenter de trouver la réponse (dans les solutions proposées) avant de l'avoir comprise réellement.* »

c) Les **régulations** sont les reconductions ou changements dans les méthodes de travail, dans les habitudes, dans l'environnement, les décisions, etc.). Elles correspondent à ce que Gombert appelle « *les capacités qu'un individu a de contrôler et planifier ses propres processus cognitifs en vue de la réalisation d'un but ou d'un objectif déterminé* ». Ex : « *Dans un examen, désormais, je commence par me demander si l'énoncé ne contient pas une absurdité, puis j'entoure au crayon les mots-clés d'un texte et je les relie via les mots qui indiquent des relations.* »

Les trois opérations seront analysées en profondeur respectivement dans les sections C, D et E.

C2. Trois moments

Opérations	PRE (avant)	PER (pendant)	POST (après)
Jugement	Sur le processus	Sur le processus	Sur le processus
	Sur le produit	Sur le produit	Sur le produit
Analyse	Sur le processus	Sur le processus	Sur le processus
	Sur le produit	Sur le produit	Sur le produit
Régulation	Sur le processus	Sur le processus	Sur le processus
	Sur le produit	Sur le produit	Sur le produit

Nous distinguons trois situations d'activité métacognitive : **avant (PRE)**, **pendant (PER)** et **après (POST)** une performance, une activité d'évaluation ou d'apprentissage repérable dans le temps.

La situation PRE est omniprésente, c'est la situation « de base » : nous sommes dans un certain « état de connaissance » ou plutôt de « représentation personnelle » du monde, de nous-mêmes, de l'apprentissage, de l'évaluation, etc. Sur cette base, nous prenons des décisions, consciemment ou inconsciemment sur des sujets AVANT d'en avoir l'expérience profonde, précise et détaillée, bref avant leur performance sur le sujet : le choix d'études, le choix de profession, le choix de tâches, le choix de collaborateurs, le choix de loisirs, etc.

Dans l'émission TV « Questions pour un champion », le candidat doit choisir entre quatre thèmes celui sur lequel il souhaite être interrogé. Le candidat estime sa « force » (à répondre aux questions d'un test) dans chacun des quatre thèmes, mais AVANT de connaître les questions.

Nous sommes rarement « testés » (épreuve standardisée, instrumentée) en situation PRE car dès qu'il y a testing, la ou les méthodes d'observation interfèrent avec les données (nous ne nous souvenons de certaines choses que si on les sollicite explicitement, nous ne fournissons certains détails que si on nous les demande), et l'observé à son tour « réagit » aux méthodes (ne pas tout dire, prétendre connaître plus (ou moins) que ce qu'il pense, etc.).

Choisir de répondre dans le domaine où l'on se sent plus performant : une manière de se mettre en confiance

Namur Pierre (2006), AESS Sciences économiques.

Lors d'une évaluation orale, j'ai proposé aux élèves de choisir une de leur question dans un des trois tas prévus. Je leur indiquais le thème général qui se dégageait de chacun des trois tas (économie de marché, droit, économie financière). Ainsi, chacun pouvait choisir d'avoir une question dans un domaine où il jugeait être plus performant (une manière de les mettre en confiance).



La situation PER. Cette situation peut durer des mois (la période de formation) ou quelques minutes : le temps d'un testing, d'une évaluation, d'une épreuve, de la remise de ses résultats ou d'un apprentissage intentionnel. C'est le plus souvent à ces occasions que sont recueillies de façon systématique et approfondie (à un degré de subtilité proche du maximum possible) des données cognitives et métacognitives. Quand les réponses correctes ne donnent pas lieu à discussion, à contestation, il est possible, après la communication des réponses correctes, de calculer des indices métacognitifs comme Confiance et Prudence (voir section H6) permettant un feedback facilitant l'évaluation des PER-dictions.

Il s'agit ci de situations où l'apprenant est confronté à des tâches précises (des **performances**) à exécuter, soit de sa propre initiative, soit sur sollicitation d'un observateur (un professeur, un employeur, etc.), soit par hasard (ranimer une personne qui s'effondre en rue), bref sous forme d'épreuve ou d'examen ou de test¹⁰ ou de résolution de problème ou encore d'apprentissage¹¹.

La différence entre les situations PRE-Test et les situations PER est que dans ces dernières, l'apprenant connaît maintenant les circonstances et/ou questions ou tâches précises, les réponses qu'il a essayé d'y apporter, les difficultés qu'il a éprouvées et les efforts qu'il a déployés pour y arriver. A la différence de la situation POST, dans la situation PER, l'apprenant ne connaît pas encore les « réponses modèles » (ou correctes) attendues par l'observateur (le testeur), ou par la situation-problème (la personne reviendra-t-elle à la vie ? Gardera-t-elle des séquelles ?).

La régulation, c'est aussi ne rien changer...

Etudiant(e) (2006) AESS chimie.

Jugement : L'élève estime qu'il réalise parfaitement la manipulation et qu'il obtiendra de bons résultats.

Analyse : L'élève constate que les résultats qu'il obtient sont logiques par rapport à ce qu'il pourrait attendre.

Régulation : L'élève continue à appliquer les mêmes principes dans les exercices suivants.



La situation POST. Dans cette situation, le sujet a pu fournir des réponses à des problèmes, puis disposer des réponses d'autrui (expert, pairs) et les comparer aux siennes, mais il y a matière à discussion. Même quand l'expert est incontestable, le jugement de « qui a raison » fait problème ; a fortiori quand les réponses d'autrui sont celles de pairs. S'engage alors un processus de **jugement comparatif** requérant d'une part de se décentrer (prendre en compte du point de vue d'autrui) et d'autre part d'être exigeant vis-à-vis de soi comme vis-à-vis d'autrui.

Les notions d'Auto-complaisance, d'auto-laxisme, d'auto-indulgence ou au contraire d'auto-sévérité sont ici centrales. C'est un défi pédagogique de développer des méthodes où l'évaluateur observe précisément ces variables-là pour en donner à l'apprenant une image la plus objective possible, même

¹⁰ Parmi les épreuves, les examens comportent la conséquence « réussite – échec » et les tests ont la particularité d'une application standardisée (les mêmes conditions pour tous), ce que ne sont pas des examens oraux par exemple.

¹¹ Pour Bereiter et Scardamalia, tout apprentissage est une résolution de problème.

si elle reste parcellaire et ponctuelle. Lire un texte écrit par un expert sur un contenu dont nous avons une connaissance partielle est une situation de type POST : nous avons des pré-représentations, des pré-notions. Elles sont confrontées à d'autres, celles de l'expert. Que croire ? Que revoir ?

La métacognition post-performance vise à permettre à l'apprenant d'adapter ses stratégies ultérieures, ses réactions futures, ses décisions d'actions, son rapport à la matière, son analyse des problèmes. Cette situation de post jugement, post analyse et post régulation est très fréquente dans la vie adulte, et doit être entraînée pour être pertinente et efficace.

La situation POST est quasi présente en permanence dans l'Événement d'Apprentissage que nous avons appelé « **Débat** »¹², ou « Confrontation de points de vues ». En effet, pour que le débat aie une « valeur ajoutée mathétique » (contribuant à l'apprentissage), il ne suffit pas que les apprenants travaillent en groupe, ni que leurs idées s'opposent, se complémentent, s'entrechoquent, encore faut-il que tout cela produise des effets dans l'esprit des apprenants. Des débats peuvent être stériles si chacun reste figé sur ses positions.

Savoir aide à changer de cap

Etudiant(e) (2006), AECS communication.

Jugement : Une jeune fille de rhétorique réalise un travail de fin d'études sur les journalistes reporters car elle souhaite exercer cette profession plus tard.

Analyse : Après avoir réalisé son travail, elle se rend compte qu'elle idéalisait cette profession et que la vision idyllique qu'elle s'en faisait est éloignée de la réalité.

Régulation : La jeune fille, ayant acquis un surcroît de connaissances sur la réalité de cette profession décide de réorienter son choix professionnel.



C3. Deux objets

Opérations	PRE (avant)	PER (pendant)	POST (après)
Jugement	Sur le processus	Sur le processus	Sur le processus
Analyse	Sur le produit	Sur le produit	Sur le produit
	Sur le processus	Sur le processus	Sur le processus
Régulation	Sur le produit	Sur le produit	Sur le produit
	Sur le processus	Sur le processus	Sur le processus
	Sur le produit	Sur le produit	Sur le produit

a) Les produits

La métacognition peut porter sur les **produits** (les résultats) :

- Ce que je suis capable de me rappeler et ce que j'ai oublié.
- Les qualités de ce que j'ai produit : ma performance est-elle correcte ? complète ? rapide ? autonome ? éthiquement défendable ? économe (peu coûteuse) ? fluide ? flexible ? originale ? détaillée ou sophistiquée ? (on reconnaît, dans les quatre derniers critères les idées de Torrance)
- Les conséquences de mes actions : quel effet (sur les lecteurs de mon texte par exemple) aura mon travail et pourquoi.

b) Les processus

La métacognition peut porter sur les **processus** (les démarches) d'apprentissage :

- La nécessité ressentie d'apprendre : pour répondre à quel besoin, à quel manque, à quel désir ?
- La sélection des contenus à apprendre et des ressources (dans quels livres ? avec quels sites web ?)
- L'autofixation de la difficulté des tâches et leur répartition dans le temps
- Les moments et leur pénibilité ou leur attrait ressentis : le démarrage, la récolte de données, la synthèse, la mise au net, etc.
- Les critères d'arrêt (quand c'est suffisamment bon pour être présenté à autrui, quand j'ai assez étudié pour m'arrêter, mes seuils de satisfaction dans diverses matières)
- L'évolution, au cours du temps, de la satisfaction ressentie de la tâche accomplie, du niveau atteint, des progrès faits.

¹² Leclercq et Poumay, 2003b

Afin de compléter la définition, voici quelques autres considérations (sections B4 à B9) sur le modèle.

Opérations	PRE (avant)	PER (pendant)	POST (après)
Jugement	Sur le processus	Sur le processus	Sur le processus
Analyse	Sur le produit	Sur le produit	Sur le produit
	Sur le processus	Sur le processus	Sur le processus
Régulation	Sur le produit	Sur le produit	Sur le produit
	Sur le processus	Sur le processus	Sur le processus

C4. Des données conscientes, affectives, et volontairement subjectives

Lors d'une activité, la métacognition peut¹³ constituer une irruption massive de la **conscience** dans des processus mentaux qui sont continus, car même la nuit, quand l'apprenant dort, son cerveau « se » gère et « organise » ses connaissances, mais de façon **inconsciente**.

Dans la métacognition, l'**affectif** est intimement mêlé au cognitif, tout spécialement dans l'opération métacognitive de « jugement » (de sa performance), car il est difficile de distinguer le jugement sur la performance du jugement sur la personne. On voit l'irruption de mécanismes de défense, de déni, de surestimation, sous-estimation, etc.

Nous nous baserons sur l'**introspection** métacognitive et plus précisément encore sur ce que nous pouvons en connaître via **son expression** par l'apprenant. L'aspect **subjectif** des données récoltées est précisément ce qui fait la richesse de la métacognition : le sujet est acteur de son questionnement.

C5. Malgré la subjectivité, des exigences de qualités métriques

La métacognition fait abondamment appel à l'auto-évaluation (subjective) et à l'auto-observation, sans exclure les allo-évaluations (subjectives elles aussi) et les mesures objectives, tout spécialement à des fins de comparaison, de confrontation. Nous suggérerons des **méthodes et techniques** visant à rendre les pratiques métacognitives entre autres plus valides (théoriquement fondées), plus fidèles (permettant une étude objective de la subjectivité), plus diagnostiques et suivies d'effet (offrant une validité conséquentielle), plus praticables, etc. Ces qualités seront développées plus loin.

C6. De multiples facettes à l'évaluation

L'importante composante « auto-évaluation » de la métacognition peut être caractérisée, comme les autres évaluations, par de multiples facettes ou dimensions telles que « critérielle » ou « normative » (se situer par rapport aux autres), « sanctionnante » ou « formative », etc. (voir D. Leclercq, 1999, chap. 1 : « La Rose des vents de l'Evaluation »).

C7. Une nécessaire contextualisation de la situation métacognitive

L'objet sur lequel porte la métacognition en influence la nature et la difficulté. Par exemple, pour toute performance cognitive considérée, on tentera de préciser tant les processus mentaux considérés que les produits. Par exemple, on recourra à la taxonomie de Bloom, pour spécifier si la métacognition porte sur la reconnaissance (ou le rappel) de connaissances, la compréhension, l'application, l'analyse, la synthèse ou encore l'évaluation. Les jugements sur la tâche et sur la consigne peuvent influencer les jugements sur sa propre performance.

Par exemple, Noël (1991, p. 12) cite l'évolution avec l'âge de la prise de conscience du caractère incomplet (Markman, 1977) et de l'incohérence (Markman, 1978) d'un message, chez des enfants. Considérant que cette capacité doit être entraînée systématiquement, même dans l'enseignement supérieur, Leclercq (1986, pp. 127-144) introduit dans les QCM des « Solutions générales Implicites » telles que « Aucune, Toutes, Manque de données pour répondre, Absurdité dans l'énoncé ». Bien que cela relève de la compréhension et non de la méta-compréhension, la nature de la question peut interférer de façon observable tant objectivement (les taux de réussite diffèrent selon les types de questions) tantôt subjectivement (la difficulté ressentie varie selon les questions)¹⁴.

¹³ Nous disons « peut » car la métacognition consciente et systématique est trop peu développée.

¹⁴ Ainsi, dans des épreuves sur la Psychologie de l'apprentissage et la Technologie de l'éducation passées par des étudiants universitaires en 2^o candidature en psychologie, Leclercq (1993, 222) constate que, par rapport à une QCM dont la Réponse Correcte (RC) est explicite (QCM-RCExpl), les QCM-RC « Aucune » nécessitent un délai de réponse plus grand et les

Sur la nécessité d'entraîner les élèves à se rendre compte qu'une question demande des précisions, citons Cohen (1963, p. 38) : « Dans nombre de nos écoles, on semble enseigner aux enfants qu'il n'existe que des questions à l'emporte pièce, auxquelles doivent être données des réponses également à l'emporte pièce. [...] Face à une question qui ne comporte pas une réponse valable unique, ils sont déconcertés, effarés, et se raccrochent à n'importe quel indice, si mal choisi soit-il [...] Il est du devoir du maître de surveiller de près l'élève qui préfère une donnée claire quoique incorrecte, à [la reconnaissance de] l'ambiguïté. »

C8. Le niveau d'implication et d'autonomie de l'étudiant dans la métacognition

On distinguera par exemple la métacognition **spontanée** de la métacognition **solicitée**. Ainsi, certains étudiants tiennent régulièrement un « carnet de bord », alors que d'autres ne pratiquent la métacognition que lorsque le formateur le propose (vous pouvez fournir des degrés de certitude) ou l'impose (vous devez justifier votre certitude).

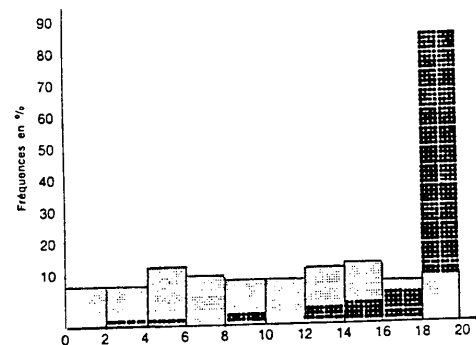
On distinguera aussi la métacognition **autonome** de la métacognition **assistée**. A côté des données subjectives fournies par l'étudiant, il est possible de recueillir des données objectives sur la performance de l'apprenant, par observation directe ou, par ordinateur, via des traces (Georges, 2003). Par exemple, dans un test, l'ordre de réponse aux questions, le nombre de modifications de réponses, les demandes d'aide, leur degré de profondeur, la nature et la durée de consultation des aides, la vitesse de réponse, les commentaires ajoutés au texte de base (points d'interrogation sur certains mots, « OK » sur d'autres, etc.).

C9. Une nécessaire réserve

La performance métacognitive peut varier selon les types de performances et de contenus. Il ne faudrait pas tirer des conclusions sur les capacités métacognitives d'un étudiant ou des étudiants au-delà des données observées en un lieu, en un temps et dans des circonstances précises. Jans (1994) a montré que des élèves de cinquième année primaire étaient capables de créer des problèmes (de géométrie) comportant des « manques » et des « absurdités » et d'y répondre (à celles fabriquées par d'autres) avec des taux de réussite élevés, alors que les Questions à Réponse Correcte « Manque » sont les moins bien réussies chez des étudiants universitaire dans un autre domaine (la psychologie). Ces différences de qualité de performances peuvent être imputées à diverses causes.

L'une d'entre elles est le **contenu** (il est facile de se rendre compte qu'il manque la hauteur pour pouvoir calculer l'aire d'un triangle). Un autre est le type **d'entraînement** dont ont bénéficié les apprenants, comme le montre l'exemple ci-après.

D'Hainaut et Michez (1979) ont utilisé une méthodologie, recourant à la création avec des élèves de 6^o primaire. Ils leur ont demandé, une fois par jour pendant 20 jours, d'inventer des problèmes (mesure de l'aire du carré et du rectangle, pourcentages et intérêts). Ils ont montré qu'ils étaient capables de résoudre beaucoup mieux les problèmes (distribution noire) que leurs condisciples formés plus traditionnellement (distribution claire des notes).



QCM-RC « Absurdité » un délai plus court. Les QCM-SGI ont été moins bien réussies que les QCM-RCEExpl (66% en moyenne), tout spécialement les QCM-RC « manque » (28% en moyenne), alors que leur indice de discrimination moyen (rpbis) a été le même (0,29). Les QCM-RC « Absurdité », aussi bien réussies (66%) que les QCM-RCEExpl, ont eu un indice de discrimination plus élevé (0,42).

D. Quelles sont les modalités possibles de collecte de jugements de vos étudiants ?

Les méthodes et consignes

D1. Les commentaires verbaux libre (auto ou allo) ou catégories nominales¹⁵

On peut demander au sujet de s'exprimer de la façon qui lui convient. Par exemple, « Que pensez-vous de votre maîtrise de la matière ? ». On obtient alors des réponses variées telles que « Je n'ai jamais été très à l'aise avec ce thème » ou « Cela m'étonnerait que je réussisse beaucoup de questions sur le sujet » ou encore « je n'en ai pas retenu grand-chose ». On voit en quoi l'analyse a posteriori par l'observateur risque d'être complexe et peu féconde tant est imprécise la consigne, donc la diversité des réponses possibles, dans leur fond et dans leur forme.

D2. Les échelles d'évaluation verbales (auto ou allo) ou ordinales¹⁶

Le sujet est invité à se positionner sur un des échelons d'une échelle qui lui est proposée (ex : « Si l'on me posait des questions sur le sujet, j'en réussirais... »), chaque échelon étant décrit avec des mots, comme « très peu voire aucune », « peu », « à peu près la moitié », « beaucoup », « toutes ». On voit en quoi cette façon de poser la question est une demi-mesure, puisqu'il existe (voir ci-après) une façon plus précise et tout aussi peu coûteuse de poser la question. On est dans un cas semblable à celui des enquêtes sur la consommation de cigarettes où il est tellement simple de demander « combien de cigarettes fumez-vous par jour ? » plutôt que de s'en tenir à des expressions vagues du genre « peu », « moyennement », « beaucoup », « excessivement ».

D3. Les échelles normatives (auto ou allo) métriques¹⁷

Le sujet se positionne par rapport aux performances d'un groupe de référence (dont les performances sont ou ne sont pas (encore) connues. Par exemple, l'étudiant positionnera sa performance parmi celles de ses condisciples, le patient parmi celles de son groupe de formation. La connaissance du groupe par le sujet est évidemment cruciale dans ce type de mesure. Les modalités de réponses peuvent être « Le meilleur », « Parmi les 20% meilleurs », « parmi les 40% meilleurs », etc. jusque « Le plus faible ». La largeur des catégories a du sens dans la mesure où les questions ou tâches très précises ne sont pas encore connues, ce qui rend difficile pour un sujet de se positionner.

Le feedback sur ces PRE-dictions peut se faire sous forme de nombres (« Vous êtes en réalité le 24^{ème} sur 100 »), de graphiques (situant la personne dans le continuum) ou les deux.

D4. Les échelles critérielles (auto ou allo) métriques pour performances isolables

Le sujet exprime son degré de maîtrise anticipé ou sa compétence, l'échelle la plus simple étant la dichotomie. Par exemple un questionnaire énumérant des capacités relatives à l'informatique (sauver un fichier dans Office, chercher un mot sur le web, etc..) où l'apprenant répond par OUI (je maîtrise) ou NON (je ne maîtrise pas).

Une forme plus nuancée consiste à demander à l'apprenant d'annoncer son pourcentage de réussite le plus probable (ex : 60%) à des épreuves sur le sujet. On trouvera dans les sections C6 et H « la connaissance partielle... » des détails sur la façon de poser cette question métacognitive et d'en exploiter les réponses.

¹⁵ Les catégories nominales ne peuvent faire l'objet de sériation : il n'y a pas d'ordre entre elles.

¹⁶ On ne peut appliquer les opérateurs arithmétiques sur les catégories et échelles nominales, tout au plus les opérateurs > et <.

¹⁷ On peut appliquer les opérateurs + et - sur ces échelles et, quand elles ont un 0 absolu (pas de valeurs négatives) les opérateurs * et /.

D5. Les échelles critérielles pour performances complexes

Dans un cours de 2^o année de candidature en Psychologie et Sciences de l'Education, nous avons proposé aux étudiants (de 19 ans) de pratiquer la méthode des PARMs (Leclercq et al., 2003, AIPU) où, grosso modo, ce sont les étudiants qui donnent le cours et non plus le professeur. Les étudiants (volontaires) étaient invités à constituer des groupes de 5 dont un ayant accès à internet depuis son domicile. C'est cet étudiant qui, via le mail, est en contact avec le professeur, notamment pour soumettre des idées, en discuter, etc., dans la phase de préparation de l'animation. Le professeur avait annoncé que la performance (animer ses collègues de façon multimédias pendant 15 minutes sur un point particulier de la matière, puis répondre aux questions de la salle) serait notée selon les 10 critères suivants.

	sur	justifications
Défi	10	
Exposé Fond	10	
Exposé Forme orale	10	
Exposé Forme Média	10	
Critique	10	
Lien	10	
Activité-fond	10	
Activité forme	10	
Réponse aux Questions fond	10	
Réponse aux Questions forme	10	
Somme	100	

On aurait pu inviter les étudiants à auto-noter leur performance avant que le professeur remette SES notes, sur chacun des 10 critères, et de justifier leur note dans la colonne de droite ; ce qui n'a pas été fait. Après la prestation, le professeur a utilisé la colonne de droite pour y mettre SES justifications. C'est sur cette base que les groupes qui le voulaient pouvaient réagir, toujours par mail, ce que certains groupes ont fait.

