

CAPAES

Evaluations en Grands Groupes

Février 2011

Marianne POUMAY & Dieudonné LECLERCQ
Assistant : Olivier DEWIT

Introduction.....	1
Évaluation et motivation (Viau).....	2
A.T.O.M.E.S.....	4
Caractéristiques de dispositifs d'évaluation (...formative).....	9
Représentation synoptique des facettes et dimensions de dispositifs d'évaluation.....	15
Leclercq, D. (2006) L'évolution des QCM.....	16
T.O.I.S.E. Recognition - Evocation.....	22
T.O.I.S.E. Évalue la compréhension à la synthèse-crédation.....	36
L'évaluation des performances complexes ou des compétences, principes et illustrations.....	50
T.O.I.S.E. L'évaluation - jugements - certitudes.....	64
Annexes.....	76

Ce syllabus est complété par des ressources en ligne :

- Le contrat de cours, incluant les consignes de présentation du travail ;
- Les syllabus des deux séances, en version pdf ;
- Des vidéos, articles ou présentations « Powerpoint » additionnelles, qui traitent de sujets abordés lors des deux séances ;
- Un forum de discussion, dans lequel vous pouvez poser vos questions liées à ce cours, par exemple en préparation à votre travail.

Pour joindre ce site, au départ du site du CAPAES, cliquez sur la gauche sur « cours M. Poumay IEGG » (ce cours est en fait pris en charge par M. Poumay et D. Leclercq), puis sur l'adresse de l'espace web dédié à ce cours « grands groupes ». On vous demande un identifiant et un mot de passe (que vous devez avoir reçu de l'ULg).

Introduction

Objectifs généraux du cours « Méthodes d'action pédagogique et d'évaluation en grands groupes »

Compétences du décret CAPAES visées dans ce cours :

- 7) « *Planifier le cours et concevoir des dispositifs d'enseignement appropriés aux adultes* »
→ en particulier dans des grands groupes
- 8) « *Maîtriser et utiliser les outils d'évaluation des apprentissages adaptés à l'enseignement dispensé et pouvoir répondre de ses choix* »
→ en particulier, dans des grands groupes
- 11) « *Porter un regard réflexif et interdisciplinaire sur ses connaissances scientifiques et son enseignement* »
→ en particulier, ses connaissances relatives à l'intervention et à l'évaluation dans des grands groupes et son enseignement en grands groupes
- 12) « *Actualiser ses connaissances et ses pratiques* »
→ en particulier, concernant l'intervention et l'évaluation en grands groupes

Objectifs spécifiques de la séance « Méthodes d'évaluation pédagogique en grands groupes », mars 2008

Au terme de la séance consacrée aux méthodes d'évaluation en grands groupes, l'enseignant participant démontrera sa capacité à adopter une posture de réflexion critique constructive par rapport à sa pratique, c'est-à-dire :

- * Analyser l'un des dispositifs de formation dont il est titulaire, au moyen des outils exposés (triple concordance et caractéristiques et qualités des dispositifs d'évaluation), pour identifier des aspects améliorables dans ses pratiques évaluatives. Choisir les difficultés qu'il souhaite pallier le plus immédiatement.
- * Sur base des outils exposés (triple concordance et caractéristiques et qualités des dispositifs d'évaluation) et en fonction des difficultés identifiées, tracer des pistes d'amélioration du dispositif précédemment analysé, en particulier, au niveau des méthodes d'évaluation utilisées.
- * Expliquer en quoi la (ou les) pistes d'amélioration choisies contribueraient à améliorer le dispositif de formation dans son ensemble et en particulier l'évaluation des apprentissages réalisés par les étudiants.

N.B. De façon à ce que le cours soit utile à chaque participant, pour chacun de ces trois objectifs, lorsque c'est possible, les encadrants tenteront de favoriser le transfert de ces réflexions à des enseignants qui n'auraient pas affaire à de grands groupes.

