

Une définition opérationnelle de la Métacognition et ses mises en œuvre

D. Leclercq (1) et M. Poumay (2)

(1) Université de Liège, Directeur du Service de Technologie de l'Education – ULg (STE-ULg) et responsable académique du LabSET
Boulevard du Rectorat, 5, Bât. B32
B – 4000 Liège
Tél. ++ 00 32 (0)4 366 20 72
Fax ++ 00 32 (0)4 366 29 53
D.Leclercq@ulg.ac.be

(2) Université de Liège, Directrice du LabSET-ULg (Laboratoire de Soutien à l'Enseignement Télématique)
Boulevard de Colonster, 2, Bât. B9
B – 4000 Liège
Tél. ++ 00 32 (0)4 366 20 79
Fax ++ 00 32 (0)4 366 29 53
m.poumay@ulg.ac.be

A. Une étonnante absence...qui s'explique

Alors qu'au terme d'une méta-analyse¹ 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 Tardiff (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 méta-cognition. 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.

Cause 1 : Un événement d'apprentissage non généralisé

¹ 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 ».

Contrairement aux autres « Evénements d'Apprentissage »² tels que la modélisation – observation ou la transmission – réception, etc., la métacognition est d'apparition très récente dans les mécanismes généralisés d'apprentissage. Nous disons « généralisés » car il est évident que des penseurs comme Socrate (Connais-toi toi-même), Descartes (Je pense donc je suis – Discours de la Méthode), Pascal (Théorie des décisions), Bernouilli (Théorie de l'utilité), Damasio (L'erreur de Descartes – Le sentiment même de soi) et bien des philosophes, anciens et modernes non seulement ont eu recours à la métacognition, mais en ont proposé des analyses et des théories.

Néanmoins, l'usage qui en a été fait 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 EPA, 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).

Cause 2 : Un flou conceptuel

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'intentionnalité³, déjà présentes mais moins explicitement chez Flavell :

² Nous avons décrit ailleurs (Leclercq, 2003) 8 « Evénements d'apprentissage » ou « Events of learning » Leclercq & Poumay (2003), FORMASUP – version anglaise. Université de Liège,

³ Nous avons mis en gras ces deux mots dans la définition de Gombert (1990, 27) : « Domaine qui regroupe (1) les connaissances introspectives conscientes qu'un individu particulier a de ses propres états et processus D. Leclercq (2003) – Méthodes de Formation et Théories de l'Apprentissage. Evénements d'apprentissage. Editions de l'Université de Liège - Chap 7 - D. Leclercq & M. Poumay (2003) - La Métacognition - p. 2/40

Flavell (1976, 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.* »

De nombreux chercheurs (Brown & Knight, 1994, Noël, 1991 ; Noël, Romainville et Wolfs, 1995) reconnaissent le manque d'opérativité de telles définitions. Nous essayons ici de les pallier.

Pour nous, la **métacognition** est

« L'ensemble des jugements, des analyses, des régulations, conscientes ou non (mais qu'il importe de rendre explicites, observables et conscientes) effectuées par l'apprenant sur ses propres performances (processus ou produits) , dans des situations de PRE, PER ou POST performance. »

Nous tenterons, dans la section B ci-dessous de montrer la fécondité potentielle d'une telle définition et en quoi elle peut être dite « opérationnelle ».

Cause 3 : Un manque d'instrumentation

La psychologie scientifique, via les approches réflexologiques, associationnistes, constructivistes, cognitivistes, instrumentalistes, 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 » dans son article « *Two dimensions of tests : Psychometric and edumetric* », on ne peut que constater que c'est la psychométrie qui a le plus (encore) développé les concepts et instruments typiques de sa métrique. C'est le cas pour le modèle de Rasch, pour la théorie des traits latents, ... destinés au testing pédagogique, mais procédant de techniques

largement inspirées de la psychométrie, entre autres d'indices de discrimination comme la corrélation point bisériale, la pente des Item Characteristic Curves, fondés sur des repères normatifs.

Le présent article renforcera et développera des pistes typiquement éduométriques fondées sur la connaissance partielle, son auto-évaluation via des degrés de certitude, via le calcul de scores de réalisme individuels et diagnostiques, à haute validité conséquentielle, via la représentation des connaissances au moyen d'une distribution spectrale dont l'idéal est la courbe en J et non la courbe de Gauss.

B. Les caractéristiques de la métacognition

1. Trois moments : nous distinguons trois situations d'activité métacognitive : **avant (PRE), pendant (PER) et après (POST)** une performance, une activité, une tâche, un épisode⁴ d'évaluation ou d'apprentissage repérable dans le temps.

2. Deux objets : La métacognition peut porter sur les **produits** (les résultats) et sur les **processus** (les démarches).

3. Des données conscientes, subjectives⁵ et affectives.

Lors d'une activité, la métacognition peut⁶ 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**.

Nous nous baserons sur l'**introspection** métacognitive et plus précisément encore sur ce que nous pouvons en connaître via **son expression par l'apprenant**, avec ce que cela comporte de subjectif, d'imbrication de l'affectif dans le cognitif. Cela justifierait de parler de **Pré-diction**, de **Per-diction** et de **Post-diction**.

Enfin, 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.

4. Des exigences de validité, fidélité et diagnosticité. En principe la métacognition fait abondamment appel à l'auto-évaluation (subjective) et à l'auto-observation, sans

⁴ Un **épisode d'évaluation** peut être un test, un examen, une situation de résolution de problème dont on peut identifier le début et la fin (dans le temps). Outre les épisodes, qui sont source d'apprentissage, un **épisode d'apprentissage** intentionnel⁴ peut être l'écoute d'un message sonore, la lecture d'un écrit, l'assistance à un cours multimédias, une discussion (cf. notre modèle des 8 événements d'apprentissage) et leurs combinaisons (scénarios de formation par exemple).

⁵ volontairement

⁶ Nous disons « peut » car, comme nous l'avons déjà dit, la métacognition consciente et systématique est trop peu développée, tout spécialement dans ces épisodes de vie.

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 spontanées, suggérées ou imposées

- plus valides (théoriquement fondées),
- plus fidèles (permettant une étude objective de la subjectivité) et surtout
- plus diagnostiques et suivies d'effet (offrant une validité conséquente),

5. De multiples facettes. L'importante composante « auto-évaluation » de la métacognition peut être caractérisée, comme les autres évaluations, par 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 »).

6. Trois types d'opérations portant sur les performances :

- les **jugements** sont des évaluations par l'apprenant lui-même du degré de qualité ou du degré de certitude 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.

- les **analyses** sont les justifications et explications (les pourquoi), c'est-à-dire 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).

- 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* ».

7. Toutes ces considérations déterminent le champ suivant des situations possibles :

	processus		
produit	PRE	PER	POST
Jugement			
Analyse			
Régulation			

Voici quatre exemples qui montrent en quoi notre définition se veut opérationnelle :
elle permet de classer sans ambiguïté des situations différentes (encadrées) :

Jugements, analyses et/ou régulations
observables ou non observable
effectués par l'apprenant sur ses propres performances
(processus ou produits d'apprentissages),
dans des situations de **PRE, PER ou POST** performance.

Il trouva que son raisonnement avait été insatisfaisant et que c'était la raison pour laquelle il s'était trompé.

Jugements, analyses et/ou régulations
observables ou non observable
effectués par l'apprenant sur ses propres performances
(processus ou produits d'apprentissages),
dans des situations de **PRE, PER ou POST** performance.

L'étudiant donne un degré de certitude à chaque réponse d'un test de 20 Questions à Réponse Ouverte Courte.