On ne parlera d'échelle d'évaluation descriptive pour un critère donné que si chacun des échelons des degrés d'excellence est décrit. Voici une telle description pour le critère « lien » de la grille ci-dessus :

- Le lien est pertinent 1. oui 2. très
- Le lien est exploité par l'étudiant 1. oui 2. assez bien 3. de façon très poussée
- Le lien est original 1. oui 2. très
- Le lien est bien présenté 1. oui 2. très bien 3. Exceptionnellement bien

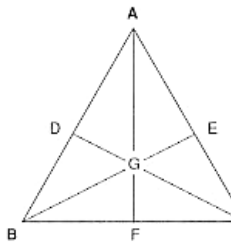
On constatera que ces quatre échelles (pertinence, exploitation, originalité et présentation) permettent d'obtenir au total 10 points pour le critère « lien » et que ces critères, tout « absolus » qu'ils soient recèlent des composantes normatives (l'originalité et le caractère exceptionnel).

D6. Les degrés de certitude quant à l'exactitude d'une réponse donnée

a) La certitude binaire. L'étudiant indique s'il est ou non sûr de sa réponse. En Hollande, en l'accompagnant ou non d'un Z (Zekerheid).

b) Les certitudes ordinales. L'étudiant choisit parmi les degrés d'une échelle verbale (Pas sûr du tout, Peu sûr, Moyennement, Sûr, Très sûr). Nous verrons que cette consigne est, comme la précédente, largement insatisfaisante car un même niveau (ex : très sûr) prend des significations différentes pour une même personne dans un même test. Alors d'une personne à l'autre ou d'un test à l'autre, plus rien n'est comparable.

c) Les degrés-zones de probabilités

<p>L'apprenant doit se positionner :</p> <ul style="list-style-type: none">• soit sur un des sommets (A, B ou C) d'un triangle équilatéral ;• soit exactement entre A et B (réponse D) ;• soit exactement entre A et C (réponse E) ;• soit exactement entre B et C (réponse F) ;• soit au centre du triangle (réponse G).	
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------

De 1970 à 1999, nous avons utilisé des zones portant des noms de code convenus. Par ex :

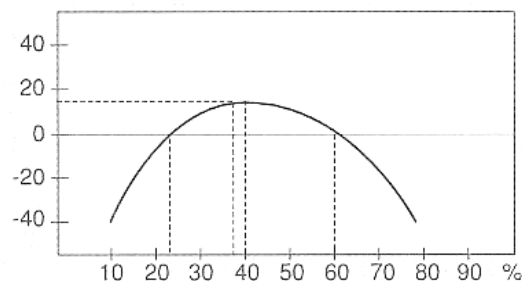
- certitude 0 = de 0% à 25%
- certitude 1 = de 25% à 50%
- certitude 2 = de 50% à 75%
- certitude 3 = de 75% à 100%

Des « coupures de l'axe » plus sophistiquées ont aussi été utilisées pour tenir compte du fait que la sensibilité humaine est plus grande aux extrêmes qu'au milieu du continuum des probabilités. Voici les « bornes » de tels intervalles : 0%, 2%, 10%, 25%, 50%, 75%, 90%, 98%, 100%.

d) Le « confidence contour ». Cette technique, due à James Bruno (UCLA) ne fonctionne que pour les QCM à 3 solutions proposées dont l'une des 3 est correcte. L'étudiant reçoit (cf. ci-dessous) un triangle où les 3 sommets sont les trois solutions proposées (A, B ou C). Choisir les positions intermédiaires revient à exprimer quelle(s) solution(s) on exclut et entre lesquelles on hésite. Ce système manque de souplesse (inapplicable avec plus de 3 solutions) et de subtilité (il n'y a que 3 valeurs de certitude : 0%, 50%, 100%)

e) La distribution de probabilités. Ce système, le plus sophistiqué, a été mis au point par Van Lenthé (1993).

Lower bound	=	20%
Best guess	=	40%
Upper bound	=	59%
True value	=	38%
Score	=	15
Total score	=	12



Il est limité à des réponses portant sur des pourcentages. Par exemple, à la question : « Quel est le % de ménages belges qui possédaient deux postes de TV au 1/1/2002 ? ». Il consiste à déplacer latéralement (en abscisse) une courbe gaussienne, de telle sorte que la moyenne corresponde à votre estimation la plus probable (dans l'exemple ci-dessous : .40). La courbe peut être aplatie (moins de risques d'erreur mais moins de points en cas de réussite) ou rendu plus « pointue » à volonté par l'étudiant, ce qui rend compte du degré de confiance dans la réponse. Les points sont visualisés en ordonnée.

f) Certaines méthodes sont limitées à des réponses sur un continuum (numérique) :

C'est le cas de la méthode des fractiles de Pitz (1974). Ainsi sa consigne par tertiles est : "Donnez deux valeurs numériques, x et y de manière que la réponse correcte ait 33% de chances d'être en-dessous de cet intervalle, 33% de chance d'être contenue dans l'intervalle, 33% d'être supérieure à l'intervalle". Quand la valeur correcte n'est pas dans l'intervalle, Pitz parle de "surprise". Des variantes beaucoup plus subtiles que ces "tertiles" existent : quintiles, heptiles, etc. Leur extrême est la méthode de Van Lenthé (1993) de positionnement sur l'axe 0-100% d'une distribution de probabilités sous la forme

d'une courbe de Gauss dont le sujet peut régler le positionnement de la moyenne (par les flèches gauche-droite du clavier) et la kurtosis (par les flèche haut-bas du clavier).

g) Certaines méthodes (consignes) exigent le **recours à un ordinateur** pour exprimer la certitude. C'est le cas de méthodes proposées par Baker, J.D. (1965). *The Uncertain Student and the Understanding Computer*, La recherche en enseignement programmé. *Tendances actuelles*, Paris : Dunod, pp. 303-319.

(1975), Dirkzwager (1993), Shuford (1993) où l'ordinateur recalcule dynamiquement les probabilités attribuées à toutes les k solutions d'une QCM dès que le sujet en modifie une (toutes étant fixées, par défaut, à 1/k au départ), de manière à ce que leur somme soit toujours 1. Ces méthodes exigent que le sujet attribue une probabilité à chacune des solutions d'une QCM et que les solutions s'excluent l'une l'autre.

D7. Pour le recueil de la certitude, plusieurs règles doivent être respectées

Non seulement de Finetti (1965), mais aussi Van Naerssen (1965) et surtout Emir **Shuford** et al. (1966) ont décrit ce que sont des « *Admissible Probability Measurement Procedures* ».



Ils démontrent qu'une consigne verbale telle que « *Etes-vous peu sûr, moyennement sûr, très sûr, extrêmement sûr ?* » est insuffisamment précise. Il faut absolument recourir à l'expression de probabilités (ou pourcentages de chances) d'avoir fourni la réponse correcte.

De Finetti recommande le « *five stars systems* » qui consiste à accompagner sa réponse d'autant d'étoiles (minimum 0, maximum 5). On peut aussi donner 3 étoiles à une solution, et une étoile à deux autres solutions, ou deux étoiles à une solution.

On voit que ce système revient à distribuer des « blocs » de probabilité de 0,2 ou des blocs de pourcentage de chance de 20%.

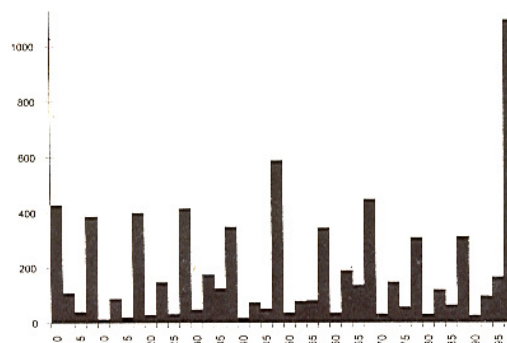
C'est fondamentalement cette formule que nous avons retenue : permettre de choisir entre l'une des 6 possibilités suivantes :

0 % 20 % 40 % 60 % 80 % 100 %

Les raisons de nous être limités à cette consigne sont exposées ci-après.

D8. La subtilité humaine a ses limites

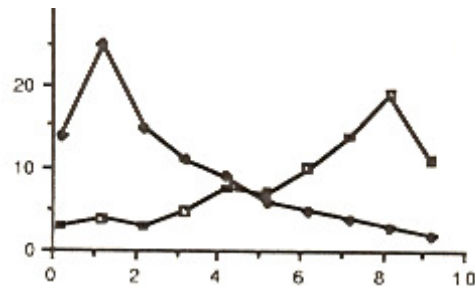
Les sujets adultes n'utilisent qu'une dizaine de degrés de certitude (allant de 0 à 100%) alors qu'ils pourraient en utiliser 40 :



Cependant, la « replication » des certitudes n'est pas garantie. Les courbes (ci-après) de replication sont obtenues en représentant à des sujets les questions et les réponses qu'ils avaient fournies deux mois plus tôt à une centaine de questions sur une matière où il n'est possible ni d'oublier ni d'apprendre en deux mois. Les certitudes, elles, ne sont pas rendues et les sujets doivent les fournir à

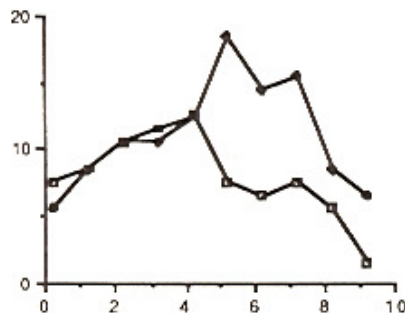
nouveau. Quelle certitude est fournie la seconde fois quand la première fois elle était 1 (entre 10% et 19%) ou 8 (entre 80% et 89%) ?

C'est ce que montre le graphique ci-dessous :



Même si la replication n'est pas parfaite, on constate que le « pic » (de replication) de 1 se produit pour 1 et le pic de 8 pour 8.

Par contre, ce n'est plus le cas au centre de l'échelle des probabilités (pour 4, 5 et 6) :



Notre étude (Leclercq, 1983) de la capacité humaine à discriminer de façon fiable des degrés de certitude entre 0 et 100 % a nous a amené à constater que le nombre de zones que l'on est capable de discriminer fidèlement ne dépasse pas 7.



Nous avons retrouvé ainsi dans ce domaine ce que **Miller** (1956) avait découvert dans le domaine de l'attention et de la perception : notre limitation à 7 éléments environ, ce que traduit le titre de son célèbre article :

« *The magical number seven, plus or minus two* ».

Notre consigne, appelée « 6 multiples de 20 % » tient compte de cette limitation.

D9. Les limitations des humains dans le traitement des probabilités.

Damasio (1994, p. 222) avance une raison d'inapplicabilité de la théorie des décisions pour expliquer les choix humains :

« [même dans ce cas...] il se trouve que les stratégies employées pour le raisonnement peuvent elles-mêmes présenter certaines failles, comme Amos Tversky et Daniel Kahneman (1973) l'ont démontré. L'une de ces dernières, parmi les plus importantes, semble bien être que les gens font généralement preuve d'une ignorance profonde en matière de statistique et de théorie des probabilités, et ne les appliquent que de façon défectueuse, comme Stuart Sutherland (1992) l'a suggéré¹⁸. »

Parmi ces « applications défectueuses, nous en avons relevé quelques unes par la pratique :

¹⁸ Le titre du livre de Sutherland, « L'ennemi de l'intérieur », est pour lui synonyme de l'irrationalité.

1. Devant une situation où plusieurs solutions sont proposées, mais s'excluent réciproquement et où on sait qu'une seule est correcte, certaines personnes distribuent leurs probabilités sur ces options, avec une somme qui dépasse 1 (ou 100%) ou qui est inférieure à 1.
2. Dans un Vrai-Faux, certaines personnes fournissent leur réponse avec une probabilité inférieure à 0,5 (par exemple VRAI à 0,2, alors qu'ils auraient dû répondre FAUX à 0,8).
3. Kahneman et Tversky ont montré que bien des personnes confondent les probabilités JOINTES (Quelle est la probabilité que A et B se produisent en même temps ?) et les probabilités CONDITIONNELLES (quelle est la probabilité que B se produise étant donné que A s'est déjà produit ? ou si A est vrai). Cette confusion explique que nous soyons sous-bayésiens dans la révision de nos probabilités. Cela signifie que nous tirons d'un événement, d'un message, des conclusions (des informations) inférieures à ce que nous autoriserait le théorème de Bayes sur la révision de probabilités.

D10. Et pourtant, seule la connaissance partielle donne sens à la mesure des connaissances



"La connaissance partielle existe. La mesurer est utile et possible. Seuls le recueil des probabilités subjectives et les scoring appropriés qui en tiennent compte peuvent donner du sens à l'évaluation des connaissances."
C'est ce qu'a démontré, dans son article de 1965 "*Methods of determining levels of partial knowledge*", le mathématicien **Bruno De Finetti** (Université de Rome).

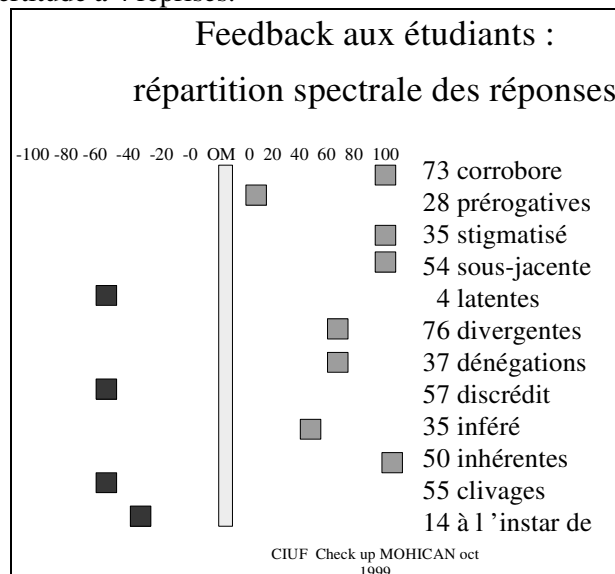
- 1.a. Dans certaines situations, notre **connaissance** est **parfaite**. Par exemple, nous connaissons sans hésitations ni doute certains mots, dates, valeurs, lieux, etc.
- 1.b. Dans d'autres cas, notre **connaissance** est **nulle** (ou notre ignorance est totale). Par exemple, à la question "Quel est le jour de la semaine où Einstein est né ?", répondre "Un mardi, avec 1 chance sur 7 d'avoir raison" peut être considéré comme l'ignorance totale sur le jour de la naissance d'Einstein, mais comme une bonne connaissance des jours de la semaine de notre calendrier.
- 1.c. Notre **ignorance** est gigantesque et en pleine expansion (à la mesure de l'explosion des connaissances). Heureusement, comme le disait Mark Twain, "Ce n'est pas notre ignorance qui nous cause des problèmes, c'est ce que nous savons, mais qui est faux".
- 1.d. Il y a pire que l'ignorance : la connaissance **erronée** qui, si l'ignorance totale est positionnée à 0 sur **l'échelle de la compétence**, devrait être alors positionnée **en négatif**.
- 1.e. Entre les situations de connaissance nulle et celles de connaissance parfaite existent bon nombre de situations, notamment celles où notre apprentissage est "en cours", où notre **connaissance** est **partielle**, ni nulle, ni totale. Par exemple, sans connaître la réponse précise, nous pouvons approximer l'heure dans la journée, l'âge d'une personne, une surface, une distance parcourue, les seules solutions possibles parmi lesquelles pourraient figurer la réponse correcte à une question.

D11. Des feedbacks diagnostiques, aux étudiants, devraient faciliter leurs décisions d'apprentissage

Voici un exemple de feedback spécifiant la connaissance partielle

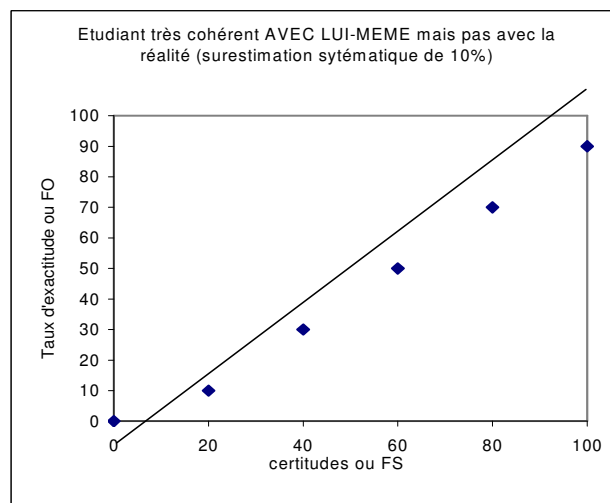
L'image ci-dessous est extraite du graphique qui explique à un étudiant ses performances au check-up MOHICAN d'octobre 1999 en vocabulaire (45 questions). L'image ne reprend que les 12 premières questions (de la question sur le mot "corrobore" à la question sur le mot "à l'instar de"). Le nombre qui précède le mot testé est le Pourcentage d'Exactitude (ou Facilité Objective ou FO) pour la population totale, par exemple 73% pour le mot "corrobore". La position du petit carré indique la position spectrale (sur le **spectre de qualité de la performance**) de SA réponse à ce mot.

On constatera que, sur ces 12 questions, s'il s'est trompé à 4 reprises, cet étudiant a eu le "bon goût" (en réalité le réalisme) de ne jamais se tromper avec 100% ni même 80% de certitude. Or il s'est exprimé avec 100% de certitude à 4 reprises.



D12. Des feedbacks globaux à l'étudiant sur son réalisme à l'épreuve permettent d'améliorer sa métacognition

Diverses mesures globales (pour toute une épreuve) du **réalisme** d'une personne existent :



- **l'Erreur Moyenne de Centration** ($EMC = FS - FO$ où $FS =$ Facilité Subjective ou Certitude Moyenne et $FO =$ Facilité Objective ou Taux d'Exactitude au Test). Quand EMC est positif, le sujet s'est surestimé et négatif, le sujet s'est sous-estimé. Cet indice est trompeur car une correspondance parfaite ($FS = FO$) pourrait résulter de compensation des surestimations par des sous-estimations.
- la **Cohérence personnelle** se calcule par la corrélation entre les divers DC utilisés (0, 20, 40, 60, 80, 100) et les taux d'exactitude observés pour les réponses fournies avec chacun de ces taux. Cet indice ne correspond pas au réalisme, car on peut être très cohérent vis-à-vis de soi-même, mais très éloigné de la réalité.

E. Comment recueillir, mesurer et expliquer la première opération ?

Les jugements métacognitifs

E1. Les jugements PRE d'auto-performances (futures)



Bernadette NOEL
FUCAM – Belgique



Ann BROWN
Univ. de Californie à Berkeley

« Brown (1978) constate que, pour les jeunes enfants, il s'avère plus difficile d'estimer leur réussite à une tâche avant qu'on la leur propose, plutôt qu'après qu'ils aient tenté d'y apporter une réponse. Brown et Lawton attribuent cette difficulté ...au niveau d'abstraction qu'une telle prédiction nécessite. Prédire avant de répondre repose sur la capacité d'imaginer un produit non encore réalisé...Quand on demande à des enfants d'âge préscolaire, de 2^o et de 4^o années, d'estimer l'étendue de leur mémorisation, on observe chez les plus jeunes une surestimation du nombre d'items dont ils pourront se rappeler...Les résultats [de Brown] indiquent que les enfants deviennent plus conscients de la difficulté d'une tâche au fur et à mesure qu'ils progressent en maturité...[...] Les jeunes enfants éprouvent beaucoup de difficultés à apprécier s'ils connaissent assez le matériel pour le restituer intégralement si on le leur demandait.» (Noël, 1991, p. 25).

Cette dernière capacité (estimer quand on est prêt à affronter une évaluation avec succès) est au cœur de l'*Individually Prescribed Instruction* (IPI) de Pittsburgh (Leclercq, 2003a, p. 11). L'IPI pratiquait le principe des « *self reliant students* » où les élèves gagnent et perdent le « privilège de l'autoévaluation » (où ils décident du moment où ils se testent sur ordinateur) selon qu'ils atteignent ou non le score de 85% de réussite.

E2. Les jugements PER de ses performances actuelles

Nous parcourons ci-après quelques approches de jugement PER et leurs méthodes de recueil.

a) Interview (écrit ou oral)

Le plus souvent, c'est l'apprenant lui-même qui se questionne, qui se « souffle » les questions stratégiques à se poser, par exemple lors de la passation d'un examen écrit. Dans l'exemple ci-dessous, nous soulignons l'aspect subjectif (et autoréflexif) et nous indiquons entre parenthèses les actions de régulation qui suivent naturellement les analyses (en italique) et les jugements (en gars) en situation PER.

Ex. lors d'un examen écrit par QCM :

« Maintenant que je vois les questions, quelles sont celles qui auxquelles je pense pouvoir fournir *rapidement* la réponse correcte *avec une certitude élevée* (pour commencer par celles-là) ? **Oui/Non pour chacune.**

Arriverai-je à faire *l'essentiel et suffisant* dans le temps disponible (pour décider quand je reviendrai sur les questions « sautées ») ? **Oui/non.**

Y a-t-il plusieurs questions pour lesquelles je soupçonne un *manque de données* (pour leur attribuer une certitude approprié) ? Lesquelles ? c.-à-d. **Oui/Non pour chacune.**

Pour quelles questions ai-je le sentiment qu'il est *utile* d'expliquer mon point de vue à l'enseignant ? (pour écrire une « justification » de ma réponse comme on me le permet) ? **c.-à-d. Oui/Non pour chacune, etc.. ».**

Un entraînement à l'analyse métacognitive se fait aussi par ce mode de questionnement « assisté » par l'enseignant qui, assis aux côtés de l'apprenant, réfléchit tout haut avec lui.

b) Le sentiment de connaître

Le « Feeling of Knowing » (FOK) consiste à « *demander au sujet un jugement après son échec en rappel sur la probabilité de reconnaître la réponse correcte* » (Nhoyvanisvong & Reder, 1998). La consigne est du type : « Vous êtes dans l'impossibilité de vous rappeler du nom de...(de vous souvenir des traits du visage de..) Quelle est votre probabilité de le(s) reconnaître parmi une liste de noms (de photos) présenté(e)s ? ». Cette consigne, on le voit, fait suite à une erreur par « omission ». elle pourrait aussi être appliquée, mutatis mutandis, après une erreur « par commission » (fournir une réponse incorrecte).