Une définition des cours de qualité (Poumay, 2003)

« Les cours de qualité fondent leurs **objectifs** sur une bonne **analyse des besoins** des apprenants ; ils incluent des composants **motivationnels** (affectifs) aussi bien que **cognitifs** et **métacognitifs**, utilisent des **méthodes** choisies sur base de théories valides et explicites, **évaluent** les **produits et processus** d'apprentissage par des méthodes et des **outils** adéquats et écologiquement **valides**, débouchant sur des indices **diagnostiques** et porteurs de sens pour les apprenants. Ils permettent par ailleurs à une **large majorité** d'entre eux d'atteindre les objectifs d'apprentissage fixés, ceci avec un **haut niveau de maîtrise**. »

Evaluation et motivation

(pour plus de détails sur le modèle motivationnel de Viau, voir syllabus partie 1, page 61 à 69)



Extrait de la conférence de Roland Viau (2004) « La motivation : condition au plaisir d'apprendre et d'enseigner en contexte scolaire ». 3^e congrès des chercheurs en éducation. Bruxelles.

Un des professeurs avec qui j'ai travaillé à l'Université de Californie à Los Angeles (UCLA) me disait : « Si tu veux que ton enfant cesse d'écouter une émission de télévision, impose-lui après chaque épisode un examen sur ce qu'il a vu ». Cette boutade illustre bien l'effet négatif que l'évaluation peut avoir sur la motivation et l'intérêt d'une personne.

Bon nombre de chercheurs considèrent que les pratiques évaluatives des enseignants démotivent l'élève à apprendre parce qu'elles sont centrées seulement sur la performance. Dans cette approche de l'évaluation, les enseignants considèrent que l'élève a appris lorsqu'il a obtenu de bonnes notes. Pour juger de sa performance, ils la comparent généralement avec celle des autres élèves : si sa note est la plus élevée du groupe, il est vu aux yeux de l'enseignant comme le meilleur ; si sa note est la plus faible, il est considéré comme le cancre de la classe. Les pratiques évaluatives centrées sur la performance de l'élève sont celles qui sont le plus souvent utilisées dans les écoles. Selon une enquête américaine, citée par Midgley (1993), 99% des professeurs des écoles secondaires américaines évaluent la performance de leurs élèves, 26% les efforts qu'ils ont fournis, et seulement 18%, les progrès qu'ils ont réalisés. On peut penser que ces résultats seraient sensiblement les mêmes au Québec si une telle enquête était menée.

Les pratiques évaluatives centrées sur la performance suscitent parfois la motivation des forts, mais nuisent généralement à celle des élèves moyens et faibles (Covington, 1992). Pour ces élèves, l'erreur est vue comme « une plaie » à éviter absolument, car elle réduit leurs chances « de passer ». Menacés encore une fois de se retrouver au bas de l'échelle, ils oublient rapidement qu'ils sont à l'école pour apprendre et que l'on peut y trouver un réel plaisir à le faire.

On connaît les effets pervers d'utiliser les notes pour motiver les élèves. Un des plus importants réside dans le fait que les élèves ont tendance à rechercher des exercices qui leur demandent peu d'effort cognitif et à se limiter la plupart du temps à mémoriser plutôt qu'à comprendre la matière à l'étude (Lepper et Hodell, 1989). De plus, craignant d'échouer, ils sont peu audacieux dans leurs idées et sont encore moins créatifs. Or, n'est-ce pas le reproche que l'on fait de nos jours aux élèves dans les écoles, de ne pas être imaginatifs et créateurs ? Lorsque l'on regarde « l'atmosphère éducative » qui règne dans les classes, peut-on sincèrement le leur reprocher ?