Jugements, analyses et/ou régulations
observables ou non observable
effectués par l'apprenant sur ses propres performances
(processus ou produits d'apprentissages),
dans des situations de PRE, PER ou POST performance.

L'étudiant accompagne sa réponse d'une justification écrite puis remet sa copie.

Jugements, analyses et/ou régulations
observables ou non observable
effectués par l'apprenant sur ses propres performances
(processus ou produits d'apprentissages),
dans des situations de PRE, PER ou POST performance.

En août, avant d'entrer en 2^o candi psycho, l'étudiante dit à sa copine qu'elle décide de travailler dans le but d'avoir une satisfaction, pas une distinction.

Le présent article vise à rendre compte de travaux portant sur plusieurs de ces zones, pas toutes. Nous partirons de la zone « jugements PER », traitée par une technique bien précise : les degrés de certitude. De là, nous étendrons nos préoccupations à d'autres zones, d'analyse et de régulation POST.

C. Les Degrés de certitude ou une forme de la métacognition en situation PER

C1. Précisions conceptuelles

La «situation» dans laquelle l'apprenant émet des jugements PER 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, Prudence et Nuance permettant un feedback facilitant l'évaluation des PER-dictions. Il s'agit ici 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 test⁷ ou de résolution de problème ou encore d'apprentissage⁸.

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 ?).

C2. Les liens entre métacognition et difficulté de la tâche

a) Données objectives indicatrices de la difficulté de la tâche.

Elles sont possibles 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

⁷ 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.

⁸ Pour Bereiter et Scardamalia, tout apprentissage est une résolution de problème.

D. Leclercq (2003) – Méthodes de Formation et Théories de l'Apprentissage. Evénements d'apprentissage. Editions de l'Université de Liège - Chap 7 - D. Leclercq & M. Pourmay (2003) - La Métacognition - p. 10/40

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.). Schraw (1997) a observé, chez des étudiants universitaires, que la corrélation entre les certitudes moyennes d'un test à l'autre était de 0,43 dans ses échantillons observés. Nous amènerons nous aussi, (en section x ci-dessous) mais par une autre approche mathématique, des données sur ce sujet.

b) Les types de questions

Les jugements sur la tâche et sur la consigne peuvent influencer les jugements sur sa propre performance. Par exemple, Noël (1991, 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, 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 métacompréhension, 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). 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-RCExp), les QCM-RC « Aucune » nécessitent un délai de réponse plus grand et les QCM-RC « Absurdité » un délai plus court. Les QCM-SGI ont été moins bien réussies que les QCM-RCExp (66% en moyenne), tout spécialement les QCM-RC « manque » (28% en moyenne), alors que leur indice de discrimination moyen (r_{pbis}) a été le même (0,29). Les QCM-RC « Absurdité », aussi bien réussies (66%) que les QCM-RCExp, ont eu un indice de discrimination plus élevé (0,42). Sur la nécessité d'entraîner les élèves à se rendre compte qu'une question demande des précisions, citons Cohen (1963,38) :

« 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. »

c) Le type d'aides demandées lors de la performance

Souvent, un coup de pouce est nécessaire. Une des façons de procéder est d'autoriser l'étudiant à consulter une référence (un livre, des notes, un site web) après avoir répondu une première fois. 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 améliorerait 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.

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

- pour les QCM-RCExpl: 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.

d) Le type d'aides offertes lors de la performance

Magain (1995) et Leclercq (1999, 4,23) 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-RCExpl 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. On rencontre ici le souci mentionné par Noël (1991, 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. »*

C3. Les Jugements PER sur ses propres performances

a) 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, 163).

b) Certitude, Connaissance et Apprentissage

« L'Homme est l'espèce qui apprend » disait 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), le débat (la confrontation des points de vue), la métacognition (le savoir sur ses savoirs). C'est de cette toute dernière façon d'apprendre que nous traiterons ici.

Au siècle « 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.

Mais qu'est-ce qu'apprendre une chose ? Evoluer de l'ignorance totale à la connaissance parfaite en passant par divers degrés de « connaissance partielle » ? C'est un peu plus subtil que cela. De Finetti (1965) a montré que la plupart des apprentissages ne consistent pas à passer de l'ignorance totale à la connaissance parfaite (donc de 0% de connaissance à 100% de connaissance), mais d'un état de connaissance partielle (donc supérieur à 0%) à un autre état plus élevé de connaissance partielle. Quand nous voulons mesurer les progrès de connaissance dûs 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.

La gestion de la connaissance est vitale dans des professions.

C'est le cas pour les pilotes d'avion, les mécaniciens d'avion, les médecins urgentistes, etc. C'est auprès ces publics que nous avons d'abord recueilli les degrés de certitude accompagnant les réponses à des questions sur leurs disciplines.

Depuis 1971, nous l'appliquons aux étudiants universitaires, et depuis deux ans aux patients diabétiques (Brutomesso et al. 2003).

La méthode de recueil de la certitude doit suivre certaines règles.

Ainsi, il faut éviter de se limiter à une consigne verbale telle que « *Etes-vous peu sûr, moyennement sûr, très sûr, extrêmement sûr ?* » Il faut absolument recourir à l'expression de probabilités ou pourcentages de chances d'avoir fourni la réponse correcte en permettant de choisir entre l'une des 6 possibilités suivantes :

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

Nous nous sommes limités à cette consigne (on peut en concevoir des centaines d'autres) après avoir étudié (Leclercq, 1983, 1993) la capacité humaine à discriminer de façon fiable des degrés de certitude entre 0 et 100 %. Nous avons constaté que le nombre de zones différentes ne dépasse pas 7. Nous retrouvons donc 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 Miller : « *The magical number seven, plus or minus two* ». Notre consigne, appelée « 6 multiples de 20 % » tient compte de cette limitation.

C4. Les niveaux spectraux de qualité de la connaissance

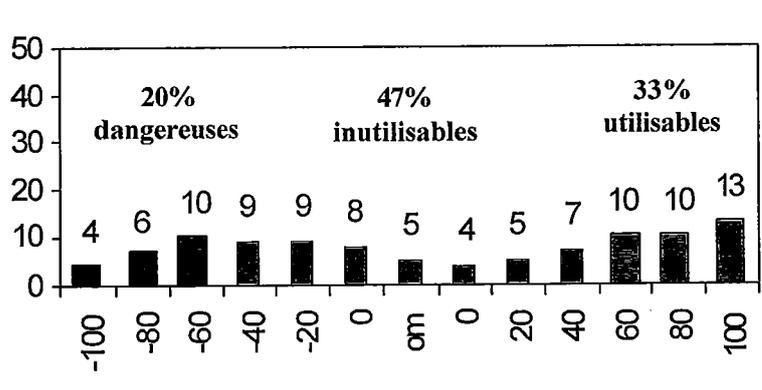
Toute connaissance est-elle utilisable pour l'action ? Non : il est des connaissances pourtant correctes mais si peu sûres que l'on ne peut se baser sur elles pour agir (Hunt, 1993). Il existe même des connaissances nuisibles, voire dangereuses : les connaissances incorrectes dont on est sûr. Il y a en effet 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. C'est ce que déplorent le philosophe anglais Bertrand Russell : « *Le problème, dans notre monde, est que les imbéciles sont sûrs de tout et les sages pleins de doutes.* » et l'humoriste américain Mark Twain : « *Ce n'est pas ce que nous ignorons qui nous nuit. C'est ce dont nous sommes sûrs, mais qui est faux.* ». On leur attribue d'ailleurs un illustre prédécesseur, Confucius (551-479 B.C., à qui l'on attribue les termes 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. »*