« *Le FOK est basé sur la quantité totale d'information partielle accessible à propos du contenu-cible et sur la facilité avec laquelle elle vient à l'esprit* » (Koriat, 1998, p.16).

Cette situation est à rapprocher très fort du phénomène du « Tip Of the Tongue ou TOT » (« je l'ai sur le bout de la langue »).

Cette consigne FOK ne devrait être utilisée systématiquement que dans les cas où seule la reconnaissance a du sens (de détection par exemple). Comme Nhoyvanisvong & Reder (1998, p.35), nous pensons que le FOK est un processus qui prend place dans les premières secondes de réponse à une question de mémoire qui amène à décider si on recherche (dans sa mémoire) ou si on répond « je ne sais pas ». Reder (1987) a observé que le délai pour émettre des jugements FOK sur la « rapellabilité » d'une question est inférieur à celui nécessaire pour retrouver la réponse elle-même. (cité par Koriat, 1998, p.19). A ce titre, il est un mécanisme de sélection de sa stratégie cognitive.

c) Les Degrés de Certitude pour les performances isolables

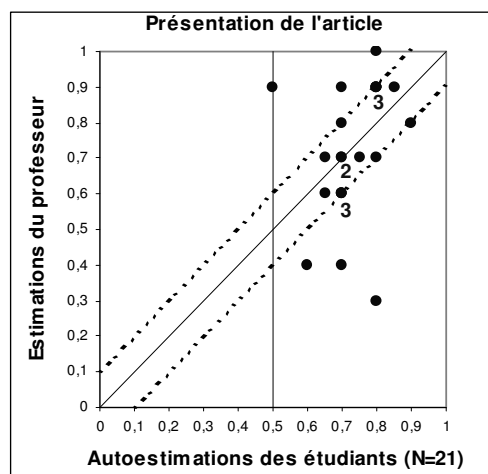
Selon Ebel (1965), « *C'est un mode de réponse spécial aux questions d'un test objectif et un mode spécial de cotation de ces réponses. En bref, le sujet doit indiquer non seulement ce qu'il croit être la réponse correcte à la question, mais aussi quelle est sa certitude dans l'exactitude de sa réponse. Au moment de la cotation, le sujet reçoit plus de points pour une réponse correcte avec certitude que pour une réponse accompagnée d'un doute. Mais la pénalisation d'une réponse incorrecte avec certitude est suffisamment lourde pour décourager les déclarations de confiance non fondées.* » (cité par Leclercq, 1983, p.163).

d) Les grilles d'évaluation de performances complexes

Voici une grille qui a été utilisée (Jans 2000, p.215) en situation PER avec les 28 étudiants d'un cours de D. Leclercq (Techno de l'Education), mené par la méthode des PARMs - Projets d'Animation Réciproques Multimédia - (Jans et al., 1998, pp. 207-241). Chacun des étudiants était invité à fournir, en plus de son travail écrit, une autoévaluation de sa performance sur chacun des critères suivants :

Cours L400 : Technologie de l'Éducation – Professeur D. Leclercq			
AUTOEVALUATION de :			
CRITERES		Cote	Sur
Contenus de la séquence			
1	<i>Pertinence :</i> Bonne adéquation avec la matière du syllabus (respect du contenu, pas de confusion avec d'autres concepts) ?		10
2	<i>Créativité :</i> • Apport de nouvelles données ? • Richesse des illustrations ?		10
Organisation de la séquence			
3	• Objectif(s) de la séquence apparu(s) clairement ? • Objectif(s) atteint(s) à la fin de la séquence ?		5
4	• Choix des activités (exposé, questions-réponses, travaux de groupe, travaux individuels, participation physique) adapté aux objectifs et aux étudiants ? • Bonne alternance des activités ?		5
5	• Rythme continu des activités ? • Gestion du temps adéquate ?		5
6	• Consignes de participation claires ? • Manière originale de faire participer les étudiants ?		5
Utilisation de Powerpoint et/ou d'autres outils (transparents, boîtiers,...)			

De la confrontation des deux parties (l'enseignant et l'étudiant), on peut représenter la relation entre leurs deux jugements. Voici, pour **un** critère (la présentation d'un article de revue lu) du genre de ceux évoqués ci-dessus, et pour 21 étudiants, en abscisse les % de qualité de leur travail **jugés** par les étudiants et, en ordonnée, les % de qualité **jugés** par l'enseignant. Certains points représentent 2 ou 3 observations (superposées sur le graphique).



De la comparaison de ces deux subjectivités, il apparaît que la corrélation est faible (0,25) mais positive.

Pratiquement, TOUS les étudiants se sont jugés « satisfaisant » (égal ou au-delà de 0,5), alors que l'enseignant donne une note inférieure à 0,5 pour 3 d'entre eux.

En outre, UN étudiant (le plus à gauche sur le graphique), s'est octroyé un jugement beaucoup plus sévère (0,5, le plus bas de toutes les auto-jugements) que le professeur (qui lui attribue 0,9).

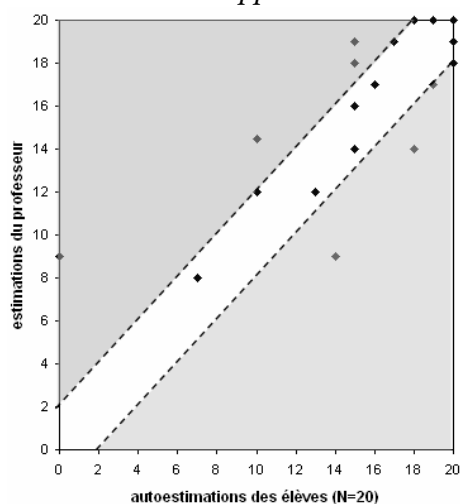
Deux autres étudiants, auxquels le professeur a octroyé respectivement 0,9 et 1 sont plus modestes que le jugement du professeur à leur sujet. tous (ils sortent des « pointillés »).

La majorité (15 sur 21) des étudiants sont dans les pointillés, ce qui indique une concordance entre le jugement du professeur et le leur. Il y a 3 exceptions vers le bas (professeur plus sévère que l'étudiant) et 3 vers le haut (étudiant plus sévère que le professeur). Comment les étudiants interpréteront-ils ces divergences quand ils en auront pris connaissance ? C'est un problème de jugement POST qui sera intimement lié à l'analyse. C'est pourquoi nous le traiterons dans l'analyse POST (section D8).

De l'utilité des auto-estimations

Matis Christelle (2006), AESS biologie.

A la suite de ce cours, j'ai demandé aux élèves (de 3^e GT) leur évaluation personnelle sur une interrogation portant sur des exercices d'application. Voici les résultats



L'analyse du graphique montre que 13 élèves sur les 20 sont en concordance avec le jugement (les points) du professeur. Quatre élèves se sous-estiment (dont un grandement) et deux élèves se sur-estiment.

A l'interrogation suivante, j'ai à nouveau demandé leur estimation personnelle de leur performance. L'analyse des résultats montre une représentation graphique très similaire à celle obtenue ci-dessus. Et le point le plus frappant : les élèves qui se sur-estimaient lors de la première interrogation se sont également sur-estimés lors de la seconde et les élèves qui se sont sous-estimés lors de la première interrogation se sont sous-estimés lors de la seconde.

En interrogeant ces derniers élèves sur « le pourquoi » de leur sous-estimation, ils m'ont répondu qu'ils n'avaient pas étudié et qu'ils pensaient que j'avais introduit des subtilités dans l'interrogation. Ayant fait cette analyse des auto-estimations dans plusieurs de mes classes, je me suis rendue compte que cette technique me permet de voir où l'élève pense qu'il en est dans son apprentissage et les raisons de son jugement.

E3. Les Jugements POST sur les performances (passées)

Sauf lorsqu'il y a contestation de l'allo-évaluation portée sur sa performance, le jugement POST est l'adoption du jugement porté par autrui ou par un système.

Par exemple, à l'université de Liège, en Faculté de Psychologie et des Sciences de l'Éducation, dans le cadre du projet MOHICAN¹⁹, s'est déroulée l'opération RESSAC (Leclercq, 2003). Elle a consisté à fournir à chaque étudiant avant la « bloque » une représentation graphique de ses performances (tel l'exemple ci-dessous) en CONNAISSANCES (de mémoire) et en UTILISATION des connaissances (ou compréhension) dans 12 épreuves : 4 de MOHICAN en octobre 1999 et 8 en janvier 2000 (leurs partiels dans 4 cours, avec, pour chaque cours, deux sous-notes). Quand le résultat de l'étudiant est supérieur à la moyenne (note Z^{20} positive), le graphique montre un bâtonnet dépassant vers la droite. Lorsque le résultat est inférieur à la moyenne (note Z négative), les bâtonnets dépassent vers la gauche.

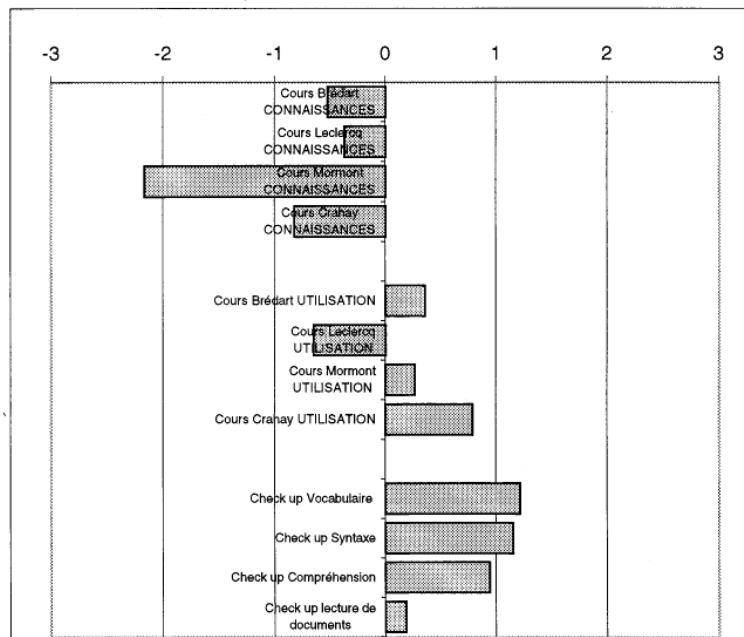
¹⁹ Monitoring HHistorique des CANDidatures ; ce projet a consisté à faire passer, en octobre 1999, 10 épreuves (dont Vocabulaire, Syntaxe, Compréhension de textes, compréhension de graphiques, tableaux et cartes de géographie) à 4000 étudiants entrant dans 8 des 9 universités de la CFWB.

²⁰ Une note Z dit de combien d'écarts-types un score est éloigné de la moyenne. Une note Z de 0 représente la moyenne. Une note Z de -2 représente un score éloigné de la moyenne (vers le bas) par une distance de 2 écarts-types.

Voici un exemple d'un tel graphique pour un étudiant :

Université de Liège - Faculté de Psychologie et des Sciences de l'éducation
Scores aux épreuves aux check-ups MOHICAN (octobre 1999) et aux partiels (janvier 2000)

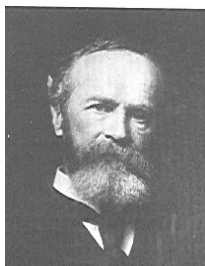
Moyennes	Ecart-type	Votre score	score Z (position dans le groupe)	Nom
57,6	20,4	47	-0,52	Cours Brédart CONNAISSANCES
71,3	17,1	65	-0,37	Cours Leclercq CONNAISSANCES
57,2	24,1	5	-2,17	Cours Mormont CONNAISSANCES
70	17,6	56	-0,82	Cours Crahay CONNAISSANCES
54	28,1	64	0,356	Cours Brédart UTILISATION
47	18,5	35	-0,65	Cours Leclercq UTILISATION
49,3	21,6	55	0,264	Cours Mormont UTILISATION
67,6	12	77	0,783	Cours Crahay UTILISATION
47,2	16,3	67	1,215	Check up Vocabulaire
71,7	16,8	91	1,149	Check up Syntaxe
46,5	20,8	66	0,938	Check up Compréhension
55,9	22,4	60	0,183	Check up lecture de documents



Voici des types de jugements POST (et de l'analyse en souligné) et régulations pour la radiographie spectrale de RESSAC obtenus en mai 2000 (Leclercq, 2003) : « *Je comprends mieux que je n'étudie* » (on constatera que pour cette étudiante « étudier » signifie « mémoriser »). « *Par rapport à d'autres étudiants de mon entourage, je m'en sors plutôt bien. Cela m'a donné encore plus confiance en moi. Je me suis dit "l'unif, c'est facile !"* ».

E4. Théories explicatives du jugement métacognitif

D4a. L'estime de soi et l'image de la tâche



Selon **William James** (1890), dans Bednar et al. (1993), « L'estime de soi globale est le résultat du rapport entre le succès (ce que l'on accomplit) et les prétentions (ce que nous pensons pouvoir accomplir) ».

$$\text{Estime de soi} = \frac{\text{Succès}}{\text{Prétentions}}$$

Il n'y a aucun impact de l'estime de soi si la personne subit un échec dans un domaine n'ayant aucune importance à ses yeux. Les individus focalisent principalement leurs compétences dans certains domaines ; en particulier dans ceux où ils aspirent à réussir.

On devrait parler de l'estime de soi dans divers domaines : l'estime de soi scolaire, l'estime de soi sportif, l'estime de soi dans les relations interpersonnelles, etc. L'estime de soi (tout court) serait la RESULTANTE de ces diverses composantes, mais selon des "formules" propres à chacun. L'un s'estimant beaucoup sur un terrain minimisera l'importance des autres terrains (où, très réalistement, il se considère nul), qui, à ses yeux, ne comptent pas; le résultat en est une très bonne estime de soi dans l'ensemble.

Ce même mécanisme (hypervalorisation d'un seul aspect) peut avoir des conséquences fâcheuses. Ainsi, dans nos pays (et plus encore au Japon), l'estime de soi scolaire pèse fortement sur l'image que les jeunes ont d'eux-mêmes. Une image scolaire négative peut "gâcher" l'image de soi totale et mener à la dépression, à la délinquance, au suicide (au Japon notamment). L'école porte d'autant plus de responsabilité dans ces cas que c'est souvent elle qui a "imposé" cette image de lui-même à l'enfant, à l'étudiant.

Albert Jacquard a déclaré à plusieurs reprises dans les médias "C'est le regard des autres porté sur nous qui nous rend aimables ou détestables à nos propres yeux".

Pour Carl Rogers, l'estime de soi résulte des interactions avec les autres et du regard porté sur soi par les autres. C'est ce qu'il appelle le SOI REEL. Mais il existe aussi un SOI IDEAL (que nous aimerions devenir), et vers quoi nous tendons. Ce besoin d'estime de soi nous pousse à rechercher un regard positif des autres, mais ce regard est souvent CONDITIONNEL : les autres ne nous l'octroient que si nous sommes conformes à LEURS attentes, à LEURS désirs, à LEURS normes, à LEURS volontés. Nous risquons de ne pas manifester ce qui est "nous même", pour obtenir ce regard positif. Pour Rogers, l'éducateur doit toujours avoir un regard positif inconditionnel permettant à l'autre (l'élève, l'enfant) de SE développer sans crainte, d'exprimer SA personnalité.

D4b. Le sentiment d'auto-efficacité



Albert Bandura (1981) a appelé « **Self Efficacy** » « la croyance d'une personne dans sa capacité d'effectuer un comportement nécessaire pour l'atteinte d'un résultat déterminé » (Sdorow, 1995). Ce concept doit être replacé dans l'idée de Bandura du « déterminisme réciproque » où les facteurs de la conduite ne sont pas déterminés essentiellement par l'environnement (comme Skinner en faisant l'hypothèse) ni essentiellement par le psychisme (comme Freud en faisant l'hypothèse pour l'inconscient), mais que ces trois facteurs s'influencent réciproquement. Le sentiment d'efficacité détermine nos choix d'activités, l'intensité de nos efforts, notre persévérance face à des obstacles, notamment en réduisant l'anxiété. Un déterminant important de la *Self efficacy* est l'expérience préalable des succès et des échecs. Un autre est l'expérience vicariante, à savoir la connaissance des performances d'autres auxquels nous nous comparons. Un troisième est la persuasion verbale (d'un pair, d'un enseignant, d'un parent...). Un quatrième est le degré d'excitation (arousal) du sujet dans une situation : ni trop ni trop peu.

D4.c. Les élaborations métacognitives et les auto-explications



Michéle CHI, de l'université de Pittsburgh

Chi et Bassock (1989) ont mis au point une méthode d'observation et de formation consistant à demander aux étudiants de faire des **commentaires** sur toutes les composantes d'un exemple, texte et graphique de physique, ce qu'elles appellent « élaboration ». Elles observent trois types d'élaborations : un tiers concernent les exemples eux-mêmes (auto-explications), un tiers concernent des phrases de « monitoring » et un autre tiers « divers ».

Dans leurs analyses des résultats, elles ont contrasté les bons étudiants, c'est-à-dire ceux qui avaient les scores les plus élevés (82% en moyenne) dans les tests de résolution de problème, des autres, appelés ci-après « faibles ».

Elles ont constaté (p. 269) que « Les bons étudiants génèrent un nombre beaucoup plus grand d'idées (15,9 en moyenne) que les élèves faibles (4,3 idées en moyenne) ».

Pour Ericsson & Simon (1984)²¹, corrélation n'étant pas causalité, cela montre qu'en répondant à l'injonction de **penser à haute voix** ou **d'écrire leurs idées**, les étudiants ne changent pas la structure de leurs processus mentaux, mais expriment de façon visible ce qu'ils auraient de toute façon pensé de façon inobservable.

Stein et Bradford (1979) ont par ailleurs montré que c'était le fait de donner SES propres explications qui était lié à une meilleure performance (et pas de répéter celles du professeur par exemple), et que ces **auto-explications** n'étaient efficaces que si elles étaient PRECISES.

D4d. Savoir que l'on ne sait pas : un atout métacognitif au service de l'apprentissage

Chi et Bassock (1989, p.272) pensent qu'une différence importante entre les « bons » et les « faibles » en résolution de problèmes est « leur capacité de gérer leur compréhension et incompréhension. ... Dans la mesure où une large proportion d'élaborations portaient sur le pilotage « monitoring » de la compréhension (39% pour les faibles et 42% pour les bons), ils doivent probablement avoir une fonction importante. C'est ce que se sont dit ces auteurs. Les bons étudiants ont produit un beaucoup plus grand nombre (9 fois plus que les faibles) de déclarations mentionnant qu'ils ne comprenaient pas, ce qui semble indiquer que les étudiants faibles ne pilotent pas bien leur compréhension, tout spécialement en ne réalisant pas qu'ils ne comprennent pas. »

Ces deux chercheuses pensent que « quand les étudiants ont peu de structures qui correspondent avec l'exemple donné, ils peuvent croire à tort qu'ils comprennent. Les meilleurs, par contre, qui ont des structures incomplètes peuvent piloter leur compréhension plus précisément en jugeant du degré avec lequel des parties de leurs structures se retrouvent dans l'exemple ou doivent être révisées ou complétées. Ils ont des schémas spécifiques à l'esprit et essayent de les confronter à l'exemple en posant des questions spécifiques (« pourquoi ce vecteur est-il négatif ? » ou « pour quelle raison la force change-t-elle ? »). Les étudiants faibles, par contre, posent des questions très générales (« Bon, et ici, que faut-il faire ? »). » (Idem, p. 273).

Les mêmes auteurs observèrent aussi que la prise de conscience des incompréhensions entraînait des efforts pour comprendre, mais à nouveau dans des proportions différentes (dans 85% des cas pour les forts et 60% pour les faibles) ». (Idem, p. 274)

En outre, elles observent que c'est essentiellement par l'augmentation des liaisons entre les éléments du texte qu'ils ont amélioré leur compréhension globale.



Pistes d'intervention

Le jugement sur ses performances étant la première opération métacognitive qui devrait déclencher les autres (l'analyse et la régulation), il importe d'entraîner les étudiants à porter SPONTANÉMENT des jugements sur leurs raisonnements, leurs productions, leur efficacité, etc.

Il importe aussi que ces situations de jugement soient le plus possible :

- pertinentes (portant sur des éléments importants),
- authentiques (correspondant à des situations réelles).

Il importe aussi que les jugements portés par les étudiants soient le plus possible :

- réalistes (proches de la réalité objective),
- discriminants (distinguant bien le correct de l'incorrect ou du douteux).

²¹ Cités par Chi & Bassock (1989, p. 270)

F. L'étudiant peut-il expliquer ses jugements sur sa connaissance ?

Les analyses métacognitives

F1. Analyses métacognitives en situation PRE

Les analyses ainsi menées sont particulièrement présentes dans les dialogues avec des conseillers d'orientation : vers quelles études me tourner ? Pourquoi ? Ai-je les prérequis en termes de produits (mes connaissances) ou/et en termes de processus (mes démarches mentales, mes méthodes de travail ?) et pourquoi ? Est-ce dû à moi ou à d'autres (le curriculum, les établissements de formation par lesquels je suis passé, les enseignants, mes parents, etc.). Est-ce changeable ? Comment ai-je réussi dans le passé ? Pourquoi ai-je échoué dans le passé ? Pour quelles raisons aurais-je des chances de succès en m'engageant dans cette voie ?

F2. Analyses métacognitives en situation PER

Les analyses sont sollicitées par exemple, en situation PER, par la présence de zones sur les documents écrits invitant à justifier ses réponses. Pour les performances isolables, des justifications peuvent être fournies par l'étudiant pour chaque réponse. Le professeur peut convenir de ne lire QUE les justifications qui accompagnent les réponses ne correspondant pas aux réponses attendues. De telles justifications portent souvent sur les produits : pourquoi je pense que ma réponse est correcte. Quand elles portent sur le processus, c'est souvent pour le réguler comme nous l'avons montré en C2.

Il y a analyse chaque fois que l'étudiant sélectionne un CRITERE de réflexion sur sa propre performance, critère à propos duquel il jugera et il se régulera.