La motivation à performer ou tout au moins à « passer l'examen » a remplacé celle d'apprendre. Les chercheurs expliquent cette substitution par l'effet de sur justification. Cet effet se produit lorsque la motivation intrinsèque à apprendre d'un élève diminue à cause d'un

usage abusif des notes ou des récompenses comme moyen de le faire travailler (Tang et Hall, 1995). En d'autres termes, le fait d'évaluer constamment un élève, et ce depuis le début du primaire, amène ce dernier à ne plus travailler pour le plaisir d'apprendre, mais pour les conséquences de l'apprentissage, c'est-à-dire pour les notes ou les récompenses annoncées. En fait, les nombreuses années passées sur les bancs d'école ont amené les élèves à comprendre que ce n'est pas tant le fait d'apprendre qui est important à l'école, mais de réussir les épreuves et d'obtenir un diplôme. Comme on peut le constater, les enseignants subissent les contrecoups des pratiques évaluatives mises de l'avant par leurs prédécesseurs. En continuant cette tradition, ils ne font qu'alimenter la perception des élèves que l'école n'est pas un lieu où l'on apprend, mais une vraie cour de justice où ils se font constamment juger pour des fautes qu'ils ont commises.

Faire de l'évaluation une composante de l'enseignement qui favorise la motivation à apprendre des élèves est un défi de taille pour un enseignant. La valeur que la société accorde à la compétitivité et à la performance, la pression des parents pour que leurs enfants soient notés et l'atmosphère évaluative qui règne dans plusieurs établissements scolaires sont des arguments de taille pour les enseignants qui ne désirent pas changer leurs pratiques. Pour ceux qui désirent être rebelles et aller contre vents et marées, une piste peut être exploitée : mettre en œuvre des pratiques évaluatives centrées sur le processus d'apprentissage plutôt que sur les performances seulement. Dans les pratiques évaluatives centrées sur l'apprentissage,

- l'accent est mis sur le processus d'apprentissage plutôt que sur les résultats. Le progrès accompli par l'élève devient alors un des principaux critères d'appréciation.

- la performance n'est qu'un des indices retenus pour évaluer l'apprentissage réalisé par l'élève. L'acquisition de stratégies d'apprentissage efficaces, le développement de l'autonomie et de la créativité deviennent des indices tout aussi importants que la performance.

- l'examen n'est plus l'outil d'évaluation par excellence. Le portfolio ou le dossier d'apprentissage, dans lequel sont colligés les travaux importants de l'élève, est considéré comme un outil plus adéquat que le relevé de notes pour évaluer le processus d'apprentissage.

Références

- Covington, M.V. (1992). *Making the grade : a self-worth perspective on motivation and school reform*. New-York : Cambridge University.
- Lepper, M.R., & Hoddel, M. (1989). Intrinsic motivation in the classroom. In R.E. Ames, & C. Ames (Eds.), *Research on motivation in education, goals and cognition*, vol. 3 (pp. 73-106). Toronto : Academic Press.
- Midgley, C. (1993). Motivation and middle level schools. In M.L. Mahler, & P.R. Pintrich (Eds.), *Advances in motivation and achievement*, vol.8 (pp. 217-274). Greenwich, CT : JAI Press.
- Tang, S.-H., & Hall, V.C. (1995). The overjustification effect : a meta-analysis. *Applied Cognitive Psychology*, 9, 365-404.

Colloque PROSE
Bruxelles 25 mars 2010

Cohérence, créativité et performance dans l'enseignement :
Que peut apporter la gestion de la qualité ?

ATOMES

Alignement Tabulaire des Objectifs, Méthodes et Evaluations en Strates
d'un curriculum de formation