**La qualité des réponses s'exprime par leur position sur un « Spectre des
qualités ».**

Les réponses des personnes testées avec ces degrés de certitude 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. 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 MOHICAN⁹ :



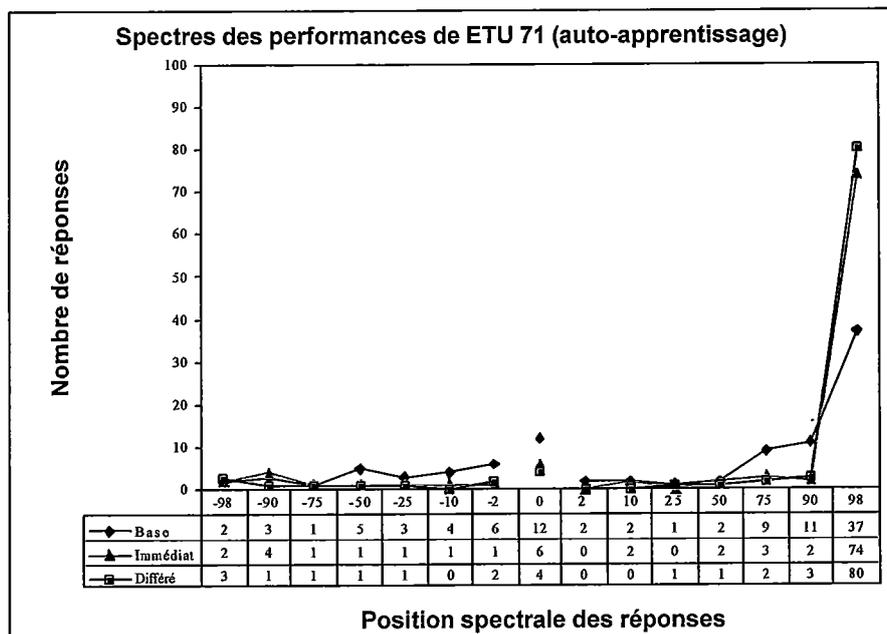
On constate que, selon les expressions inspirées de Hunt (1993), 33 % seulement
des connaissances sont utilisables (correctes avec plus de 50 % de certitude), que
20 % sont dangereuses (incorrectes avec plus de 50 % de certitude) et que 47 %
sont inutilisables (certitude inférieure à 50 %). Cette dernière catégorie peut être
subdivisée en ignorance reconnue (26 %) et connaissance peu assurée (16 %).
L'omission (5 %) constitue la « tâche aveugle » de tout testing, une « non
information ».

**C5. La distribution recherchée par les éducateurs est une courbe en J tant pour
les réponses correctes (hémispectre droit) que pour les réponses incorrectes**

⁹ Monitoring Historique des CANDidatures (Leclercq, 2003).

(hémispectre gauche). Dans d'autres professions (par exemple en psychologie différentielle), on cherche plutôt à obtenir une courbe de Gauss, en cloche). L'amélioration d'une performance (après apprentissage) ne tient donc pas seulement dans la diminution des réponses incorrectes au profit de réponses correctes. A nombre de réponses incorrectes égal entre un pré et un post-test, l'amélioration tient aussi dans la diminution des réponses incorrectes dangereuses au profit des réponses incorrectes douteuses. Il en va de même pour les réponses correctes : à nombre égal, c'est une amélioration de voir leur évolution vers des certitudes plus élevées.

Les résultats de Jans (200) montrent de telles évolutions entre le PRETEST (appelé « Base »), le POST-TEST immédiat et le POST-TEST différé (les mêmes 100 questions à chaque fois) dans une expérience d'apprentissage du vocabulaire anglais chez des étudiants universitaires. Voici les résultats –typiques- d'un de ces étudiants :

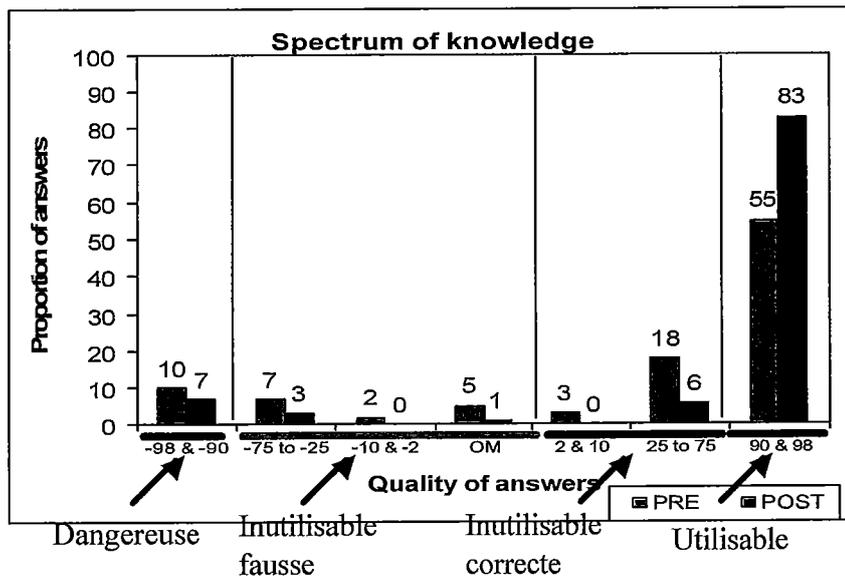


Leclercq & Poumay (2004) Objective assessment of subjectivity. EARLI SIG on Metacognition. Amsterdam.

C6. Degrés de certitude et éducation du patient à sa maladie

Le même genre de courbe apparaît avec des patients diabétiques adultes testés avant et après une formation. Voici la courbe globalisée pour les 38 apprenants confondus ayant répondu à 39 questions. :

Brutomesso et al. 2003 Diabetic Patient Education



Leclercq & Poumay (2004) Objective assessment of subjectivity. EARLI SIG on Metacognition. Amsterdam.

7

Les 7% de connaissances dangereuses APRES formation sont évidemment un progrès par rapport aux 10% AVANT formation, mais indiquent un reliquat inquiétant dont il faut s'occuper.

Dans une autre expérience menée par A. Rinaldi, une infirmière tient en mains un questionnaire « spectral » (Leclercq, Rinaldi et Ernould, 2003) en ce sens que sont imprimées, de part et d'autres des questions (qui sont dans la colonne centrale) les positions possibles des réponses jugées à leur qualité combinant exactitude et degrés de certitude (DC). Les questions étant toutes du type Vrai-Faux, les seuls degrés de certitude permis sont 50%, 60%, 80% et 100%. Les réponses correctes (tantôt V tantôt F), que seule l'infirmière voit, sont indiquées dans la colonne marquée d'une flèche noire verticale.

L'infirmière pose la question oralement et indique sur la feuille (qui sera directement remise au médecin) la réponse du patient, ainsi que sa certitude en entourant la position adéquate sur le spectre de qualité. Voici 6 des questions et des positions en qualité des réponses d'un(e) patient(e) :



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%	F	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 voit que c'est la réponse à la dernière de ces questions qui pose le plus de problèmes puisque le patient croit fermement (à 80%) à une réponse fausse.