En situation PRE, en vue de fournir un feedback, sur les jugements et analyses, l'enseignant (ou le conseiller en méthodes de travail ou en choix d'études)

- examine le dossier scolaire des performances antérieures ou les résultats à un test (d'intelligence par exemple), ou consulte des enseignants ayant eu cet étudiant dans leur classe.
- fournit un feedback d'optimisme ou de pessimisme, étant donné les statistiques dont il dispose... sans toutefois oublier le phénomène de résilience²² toujours possible.

F3. Les analyses métacognitives en situation POST

Voici des exemples de post-analyses d'étudiants de première année de candidature ainsi que les commentaires faits par le professeur (D. Leclercq) et une étudiante en mémoire (L. Wislez) qui joue le rôle d'assistante. Les étudiants sont invités à répondre sur des Questionnaires spectraux (Leclercq et Poumay, 2003) en version « papier ». Ces questionnaires facilitent la mise en évidence des situations problématiques, et de connaissance partielle, ainsi que le calcul des scores de Confiance et Prudence (Leclercq et Poumay, 2003) :

L'indice de Confiance est la moyenne des degrés de certitude qui ont accompagné les réponses correctes.

L'indice d'Imprudence est la moyenne des degrés de certitude qui ont accompagné les réponses incorrectes. La prudence est évidemment d'autant plus grande que l'indice d'imprudence est faible.

²² Capacité de certains individus à surmonter les circonstances défavorables et à réussir malgré des prédictions pessimistes.

Un formulaire spectral est une technique de questionnement favorisant l'analyse. Voici un tel questionnaire spectral.

24/09/2003

PPUQ Chap 3 et 4 p. 66 à 80

6. Aucune 7. Toutes

						Q	C						
100	80	60	40	20	0	1	T.S. Eliot a dit "Where is information lost in....." 1. affaires 2. knowledge 3. life 4. wiseness 5. men	0	20	40	60	80	100
100	80	60	40	20	0	2	L'unité d'information proposée par Shannon est 1. le volt 2. Le mot 3. le caractère 4. la syllabe 5. la lettre	0	20	40	60	80	100
100	80	60	40	20	0	3	Utiliser un ordinateur pour produire des textes touche des compétences 1. spécifiques (disciplinaires) 2. démultiplicatrices (instrumentales) 3. stratégiques (auto-cognitives) 4. dynamiques (motivacionnelles)	0	20	40	60	80	100
100	80	60	40	20	0	4	L'architecture des compétences en 4 paliers superposés est comparée par Leclercq à 1. un véhicule (un avion) 2. un animal (un insecte) 3. un outil (une mâche de foreuse) 4. Un jouet (un cerf-volant)	0	20	40	60	80	100

Prudence :

Confiance :

Un tel questionnaire favorise l'analyse pour les raisons suivantes :

- En analyse POST (parce que les étudiants disposent des réponses correctes prévues par le professeur), chaque étudiant va entourer la position spectrale de sa réponse à la question, donc à gauche en cas d'erreur (et d'autant plus à gauche que son degré de certitude était élevé) et à droite en cas de réussite (et d'autant plus à droite que son degré de certitude était élevé).
- Le libellé de la question reste sous les yeux de l'étudiant pour interpréter les mécanismes qui l'ont amené à fournir soit une réponse incorrecte avec une certitude élevée, soit une réponse correcte, mais avec une certitude faible.

Dans son cours de premier baccalauréat, Leclercq propose de tels tests lors de diverses séances. Ces tests portent comme nom les lettres de l'alphabet (A, B, C, D, etc.), dans l'ordre de leur passation. Les étudiants gardent le questionnaire spectral avec leurs réponses, mais envoient à l'assistant(e) une feuille de « **Dialogue métacognitif** ». L'assistant(e) transmet ce même document au professeur, mais sans les lignes 1 et 2 où figurent les nom et prénom de l'étudiant. Le professeur et l'assistant(e) renvoient leurs commentaires (d'où l'expression « dialogue métacognitif »).

Dans les 4 dialogues métacognitifs dans l'encadré ci-dessous, où l'intervention des encadrants est en gras, on retrouve à nouveau entremêlés des éléments d'analyse et des décisions de régulation. Nous avons souligné ces dernières pour les distinguer de l'analyse et pour mettre en évidence que souvent, l'étudiant saute du jugement (systématisé dans le formulaire) à la régulation sans passer par la phase d'analyse.

E16 (a eu 11 RC mais 9 RI qui lui ont fait perdre 4,5 points donc son score classique est 6,5 / 20). Jusqu'à présent c'est pas mal mais le problème c'est que malgré ces points de plus (de Confiance, de Prudence et de Nuance), je ne dépasse pas la moitié (10) sur 20. Le mieux c'est de ne pas répondre du tout aux questions où on hésite.

Vous obtenez en effet 9,5 sur 20. Quand des points sont retirés par RI, il vaut mieux omettre, mais seulement quand on est TROP PEU sûr(e).

Qu'est-ce qui aurait pu, selon vous, améliorer le nombre de bonnes réponses ?

E25

(Test D) Trop faible confiance: 3 réponses dont je n'étais pas sûre et auxquelles j'ai mis un coefficient de 0 se sont avérées exactes

Ce qu'il faudrait, c'est que vous compreniez pourquoi vous les avez crues incorrectes.

(Test D) : J'ai appris à faire un choix rapide entre diverses propositions

Oui, mais avec une trop faible confiance ! La rapidité, était elle le but recherché lors de ce test ? Ou l'exactitude ou la métacognition ?

E26 (Test D)

J'étais heureuse que j'avais regardé dans le dictionnaire quand j'ai lu le premier chapitre

Vous aviez eu là un réflexe TRES précieux quand on étudie...et il a été payant. Continuez comme cela !!

Avez-vous l'impression que vous auriez dû le consulter encore plus, le dictionnaire ?

E31 (Test D : n'a que 7 RC sur 20 Q).

J'aurais dû aller voir au dictionnaire lors de ma lecture

Exact : il faudra le faire pour tous les chapitres.

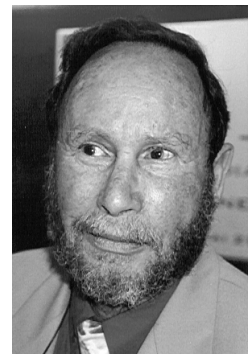
Ce qui est intéressant dans votre jugement, c'est que vous entrevoyez des modifications dans la façon de préparer un test et pas seulement dans la façon d'y répondre.

F4. Théories explicatives des analyses métacognitives

Certaines de ces théories explicatives permettent d'expliquer particulièrement bien les jugements, d'autres les analyses, et d'autres enfin les décisions de régulation. Certaines théories sont utiles pour interpréter plusieurs de ces opérations.

E4a. Les conceptions sur les médias

Ces analyses sont influencées par des conceptions spécifiques comme celles qui concernent les médias. **Gavriel Salomon**, 1983 et Cennamo, 1983 ont testé l'hypothèse selon laquelle les élèves pensent « *Reading is tough, TV is easy* ». Dès lors, on peut expliquer le constat de réussite moins bonne de la TV par l'implication moindre des apprenants, qui pensent que cela va « entrer tout seul ». De telles représentations (« préconceptions ») existent aussi sur les techniques d'évaluation : « Si je suis interrogé par QCM, je réussirai beaucoup moins bien que quand je peux me défendre à l'oral ». Donc, la précision de la consigne (type de questionnement, temps limité ou non, etc.) facilite la métacognition. Les mesures prises par certaines universités qui obligent les enseignants à préciser les modalités d'évaluation dès le début de l'année ont donc un effet bénéfique sur la métacognition.



E4b. Les attributions causales...



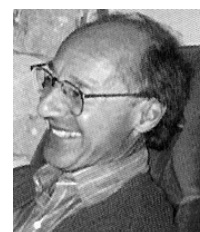
Rotter (1966) a défini le concept « locus of control » par le fait qu'une personne peut attribuer à ce qui lui arrive des causes internes (c'est de ma faute, c'est à moi qu'en revient le mérite) ou par des causes externes (je n'y suis pour rien, c'est dû aux circonstances ou à autrui).

Julian Rotter

Weiner (1985) distingue les causes « changeables » des causes « non changeables ».

La combinaison des théories de ces deux auteurs amène 4 situations illustrées par les explications que peut donner un étudiant pour expliquer son échec à un examen (de psychologie par exemple) :

J'ai échoué parce que...		Attribution causale	
		Interne	externe
changeabilité	changeable	...je n'avais pas étudié suffisamment	...j'ai tiré au sort la question la plus difficile de toutes
	Non-changeable	...je ne suis pas fait pour cette matière-là	...le professeur est très sévère (une « peau de vache »).



Bernard Wiener

Weiner (1984) a montré que ce n'est que lorsqu'un apprenant considère les causes d'une situation comme changeables et internes qu'il s'engage dans un processus de prise en charge de ses difficultés en vue de les surmonter.

...et leurs justifications

Le « je ne suis pas fait pour cette matière-là » peut être à son tour

- expliqué : « personne dans la famille n'a d'orientation vers cette matière » ou « mes études antérieures ne m'y ont pas préparé » ou « ces études conviennent à un autre type de personnalité »)
- argumenté : « la preuve, c'est que je réussis bien dans les matières x, y et z et non dans les matières a, b et c » ou « d'ailleurs, Mme D. pense comme moi ».

Faire parler l'étudiant des raisons de son échec ou de sa réussite et lui faire justifier ses jugements permet à l'enseignant de déterminer le type d'attribution causale et d'en faire prendre conscience à l'étudiant.

Pourquoi soudain elle ne comprend plus ?

Mignolet Benoît (2006), *AESS Sciences sociales*.

Une de mes élèves de 5^e secondaire était très inquiète quant à ses faibles résultats en sciences. Ayant discuté avec moi à plusieurs reprises des causes qu'elle attribuait à ses difficultés, cette élève m'indiquait qu'elle travaillait énormément à domicile, mais qu'elle étudiait fréquemment des notions qu'elle ne comprenait pas. Elle a dès lors pris la résolution de poser davantage de questions pendant le cours.

→ Ici, il s'agit d'un élément changeable et interne : « Je ne pose pas suffisamment de questions au professeur lorsque je ne comprends pas une notion ou un exercice ».

Cette élève me signalait également que les années précédentes ses résultats en sciences étaient satisfaisants. Elle attribuait ce changement au fait que le professeur actuel était beaucoup plus exigeant que ceux avec lesquels elle avait eu cours auparavant.

→ Ici, il s'agit d'un élément externe et non changeable (sévérité du professeur).



F5. Facteurs et méthodes favorisant l'analyse

E5a. Les méthodes de B. Noel

« Dans l'étude clinique, nous avons utilisé une méthode dite « concourante », basée sur les commentaires des sujets durant une activité de conceptualisation ou d'application. Nous avons essayé d'allier à la rigueur du questionnaire écrit l'observation du sujet en train de répondre ainsi que la souplesse et l'interactivité de l'interview du sujet immédiatement après sa réponse. Les questionnaires écrits étaient constitués d'échelles à choix multiples pour l'expression du jugement métacognitif et de la confiance dans les réponses. ». On reconnaît dans cette méthode de B. Noel (1991, p.23) la méthode clinique de style piagétien.

« Nous avons, dans les interviews, évité au maximum les questions de type « pourquoi ? », en faisant de préférence décrire par le sujet les éléments qui conduisent à la métacognition (« Qu'est-ce qui te fait dire ... ? Quelle est la différence pour toi entre ... et ... ? Peux-tu me donner un autre exemple de... ? » [...] « Peux-tu un peu mieux m'expliquer comment tu as fait [...] ce que tu veux dire [...] comment tu as trouvé cela. ». (Noël, 1991, p.23).

Noël (1991, p.144) a par ailleurs étudié la corrélation entre le réalisme (conformité entre jugement et réalité objective) et diverses variables :

Nom du test et auteur	Trait mesuré	Corrélation entre réalisme et score au test
GEFT (Wintkin)	Dépendance / indépendance du champ	0,159
PACT (Rotter)	Locus of control	-0,237
TASC (Sarrason)	Anxiété	-0,249
PMT-K (Hermans)	Motivation d'accomplissement	0,157
SW (Chaballe)	Désirabilité sociale	-0,106
Opin. M (Noel)	Opinion du maître (est-ce un bon élève ?)	0,275

Lors de ses entretiens cliniques, B. Noël (1991) recueille des déclarations telles que celles –ci :

« Je suis moyennement sûre parce que je ne suis pas forte habituellement en mathématique » (p. 83)

Elle relève des « métacognitions négatives et justes » (p. 89), des « négatives et fausses » (pessimiste) et des « positives et fausses (optimiste) » (p. 90).

Elle propose aussi des réflexions (p. 172) sur le constat « La métacognition optimiste est nettement plus fréquente que la métacognition pessimiste », constat maintes fois fait (Leclercq, 2003b, pp.60-91.)

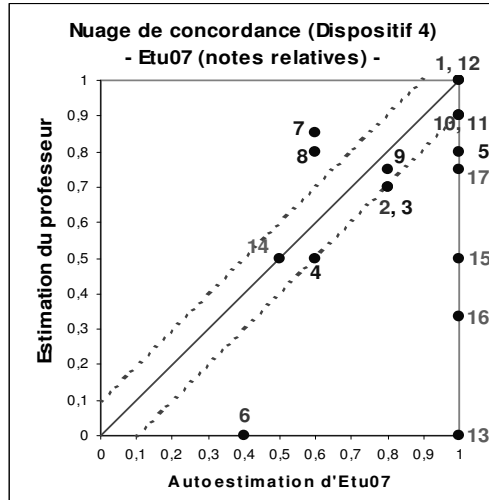


E5b. L'expérience de confrontation systématique de Jans (2000)

Jans et al. (1998, pp. 213-215) a organisé des confrontations dialectiques des jugements et des analyses entre chaque étudiant et le professeur (comme annoncé en section C2d) sur des PARMs (Projets d'Animation Réciproques Multimédia). Les notes d'auto-évaluation ont été remises **sous pli scellé** afin que le professeur ne puisse pas en prendre connaissance avant de mettre ses propres notes sur le même travail. Le professeur a annoncé que l'entretien se déroulerait comme suit :

« Les 17 critères seront passés en revue, mais nous discuterons principalement ceux pour lesquels existe un écart important entre votre note et la mienne, et donc une discordance de vues entre vous et moi. »

Pour les performances complexes, voici comment Jans (2000) a présenté les auto-notations ou jugements (en abscisse) comparées aux allo notations ou jugements (en ordonnée), à gauche pour un étudiant sur les 17 critères (numérotés), et à droite pour les 18 étudiants pour un critère précis « Présentation de l'article ». On constate que l'étudiant 07 est plus optimiste que le professeur. Une confrontation professeur / étudiant a eu lieu, créant ainsi une situation POST. Sans une telle confrontation et discussion, les discordances sont considérées comme des erreurs d'auto-évaluation de l'étudiant.



Le graphique individuel (de gauche) de chaque étudiant lui a été remis pour qu'il prépare sa rencontre-discussion (confrontation dialectique) avec le professeur.

Lors de la séance de confrontation, l'étudiant et le professeur sont munis des notes (et justifications de ces notes) de l'un et de l'autre. « L'étudiant et le professeur ont eu l'occasion de se poser mutuellement des questions, de demander des explications quant aux points accordés ou aux justifications écrites fournies. » (p. 345). Ces entretiens ont été filmés.

Jans (2000, p.336) rapporte la phrase de Boud (1995, pp.204-205) sur le « feedback réussi » :

« Si vous souhaitez donner un feedback efficace, vous devriez : être réaliste, être précis, être sensible aux

objectifs de la personne, répondre en temps opportun, être descriptif, ne pas porter consciemment de jugements, ne pas comparer, vous montrer appliqué, être direct, être positif, être conscient de votre état d'esprit. ». Jans continue : « Et au récepteur du feedback, Boud lui conseille d'être explicite²³, attentif, conscient de son état d'esprit²⁴, silencieux. ».



Pistes d'intervention

Se limiter à recueillir et évaluer les seuls jugements métacognitifs risque d'avoir bien peu d'impact sur les comportements (d'apprentissage ou de réponse à des tests). L'étape de l'analyse est cruciale pour décider de la troisième opération, à savoir les régulations. C'est l'analyse qui indique ce qu'il faut changer et pourquoi.

Cette analyse se fait notamment² en termes d'attributions causales. Le professeur aidera l'étudiant à en apprécier l'exactitude : une cause jugée externe l'est-elle vraiment ? Et une cause jugée inchangeable ? Etc.

L'analyse porte aussi sur la nature des forces et des faiblesses des processus ou des produits. A quelle étape de la démarche me suis-je trompé(e) ? Comment pourrais-je m'en rendre compte ? Quel est l'aspect le plus faible de ma réponse ? De quelle nature est cette faiblesse ? Quelle est sa gravité ? Etc.

²³ Make it clear what kind of feedback you are seeking

²⁴ Notice your own reactions, both intellectual and emotionnal

G. La métacognition sert-elle l'action ?

Les Régulations métacognitives

G1. Les Régulations PRE

Ce sont les décisions que l'on prend avant qu'une performance puisse être produite.

L'auto-fixation de la difficulté des tâches en est un bel exemple

On se référera à la théorie d'Atkinson exposée au chapitre 4 section F4.

Voici un exemple de dialogue professeur / étudiant sur ce point, après les examens partiels du mois de janvier :

Professeur : « Maintenant que nous sommes six mois AVANT les examens de fin d'année, que visez-vous comme résultat ? ».

Etudiant : « Réussir en deuxième session, avec une satisfaction »

Professeur « Pourquoi pas en première session ? »

Etudiant : « Maintenant que je connais le niveau d'exigence, je sais que je n'y arriverais pas en première session sans gâcher la qualité de ma vie ».

L'adoption préventive d'habitudes d'étude relève aussi de régulations PRE (ici PRE-examen).

« Une des raisons pour lesquelles les étudiants développent une propension à apprendre par cœur des sujets potentiellement significatifs est qu'ils ont appris par de mauvaises expériences que des réponses fondamentalement correctes mais non exprimées dans une correspondance « verbatim » avec ce qui leur avait été expliqué sont considérées comme incorrectes par certains professeurs. Une autre raison vient d'un niveau général élevé d'anxiété ou d'une expérience d'échec chronique dans une matière ... Ils manquent de confiance dans leur capacité d'apprendre de façon significative, et dès lors ne voient pas d'autre alternative à la panique que d'étudier par cœur » (Ausubel et al., 1968, pp.41-43).

G2. Les Régulations (ou décisions) PER

Pendant la production du travail (ou la réponse aux questions du test), les étudiants ont pris une série de décisions : Omissions, corrections de réponses, approfondissement de certaines questions et pas d'autres, demande d'aide. Dans ces décisions, l'influence du payement (le barème de cotation annoncé) est omniprésente. On rappellera ici que l'omission constituant une zone aveugle de l'évaluation, on tentera de l'éviter le plus possible par un barème approprié. On consultera à ce sujet l'article de Leclercq et Poumay (2003) « La connaissance partielle ».

Faisons ici un gros plan sur l'accès à des aides lors de la performance

a) On peut autoriser l'étudiant à consulter une référence (un livre, des notes, un site web) après avoir répondu une première fois.

Exemple 1 :

Dans une expérience de testing en deux temps, Leclercq et Boskin (1990) ont montré que la consultation (par des étudiants universitaires) des pages d'un hypermédia améliorait la performance au post-test chez 43 des 50 étudiants. Ils ont aussi mis en évidence que le gain dépendait du degré de familiarité de l'étudiant avec les QCM, avec l'informatique et avec la connaissance préalable du contenu.

Exemple 2 :

Dans une autre épreuve universitaire en deux temps (avec consultation du livre de référence entretemps), Leclercq (1993, p. 223) a observé les améliorations suivantes :

-pour les QCM-RCEexpl : de 50% à 54% (+4%) et les QCM-RC « Toutes » : de 55 à 60% (+5%)

-pour les QCM-RC « Aucune » : de 40 à 48% (+8%) et les QCM-RC « Absurdité » : de 58 à 65% (+7%)

-pour les QCM-RC « Manque » : de 33 à 35% (+2%)

Il n'est pas difficile d'imaginer les raisons qui amènent à ces différences de résultats.

b) On peut aussi organiser le testing de sorte à ce que l'on obtienne systématiquement la réponse d'un étudiant d'abord sans aide, puis avec aide.

Exemple :

Magain (1995) et Leclercq (1999, chap. 4, p.23) recourent à des questions PRIM-BIS, où la deuxième question commence par dévoiler une partie de la réponse puis pose une question plus simple que la première. Par exemple, : « Dans la question PRIM il y avait une absurdité. Voyez-vous laquelle ? » . Suit alors la place pour une réponse rédigée (QRO) ou une liste de solutions (QCM), la solution correcte de la BIS pouvant être une QCM-RCExpI ou « Aucune » ou « Toutes » ou « Manque ». Ces auteurs ont observé qu'alors que les questions PRIM entraînaient 35% de réponses incorrectes, 20% subsistaient même après la BIS, 15% seulement étant « corrigées ». A l'inverse, parmi les 65% de réponses correctes à la PRIM, seules 50% étaient correctement justifiées lors de la BIS, 15% ne l'étant pas.

Ce **questionnement en deux temps** permet à D. Leclercq de distinguer la capacité d'analyse (vigilance cognitive) de la compréhension. La deuxième situation (b) ne relève PAS de la régulation, mais en constitue un entraînement, car, comme le dit Noël (1991, p.13) : « *Ce n'est que lorsqu'ils sont avertis de l'existence d'un problème ...que les enfants améliorent leurs résultats aussi bien lorsque l'incohérence est présente de manière explicite que lorsqu'elle n'apparaît qu'implicitement.* »

G3. La régulation en situation POST

1. De possibles effets pervers

On aurait pu encore distinguer la RETRO-diction qui consiste à répondre à la question « A refaire, maintenant que je sais ce que je sais, qu'aurais-je dû répondre ? » et en faire une composante de la situation de POST diction. Nous ne suivons pas ce chemin car il pourrait entraîner un effet pervers important, à savoir véhiculer l'idée que les expériences d'échec sont néfastes et à éviter à tout prix, alors que nous pouvons apprendre beaucoup de nos erreurs. En outre, le temps ne se « rebobine » pas et il ne nous semble pas souhaitable d'encourager les évalués à reconstruire leur propre histoire comme si elle ne s'était pas déroulée.