Dieudonné Leclercq
d.leclercq@ulg.ac.be
<http://orbi.ulg.ac.be>

Introduction

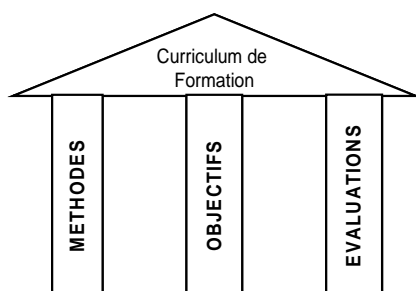
Les professionnels qui s'attaquent à des **tâches complexes** se donnent les moyens conceptuels et communicationnels sous forme de **codes** verbaux (des mots), graphiques : des images), symboliques (des signes) pour **concevoir et gérer** ces tâches. Les électriciens construisent des schémas, les géographes des cartes, les architectes des croquis, etc. Chacun a ses codes, souvent rappelés dans des légendes. En outre, chaque spécialiste, dans son domaine, travaille par **strates** et superpose ou juxtapose diverses représentations. Ainsi le géographe construit plusieurs types de cartes : politiques (avec les frontières des états), économiques (avec des sites d'activités humaines), du relief (avec les courbes de niveau), etc. car il est impossible d'indiquer tous ces renseignements sur une seule carte. L'architecte dessine, en plus d'une épure pour le client, un plan de construction pour l'entrepreneur et un plan de situation pour le cadastre. C'est la multiplicité de ces représentations qui permet aux divers acteurs de **se construire** (à défaut de la recevoir toute faite) **leur vue d'ensemble** des éléments et de **leurs relations**, et ce en vue de les gérer. Nous pensons qu'il ne sert à rien de traiter une situation complexe comme si elle était simple. Des accidents entraînant des dizaines de morts¹ en sont de malheureux exemples.

La complexité d'un programme d'enseignement impose de prendre en considération des « enjeux » que nous proposons de représenter comme un immeuble en trois dimensions. La façade permet de visualiser deux dimensions : la largeur et la hauteur (les étages). La troisième dimension est la profondeur.

¹ (comme celui de Ghislenghien en Belgique, qui, en 2004, fit 24 tués et 132 mutilés ou brûlés)

Dimension 1 (axe horizontal) : Les trois piliers de la façade : OME

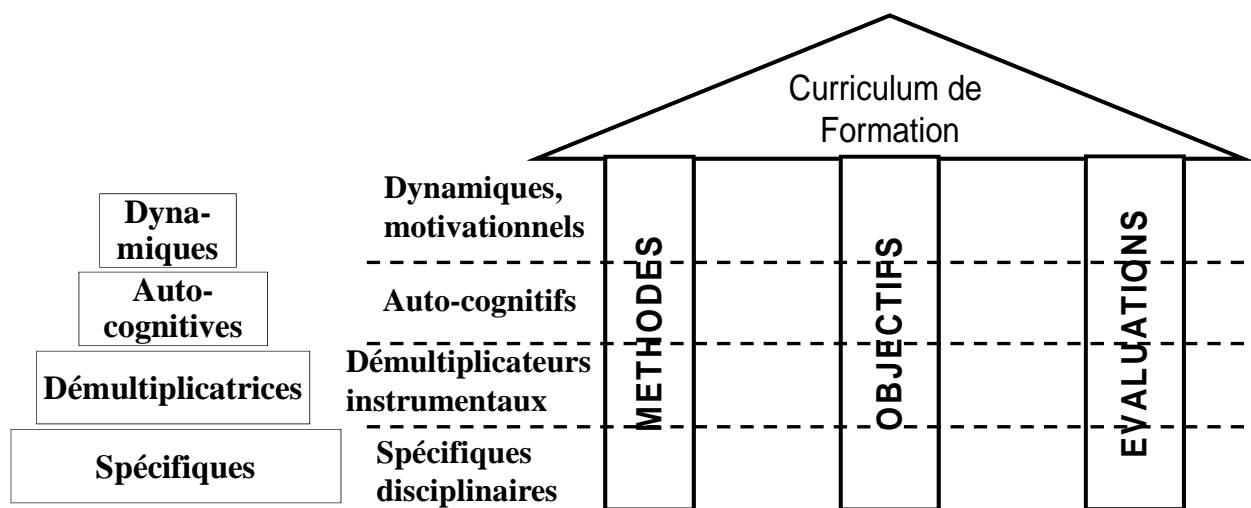
Les éléments « de construction » les plus visibles de l'édifice pédagogique sont au nombre de trois : (1) les Objectifs, éléments centraux (y compris dans notre dessin) car ils conditionnent les deux autres ; (2) les Méthodes de formation et (3) les Evaluations.