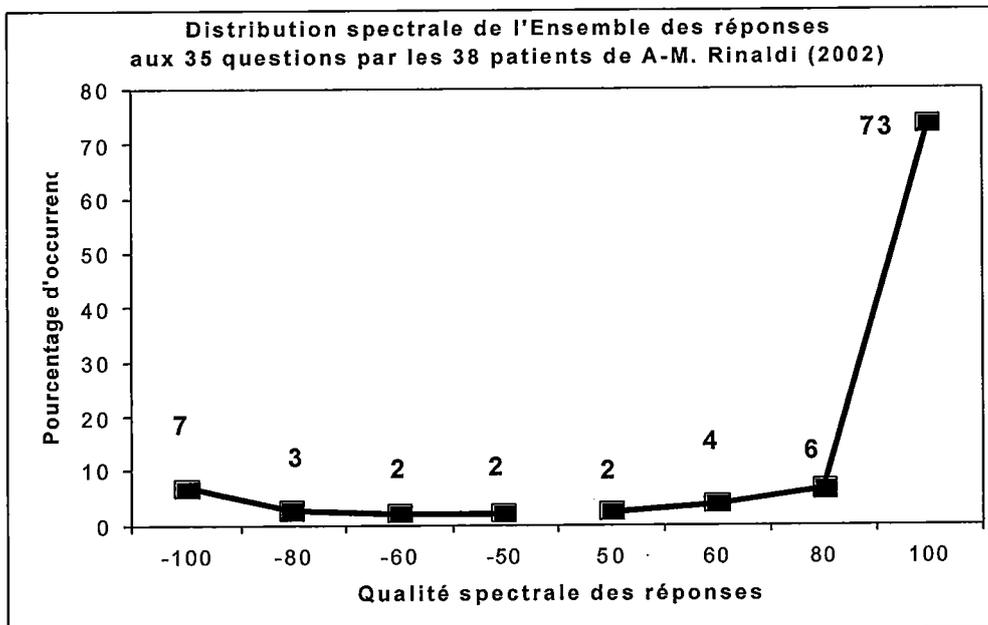
Pour les questions 1, 3 et 5, il y a encore des progrès à faire, non en connaissance, mais en certitude. A la question 2, le patient a fait aveu d'incompétence (ignorance totale) en répondant 50%. Seule la question 4 a donné lieu à une réponse parfaite.

Si l'on considère un groupe de patients et que l'on rassemble leurs réponses question par questions, les positions spectrales permettent de situer immédiatement les questions « posant problème ». Ici, c'est la question commençant par « Si vous

perdez totalement connaissance... » qui a donné lieu au plus de réponses (10 personnes différentes) erronées avec un degré de certitude (DC) maximal (100%). C'est donc la matière à réexpliquer en priorité pour ce groupe.

1	1	0	4	1	0	2	23	Lors d'une perte de conscience
0	0	0	0	0	0	1	37	Lors d'une hypoglycémie
4	4	1	1	4	3	1	19	En présence de nausées
0	0	0	0	0	1	0	37	L'hypoglycémie
1	0	1	2	0	1	1	32	Lors d'une hypoglycémie
10	3	0	0	1	2	2	18	Si vous perdez totalement
1	0	0	0	1	0	1	35	Lors d'une hypoglycémie
5	1	2	0	1	2	4	22	On parle d'hyperglycémie
4	0	3	1	0	1	3	26	En présence d'une blessure
0	1	1	0	2	1	2	18	La présence d'acétone

La distribution spectrale des 38 patients de Mme Rinaldi à ses 35 questions est typique d'une courbe en J (recherchée par les formateurs)...mais pour les réponses correctes seulement, pour le seul hémispectre de droite. Par contre, pour les réponses incorrectes, on constate une inquiétante courbe en i due à 7 % des réponses, incorrectes avec la certitude maximale, bref ce que les anglo-saxons appellent « misconceptions » et que nous appellerions « idées fausses ».



C7. Degrés de certitude et docimologie

Le réalisme des testés dans l'évaluation de leur auto-connaissance doit avoir un impact sur leur note scolaire. Nous calculons les notes des étudiants de façon à faire peser la métacognition (ou autocognition) pour 3 points sur 20, selon la procédure suivante :

Une réponse correcte = + 1 point
Une omission = 0 point
Une réponse incorrecte = -0,25 point
Somme de toutes les notes ; Expression du score total (appelé **score classique**) sur 20

Si l'étudiant a accompagné CHACUNE de ses réponses d'un Degré de Certitude¹⁰, Ajout (éventuel) de points métacognitifs ou « Plus » métacognitifs » :

Si le Degré de Certitude moyen

avec les Réponses Correctes est > à 50 % : + 1 point de **Confiance**

avec les Réponses Incorrectes est < 50 % : + 1 point de **Prudence**

Si ces deux points sont acquis et que la différence entre ces deux certitudes moyennes est > 20 % : + 1 point de **Nuance**

L'évaluation de la métacognition doit être un « plus » pour les étudiants.

Les tableaux suivants présentent les résultats de quelques uns des 400 étudiants à l'examen qu'ils ont subi pour l'un de nos cours. La moyenne des scores de ces 400 étudiants est passée de 9,3 points (sur 20) à 11,3 points sur 20. Dans ce tableau, la signification des colonnes est la suivante (de gauche à droite) :

CMRC = Certitude Moyenne avec Réponses Correctes.

CMRI = Certitude Moyenne avec Réponses Incorrectes.

DIFF = différence entre CMRC et CMRI.

C = point supplémentaire de Confiance.

P = Point supplémentaire de Prudence.

D = Point supplémentaire de Discriminance. Ou Nuance

CLASS = score classique (sans degré de certitude).

+Méta = score incluant l'utilisation des degrés de certitude.

¹⁰ Cette exigence vise à empêcher les étudiants de recourir à des stratégies de gains des points métacognitifs en ne fournissant que deux degrés de certitude, l'un (100%) à une réponse dont ils sont absolument sûrs de l'exactitude et l'autre (0%) à une réponse dont ils sont absolument sûrs qu'elle est incorrecte.

Notons que quand l'indice de Confiance (ou Facilité Subjective des Réponses Correctes ou FSC) ou celui de Prudence (ou Facilité Subjective des Réponses Incorrectes ou FSI) vaut exactement 50%, le point métacognitif n'est PAS accordé. C'est aussi le cas lorsque l'indice de Nuance (ou Discriminance) vaut exactement 20%. Les étudiants obtiennent généralement plus souvent le point de Confiance que celui de Prudence et le « plus métacognitif » le plus rarement obtenu est le point de Nuance.

Matricule	FSC	FSI	FSC-I	C	P	D	CLASS	+méta
							9,3	11,3
587	58	40	18	1	1	0	7,9	9,9

Dans l'exemple de l'étudiant au matricule 587, le point de Confiance est gagné (58% de certitude moyenne avec réponses correctes) ainsi que le point de Prudence (40% de certitude moyenne avec réponses incorrectes) mais pas le point de Nuance (18% de différence seulement, c'est-à-dire en dessous du seuil requis de 20%).

Voici les résultats d'une série de 6 étudiants :

Matricule	FSC	FSI	FSC-I	C	P	D	CLASS	+méta
							9,3	11,3
587	58	40	18	1	1	0	7,9	9,9
597	64	53	10	1	0	0	15,3	16,3
1509	65	40	25	1	1	1	10,5	13,5
1557	73	43	30	1	1	1	8,9	11,9
1864	54	47	7	1	1	0	15,3	17,3
10503	75	55	20	1	0	0	13,7	14,7

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 qu'aucun n'y renonce !).