2. Situation métacognitive POST et Débat

« *La seule expression du point de vue de l'apprenant est une condition nécessaire mais non suffisante de la **mise en relation** de son point de vue avec d'autres. Encore faut-il que cette mise en relation elle-même soit systématiquement organisée. Cela suppose que le temps nécessaire à une telle **confrontation** soit dégagé dans l'espace de formation ; cela suppose aussi la mise en place de dispositifs pédagogiques qui organisent efficacement cette confrontation des points de vue.* » (Bourgeois et Nizet, 1997, p.148).

« *Un des exercices proposés aux étudiants a pour but, par une réflexion sur leurs propres habitudes de prise de notes et par la **confrontation** de leur pratique à celles de leurs pairs, de les amener à prendre conscience que c'est à chacun de se construire sa propre méthode selon une approche personnelle, selon le contexte,...* » (Parmentier et Romainville, 1998, p.79).

Delhaxhe, Lanotte et Duchateau (2003) adoptent ce même principe. Dans leur cours à distance « Méthodes en ligne », ils invitent l'étudiant à prendre des notes sur un exposé (vidéo) puis à **comparer** ses notes avec celles d'autres et à en tirer des conclusions sur les stratégies à adopter.

3. Fournir les critères ou non ?

La tâche peut consister non seulement à produire de la qualité, mais aussi à définir les critères mêmes de cette qualité, ce qui constitue une performance d'un niveau supérieur. On risque de ne pas l'entraîner et même de l'atrophier si on fournit toujours à l'étudiant les critères par lesquels on va l'évaluer ou par lesquels il va s'auto-évaluer. La confrontation des points de vue sur les critères eux-

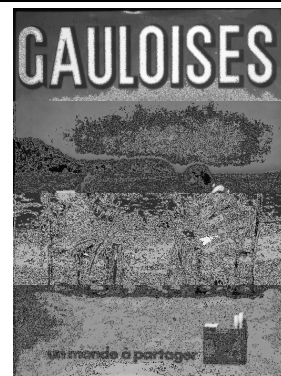
mêmes permettra à l'apprenant de tirer parti des différences et sa **capacité de s'enrichir ainsi de l'apport d'autrui** relève de la post-diction.

4. Un exemple de dispositif favorisant le jugement, l'analyse et la régulation POST

Voici un exemple de dispositif d'entraînement et d'évaluation sur cette capacité de se fixer des critères et les défendre :

Le cours de Pédagogie & Audio-Visuel est donné par D. Leclercq et F. Georges, essentiellement en Enseignement à Distance (EAD). Les étudiants disposent d'une version écrite du cours (200 pages) et, pour plusieurs chapitres, d'une version « en ligne ». Pour diverses épreuves, un certain flou est laissé dans les consignes. Ainsi, dans l'analyse d'un message publicitaire (une page dans un magazine) la consigne est

«Pour faire comprendre à des adolescents les mécanismes par lesquels les publicitaires manipulent les lecteurs - consommateurs, souvent de façon subliminale (sans que le lecteur en prenne conscience), analysez la page publicitaire ci-dessous, en appliquant, entre autres, les principes du chapitre 1 section K que vous jugez les plus pertinents pour atteindre l'objectif ».



Les étudiants sont alors invités à échanger leurs analyses et à discuter la pertinence de leurs choix et celle des choix des autres, par exemple en produisant une réponse « consolidée » par les apports du groupe (ce qui vient d'autrui étant dans une couleur différente).

5. Exemples d'analyses et de régulations POST sur les performances

Jans (2000) a organisé des sessions de confrontation métacognitive entre l'étudiant et le professeur (D. Leclercq). A l'issue de cette rencontre, l'étudiant a répondu par écrit à deux questions :

1. Qu'est-ce que cette discussion avec le professeur t'a appris (à tous points de vue) ?
2. Quelle(s) décision(s) as-tu prise(s) pour tenter d'améliorer tes auto-évaluations à l'avenir ? A quoi vas-tu être plus attentif(ve) ?

Jans (2000, p.379) a fait une synthèse des réponses à la question 1. L'apprentissage le plus souvent cité (par 12 sur 18 étudiants) concerne sa capacité d'autoévaluation. Par exemple : « J'ai appris que j'avais l'habitude, quel que soit le critère, de me surestimer ». L'utilisation du passé (j'avais) est prometteuse.

Voici quelques réponses à la question 2 :

« Je vais essayer de prendre plus de recul par rapport au niveau d'exigences parfois exagéré que je me fixe. »

« Je vais tenter d'obtenir plus d'information sur la signification et la précision de certains critères d'évaluation qui me semblent vagues ou mal définis. »

« Je vais prendre plus de temps pour m'auto-évaluer. »

« Je veillerai dorénavant à vérifier plus profondément mon travail avant de m'évaluer ». »

« Je vais donc toujours m'atteler à justifier mes observations, mes critiques, puis ensuite seulement me mettre les points. »

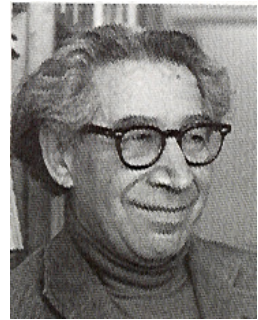
G4. Théories explicatives des régulations métacognitives

F4a. La réduction de la dissonance cognitive : une non-régulation ?

Selon Heider (1958), nous cherchons la COHERENCE entre nos diverses actions, et l'EQUILIBRE COGNITIF avec nous-mêmes et les autres. En conséquence, nous avons tendance à éviter les situations de contradiction cognitive, de rechercher les personnes qui "pensent comme nous". Nous ne supportons pas bien la contradiction entre diverses utilités et, pour réduire cet inconfort, nous sommes amenés à prendre des décisions non optimales ou à les justifier.

Ainsi imaginons que nous venions d'acheter une voiture (X) conformément aux critères auxquels nous accordons le plus d'importance, c'est-à-dire, dans l'ordre : confort, faible consommation, robustesse, rapidité. Imaginons maintenant que quelqu'un nous démontre que ce n'est pas le meilleur choix que nous aurions pu faire : une autre voiture (Y) remplit mieux toutes les conditions.

Cette dissonance cognitive (entre ce que nous aurions souhaité faire et ce que nous avons vraiment fait) est d'autant plus pénible que la différence est grande. Selon **Léo FESTINGER** (1957), pour réduire cette dissonance, nous utiliserons diverses stratégies, comme, par exemple,



Léo Festinger

- a) "réviser" nos critères : nous nous disons que, tout compte fait, le confort n'est pas si important, ni la rapidité, ni la réussite scolaire (rationalisations a posteriori)
- b) introduire de nouveaux critères (X est une voiture européenne, je connais le concessionnaire, etc.).

Quand on n'a pas ce que l'aime, il faut aimer ce que l'on a

Vervaeke Hilke (2006), AESS Arts.

Pendant l'évolution d'un projet de design, l'étudiant doit définir ces choix : les matériaux, les fonctionnalités, l'esthétiques, ... Pendant le cheminement du projet, nous essayons que l'étudiant sait argumenter correctement ces choix. Souvent, quand on pose quelques questions, l'étudiant change son discours et ses valeurs. Par exemple : le projet d'une fourchette. Au début le plus importance était l'aspect ergonomique, comme il n'arrivait pas à répondre complètement à cet aspect, il a changé sa stratégie et indique que l'aspect le plus important de son projet et le mode de production.



F4b. Réguler par auto-fixation de la difficulté des tâches

John Atkinson (1974) nous propose un modèle intéressant de régulation, modèle déjà abordé au chapitre 4, section F4. Rappelons qu'un modèle ne représente pas forcément la réalité, mais peut viser à donner des prédictions proches de la réalité. Ici, ce modèle est du type « cela se passe comme si... ». Pour Atkinson, le plaisir retiré de la réussite (ou Utilité de la réussite) d'une tâche est une fonction de la difficulté de cette tâche, difficulté que l'on peut définir comme $1-p$ (où p est la probabilité de réussite) :

$$\text{Plaisir (ou U)} = 1 - p$$

Si le modèle s'arrêtait là, les individus se fixeraient des tâches très difficiles, puisque ce sont elles qui produisent le plus de plaisir quand on les réussit. Or nous savons que la majorité des personnes ne fait pas ce choix. Notamment parce que les tâches difficiles, par définition, sont rarement réussies. A quoi bon dès lors entreprendre des tâches qui donneraient une grande satisfaction SI on les réussissait, mais qu'en réalité on ne réussit jamais. Le modèle d'Atkinson prévoit cela puisqu'il considère que le choix d'une tâche n'est pas fait sur base du Plaisir, mais du plaisir probabilisé :

$$\text{Plaisir probabilisé} = (1-p) p$$

En remplaçant p par toutes les valeurs allant de 0 à 1, on peut montrer (Leclercq, 1999) que le plaisir probabilisé maximal se produit quand p vaut 0,5, donc pour une tâche dont les chances de succès sont d'une sur deux, de 50%, soit un compromis : une tâche suffisamment difficile pour procurer du plaisir en cas de réussite et suffisamment facile pour que cette réussite survienne...assez souvent.

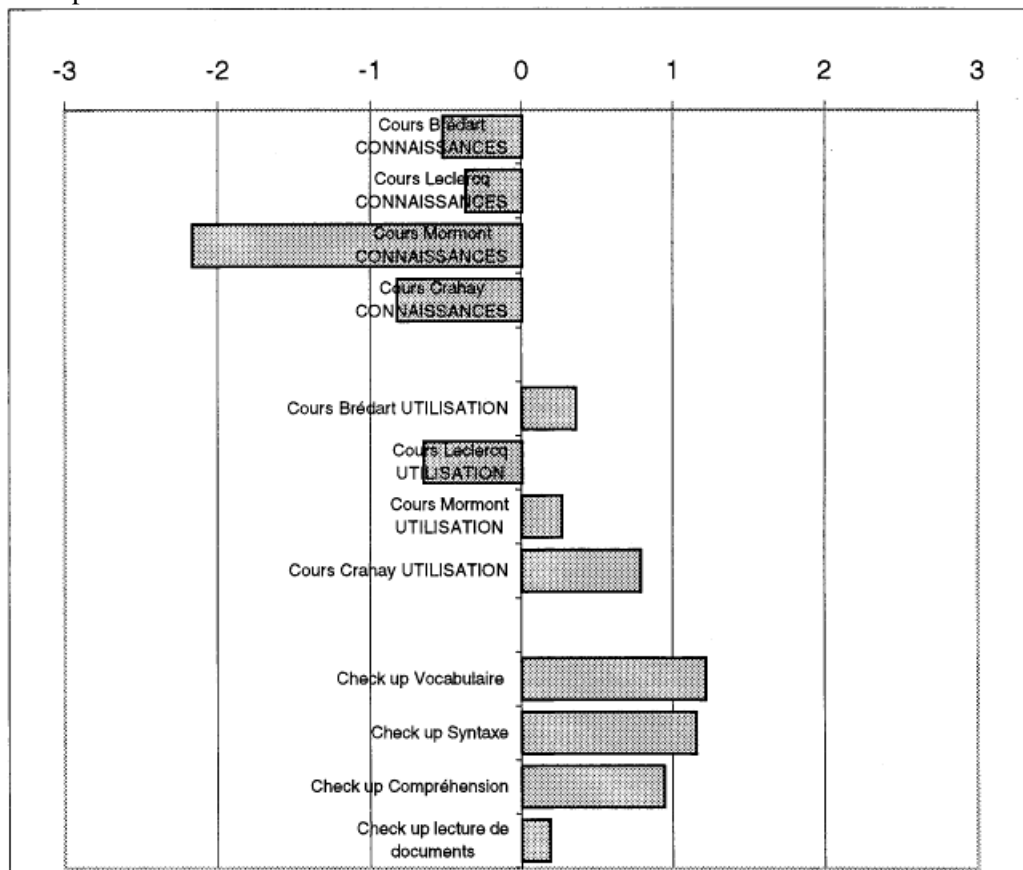
L'estimation de p est une affaire personnelle et c'est sur cette base que la plupart des personnes, selon Atkinson, se fixent la difficulté des tâches qu'ils entreprennent. Ainsi, le type d'études que l'on entreprend, à l'intérieur de ce type d'études, le niveau de réussite (S, D, GD, PGD) que l'on vise.

G5. RESSAC : un exemple de régulation portant sur les méthodes d'étude

L'expression RESSAC signifie « Résultats d'Epreuves Standardisées au Service des Apprentissages en Candidature » (Leclercq et al., 2003, pp.155-170). Cette opération a concerné les étudiants de 1^{re} candidature en Faculté de Psychologie et des Sciences de l'Education (FAPSE) de l'ULg de l'année académique 1999-2000. La première semaine d'octobre, ils avaient été testés par 4 épreuves mesurant, chacune à leur manière la « compréhension » : Vocabulaire, Syntaxe, Compréhension de textes, compréhension de cartes, tableaux et graphiques à contenu géographique. Ces étudiants avaient subi des interrogations dispensatoires en janvier dans 4 cours différents (Professeurs Brédart, Crahay, Mormont et Leclercq) et avaient la possibilité de repasser ces examens en juin ou août en cas d'échec.

Ces quatre professeurs ont calculé, pour leurs 4 interrogations dispensatoires de janvier 2000, deux scores (en %), l'un portant sur les questions de "connaissances" (C) et l'autre portant sur les questions "d'utilisation des connaissances" (UC) de manière à produire 8 notes (exprimées en % de réussite et en notes Z par rapport à la moyenne de leur cohorte).

L'opération RESSAC a consisté à fournir aux étudiants un diagnostic métacognitif avant la période de bloque (mai), de manière à ce qu'ils puissent, au vu des résultats, changer éventuellement leurs méthodes d'études avant les examens de juin. Ce diagnostic avait la forme d'une « radiographie » séparant les 8 scores de « Compréhension » des 4 scores de « Mémorisation » ou « Connaissances ». En voici un exemple :

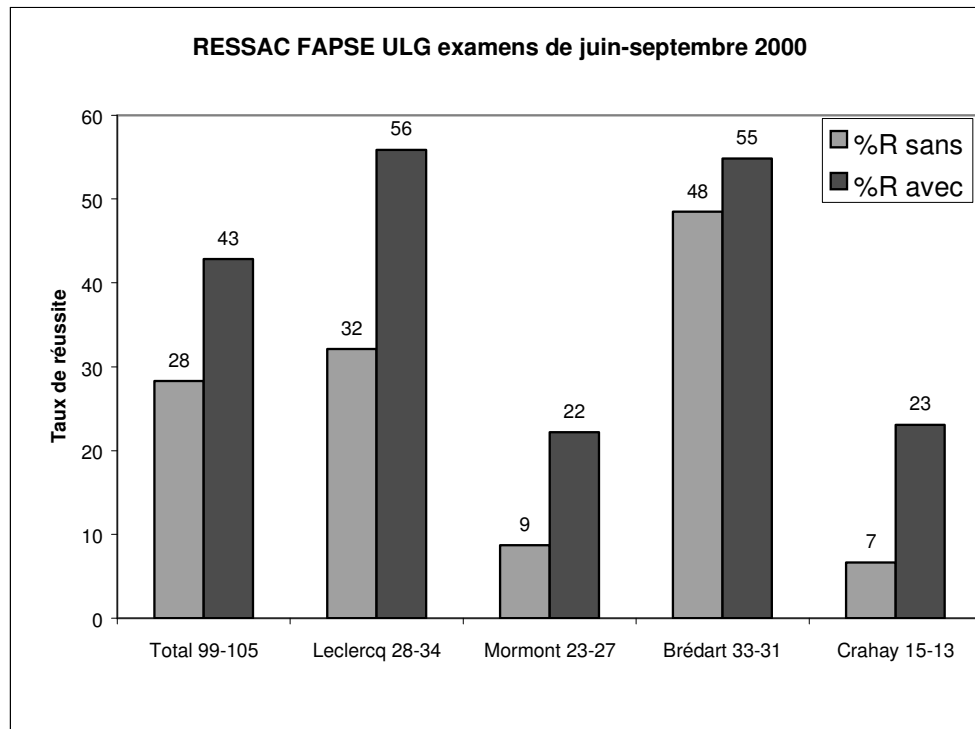


On constate que pour cette étudiante, la compréhension aux check-up est assez bonne, la compréhension aux partiels est moyenne et la Mémorisation aux partiels est faible.

Pour les analyses statistiques qui suivent, nous n'avons gardés, pour chacun des 4 cours, que les étudiants qui **A LA FOIS** y avaient échoué au partiel de janvier 2000 (leur "prétest"), reçu la radiographie RESSAC, obtenu une note pour juin ou septembre 2000, et donné au questionnaire d'avis de fin juin 2000 une réponse que l'on peut classer SANS AMBIGUITE dans l'une des catégories SANS et AVEC ci-dessous. Parmi ces étudiants (en partie les mêmes et en partie différents pour chaque cours), deux groupes ont été contrastés : ceux

Groupe 1 : **SANS** décision de changement de méthode d'étude,

Groupe 2 : **AVEC** déclaration (en juin après les examens) de changement de méthode d'étude, soit d'étudier en profondeur, de ne plus mémoriser sans comprendre, soit de mémoriser plus (là où le diagnostic était "faiblesse en connaissances").



Dans ce tableau, les nombres sous les histogrammes sont les nombres d'examens (et donc d'étudiants) concernés, sachant que les mêmes étudiants peuvent se retrouver dans plusieurs cours. Il y a donc plus d'examens (204, repris dans le graphique à côté de l'étiquette « total ») que d'étudiants (88, non précisés dans le graphique)

Sur ces 204 examens partiels ratés en janvier 2000, 73²⁵ se sont transformés en une réussite en juin-septembre 2000 à ces mêmes cours, soit 35% de réussite moyenne, mais avec 28% chez les SANS (modification de stratégie) et 43% chez les AVEC (modification de stratégie), soit 15% de différence moyenne, et **la différence moyenne va dans le même sens pour LES 4 COURS.**

Que les étudiants reçoivent un feedback serait donc une condition nécessaire mais non suffisante pour que se produise la décision d'étudier en profondeur. Encore faut-il

1. que ce feedback soit suffisamment "diagnostic" pour déclencher une prise de conscience (la distinction entre CONNAISSANCE - de mémoire- et UTILISATION ou COMPREHENSION semble un minimum suffisant pour produire un effet chez certains).
2. que l'étudiant interprète ce feedback de façon à prendre une décision de modification de ses méthodes de travail vers l'étude en profondeur. Nous pourrions l'y aider (avec l'aide de Guidance-Etude par exemple).

²⁵ Soit 28% de 99 examens, plus 43 % de 105 examens.

Nous avons décrit ailleurs (Leclercq et al. 2003, pp.166-167) les réserves à prendre en compte dans l'analyse de ces résultats.

Ces résultats sont très encourageants, non pas par leur ampleur, limitée à quelques étudiants, mais par les hypothèses d'amélioration qu'ils permettent maintenant de formuler. Ces résultats se reproduiraient-ils une autre année ? Rien ne permet de penser le contraire. Nous en faisons donc l'hypothèse. En gardant à l'esprit cependant que la relation causale "Je constate, donc je décide, donc je change, donc je réussis" puisse être fallacieuse. Il se pourrait que nous soyons devant un phénomène de **rationalisation a posteriori** pour bon nombre d'étudiants. Seules des analyses ultérieures plus approfondies dans des répliques de l'expérience pourront lever cette ambiguïté.

Peut-on espérer améliorer ces résultats ? Nous pensons que oui, et de plusieurs façons, que nous avons décrites ailleurs (ibid), en comptant sur le phénomène de résilience (Born et al.), cette capacité de certaines personnes à surmonter les obstacles et sur la possibilité de « faire mentir les prédictions » quand elles sont connues (Boxus, 1971).

Les enseignants se trouvent en effet dans une situation comparable à celle d'un médecin qui, s'il n'intervenait pas, trouverait des corrélations quasi parfaites entre le fait d'être infecté et le fait de décéder. Mais voilà, les médecins soignent, faisant (heureusement) se dégonfler les mortelles corrélations : désormais on survit, même si on a été infecté...à condition d'intervenir à temps, en suffisance et de façon adéquate !!!!



Pistes d'intervention

Juger de la qualité (insuffisante) de sa performance et en analyser correctement les causes sans porter remède est une entreprise peu rentable. Le professeur encouragera les étudiants à opérer les changements suite aux diagnostics qu'ils auront eux-mêmes portés sur leurs processus ou sur leurs productions.

Encore faut-il que le message (le feedback) **soit intelligible, systématique, et répété**. C'est le cas du tout dernier exemple (RESSAC) : le feedback est simple (il n'y a que deux catégories de modalités d'étude, mémorisation pure et compréhension en profondeur) et il est répété, dans les mêmes termes, à quatre reprises (les 4 cours différents). Même dans de telles conditions, on voit qu'à peine 1 étudiant sur 2 modifie sa stratégie d'étude. Il est vrai que le message (le feedback) n'était pas de la même limpidité pour tous les étudiants.

H.Y a-t-il des styles cognitifs et métacognitifs ?

Différences interindividuelles dans la gestion de soi et des apprentissages

L'aspect subjectif de la métacognition est souvent plus important que l'aspect objectif. Bien sûr l'individu dispose de données objectives (exemples sur sa taille, son poids, ses performances scolaires ou sportives, etc.). Néanmoins le jugement de satisfaction ou d'insatisfaction qu'il porte sur ces données objectives est lui subjectif. Il en va de même quand il s'agit de prédire sa propre performance. C'est le cas notamment des degrés de certitude accompagnant les réponses à un test. On peut facilement imaginer que certaines personnes auraient tendance à se sur-évaluer alors que d'autres auraient tendance à se sous-évaluer, et ce, quel que soit le domaine où l'on est évalué. Cette question a fait l'objet de vérifications expérimentales. D'autres aspects de la métacognition peuvent aussi varier d'une personne à l'autre, de façon systématique, quelle que soit la situation. Ces tendances systématiques ont été étudiées sous les termes « style cognitif » ou « style d'apprentissage ». C'est cette même question que l'on peut se poser à propos de la métacognition.

H1. Les styles cognitifs et préférences de styles d'apprentissage

Il existe plusieurs propositions, théories, hypothèses quant aux styles cognitifs : certains seraient plus « iconiques » et d'autres plus « verbaux » (Riding, 1976), certains seraient dépendants du champ et d'autres indépendants du champ (Witkin, 1950, 1959), certains seraient sérialistes et d'autres globalistes ou holistes (Pask, 1976), certains étudieraient de façon superficielle, d'autres de façon approfondie (Entwistle et Marton, 1985), etc. Approfondissons ici un seul de ces « styles », celui que Jérôme Kagan (1965) appelle impulsifs vs réflexifs ou **tempo conceptuel de KAGAN**.