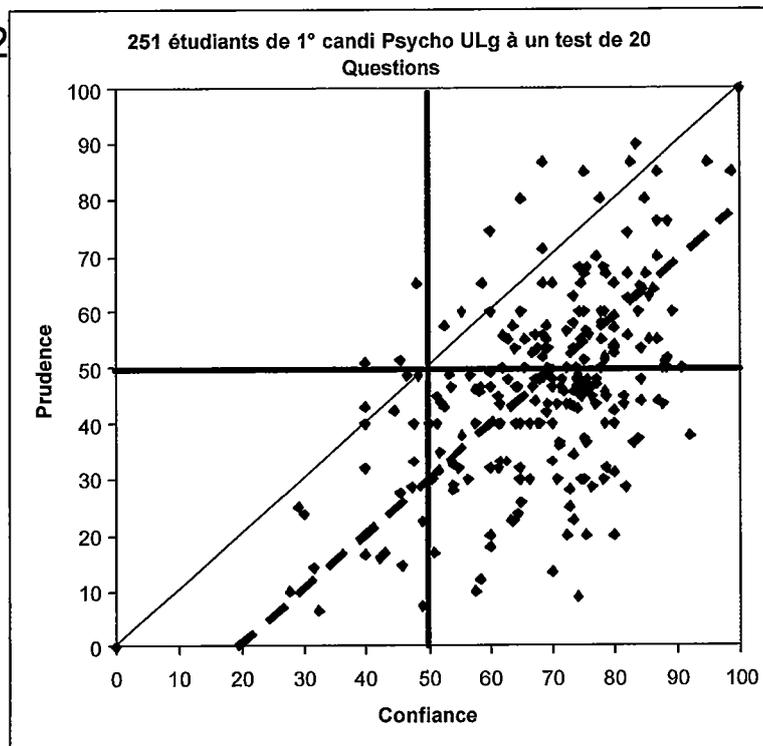
Le bénéfice éducatif et formatif tient en la qualité du feedback que l'on peut donner à l'apprenant, non seulement désormais sur l'exactitude de ses réponses, mais aussi sur sa métacognition. Et ces feedbacks sont assez nuancés puisqu'ils distinguent Confiance, Prudence et Nuance, trois aspects sur lesquels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

Enfin, les indices proposés sont faciles à calculer tant par le professeur que par l'apprenant.

Voici le nuage de points réalisé en plaçant les 251 étudiants qui, en octobre 2002, ont passé un test sur les Sciences de l'Education, à l'horizontale de leur indice de Confiance dans ce test et à la verticale de leur indice de Prudence/imprudence dans ce même test. La confiance moyenne était de 68%. La Prudence moyenne de 46% et la nuance moyenne de 22%.

Octobre 2002

- **C = 68**
- **P = 46**
- **D = 22**



D. Leclercq et M. Poumay, Trois nouveaux Indices de réalisme, 16^e Colloque de l'ADMEE, Liège, sep. 2003

15

Tous les étudiants placés à droite de la barre verticale ont reçu le point de Confiance. Tous ceux qui sont placés en dessous de la barre horizontale ont en outre reçu le point de Prudence. Tous ceux qui sont dans ce carré en bas à droite et qui en outre sont sous l'oblique en pointillés (située à 20% sous la diagonale) ont en outre reçu le point de Nuance.

D. Dialogues métacognitifs : analyses et régulations POST

Partant de ces indices, nous avons mené des dialogues métacognitifs par e-mails avec les étudiants volontaires de première année de candidature dans ce même cours, mais en 2003. On trouvera aussi les commentaires faits par le professeur (D. Leclercq) et une étudiante en mémoire (L. Wislez) qui joue le rôle d'assistante.

D1. Questionnaires spectraux formatifs

D'abord, à la fin de chaque cours (pendant 15 minutes), les étudiants sont invités à répondre à 10 questions sur un Questionnaire spectral (Leclercq, Rinaldi et Ernould, 2003) en version « papier ». Ce genre de disposition facilite la mise en évidence des situations problèmes, et de connaissance partielle, ainsi que le calcul des scores de Confiance, Prudence et Nuance (voir Leclercq et Poumay, 2003). Voici un extrait (les 4 premières questions) d'un tel questionnaire (de 10 questions) :

24/09/2003

PPUQ Chap 3 et 4 p. 66 à 80

Pré :

6. Aucune 7. Toutes

						Q	C						
100	80	60	40	20	0			0	20	40	60	80	100
100	80	60	40	20	0	1	T.S. Eliot a dit "Where is information lost in....." 1. affairs 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 (motivationnelles)	0	20	40	60	80	100
100	80	60	40	20	0	4	L'architecture des compétences en 4 pailers superposés est comparée par leclercq à 1. un véhicule (un avion) 2. un animal (un insecte) (une mâche de foreuse) 3. un outil 4. Un jouet (un cerf-volant)	0	20	40	60	80	100

Prudence :

Nuance :

Confiance :

D2. Dialogues métacognitifs sur formulaires et par mail

En moyenne, les étudiants reçoivent un test par cours (donc un par semaine), ces tests portant comme nom les lettres de l'alphabet (A, B, C, D, etc.), dans l'ordre de leur passation. Ils gardent le questionnaire spectral avec leurs réponses, mais envoient à l'assistante une feuille de « Dialogue métacognitif » comportant 22

? L'AIPU : 20 ans de Recherches et d'Actions Pédagogiques ; Bilan et Perspectives ?
 21^{ème} Congrès de l'AIPU 3-7 Mai 2004, Université Cadi Ayyad, Marrakech, Maroc

lignes. L'assistante 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'assistante renvoient leurs commentaires (d'où l'expression « dialogue métacognitif »). Voici un tel questionnaire « rempli » pour l'étudiant 49 pour deux tests différents, le premier le 1/10/2003, le deuxième le 22/10/2003. On constatera les progrès métacognitifs. :

3	Code = 49	Test A 1 octobre 2003	Test D 22 octobre 2003
4	Activité (nom du test)	Compétence et information	Vocabulaire du chapitre 1
5	De l'année	2003	2003
6	Du mois	Octobre	Octobre
7	Du jour	1	21
8	Nbre de question	10	20
9	Nbre d'Omission	0	0
10	Jugement et le Pourquoi de vos omissions	Je n'aime pas ne pas avoir de réponse à une question alors je vais chercher au plus profond de ma mémoire mais apparemment cela ne marche pas. OK, mais alors le degré de certitude doit être très faible (ce que vous avez fait, en effet).	Etant donné que c'était des mots que j'avais lus (le chapitre 1 chez moi) et analysés dans le contexte d'une phrase je pensais pouvoir donner un synonyme à chacun.
11	Nbre de RC	3/10	16/20
12	Confiance	24%	75%
13	Jugement et Pourquoi	Je ne pense ne pas avoir approfondi suffisamment ma lecture des questions. Sur quel type de question vous êtes-vous trompé ? Faites cette analyse ; elle est cruciale !!	Je pense avoir bien compris ce que j'ai lu et m'être mieux estimé. En effet
14	Nbre de RI	7	4
15	Prudence	40%	40%
16	Jugement et Pourquoi	Je me suis trop souvent surestimé pour des questions dont les réponses étaient incorrectes. Heureusement, malgré cela votre Prudence est inférieure à 50%. Au moins, vous savez que vous ne savez pas. TB !	Je ne pense plus m'être trop estimé. Pourtant votre Prudence est la même qu'en A !
17	Nuance	16%	35%
18	Jugement et pourquoi	Je pense m'estimer très mal.	Je pense m'être mieux estimé dans la globalité. En ce qui concerne le taux de réponses correctes, OUI. Et la Confiance, OUI (donc en Nuance). Mais pas en Prudence.
19	Qu'avez-vous appris à cette occasion ?	Que je ne suis pas encore prêt à ce genre de test	J'ai l'impression de m'améliorer et cela me redonne courage. Tant mieux. C'est le but de l'opération.
20	Autres Remarques		Je pense que ce genre de test m'est très utile.
21	Feedback Ludivine Wislez	Que pensez-vous qu'il serait intéressant de faire pour	Vos pourcentages (indices cognitifs et métacognitifs) sont en effet

		améliorer vos résultats? Qu'est ce qui vous semble difficile dans ce genre de tests ?	meilleurs que ceux du test précédent: Savez-vous pourquoi vous avez obtenu de meilleurs résultats ? Avez-vous modifié votre fonctionnement (dans votre façon de répondre? d'étudier ?) ?
22	Feedback D. Leclercq		

D3. Les Régulations (ou Décisions) POST

Dans les dialogues métacognitifs, apparaissent des « Conclusions » ou « Décisions » de certains étudiants, que les encadrants remettent parfois en cause, soit directement, soit sous forme de question ! En voici (soulignés par nous, les commentaires des encadrants étant en gras :)

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.