Le degré de satisfaction qu'un étudiant éprouve pour la qualité de sa réponse pendant un test a un impact sur le temps qu'il met pour répondre au test. Certains étudiants « ne sont jamais satisfaits » ou « placent la barre très haut » (ou se sous-estiment) et mettent beaucoup de temps pour compléter une épreuve. D'autres au contraire, sont vite satisfaits (ou se sur-estiment) et donc rendent leur copie beaucoup plus vite. C'est en comparant le temps de réponse avec la qualité objective de ces réponses que l'on peut approcher cette composante métacognitive et la distinguer de l'aspect cognitif.

L'épreuve typique de Kagan est le *Matching Familiar Figures Test* (MFF de Kagan *et al.*, 1964). Un modèle est présenté, ainsi que 6 dessins fort proches du modèle, mais dont un seul lui est totalement identique. L'examineur chronomètre le temps de latence à la demi-seconde près et le nombre d'erreurs.

Les **IMPULSIFS** sont les enfants dont le temps de latence moyen est inférieur à la moyenne du temps du groupe d'âge et dont le nombre d'erreurs est supérieur à la moyenne des erreurs pour le groupe d'âge.

Les **REFLEXIFS** sont définis de manière inverse. Notons que l'on détermine ainsi 4 catégories, et que l'on en ignore deux (les sous-groupes a et d) :

		Latence (temps de réponse)	
		Faible	Elevée
Taux d'erreur	Faible	a	REFLEXIFS
	Elevé	IMPULSIFS	d



Jérôme Kagan

Kagan (1965) et Yando et Kagan (1968) observent une assez bonne stabilité test-retest (dans les mêmes conditions) du trait « réflexivité-impulsivité », aussi appelé « tempo conceptuel ».

Messer (1970), lui, a montré que si l'on augmente les conditions d'anxiété, les enfants « rapides et inefficients » (les impulsifs) augmentaient leur temps de latence ... et diminuaient leur nombre d'erreurs (rejoignent les réflexifs), ce que l'auteur explique par une attention accrue à la tâche en raison d'un accroissement de la crainte de l'échec.

Bien entendu, la consigne (temps libre ou temps limité) a un impact sur les comportements, donc les observations.

Dans l'exemple qui précède, on peut avoir l'impression que l'étudiant est victime de lui-même (par sur-estimation ou sous-estimation par exemple). La métacognition peut amener l'étudiant à se gérer volontairement, et ce de diverses façons.

H2. Auto contrôle et gestion de soi

Chacun d'entre nous peut SE manipuler. Tout d'abord en s'assignant (en pleine conscience) des buts dont la difficulté est calculée de telle sorte que leur atteinte procure un plaisir suffisamment intense et suffisamment fréquent (voir compromis au centre de la théorie d'Atkinson).

Ensuite en s'octroyant des renforcements pour des efforts. C'est le cas de l'étudiant "en bloqué" qui s'octroie quelques pages de bandes dessinées ou dix minutes de jogging ou de musique, chaque fois qu'il a terminé l'étude d'un chapitre. Les personnes qui tentent d'arrêter de fumer ou de perdre du poids ou de cesser de se ronger les ongles, etc. utilisent aussi parfois ce principe (de compensation, d'auto-récompense pour efforts fournis). La plupart de ceux qui acceptent de "souffrir" durant leurs études en espèrent des compensations (différées).

Une autre technique consiste, pour **protéger sa décision** à "arranger son propre environnement" en éliminer les stimuli-déclencheurs ou simplement les possibilités matérielles de passage à l'acte. Le fumeur n'achètera plus de cigarettes, évitera même de passer en face du magasin où on vend, mettra au grenier le fauteil-dans-lequel-juste-après-le-repas-il-ETAIT-si-bon-de-fumer-une-(bonne) cigarette.

Pour plus de détails, on consultera l'ouvrage de Van Rilaer (1992) "La gestion de soi".

Ayant compris que plus un contenu est travaillé en profondeur, mieux il est retenu (Cf. Craik et Lockart, Chap. 2 section C), les étudiants ont intérêt à se **CONSTRUIRE** des résumés de cours plutôt qu'à les lire simplement ou à adopter les résumés de condisciples (il faudra qu'ils prévoient le temps ad hoc).

Ayant compris que leur mémoire verbale à court terme est limitée (Cf. Miller, chap. 2, section D) ils étudieront dans un lieu où nul autre ne parle.

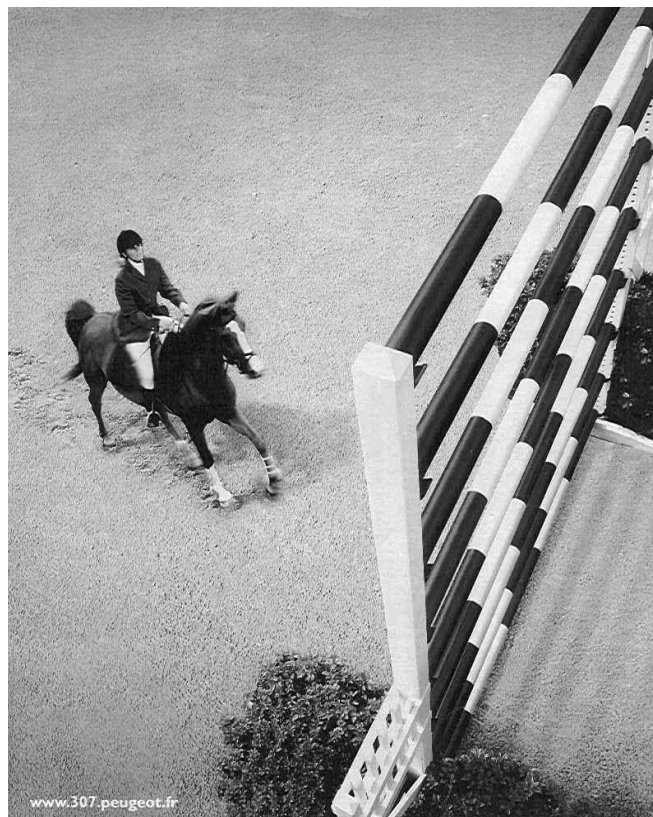
Ayant compris que l'image (Paivio, 1971) et le son (Brooks, 1968 & 1970) peuvent apporter des suppléments à ses capacités mnémoniques verbales, l'étudiant "agrémentera" son texte de dessins, de schémas, de photos. Il ira même jusqu'à enregistrer certaines séquences de raisonnement de narrations ou de descriptions pour les réentendre (en voiture par exemple).

I. Utiliser les Degrés de Certitude dans l'évaluation des performances des étudiants : pourquoi et comment ?

Connaissance partielle²⁶, auto-évaluation et docimologie



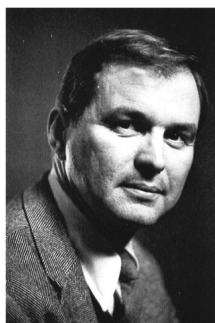
(Quino, 1983, p.107)



²⁶ Cet article est inspiré de celui que les mêmes auteurs ont présenté à l'IPCEM (Institut de Perfectionnement en Communication et Education Médicales) lors de la 13^e Journée de l'IPCEM 2003, **L'évaluation de l'éducation thérapeutique du patient**, Paris, 20 juin 2003.

I1. Apprendre

H1a. « L'Homme est l'espèce qui apprend » écrit David Kolb (1984).



Il est vrai que ce trait distinctif entre l'animal et l'homme est sans doute un des plus fondamentaux. L'espèce doit sûrement sa survie à cette formidable capacité. Il est fort probable que la majorité des 8 façons d'apprendre, des 8 événements d'apprentissage (Leclercq, 1998 ; Poumay, 2001), sont pratiquées depuis des millénaires : l'observation (et imitation de modèles comportementaux), la réception (de messages émis par d'autres humains dans un code, par exemple le langage), la pratique (accompagnée du guidage d'une personne plus compétente, idéalement un expert), l'exploration (ou recherche pour répondre à SES questions), l'expérimentation (ou vérification de SES hypothèses), la création (de SES modèles), la métacognition (le savoir sur ses savoirs) et le débat (la confrontation des points de vue). C'est de la métacognition que nous traiterons ici.

Dans notre « société de la connaissance », on est étonné du peu de considération portée aux sentiments de doute et de certitude accompagnant les réponses de personnes confrontées à des questionnaires de connaissance.

Pourtant, dans ses « Règles pour la conduite de l'esprit (1628) », **Descartes** lui-même insistait sur le doute systématique. C'est parce que nous doutons que nous vérifions (en relisant la notice du médicament), que nous demandons à un expert (le médecin par exemple), que nous prenons des précautions (par exemple agir en présence d'un tiers qui pourrait intervenir). Or nous doutons souvent. C'est même une étape normale de l'apprentissage, étape trop peu entraînée.



H1b. La connaissance partielle

Apprendre une chose est-ce passer de l'ignorance totale à la connaissance parfaite (donc de 0% de connaissance à 100% de connaissance) ?



Bruno **De Finetti** (1965) a montré que c'était plus subtil que cela en introduisant le concept de « **connaissance partielle** ».

L'apprentissage de tous les jours est-ce dès lors évoluer de l'ignorance totale à la connaissance parfaite en passant par divers degrés de « connaissance partielle » ?

Pour De Finetti, la plupart des apprentissages consistent à passer d'un état de connaissance partielle (donc supérieur à 0% car nous avons très souvent des pré-représentations ou des pré-acquis) à un autre état de connaissance partielle (souvent plus élevé mais n'atteignant pas forcément 100%).

Quand nous voulons mesurer les progrès de connaissance dus à l'apprentissage, nous devons donc nous donner les moyens conceptuels et techniques (les instruments de mesure) appropriés au degré de finesse, de granularité de l'objet observé, ici les processus mentaux. Dans ce domaine, les évaluateurs en sont trop longtemps restés à utiliser des moyens trop peu nuancés, un peu comme des biologistes qui travailleraient sans microscope ou des chimistes qui travailleraient à la pelle à charbon.

Tout le texte qui suit est marqué du sceau des phrases de De Finetti (1965, p. 109) :

« *Partial knowledge exists ; to detect it is interesting, necessary and feasible* ».

I2. La connaissance et sa gestion

H2a. La connaissance fausse

De tout temps, les philosophes ont attiré l'attention sur l'importance de la métacognition.

On attribue à **Confucius** (551-479 B.C.) les propos suivants :

Le maître dit : « Yu, veux-tu que je t'enseigne ce qu'est la connaissance ? »

« Quand vous connaissez une chose, c'est d'être conscient que vous la connaissez.

Et quand vous ne connaissez pas une chose, de reconnaître que vous ne la connaissez pas. »

Un **proverbe arabe** dit :

«Celui qui sait et sait qu'il sait est sage ; suis-le.

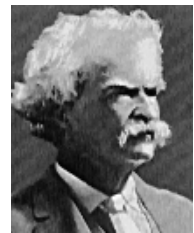
Celui qui ne sait pas et ne sait pas qu'il ne sait pas est un fou ; fuis-le.

Celui qui ne sait pas et sait qu'il ne sait pas est un enfant, enseigne-lui.

Celui qui sait et ne sait pas qu'il sait est endormi ; réveille-le». (Leclercq & Bruno, 1993, p.1).



Plus près de nous, le philosophe anglais **Bertrand Russel** fait la même distinction :
« Le problème, dans notre monde, est que les imbéciles sont sûrs de tout et les sages pleins de doutes. »

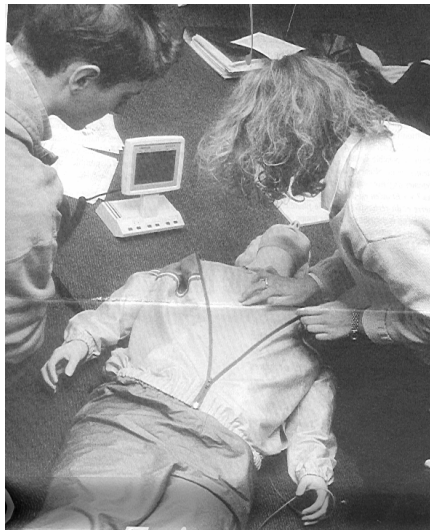


Mark Twain va dans le même sens : *« Ce n'est pas ce que nous ignorons qui nous nuit. C'est ce dont nous sommes sûrs, mais qui est faux. »*

H2b. La gestion de la connaissance est vitale dans de nombreuses professions

Certaines professions plus que d'autres amènent à prendre des risques « calculés ». Par là, on veut dire, entre autres, que l'on a estimé les probabilités avec soin. Ces professions sont celles où les risques vitaux sont les plus évidents : chirurgie, réanimation, aviation, ... C'est à des pilotes et des mécaniciens d'avion que nous avons d'abord demandé d'exprimer leur certitude dans leurs réponses à des tests, avec l'exigence « 100% correct à 100% de certitude ». (Leclercq, 1983).

Nous avons ensuite travaillé en médecine d'urgence. On voit dans le schéma ci-après diverses phases de la RCP (Réanimation Cardio-Pulmonaire) exigeant chacune des connaissances très sûres : connaître le n° d'appel téléphonique des urgences, connaître les conditions nécessaires pour l'application de la RCP (non représenté ci-dessous), connaître les gestes qui sauvent (compression, utilisation du défibrillateur, perfusion, ...)



Des systèmes de simulation tel que celui ci-contre permettent de s'exercer sans risque (on ne peut pratiquer la RCP sur une personne vivante) et de gagner progressivement de la certitude.

Nous recourrons aussi aux degrés de certitude dans l'évaluation des étudiants universitaires. Plus récemment, nous l'avons appliqué aux patients diabétiques (Bruttomesso et al. 2003 ; Leclercq et al., 2003 ; Reach et al., 2005).

I3. La qualité spectrale des performances

La qualité spectrale des performances peut être visualisée POUR CHAQUE QUESTION en temps réel par l'apprenant lui-même au moment de la correction d'une épreuve présentée sous forme d'un questionnaire spectral (section E3 ci-avant). On peut aussi visualiser la distribution spectrale d'un ensemble de réponses grâce à l'ordinateur : en Excel par exemple lors de la correction. Ces visualisations peuvent aider à prendre des décisions de régulation.

H3a. Connaissances utilisables ou non pour l'action



Darwin Hunt (1993) a suggéré que certaines connaissances pourtant correctes sont si peu sûres que l'on ne peut se baser sur elles pour agir.

Il suggère de distinguer les connaissances **utilisables** (correctes et suffisamment sûres), **inutilisables** et les **nuisibles**, voire dangereuses (les connaissances incorrectes dont on est sûr).

En effet, il y a pire que ne rien savoir, c'est être convaincu d'une chose fautive, et ce avec une certitude élevée, si bien que l'on passera à l'action sur base de cette « méconnaissance », avec les conséquences graves que cela peut éventuellement avoir.

Shuford appelle « *misconception* » les connaissances dangereuses.

On verra (exemples en éducation des patients) que l'enseignement peut avoir des effets paradoxaux, en diminuant le taux de réponses incorrectes, bien sûr, mais en augmentant la certitude des incorrectes qui demeurent.

Selon les contenus, et selon la « force » des étudiants, leur tendance à prendre ou non des risques, les distributions spectrales (cf. ci-après) individuelles reflètent des attitudes métacognitives, entraînaibles et modifiables.

H3b. La distribution spectrale de la qualité des réponses

Les réponses des personnes testées avec les 6 degrés de certitude (0, 20, 40, 60, 80, 100%) peuvent être situées le long d'un continuum horizontal en 13 positions allant de -100 % (erreur avec certitude 100 %), soit la pire des performances, à +100 % (réponse correcte avec certitude 100 %), soit la performance idéale, avec au centre « om » (l'omission).

Voici la distribution spectrale de la qualité des réponses de 3905 étudiants entrant dans 8 universités de la Communauté française Wallonie-Belgique, à un test de vocabulaire de 45 questions (soit plus de 175.000 réponses) lors de l'opération MOHICAN²⁷ :

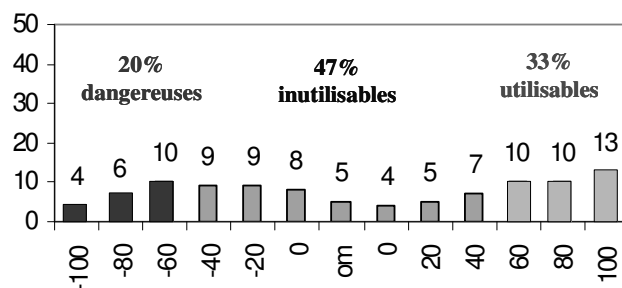
On constate que,

-33 % seulement des connaissances sont **utilisables** (correctes avec plus de 50 % de certitude),

-20 % sont **dangereuses** (incorrectes avec plus de 50 % de certitude)

-47 % sont **inutilisables** (certitude inférieure à 50 %).

Cette dernière catégorie peut être subdivisée en *ignorance reconnue* (26 % c'est-à-dire 9+9+8) et *connaissance peu assurée* (16 % c'est-à-dire 4+5+7).



²⁷ Monitoring Historique des CANDidatures (Leclercq, 2003).

L'omission (5 %) constitue la « tâche aveugle » de tout testing, une « non information », d'où notre invitation aux étudiants à fournir des réponses en les accompagnant de la certitude 0%, reconnaissant par là leur ignorance.

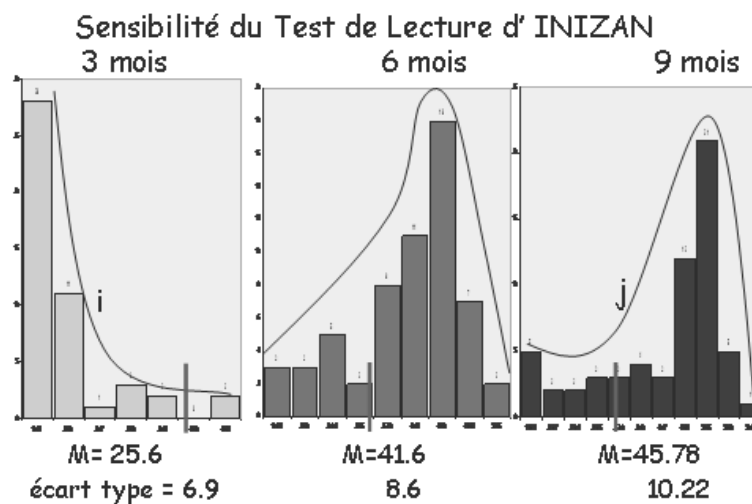
I4. La validité théorique des concepts de certitude et de qualité spectrale

H4a. La distribution recherchée par les éducateurs est une courbe en J

On sait que les psychologues font un usage considérable de la courbe en cloche aussi appelée « Courbe de Gauss », notamment en psychologie différentielle où il importe de mesurer les différences entre personnes (en QI par exemple). La courbe de Gauss exprime des phénomènes distribués aléatoirement comme le sont bien des phénomènes naturels (la taille, le poids, l'intelligence, etc.). Ses propriétés mathématiques sont bien connues des statisticiens et des constructeurs de tests psychologiques.

Certaines personnes se sont imaginé que la distribution idéale des résultats scolaires ou consécutifs à un apprentissage était la courbe de Gauss. Il n'en est rien et G. de Landsheere (1983) a dénoncé cette erreur par l'expression « le dangereux mythe de la courbe de Gauss ».

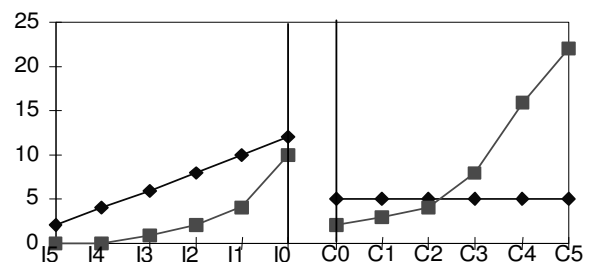
La courbe recherchée par les pédagogues est la courbe en J, indicatrice d'une concentration des résultats vers le plafond de l'excellence.



Le graphique ci-dessus (Boxus, 1971) montre les 3 distributions des scores d'une soixantaine d'élèves de 1^{ère} année primaire ayant passé à trois reprises le même test de lecture (d'Inizan) pour lequel le seuil de réussite est à 38 points (position des fines barres verticales dans les 3 graphiques). On constate que ces élèves progressent régulièrement et que les 3 distributions de leurs scores totaux passent d'une courbe en « i » à une courbe en « j », en passant par une courbe proche de la courbe de Gauss.

Dans une distribution spectrale, la forme de la distribution peut être observée tant pour les réponses correctes (hémispectre droit) que pour les réponses incorrectes (hémispectre gauche).

L'amélioration d'une performance (après apprentissage) ne tient pas seulement à la diminution du nombre de réponses incorrectes, mais aussi à la forme des deux distributions hémispectrales. Dans le graphique ci-dessous, les losanges représentent les valeurs du prétest et les rectangles celles du post-test (plus en J que le prétest). (Jans, 2000).



H4b. Trois modèles mentaux sous-tendant les formules de notation

Quand on utilise des QCM classiques (1 des k solutions proposées est correcte), l'étudiant a des chances de fournir la réponse correcte en répondant au pur hasard. On a, depuis longtemps (Henmon, 1911) imaginé des formules « de correction pour divination » (en anglais « *correction for guessing* ») qui consistent à retirer des points en cas d'erreur, mais pas en cas d'omission. Dans la correction for guessing classique, on retire $1/(k-1)$ point par erreur si l'on a accordé 1 point par réponse correcte (k = nombre de solutions proposées).



Bruce Choppin (1970) a montré que cette formule de correction est inspirée par un modèle que l'on s'est fait de l'activité mentale de l'étudiant en train de répondre à une QCM.

Il définit 3 modèles d'activités mentales. Le **modèle 1** est « Quand l'étudiant sait, il choisit la Réponse Correcte. Quand il ne sait pas, il tape au hasard parmi les k solutions ».

Or, montre Choppin, d'autres modèles sont plus valides. Par exemple le **modèle 2** : « Quand l'étudiant sait, il choisit la solution correcte. Quand il ne sait pas, il commence par éliminer les solutions qu'il sait être incorrectes et tape au hasard dans celles qui restent, ce qui amène à poser les questions

autrement (Leclercq, 1986).