E. Des Jugements POST alimentent des Analyses et Régulations PER

En octobre 1999 (lors de la première semaine de cours universitaires), le Groupe de Travail « réussite » instauré par le CIUF a réalisé le testing de près de 4000 étudiants entrant dans 8 universités de la Communauté Française de Belgique. Ces étudiants, de toutes les facultés, ont été invités à répondre à 10 épreuves (des check-up car à visée entièrement formative) différentes (math, Physique, biologie, etc.), avec degrés de certitude. Un feedback immédiat leur a été donné confidentiellement et ils ont été « suivis » pour établir le lien entre ces performances à l'entrée et leur succès scolaire en fin de première année universitaire (1^o et 2^o session). Cette opération a été appelée MOHICAN : MOntoring Hlstorique des CANDidatures.

En mai 2000 (juste avant le mois de bloque préparatoire aux examens, dépourvu de séances de cours), nous avons procédé à une opération « Ressac » (Résultats d'Épreuves Standardisées au Service de l'Apprentissage en candis » en fournissant à chacun des étudiants de la Faculté de Psychologie et Sciences de l'Éducation de l'université de Liège une « radiographie Z » de certaines de leurs performances. Certaines de ces performances étaient leurs résultats dans 4 des épreuves MOHICAN (Vocabulaire général de la langue française, Syntaxe, Compréhension de Texte, Compréhension de tableaux et de Graphiques) parce que ces épreuves nous paraissaient couvrir la « compréhension ».

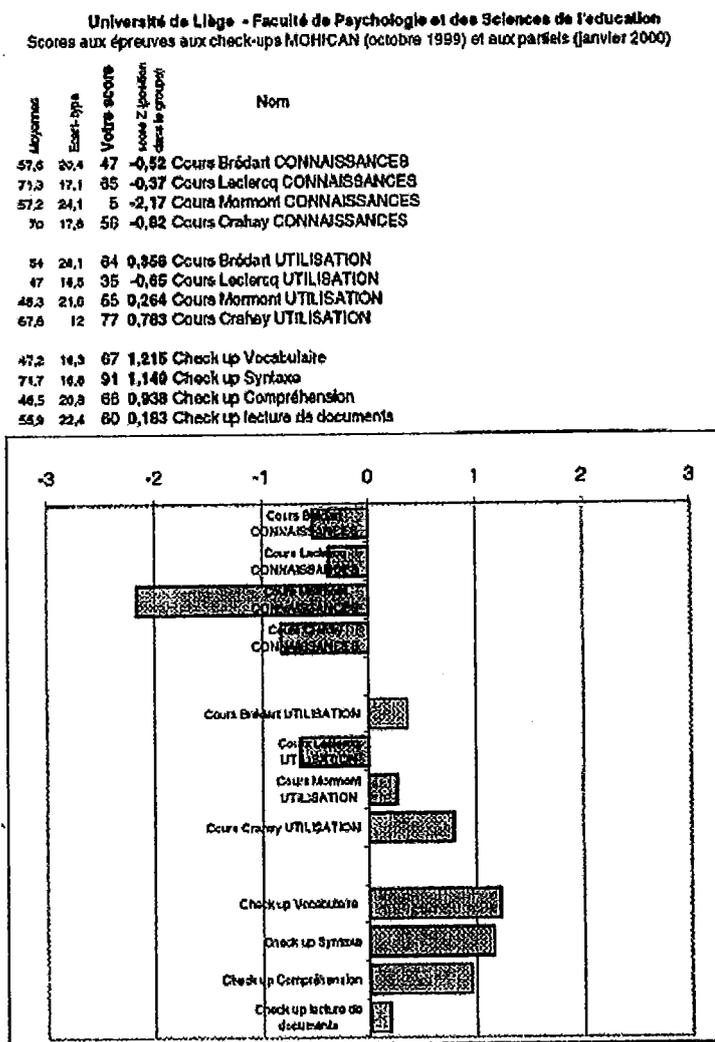
En outre, ces étudiants ont reçu leurs scores à 4 épreuves qu'ils avaient passées en janvier dans 4 cours différents. Les professeurs (Bredart, Crahay, Leclercq et Mormont) avaient calculé pour leur épreuve 2 scores : l'un de « Mémorisation pure » (ou « Connaissance » et l'autre de « Utilisation de connaissances, ou Compréhension » (souvent à livres ouverts). Les scores des étudiants étaient exprimés en valeurs Z de leurs performances par rapport à celle du groupe (plus de 300 étudiants), d'où le nom « radiographie Z ».

Cette représentation graphique avait pour but de faire apparaître les points forts et les points faibles dans deux aspects différents de l'étude : l'étude « de mémoire » et l'étude « en profondeur » ou de Compréhension.

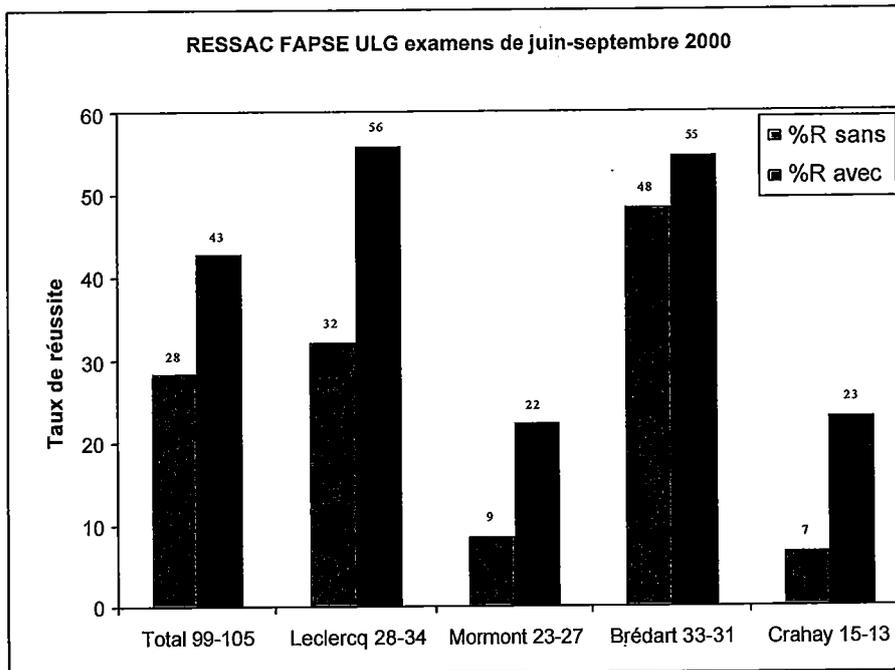
Nous avons recueilli les réactions des étudiants : ont-ils changé leur méthode d'étude durant le dernier mois avant les examens à la lumière de ces informations ?

? L'AIPU : 20 ans de Recherches et d'Actions Pédagogiques ; Bilan et Perspectives ?
 21^{ème} Congrès de l'AIPU 3-7 Mai 2004, Université Cadi Ayyad, Marrakech, Maroc

Nous ne nous sommes intéressés qu'aux étudiants qui avaient échoué à un (au moins) des 4 cours en janvier et qui, donc, devaient le repasser en juin. La moitié environ des étudiants ont déclaré avoir changé leurs méthodes de travail, l'autre non. Les résultats (en termes de réussite) sont dramatiquement différents pour les deux groupes. Ceci est un exemple de régulation POST. Portant sur les processus (d'étude). Nous avons décrit ailleurs (Leclercq, 2003) les analyses auxquelles ces étudiants se sont livrés pour déboucher sur des changements de stratégies d'apprentissage.



Voici les résultats où les nombres sous les bâtons signifient le nombre d'étudiants sur lesquels les pourcentages ont été calculés dans chaque groupe et où les nombres au dessus des bâtons indiquent les pourcentages de réussite dans chacun des cours et dans l'ensemble de ces 4 cours.



Voici de tels jugements, analyses et régulations pour la radiographie spectrale de RESSAC 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 !* ».

F. Les validités des mesures 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.

F1. La validité théorique (ou de construct) 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).

Rappelons encore (Leclercq, 1983, 1993) que la note qui combine les mesures de la cognition et de la métacognition n'est pas elle-même une mesure, mais une « combinaison » de mesures, comme un cocktail n'est pas un alcool mais une combinaison d'alcools où le « dosage » est important et souvent arbitraire. Dans notre formule, Permettre de gagner 3 points sur 20 (et non 2 ou 4) est un compromis. Moins de poids ferait que les étudiants ne s'y intéresseraient pas au point de ne pas s'en servir. Plus de poids serait jugé inacceptable par les collègues. En particulier en Belgique où une Insuffisance est de 8 et 9 points alors qu'une faiblesse est de 10 et 11 points sur 20. Passer de 9 à 12 (satisfaction) grâce aux seules certitudes paraîtra indécent à certains. Passer de 8 à 12, intolérable. Nous tenons donc compte de la sensibilité du moment et du lieu, qui pourra ou devra être revue avec les nouvelles règles consécutives aux réformes de Bologne (par exemple si la note de satisfaction est fixée à 10/20). La note incorporant du métacognitif est donc un payement et non une mesure.

F2. La validité diagnostique concerne le pouvoir informatif 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, ainsi que les conseils pour s'améliorer. Parce qu'ils apportent des données nouvelles, sur un aspect négligé jusque là (la métacognition), et que ces indices portent sur des processus transversaux aux contenus, les indices de réalisme par triangulation (Confiance, Prudence, Nuance) ont une forte validité diagnostique, ce que n'ont pas des indices que nous avons pratiqués pendant des années, les indices de réalisme par calibration (dans la foulée de Adams et Adams, 1950, et Lichtenstein et al., 1977) ni l'indice de Centration calculé par l'Erreur Moyenne de Certitude (EMC) décrite par Leclercq (2003, ..).

Ainsi, sous la direction de J.L. Gilles (SMART-ULG), Deogratias Baragabirije (2003) a testé, sur 27 étudiants entrant à l'université de Bujumbura (Burundi) les représentations erronées de la physique qu'ont un grand nombre d'étudiants du début de l'enseignement universitaire. Il a restreint son étude à des concepts de cinématique liés à l'interprétation et à la connexion entre les graphiques de position, de vitesse et d'accélération en fonction du temps. Il a comparé ses résultats avec ceux d'autres enseignants universitaires (Beichner, 1994 ; Mc Dermott et al. (1987) ; Clement, 1993). Pour ce faire, il a construit une épreuve de 8 questions portant toutes sur des concepts mal compris (selon l'expérience des enseignants de 1^o année universitaire), mettant en évidence ce qu'il appelle pudiquement des « conceptions alternatives » (Onlick, 1999) ou encore « des explications en désaccord avec la vérité scientifique ». Afin d'y remédier, il a mis à la disposition de ses étudiants des logiciels de simulation réalisés grâce à « Interactive Physics »¹¹. Au prétest, le taux moyen de réussite était de 45% et au post-test de 67%, ce qui montre à la fois la nécessité d'intervenir et l'utilité de l'intervention (mais avec un Gain relatif assez modeste : quasiment 50% puisque ne réalisant que 22% de gain sur 55% de gain possible¹²).

Typique des « misconceptions » ou représentations erronées, au prétest, la Confiance moyenne dans les réponses (correctes) était de 40% seulement, c'est-à-dire en dessous du seuil que nous exigeons pour octroyer le point de confiance, tandis que l'imprudence était de 49%, c-à-d frôlant la limite à partir de laquelle nous n'octroyons plus le point de prudence. Et, surtout, l'Imprudence était de 9% supérieure à la Confiance !

Au post-test, les résultats sont aussi intéressants. La confiance moyenne (sur 22% de réponses correctes supplémentaires !) est passée de 40% à 71% et ce bond doit être jugé comme satisfaisant, même s'il montre qu'il existe encore, même pour les réponses correctes, une large marge de doute...dans ces FONDLEMENTS de la physique, et donc encore du travail !

¹¹ <http://www.interactivephysics.com/description.html>

¹² On sait que la formule du Gain Relatif vaut le rapport entre le Gain et le Gain Possible (Mc Guigan, 1967, cité par Leclercq et al., 1977).

Par contre, les réponses erronées du post-test, bien que moins nombreuses de 22% (il n'en reste que 33%) sont accompagnées en moyenne d'un indice d'imprudence qui a augmenté : il est passé de 49% à 52%.

Un peu comme si l'intervention –antibiotique avait bien tué certaines fausses représentations-microbes, mais que ceux (et celles) qui restent sont devenu(e)s plus résistant(e)s encore.

Voilà un éclairage quantitatif que des observations cliniques (interviews sur les conceptions) devraient compléter pour trouver les raisons de ces consolidations de représentations fausses suite à des interventions remédiatives. Nous mettons des s car ce genre de phénomène a pu être aussi observé, *mutatis mutandis*, avec des patients diabétiques.

En outre, il est fréquent (Lucassen, 1984) que, quand ils ont l'initiative de la remédiation, les étudiants commencent par le plus facile et non pas par le plus grave. Un peu comme si des secouristes intervenaient sur les lieux d'une catastrophe en soignant d'abord les blessures bénignes plutôt que les cas en danger de mort. Ce genre de comportement a été expliqué par Atkinson dans sa théorie de l'auto-fixation de la difficulté des tâches et par Maslow dans sa distinction entre deux types de personnes, celles qui sont surtout mues par le « hope of success » et celles qui le sont par le « fear of failure ».(Leclercq, 2003a, chap. 6, 21).

F3. 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 (1997) souligne l'importance de ce qu'il appelle « *practical significance* » (ou Ampleur de l'Effet dans le sens de Glass, 1971) qu'il oppose à « *statistical significance* ». En somme, plutôt que de distinguer le « significatif » du « non-significatif », il propose de distinguer entre le signifiant et l'insignifiant. Nos trois indices permettent de mesurer très directement l'ampleur des gains métacognitifs qui peuvent se produire dans un des indices et non dans les autres, comme on vient de le montrer dans l'expérience de Baragabirije.

Les observations de nos recherches (notamment RESSAC) rejoignent, au niveau de l'enseignement supérieur, celles de Chi et Bassock (1989, 272) au niveau du primaire et du secondaire. Ces deux auteurs 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)

F4. 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, 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.

Leclercq (2003) a montré, sur les données de l'opération MOHICAN (chapitre 7) que les scores incorporant les degrés de certitude étaient plus prédictifs de la réussite que les scores ne les incorporant pas.