Le **modèle 3**, auquel nous nous rallions, ne fait plus cette distinction entre « Quand il sait » et « quand il ne sait pas », distinction tellement radicale qu'elle est illusoire. Il dit « Quand un étudiant est devant les solutions d'une QCM, il les range par ordre de plausibilité et, si la consigne lui impose de ne fournir qu'une seule réponse, alors il fournit celle dont la probabilité est la plus élevée. C'est sur ce modèle 3 que sont basées les approches recourant à la certitude.

I5. Le remplacement de la *correction for guessing* classique

La *Correction For Guessing* (CFG) classique pourtant universellement répandue comporte, en comparaison avec les degrés de certitude, bien des faiblesses. Les deux points qui suivent tentent de le montrer.

H5a. Les défauts de la CFG

La "rectification classique pour devinette" (traduction de "*correction for guessing*"), consiste à appliquer aux QCM un Tarif positif en cas de Réponse Correcte (TC) et un Tarif négatif en cas de réponse Incorrecte (TI). Si TC vaut + 1 point, TI est calculé en fonction du nombre (k) de solutions (proposées dans la QCM) de manière à ce que l'espérance mathématique d'une réponse au hasard soit nulle. Ce tarif vaut : $-1/(k-1)$, le signe moins indiquant un RETRAIT de points.

Cette procédure doit être abandonnée pour les cinq raisons fondamentales qui suivent :

1. Elle est basée sur un modèle théorique faux (le modèle 1 dans la classification de Choppin ignorant la théorie des décisions)²⁸.
2. Elle est injuste : elle sur-rectifie et sous-rectifie selon les cas, et ce de façon incontrôlable, parce qu'elle est aveugle²⁹.
3. Elle n'est pas informative pour le professeur : il n'apprend rien de plus sur l'étudiant.³⁰

²⁸ (Voir dans Leclercq (1986), les expériences directes sur la *correction for guessing* de Sabers et Feldt (1968, Cross et Frary (1977), décrites dans Leclercq, 1987, 122-130), qui montrent que les étudiants qui omettent ont en fait une connaissance partielle supérieure à l'ignorance totale.

²⁹ Ce qu'ont démontré Davis (1952, Hammerton (1965), Chernoff (1962) dont les travaux sont décrits dans Leclercq, 1987, pp.114, 116 et 134.

³⁰ Ce qu'ont démontré Slakter (1968) et, Ziller (1957), décrits dans Leclercq (1987, pp. 118-119).

4. Elle n'est pas formative pour l'évalué : il n'apprend rien de plus sur son réalisme quant à sa connaissance ; il n'apprend pas à gérer le doute, à la fois moteur dans l'élaboration de la connaissance, respectable moralement et au coeur même de la connaissance partielle (Descartes).

5. Elle est restreinte aux QCM (par définition), ce qui limite son intérêt à un coefficient de correction (dans le sens "rectification") sans généralité.

H5b. Il y a beaucoup mieux que la CFG !

Jusqu'il y a peu, on continuait à appliquer cette correction par résignation entre deux maux (ne pas rectifier ou rectifier), on choisissait le moindre car on avait le sentiment de ne pas disposer d'alternative ! Or on dispose désormais d'une procédure qui consiste à demander à l'étudiant de fournir, en plus de sa réponse, un **degré de certitude** (sur une échelle de probabilités d'exactitude), et à noter cette réponse à l'aide d'un barème de tarifs tenant compte à la fois de la qualité de la réponse et du degré de certitude de l'évalué.

Cette procédure présente les avantages suivants :

1. Elle est basée sur un modèle théorique (le modèle 3 de Choppin conforme à la théorie des décisions et aux probabilités subjectives).

2. Elle est juste car elle récompense chacun en fonction du degré (subtil) de sa connaissance et de sa compétence bref de son mérite.

3. Elle est informative, pour le professeur (et l'étudiant) par des données nouvelles : son réalisme (se sur ou sous-évalue-t-il ?)

4. Elle est formative, car elle fait vivre à l'apprenant la gestion de son doute, de sa connaissance partielle, éléments accompagnant tout processus d'apprentissage.

5. Elle est générale, car elle s'applique autant aux questions ouvertes qu'aux QCM (pour lesquelles elle résout en passant le problème du "guessing").

D. Leclercq y a consacré plusieurs ouvrages, en 1982, 1993 et 2003.

I6. Métacognition et Docimologie

Devons-nous attribuer des points aux étudiants en fonction de leurs degrés de certitude, de la même façon que nous en attribuons en fonction du pourcentage d'exactitude de la réponse ? Nous pensons que oui pour deux raisons. La première est que « seule la connaissance partielle donne un sens au mode de recueil des réponses et au mode de scoring ». Cette paraphrase du postulat de De Finetti est une option épistémologique qui considère que la certitude dans la connaissance EST la connaissance. La seconde raison est que le degré de certitude est un micro jugement relevant du 6^{ème} processus mental de la taxonomie des objectifs cognitifs de Bloom (voir H6a ci-dessous).

H6a. Part du réalisme et de la métacognition dans la notation



Benjamin Bloom (1956) a suggéré une taxonomie d'objectifs cognitifs afin d'encourager les enseignants à varier les objets de questionnement : pas seulement la restitution (ou la reconnaissance) de **mémoire**, ce qu'il appelle

-la **connaissance**

mais aussi de mesurer

-la **compréhension** (par des épreuves à livres ouverts par exemple),

-l'**application** (de la théorie à des cas particuliers),

-l'**analyse** (anomalies, pièges sollicitant la vigilance cognitive),

-la **synthèse** (formulation personnelle, créativité)

-l'**évaluation** (le jugement selon des critères personnels).

Or ce dernier niveau n'est quasiment jamais exercé systématiquement. Le recueil de degrés de certitude est une technique visant à entraîner et évaluer cette capacité d'évaluation, qui ne porte, il est vrai, que sur un aspect particulier (mais important) : la propre connaissance de l'apprenant.

H6b. « Plus » métacognitifs dans les scores

Dieudonné Leclercq et **Marianne Poumay** (photo ci-contre) suggèrent les procédures ci-dessous concernant la métacognition PER (réflexions de l'apprenant pendant la tâche, la performance). Ils calculent les notes des étudiants de façon à faire peser la métacognition (ou autocognition) pour 3 points sur 20, selon le barème ci-après. Il a été convenu avec les étudiants que lorsqu'ils ont envie d'omettre (parce qu'ils ne savent pas) doivent le signifier en donnant la certitude 0. Ils sont invités en outre à répondre quand même « en devinant ». Les recherches de Cross et Frary (1977) ont montré que la connaissance des étudiants sur un sujet leur permet souvent d'obtenir à des questions à choix multiples des scores supérieurs au hasard, même quand ils ont l'impression de ne pas connaître la matière (modèle 2 de Choppin).



Dans les tableaux ci-après,

MAT = Matricule (code donné à l'étudiant)

DC signifie « Degrés de Certitude »,

RC signifie « Réponses Correctes »

RI signifie « Réponses Incorrectes »

Confiance = Certitude Moyenne des RC.

Imprudence = Certitude Moyenne des RI.

C = Point supplémentaire de Confiance.

P = Point supplémentaire de Prudence.

CLASS = score classique (sans DC).

+Méta= score incluant l'utilisation des DC.

H6c. Le barème proposé

(1) Une réponse correcte = + 1 point

Une omission ou une réponse incorrecte = -0,25 point

Ce retrait de point « forfaitaire » (non égal à la formule de la CFG « $-1/k-1$ ») facilite les calculs de la note pour les étudiants. Et ne leur est défavorable que lorsque k est supérieur à 5.

(2) On somme toutes les notes et on exprime le **score classique** sur 20 (par règle de trois).

(3) Si l'étudiant a moins de 20% d'omissions ou de certitudes 0, on calcule les points métacognitifs qu'il peut gagner (par sa confiance et sa prudence). Au-delà de ces taux d'omissions et de certitudes 0, l'étudiant perd son droit aux « **Plus** » **métacognitifs** » qui sont calculés comme suit :

Si la **Confiance** est > à 50 %, + 0,5 point

> à 60%, +1 point

> à 70%, +1,5 point

Si l'**Imprudence** est < à 50 %, + 0,5 point

< à 45%, +1 point

< à 40%, + 1,5 point

Par contre, si l'Imprudence est

> à 50%, - 0,5 point

> à 60 %, - 1 point

> à 70 %, -1,5 point

Cependant, si le total de tous les points métacognitifs est négatif, celui-ci n'entre pas en ligne de compte (seul le score classique est pris en considération).

(4) Si un étudiant obtient TOUTES les réponses correctes, il est évident qu'il ne peut obtenir aucun bonus de prudence (puisque'il n'y a aucune erreur). Il peut gagner :

+ 1 point si sa Confiance >50%,

+ 2 points si sa Confiance > 60%,

+ 3 points si sa Confiance > 70%.

On constate que :

-La métacognition ne fait QUE favoriser les étudiants, chacun d'entre eux restant libre d'utiliser ou non les degrés de certitude. Jusqu'à présent, nous n'avons rencontré que très peu d'étudiants renonçant à ces « plus » potentiels.

-Si des étudiants ont des points supérieurs à 20, c'est cette note qui sera prise en compte dans les additions avec les autres sous-scores d'un même enseignant. S'il n'a pas d'autres sous-scores, la note est plafonnée à 20/20 en délibération.

-Avec un score classique de 8/20, il n'est pas possible d'obtenir 12/20 grâce aux « plus » de métacognition.

Voici les scores d'un étudiant (codé A) :

MAT	Conf	Impr	C	I	Class	+Méta
A	58	39	0,5	1,5	7,9	9,9

Séville en Catalogne ?!

Riga Julie (2006), AESS Communication.

Lors d'un cours d'arts d'expression, je donne une séquence de leçon sur l'œuvre de Gaudi. J'explique aux élèves que ce personnage est originaire de Catalogne. Je demande à mes élèves s'ils savent où se situe la Catalogne. L'un d'entre eux me soutient que la Catalogne se situe dans le sud de l'Espagne. Je rétorque à l'élève qu'il se trompe et que ça se situe dans le nord-est. Celui-ci s'énerve et prétend qu'il connaît sa géographie, qu'il est déjà allé en Catalogne et que par conséquent il ne peut pas se tromper. Cet élève a une confiance aveugle de ce qu'il avance et par-là, n'émet aucune prudence. Cet exemple relate un cas de connaissance nuisible voire dangereuse dont parle Darwin Hunt, c'est-à-dire les connaissances incorrectes dont on est sûr.



I7. Les philosophes et les poètes l'ont dit...

Nous pourrions conclure comme Anthistènes (445-365 B.C.), fondateur du cynisme en philosophie :
“The most useful piece of learning for the uses of life is to unlearn what is untrue.” (cité par Shuford 1993, p.76).

Nous pensons avoir démontré par des exemples pratiques dans divers domaines le bien fondé de la **position épistémologique** (définition de la connaissance) de De Finetti (1965) : « *Seule la probabilité subjective peut donner une signification objective à toute réponse et toute méthode de notation* » (p. 111) et nous ajouterons « ... *et de mesure des gains d'apprentissage.* »

C'est donc le statut lui-même du savoir (et de la sagesse) qui est en question dans les travaux relevant de la métacognition :



T.S. Eliot :

*Where is information lost in data?
Where is knowledge lost in information?
Where is wisdom³¹ lost in knowledge?*

³¹ pour nous, la métacognition fait partie de cette “wisdom”.

J. Quelles qualités peut et devrait avoir une approche métacognitive ?

J1. Les qualités des données métacognitives

Les méthodes et instruments de recueil de données et de traitement et d'interprétation de celles-ci visent à obtenir des mesures ayant des qualités de **validité**, de **fidélité** et de **concordance** sous plusieurs aspects, que nous avons synthétisés sous l'expression ETIC PRAD (Leclercq, 2006).

1. La validité Ecologique porte sur la proximité entre la situation de recueil des données et la vie réelle, la situation professionnelle dans toutes ses dimensions, y compris l'incertitude et les risques pour soi et pour autrui. Poumay et al. (1998) ont décrit les niveaux de multidimensionnalité de l'expérience, en faisant référence au « cône de Dale » (Poumay et al, 1998, pp. 112 et suivantes.).

2. La validité Théorique

a) de construct. Elle porte sur l'argumentation conceptuelle qui fonde les méthodes, les instruments de mesure, etc... de mesure. Rappelons ici deux phrases de De Finetti (1965) :

« *La connaissance partielle existe ; la détecter est intéressant, nécessaire et faisable.* » (p. 109).

« *Seule la probabilité subjective peut donner une signification objective à toute réponse et toute méthode de notation.* » (p.111).

Schraw (1997) a montré une certaine stabilité de la tendance (surestimation, sous-estimation) du réalisme dans l'auto-évaluation chez une même personne à travers des épreuves de nature différente. Nous avons fait la même observation (Leclercq et Detroz, 2003).

b) concurrente : la mesure est-elle bien corrélée avec une autre jugée plus valide ? On en verra une illustration ci-après (section I2).

3. La validité Informatrice ou diagnostique concerne le pouvoir informatif et surtout explicatif du feedback délivré à l'apprenant : quels sont ses points forts et ses points faibles. Elle concerne aussi la précision, la sensibilité, la granularité dans les renseignements fournis (ex : indices de Confiance et d'Imprudence), ainsi que les conseils pour s'améliorer.

4. La validité Conséquentielle porte sur les effets, l'impact des mesures : sont-elles suivies d'effets ou non. On accordera aussi de l'importance à l'ampleur de cet impact, de la même façon que Kirk (1996) souligne l'importance de ce qu'il appelle « *practical significance* » (quand le résultat est utilisable dans le monde réel) qu'il oppose à « *statistical significance* ». **Plutôt que de distinguer le « significatif » du « non-significatif »³², Kirk propose de distinguer « le signifiant » de « l'insignifiant »³³.** Nous reviendrons sur la validité conséquentielle précisément dans le cadre de la formation dans la section I2 ci-après.

5. La validité Prédictive et/ou explicative porte sur les liens (notamment explicatifs) que des mesures entretiennent les unes avec les autres. Ainsi, Leclercq et Boskin (1990) ont montré expérimentalement que plus un apprenant doute de sa réponse, plus il consulte les aides mises à sa disposition. De même, Jans (2000, p.144) a montré expérimentalement que la compétence de départ dans une matière était en relation curvilinéaire avec le gain relatif dans cette matière lors d'une séquence d'apprentissage par un hypermédia. Brutomesso et al (2006) ont montré (chez des patients diabétiques) que les « *misconceptions* » (réponses erronées avec une certitude élevée) étaient très résistantes aux remédiations.

6. La Reproductibilité ou fidélité ou stabilité dans le temps. Exemple : Si la méthode de l'association libre (citer tous les mots qui vous viennent à l'esprit quand on énonce un mot, un concept) est

³² Par cette opposition, Kirk fait référence aux repères statistiques qui sont atteints (indices significatifs) ou non (indices non significatifs).

³³ Par cette opposition, Kirk insiste sur la différence entre « ce qui fait le poids » et « ce qui ne fait pas le poids », qui ne mérite pas d'être pris en considération.

répliquée, sur le même thème après un certain délai, les termes évoqués par le sujet seront-ils les mêmes ? dans le même ordre ? Et les valeurs numériques ? Voilà qui est facile à vérifier.

7. La validité d'Acceptabilité concerne l'adhésion (des professeurs et des élèves) aux pratiques d'évaluation, ainsi que la facilité d'usage. La plupart des traitements que nous proposons peuvent se faire par l'apprenant lui-même, sauf les allo-corrections dont le principe-même est d'être menées (ou orchestrées) par un expert ou par un pair.

8. La validité Déontologique. L'aspect le plus évident est l'équité. Tous les évalués doivent être traités de façon équitable. La transparence des procédures de notation relève aussi de cet aspect.

J2. Validité conséquentielle des indices métacognitifs à l'université

Les étudiants obtiennent généralement plus souvent le point de Confiance que celui de Prudence. Voici les résultats d'une série de 10 étudiants fictifs (nommés de A à J). En 2002³⁴, la moyenne générale des 250 étudiants de 1^o candi universitaire en psycho ULg est passée de 9,3 à 11,3.

matr	Conf	Impr	C	P	Class	+Méta
					9,3	11,3
A	58	38	0,5	1,5	7,9	9,9
B	64	53	1	-0,5	15,3	15,8
C	65	40	1	1	10,5	12,5
D	73	43	1,5	1	8,9	11,4
E	54	47	0,5	0,5	15,3	16,3
F	75	53	1,5	-0,5	13,7	14,7
G	49	8	0	1,5	7,4	8,9
H	64	47	1	0,5	4,7	6,2
I	72	46	1,5	0,5	8,9	11,9
J	75	37	1,5	1,5	10,5	13,5
Etc.						

On peut constater que ce système avantage considérablement les étudiants et n'a aucun inconvénient pour eux. Libre à eux de renoncer à cet avantage en n'utilisant pas les degrés de certitude (nos observations montrent que très peu y renoncent !).

Le bénéfice éducatif et formatif tient en la qualité du feedback que l'on peut donner à l'apprenant, non seulement désormais sur l'exactitude de ses réponses, mais aussi sur sa métacognition. Et ces feedbacks sont assez nuancés puisqu'ils distinguent Confiance, et Prudence deux aspects sur lesquels chaque étudiant peut envisager de s'améliorer.

Enfin, les indices proposés sont faciles à calculer tant par le professeur que par l'apprenant, ce qui devrait faciliter la validité conséquentielle de cette évaluation : qu'elle soit suivie d'effets (par exemple que l'étudiant change sa façon d'étudier, de répondre à des tests, d'agir).

J3. Validité informative ou diagnostique des indices métacognitifs en éducation du patient

I3a. Questionnaire spectral en éducation du patient

Nous avons conçu un questionnaire permettant de positionner (ici sur feuille, mais cela pourrait être sur écran d'ordinateur) les réponses d'une personne de telle sorte que soient visualisés automatiquement ses points forts et ses points faibles.

Nous l'avons appliqué entre autres au questionnement de patients diabétiques sur leurs connaissances de la maladie et de son traitement. Les questions sont des Vrai-Faux (il n'y a donc pas de degré de

³⁴ Le barème de notation de l'époque était légèrement différent de celui que nous proposons ici.

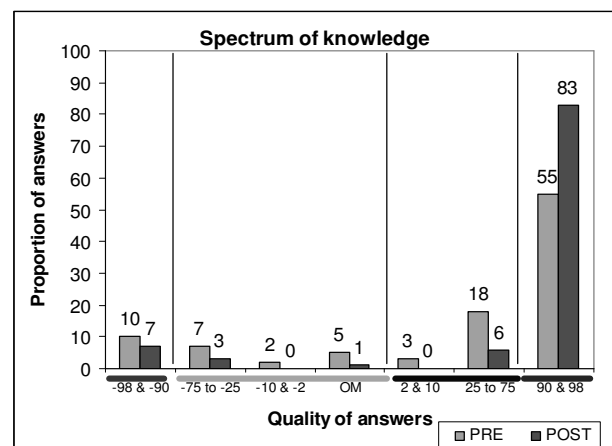
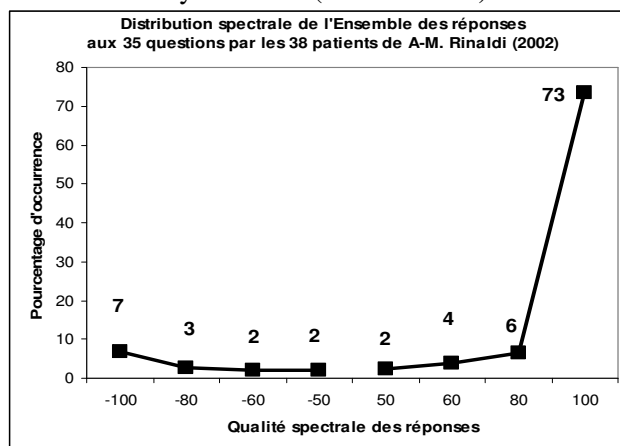
certitude inférieur à 50%). La soignante lit les questions tout haut. Selon la réponse orale du patient, la soignante positionne le degré de certitude à gauche (réponse incorrecte) ou à droite (réponse correcte). Voici le résultat obtenu pour un patient (Leclercq, Rinaldi et Ernould, 2003) :

100%	80%	60%	50%	V	30C	En cas de perte de conscience <u>brutale</u> , il faut que l'entourage fasse <u>immédiatement</u> 2 à 4 unités d'insuline.	F	50%	60%	80%	100%
100%	80%	60%	50%	V	31C	Un effort physique intense ou prolongé peut provoquer une hypoglycémie (hypo).	V	50%	60%	80%	100%
100%	80%	60%	50%	F	32D	En présence de nausées, maux de ventre, vomissements il est utile de rechercher l'acétone dans les urines.	V	50%	60%	80%	100%
100%	80%	60%	50%	F	33E	L'apparition d'une plaie au niveau du pied nécessite une surveillance stricte de la plaie, de l'équilibre glycémique et une visite chez le médecin.	V	50%	60%	80%	100%
100%	80%	60%	50%	F	34E	Le dosage de l'hémoglobine glyquée (HBA1C) donne une idée assez exacte de l'équilibre de votre diabète.	V	50%	60%	80%	100%
100%	80%	60%	50%	V	35C	Lors d'une hypoglycémie (hypo) on peut se « resucrer » avec une boisson light.	F	50%	60%	80%	100%

On peut dresser la distribution spectrale de l'ensemble des réponses d'un groupe. Voici celle des 38 patients aux 35 questions de Mme Rinaldi. Les 7% d'erreurs avec certitude 100% sont inquiétants et il va falloir les identifier pour remédier à ces « *misconceptions* » (idées fausses).

Les deux distributions spectrales (pre-post) du graphique de droite ci-dessous représentent les qualités de réponses au questionnaire de Brutomesso et *al.* (2003) où 39 questions (à choix multiple et non Vrai-Faux) avaient été posées à 38 patients diabétiques à Padoue avant et après une formation. (NB : Les échelles de degrés de certitude différaient légèrement de celles utilisées à Liège).

On constate (par hasard ?) le même taux de *misconceptions* (7%) qu'à Liège (graphique de gauche), la formation les ayant réduit (de 10% à 7%).



J4. Un exemple de non validité concurrente démontrée par confrontation

On parle de validité concurrente si la référence est fiable. Une première possibilité de confrontation des estimations métacognitive est avec les représentations d'autres et notamment des experts. Les allo-représentations que d'autres (par exemple les parents ou les enseignants) ont sur les représentations du sujet peuvent en être très différentes de celles du sujet lui-même. Ainsi, Peeters et Debled (1987)³⁵ ont demandé à 481 élèves (garçons et filles) quels étaient LEURS intérêts pour divers thèmes (dont **l'informatique**). Ils ont obtenus, pour les **filles**, les deux classements de gauche. Ils ont par ailleurs demandé aux parents et aux enseignants quel classement ces mêmes élèves donneraient. Le classement de droite est celui fourni par les enseignants. La flèche oblique montre à quel point les enseignants se sont trompés en ce qui concerne le sujet « informatique » (vu par les filles).

³⁵ Peeters, R. & Debled Ch. (1989) Intérêt et participation des filles lors des activités scolaires liées aux NTI, Bruxelles : CCE et DGOE CFWB, p. 156.

K. Conclusions et perspectives

K1. La métacognition : l'envol, le déclin, le retour

Le concept de métacognition a connu un succès croissant dans la littérature qui a culminé en 1988 selon le relevé de Tochon, 1991 ; cité par Noël et al, 1995, p.48) pour décroître (légèrement) ensuite.

Pour Noël, Romainville et Wolfs (1995), ce déclin est dû à plusieurs ambiguïtés quant aux composantes du concept. Nous n'en reprenons que quelques-unes (en italique), pour illustrer à quelles difficultés le présent article essaye de se confronter.

1. La métacognition porte-t-elle sur des composantes introspectives conscientes ou sur des régulations, éventuellement inconscientes, de l'apprentissage ?

Nous avons déjà répondu aux deux questions posées ici : pour nous la métacognition porte sur les introspections ET sur la régulation et s'intéresse aux mécanismes conscients ET inconscients, ces derniers étant « révélés » par des traces d'actions significatives, comme on le verra plus loin).

2. Quelles différences faut-il faire entre métacognition et des concepts voisins tels que la cognition de la cognition, l'abstraction réfléchissante (chez Piaget), l'internalisation du contrôle social de la cognition (chez Vygostky), la métamémoire (chez Flavell), les stratégies d'auto-questionnement (chez Wolfs, 1991) ? Doit-on mettre dans le même panier description, analyse et jugement métacognitif ?

Pour sortir de ces ambiguïtés, Noël et al (1995) proposent une taxonomie (p. 51), assortie d'un « continuum de degrés d'intervention de la métacognition dans le fonctionnement cognitif d'un apprenant » (pp. 53-55). Nous proposerons nous aussi une taxonomie, assortie de méthodes et d'instruments de mesure de l'auto-évaluation.

3. Comment opérationnaliser le concept ? Comment accéder à des processus internes ? Quelle est la fiabilité des « rapports introspectifs » ? Quels liens entre les processus métacognitifs sollicités (par exemple les Degrés de Certitude accompagnant chaque réponse à un test) et les processus spontanés ? Quels impacts des contenus et des circonstances ? Quelles variabilités inter et intra-individuelles ?

On se rend bien compte qu'il y a là de quoi remplir l'agenda de dizaines d'équipes de chercheurs pendant des années. Notre apport consistera à proposer avec quels concepts et avec quels instruments on devrait s'attaquer à ces questions car rien ne sert d'aller vite, si c'est dans la mauvaise direction. Nos propositions sont souvent le résultat d'explorations expérimentales de pistes diversifiées durant plusieurs décennies (depuis 1971).

Pour en revenir au titre de cette section, nous avons l'espoir que ces précisions assureront un « retour » à l'avant scène de pratiques hautement fécondes dans les processus d'apprentissage et de formation.



Pistes d'intervention

L'évaluation de la métacognition doit être un « plus » pour les étudiants. Si le barème de notation, associé aux degrés de certitude, par exemple débouche sur des scores totaux au test plus faibles que la notation classique, les étudiants vont détester cette méthode de scoring. Peut-on leur jeter la pierre ? Par contre, un barème qui peut favoriser sans jamais défavoriser risque de séduire la majorité des étudiants.

Ce doit être « un plus » en matière diagnostique aussi. Les degrés de certitude, par exemple, doivent indiquer les zones « de doute », d'idées fausses, de connaissances inutilisables.

Un des buts importants de la formation est que les apprenants deviennent des apprenants autonomes. Pour ce faire, ils doivent devenir capables de détecter seuls

- ce qu'ils ont besoin d'apprendre (on leur demandera donc d'estimer les contenus où ce besoin leur paraît le plus urgent),
- avec quels moyens (on leur demandera donc d'estimer ce qu'apporteront en plus certaines ressources comme le dictionnaire, le web, un pair ou un expert, etc.)
- et jusqu'à quand (quand ils ont atteint un niveau suffisant de maîtrise, ce qu'on leur demandera aussi d'estimer).

Bibliographie

- Atkinson, J.W. & Raynor, J.O. (Eds) (1974). *Motivation and achievement*. Washington DC: Winston.
- Ausubel, D. (1968). *Educational psychology, a cognitive view*. New York: Grune & Stratton.
- Baker, J.D. (1965), *The Uncertain Student and the Understanding Computer, La recherche en enseignement programmé, Tendances actuelles*. Paris : Dunod.
- Bandura, A. (1980). *L'apprentissage correctif*. In *L'apprentissage social* (pp. 77-83). Bruxelles : Mardaga.
- Bednar, R.L. et al. (1993). *Self esteem : paradox and innovations in clinical theory and practice*. Washington: American Psychology Association,
- Bereiter, C., & Scardamalia, M. (1989). Intentional Learning as a goal of instruction. ? In Resnick L.B. (Ed.), *Knowing, Learning and Instruction. Essays in honor of Robert Glaser* (pp. 361-392). New Jersey – Hillsdale : Lawrence Erlbaum Associates, Publishers.
- Bernouilli, D. (1954). Exposition of a new theory on the measurement of risk (English translation of « Specimen theoriae novae de mensura sortis », Commentarii academiae scientiarum imperialis Petropolitanae, 1.0 et 1573, 5, pp. 175-192), by Louise SOMMER). *Economica*, 22 23-26.
- Bloom, B.S., Engelhart, M.D., Forst, E.J., Hill, W.H. & Krathwohl, D.R. (1969). *Taxonomie des objectifs pédagogiques (Tome I, Domaine cognitif)*. Montréal : Education Nouvelle, (original 1956, New York : Longman).
- Born, M., Chevalier, V. & Humblet, I. (1997). Resilience, desistance and delinquent career of adolescent offenders. *Journal of Adolescence*, 20, 679-694
- Boud, D. (1995a). *Enhancing Learning through Self Assessment*. London : Kogan Page.
- Boud, D. (1995b). Assessment and learning : contradictory or complementary ? In P. Knight (1995), *Assessment for Learning in Higher Education* (pp. 35-48). London: Kogan.
- Bourgeois, E. & Nizet, J. (1997). *Apprentissage et formation des adultes*. Paris : Presses Universitaires de France.
- Boxus, E. (1971). Une méthode de prédiction du rendement en lecture en première année. *Education - Tribune Libre*, 127, 63-78.
- Brooks, L.R. (1968). Spatial and Verbal Components of the act of recall. *Canadian Journal of Psychology*, 2, 349-368.
- Brooks, L.R. (1970). *Visual and Verbal processes in internal representation*. Communication à l'Institut SALK, La Jolla, Ca.
- Brown, S. & Knight, P. (1994). *Assessing Learners in Higher Education*. London: Kogan Page.
- Bruno, J. (1993). Using Testing to Provide Feedback to Support Instruction : A Reexamination of the Role of Assessment Organizations. In: Leclercq D. & Bruno J. (1993), *Item Banking : Interactive Testing and Self-Assessment*, NATO ASI Series, F 112, Berlin: Springer Verlag, 190-209.
- Bruttomesso D., Leclercq D., Gagnayre R., Crazzolaro D., Busata E., d'Ivernois J-F., Casiglia E., Tiengo A., & Baritussio A. (2003). Confidence degrees to evaluate knowledge in patients with Type 1 diabetes. *Patient Education and Counseling*.
- Bruttomesso, D., Costa, S., Dal Ps, M. Crazzola, D., Realdi, G.; Tiengo, A., Britussio, A. & Gagnayre, R. (2006). Educating diabetic patients about insulin use: changes over time in certainty and correctness of knowledge. *Diabetes Metabolism*, 32, 256-261.
- Carver, R.P. (1974). Two dimensions of tests : Psychometric and edumetric. *American Psychologist*, 29, 512-518.
- Cennamo, K.S., Savenye, W.C., & Smith, P.L. (1991). Mental effort and video-based learning : the relationship of preconceptions and the effects of interactive and covert practice. *ETR&D*, 39(1), 5-16.
- Cennamo, K.S. (1993), Learning from video: factors influencing learners preconceptions and invested mental effort. *Educational Technology Research and Development*, 41(3), 33-45.
- Chi, M. T. H., Bassock, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13, 145 - 182.
- Chi, M., T.H. (1989). Learning from examples via self-explanations. In Resnick L.B. (Ed.), *Knowing, Learning and Instruction. Essays in honor of Robert Glaser* (pp. 251-282). New Jersey – Hillsdale: Lawrence Erlbaum Associates, Publishers.

- Choppin, B. (1970), *An IEA Study of Guessing. A Proposal*, Stockholm : International Association for the Evaluation of Educational Achievement, unpublished memorandum, IEA/TR/9.
- Cohen, J. (1963). *Hasard, adresse et chance*. Paris : PUF
- Cross, L. & Frary, (1977), An empirical test of Lord's theoretical results regarding formula scoring of multiple choice tests, *Journal of Educational Measurement*, vol. 14, 313-321.
- Damasio, A. R. (2001). *L'erreur de Descartes*. Paris : Odile Jacob.
- De Finetti, B. (1965). Methods of discriminating levels of partial knowledge concerning a test item. *British Journal of Math. & Statist. Psychol.*, 18, 87-123.
- de Landsheere, G. (1982). *Introduction à la recherche en éducation* (5^e Ed.). Liège : Thone.
- Décret définissant les missions prioritaires de l'enseignement fondamental et de l'enseignement secondaire et organisant les structures propres à les atteindre. D.24-07-1997 – M.B. 23.09.1997. Communauté française de Belgique.
- Descartes, R. (1636). *Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences*. Paris : Plon (1945).
- D'Hainaut, L. et Michez, R. (1979). Une méthode récurrente pour enseigner la résolution de problèmes. *Education - Tribune Libre*, 173, 1-19.
- Dirkzwager, A. (1993), A Computer Environment to Develop Valid and Realistic Predictions and Self-Assessment of Knowledge with Personal Probabilities. In: Leclercq D. & Bruno J. (1993), *Item Banking : Interactive Testing and Self-Assessment*, NATO ASI Series, F 112. Berlin: Springer Verlag, 146-166.
- Ebel, R.L. (1965). Confidence Weighting and test reliability. *Journal of Educational Measurement*, 49-57
- Entwistle, N. & Marton, F. (1984). Changing Conceptions of Learning and Research. In F. Marton, D. Hounsell & N. Entwistle (Eds), *The Experience of Learning* (pp. 211-228). Edinburgh: Scottish Academic Press.
- Ericsson, K. & Simon, H.A. (1984). *Protocol analysis : verbal reports as data*. Cambridge, MA : MIT Press.
- Festinger, L. (1957). *A theory of Cognitive Dissonance*. Evanston, Illinois: Row and Peterson.
- Flavell, J. (1976). Metacognitive aspects of problem solving. In Resnick, L. (Ed.) *The Nature of Intelligence*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Flavell, J. H. (1982). Structures, stages, and sequences in cognitive development. In W. A. Collins (Ed.), *The concept of Development. Minnesota symposia on child psychology* (Vol. 15). Hillsdale, N.J.: Lawrence Erlbaum Associates Inc.
- Georges, F. (2003). *EAD et suivi des étudiants*. ULg-LabSET : Journée Formadis.
- Gombert, J.E. (1990). Le développement étalinguistique. Paris : PUF.
- Hunt, D. (1993). Human self-assessment : theory and application to learning and testing. In D. Leclercq & J. Bruno. *Item banking : Interactive testing and self assessment* (pp. 177-189). NATO ASI Series, F112, Berlin: Springer verlag.
- Henmon, V.A.C. (1911). The relation of the time of a judgment to its accuracy. *Psychological Review*, 18, 186-201.
- Jacquard, A. (1982). *Au péril de la science ? Interrogation d'un généticien*. Paris : Seuil
- James, W. (1890). *The principles of psychology*. New York: Holt.
- Jans, V. (1994). *Développer l'auto-estimation et la vigilance cognitive à l'école primaire*. Mémoire de licence en Sciences et techniques de la formation continue non publié, Université de Liège, Liège, Belgique.
- Jans, V. et al. (1998). Projets d'Animations Réciproques Mutimédias (PARM). In D. Leclercq (Ed.), *Pour une pédagogie universitaire de qualité* (pp. 207-241). Liège : Mardaga.
- Jans, V. (2000). *Confrontations instrumentées et dialectiques des jugements auto/alloévaluatifs*. Thèse de doctorat non publiée, Université de Liège, Liège, Belgique.
- Kagan, J. (1965c). Impulsive and reflective children : significance of conceptual tempo. In Krumboltz (Ed.), *Learning and the Educational Process* (pp. 133-161). Chicago, Rand Mc Nally.
- Kagan, J., Rosman, B.L., Day, D., Albert, J., & Philips, W. (1964). Information processing in the child : significance of analytic and reflective attitudes. *Psychological Monographs*, 78(1), Whole n° 587.

- Kirk, R.E. (1997). Practical significance : a concept whose time has come. *Educational and Psychological Measurement*, 56 (5), 746-759.
- Kolb, D. (1984), *Experiential Learning : Experience as the source of Learning and Development*. Englewood Cliffs: Prentice Hall Inc.
- Koriat, A. (1998). Illusions of knowing: The link between Knowledge and Metaknowledge. In Yzerbyt, Lories & Dardenne, *Metacognition* (pp. 16-34). London: SAGE.
- Leclercq D. (1982). Confidence marking. Its use in Testing. In Postlethwaite & Choppin (Eds) , *Evaluation in Education* (6(2), 161-287). Oxford: Pergamon.
- Leclercq, D. (1986). *La conception des QCM*. Bruxelles : Labor.
- Leclercq, D. (1987). *Qualité des Questions et Signification des Scores*. Bruxelles : Labor
- Leclercq D. (1993). Validity, Reliability and Acuity of Self-Assessment in Educational Testing. In Leclercq D., & Bruno J., *Item Banking : Interactive Testing and Self-Assessmen* (pp. 114-131).. NATO ASI Series, F 112, Berlin: Springer Verlag.
- Leclercq, D. (1998). *Pour une pédagogie universitaire de qualité*, Sprimont : Mardaga.
- Leclercq (1999). *Edumétrie et Docimologie*. Liège : STE- Université de Liège.
- Leclercq, D. (2003). *Un diagnostic cognitif et métacognitif au seuil de l'université. Le projet MOHICAN mené par les 9 universités de la Communauté Française Wallonie Bruxelles*. Liège : Editions de l'université de Liège.
- Leclercq, D. (2006). L'évolution des QCM. In G. Figari, & L. Mottier Lopez, *Recherches sur l'évaluation en Education* (pp. 139-146). Paris : L'Harmattan.
- Leclercq D. & Boskin A.(1990). Note taking behavior studied with the help of hypermedia. In Estes, Heene & Leclercq (Eds), *Proceedings of the 7th ICTE* (vol 2, pp. 16-19). Edimburgh: CEP Consultants.
- Leclercq, D., & Detroz, P. (2003). Liens entre caractéristiques de départ (dont les résultats aux check-up) et les réussites en première candidature. In in D. Leclercq (Ed.) *Diagnostic cognitif et métacognitif au seuil de l'université* (pp. 127-154). Liège : Editions de l'ULg.
- Leclercq, D et Poumay, M. (2003) « La connaissance partielle chez l'apprenant : pourquoi et comment la mesurer ». In Gagnayre et al. *Actes de la 13^e Journée de l'IPCCEM (Institut de Perfectionnement en Communication et Education Médicales)* « L'évaluation de l'éducation thérapeutique du patient (pp. 27-30).
- Leclercq, D., & Poumay, M. (2005). *The 8 Learning Events Model and its principles. Release 2005-1*. Retrieved April 29, 2005, from <http://www.labset.net/media/prod/8LEM.pdf>.
- Leclercq D., Rinaldi A.M. et Ernould C. (2003). Un questionnaire spectral pour l'évaluation des connaissances chez le patient diabétique. In Gagnayre et al. (Eds), *L'évaluation de l'Education Thérapeutique du Patient*. Paris : IPCCEM.
- Magain, C. (1995). Création et expérimentation d'une banque de questions Double Check, mémoire de licence en sciences de l'éducation non publié, Université de Liège, Liège, Belgique.
- Markman, E. (1977). Realizing that you don't understand : A preliminary investigation. *Child Development*, 450, 643-655.
- Messer, S.B. (1970a). The effect of anxiety over intellectual performance on reflection-impulsivity in children. *Child Development*, 41, 723-735.
- Messer, S.B. (1970b). Reflection-impulsivity : stability and school failure. *Journal of Educational Psychology*, 61, 6, 487-490.
- Miller G.A. (1956). The Magical Number Seven, plus or minus two. *Psychological review*, vol.63, 81-97.
- Nhouyvanisvong, A., & Reder, L.M. (1998). Rapid feeling-of-knowing : A strategy selection mechanism. In V.Y. Yzerbyt, G. Lories, & B. Dardenne (Eds.), *Metacognition : Cognitive and social dimension* (pp. 35-52). London: Sage.
- Noel, B. (1991). *La métacognition*. Bruxelles : De Boeck.
- Noel, B., Romainville, M. & Wolfs, J.L. (1995). La métacognition : facettes et pertinence du concept en éducation. *Revue Française de Pédagogie*, n°112, 47-56.
- Paivio, A. (1971). Imagery and language. In S.J. SEGAL (Ed.), *Imagery : Current Cognitive Approaches*. New York: Academic Press.

- Parmentier, P. & Romainville, M. (1998). Les manières d'apprendre à l'université. In M. Frenay et al., *L'étudiant-apprenant. Grilles de lecture pour l'enseignant universitaire* (chap. 4, pp. 63-80). Bruxelles : De Boeck & Larcier,.
- Pask, G., (1976b). Styles and strategies of learning. *British Journal of Educational Psychology*, 46, 128-148.
- Pask, G., (1976c). *Conversation theory : Applications in Education and Epistemology*. Amsterdam : Elsevier.
- Peeters, R. et Debled, CH. (1989). *Intérêt et participation des filles lors des activités scolaires liées aux nouvelles technologies de l'information* (271 pages). Rapport de la recherche : Action "Egalité des Chances et NTI" de la CCE et de l'Organisation des Etudes de la Communauté Française de Belgique.
- Pitz, G. (1974), Subjective probability distribution for imperfectly known quantities, in L. W. Gregg (Ed.), *Knowledge and Cognition*, New York : Wiley.
- Poumay M. (2001). *L'utilisation de cas concrets en pédagogie. Modèles pour décrire et analyser des cas et leurs usages didactiques*. Mémoire de DES en Technologie de l'Education et de la Formation non publié, Universités de Liège (ULG) et de Namur (FUNDP).
- Quino (1993). *Gente en su sitio* (3^a édition). Barcelona : Edicion Lumen,.
- Reach, G., Zerrouki A., Leclercq, D., & d'Ivernois, JF. (2005). Adjusting insulin doses : from knowledge to decision. *Patient Education and Counseling*, 56, 98-103.
- Reder, L.M. (1987). Strategy selection in question answering. *Cognitive Psychology*, 19, 111-138.
- Riding, R.J. & Taylor, E.M. (1976). Imagery performance and prose comprehension in seven-year-old children. *Educational Studies*, 2, pp. 21-27.
- Rogers, C. (1973). *Liberté pour apprendre*. (original 1969; traduction française par Le Bon). Paris : Dunod..
- Rotter, J. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 80 (1, whole, 609).
- Salomon, G. (1974). Internalization of filmic operations in relation to individual differences. *Journal of Educational Psychology*, 66, 4, 499-511.
- Salomon, G. (1983). The differential investment of mental effort in learning from different sources. *Educational Psychologist*, 18, 42-50.
- Schön, D. A. (1983). *The reflexive Practitioner – How professionals think in action*. U.S.A: Basic Books Inc.
- Schön, D. (1994). *Le praticien réflexif: A la recherche du savoir caché dans l'agir professionnel*. Montréal: Editions Logiques.
- Schraw, G. (1997). The effect of generalized Metacognitive Knowledge on Test Performance and Confidence Judgements. *Journal of Experimental Education*, 65 (2), 135-146.
- Sdorow, L.M. (1995). *Psychology* (3rd ed.). Dubuque: Wm C. Brown Communication, Inc.
- Shuford, E.H. (1993). In Pursuit of the Fallacy : Resurrecting the Penalty. In D. Leclercq & J. Bruno (Eds), *Item Banking : Interactive Testing and Self-Assessment* (pp. 76-98). NATO ASI Series, Berlin: Springer Verlag.
- Stein, B.S. & Bransford, J.D. (1979). Constraints on effective elaboration : Effects of precision and subjects generation. *Journal of Verbal Learning and Verbal Behavior*, 18, 769-777.
- Tardif, J. (1992). *Pour un enseignement stratégique – l'apport de la psychologie cognitive*. Montréal: Les Editions Logiques inc.
- Tochon, F.V. (1991). *L'enseignement stratégique*. Toulouse : EUS.
- Van Lenthe, J. (1993). The Development and Evaluation of ELI, an Interactive Elicitation Technique for Subjective Probability Distributions. In D. Leclercq & J. Bruno (Eds), *Item Banking : Interactive Testing and Self-Assessment, NATO ASI Series* (pp. 132-145). Berlin, Springer Verlag.
- Van Naerssen, R.F (1962), A scale for the measurement of subjective probability. *Acta Psychologica*, 20, 2, 159-166.
- Van Rilaer (1992). *La gestion de soi*. Liège : Mardaga.
- Wang, M.C., Haertel, G.D. & Walberg, H.J. (1990). What influences learning ? A content analysis of review literature. *Journal of Educational Research*, 84 (1), 30-44.

- Witkin, H.A. (1950). Individual differences in ease of perception of embedded figures. *Journal of Personality*, 19, 1-15.
- Witkin, H.A. (1959). The perception of the upright. *Scientific American*, 700, 50-56.
- Yando, R.M. et Kagan, J. (1968), The effect of teacher tempo on the child. *Child Development*, 39, 27-34.