F5. La validité opérationnelle concerne 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 par un expert ou par un pair.

G. 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, 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 » (p. 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).

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.

H. Conclusion

Nous concluons par deux citations de De Finetti (1965) :

« La connaissance partielle existe. La détecter est nécessaire et faisable. » (p. 109)

« Seule la probabilité subjective peut donner une signification objective à toute réponse et toutes méthode de notation » (p. 111) et nous ajouterons « ... et de mesure des gains d'apprentissage. »

K. Références

- Baragabirije D. (2003). Contribution à la réalisation et à la mise en ligne d'activités de remédiation en Physique Générale. Mémoire de DES en Technologie de l'Education et de la Formation. Université de Liège.
- Beichner, R. (1994). Testing students' interpretations of kinematics graphs. *American Journal of Physics*, 62, 750-762
- Bereiter, C. et Scardamalia (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.
- Brown, S. & Knight, P. (1994), *Assessing Learners in Higher Education*, London : Kogan Page.
- Brutomesso, D., Leclercq, D., Gagnayre, R., Crazzolara D., Busata, E., d'Ivernois, J-F., Casiglia, E., Tiengo, A. & Britussio, A.. (2003). Confidence degrees to evaluate knowledge in patients with type 1 diabetes, in *patient Education and Counseling*.
- Carver, R.P. (1974), Two dimensions of tests : Psychometric and edumetric, *American Psychologist*, 29, 512-518.
- Chi, M., & Bassock 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.
- Cohen, J. (1963). *Hasard, adresse et chance*, Paris : PUF
- De Finetti, B. (1965). Methods of discriminating levels of partial knowledge concerning a test item. *British Journal of Math. & Statist. Psychol.*, 18, 87-123.
- Descartes, R. (1636), *Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences*.
- Ebel, R.L. (1965). Confidence weighting and test reliability, *Journal of Educational measurement*, 2, 49-57.
- Flavell, J. (1978). Metacognitive aspects of problem solving. In Resnick (Ed.). *The nature of Intelligence*, pp. 231-255. Hillsdale : NJ : Erlbaum Associates.

- Georges F. (2004). L'utilité des tableaux de bord et des traces dans les cours à distance. LabSET Université de Liège.
- 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. NATO ASI Series, F112, Berlin : Springer verlag, 177-189.
- Jans, V. (2000). Confrontations instrumentées et dialectiques des jugements auto et allo-évaluatifs. Contributions conceptuelle et méthodologique à l'étude de l'autoévaluation réflexive chez des étudiants universitaires, thèse de doctorat, Université de Liège.
- Kirk, R.E. (1997) Practical significance : a concept whose time has come. Educational and Psychological Measurement, 56 (5), 746-759.
- Kolb, D.A. (1984) Experiential learning. Englewoods Cliffs : Prentice Hall.
- Koriat, A. (1998). Illusions of knowing : The link between Knowledge and Metaknowledge. In Yzerbyt, Lories & Dardenne, Metacognition, London : SAGE, 16-34.
- Leclercq, D. , Donnay, J. et DeBal , R. (1977). Construire un cours programmé. Bruxelles : Labor
- Leclercq D. & Boskin A.(1990), Note taking behavior studied with the help of hypermedia, in Estes, Heene & Leclercq (Eds), Proceedings of the 7th ICTE , Brussels, vol 2, 16-19, Edimburgh : CEP Consultants.
- Leclercq et Detroz (2003)
- Leclercq, D. (1983) Confidence Marking, its use in testing, in B. Choppin B. & N. Postlethwaite, Evaluation in Education : An International Review Series, Oxford : Pergamon Press, Vol. 6, number 2, 1983, 161-287.
- Leclercq, D. (1993). Validity, reliability and acuity of self-assesment in educational testing. in D. Leclercq, & J. Bruno. Item banking : Interactive testing and self assessment. NATO ASI Series, F112, Berlin : Springer verlag, 114-131.
- Leclercq, D., Boxus, E., DeBrogniez, P., Wuidar, H & Lambert F. The TASTE approach : General implicit Solutions in MCQs, Confidence Marking, Open books exams and Interactive testing. in D. Leclercq & J. Bruno. Item banking : Interactive testing and self assessment. NATO ASI Series, F112, Berlin : Springer verlag.

- Leclercq, D. (Ed) (1998). Pour une Pédagogie Universitaire de Qualité. Sprimont : Mardaga.
- Leclercq, D. (1999). Edumétrie et Docimologie. STE- Université de Liège.
- Leclercq, D. (2003a) Psychologie éducationnelle. Editions de l'Université de Liège
- Leclercq, D. (2003b). Méthodes de Formation et Théories de l'Apprentissage. Editions de l'Université de Liège
- Leclercq, D. (2003c). Un diagnostic Cognitif et Métacognitif au seuil de l'université. Le projet MOHICAN mené par les 9 universités de la Communauté française de Belgique. Groupe de travail « réussite » du CIUF et Editions de l'Université de Liège
- Leclercq, D et Poumay, M. (2003) Le curriculum FORMASUP : <http://www.ulg.ac.be/LabSET>
- Leclercq, D., Rinaldi A.M. et Ernould Ch. (2003). Un questionnaire spectral pour l'évaluation des connaissances chez le patient diabétique. In Gagnayre et al. (Eds). Les cahiers de l'IPCEM 2003 : L'évaluation de l'éducation thérapeutique du patient. Paris : IPCEM.
- Leclercq, D. et Poumay, M. (2003). La connaissance partielle chez l'apprenant : pourquoi et comment la mesurer, In Gagnayre et al. (Eds). Les cahiers de l'IPCEM 2003 : L'évaluation de l'éducation thérapeutique du patient. Paris : IPCEM.
- Lucassen, J.C. (1984). Diagnostic et remédiation assistés par ordinateur dans l'enseignement primaire spécial de type 8. Mémoire de licence en Sciences de l'Education. Université de Liège.
- Markman, E. (1977). Realizing that you don't understand : A preliminary investigation. Child Development, 450, 643-655.
- Maslow, A.(1970), Motivation and personality, New York : Harper and Row, 1954 (1^{re} édition), trad. française : « La psychologie de l'être ». Paris : Fayard.
- Mc Dermott, L; et al. (1987). Students difficulties in connecting graphs and physics. Examples from kinematic. American Journal of Physics, 55, 503-513
- Mc Guigan, F.J. (1967). The G statistics, an index of amount learned. NSPI Journal, 69, 14-16.
- Miller, G.A. (1956). The magical number seven, plus or minus two. Psychological review, 63, 81-97.

- Noel, B. (1991), La métacognition, Bruxelles : De Boeck.
- Noel, B., Romainville, M. & Wolfs, J.L. (1995). La métacognition : fazcettes et pertinence du concept en éducation. *Revue Française de Pédagogie*, n°112, 47-56.
- 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, Université de Liège.
- Schön, D. (1994). Le praticien réflexif: A la recherche du savoir caché dans l'agir professionnel. Montréal : Logiques.
- Schraw, G. (1997) The effect of generalized Metacognitive Knowledge on Test Performance and Confidence Judgements, Journal of Experimental Education, 65 (2), 135-146.*
- Tardif, J. (1992). Pour un enseignement stratégique – l'apport de la psychologie cognitive. Montréal : Les Editions Logiques inc.
- Tochon 1991
- Wang, M.C., Haertel, G.B. et Walberg, H.J. (1990). What influences learning. A content analysis of review literature. *Journal of Educational Research*, 84 (1), 30-44.
- Wolfs, J.-L. (1991), Analyse de l'anticipation de questions comme indicateur métacognitif, thèse de doctorat, Université Libre de Bruxelles.