Dimension 2 (axe vertical) Les quatre niveaux d'une pyramide (maya) des compétences

Comme dans une maison, on doit considérer, pour les objectifs, des **étages** qui constituent une architecture des compétences. Deux approches, celles de Leclercq (1987) et de Knight (2000) convergent à distinguer **4 étages de compétences** :

- (1) les compétences spécifiques ou disciplinaires (d'un champ de connaissances ou de professions précises ; pour Knight « *Understanding* »),
- (2) les compétences transversales démultiplicatrices (pour Knight « *Skills* » : des savoirs et savoir-agir généraux tels que la lecture, les langues étrangères, les habiletés relationnelles, informatiques, etc.),
- (3) les compétences auto-cognitives (pour Knight « *Self* » : la connaissance de soi, de ses forces, de ses limites, de ses modes préférentiels de travail, etc.) et
- (4) les compétences dynamiques, motivationnelles, d'engagement (pour Knight « *Efficacy belief* »). Les différences entre ces deux auteurs viennent de ce que Knight, préoccupé par l'*employability* des diplômés du supérieur, a extrait ses quatre niveaux de compétence (USES) d'une enquête auprès d'employeurs à qui il a demandé quelles qualités ils recherchaient lors des entretiens d'embauche.

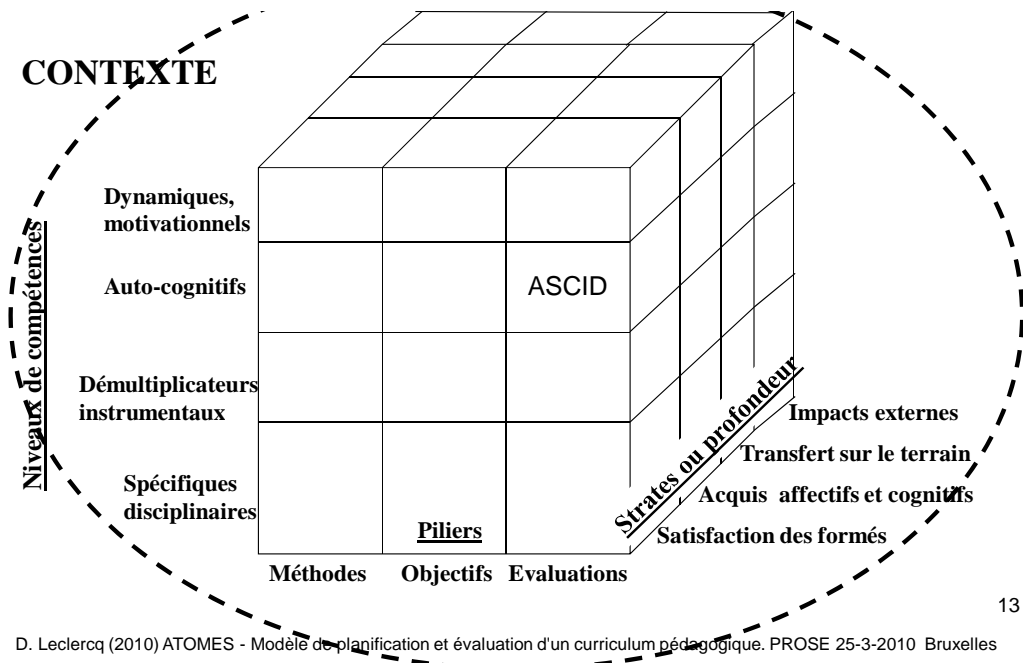


Dimension 3 : La profondeur (ou la portée dans le temps et l'espace) ou strates

La profondeur de la construction servira à représenter grosso modo les types selon la classification de Kirkpatrick (1959), en 4 degrés « de portée » (dans le temps) :

- (1) la **satisfaction** (subjective) des bénéficiaires pendant et juste après ou longtemps après la formation, satisfaction souvent mesurée par un questionnaire.
- (2) les **acquis au terme de la formation**, souvent mesurés par un examen.
- (3) le **transfert sur le terrain** : les comportements, conduites, habitudes, qui ne peuvent être mesurés que sur une période de quelques jours, mois, semaines, voire années.
- (4) Les **impacts externes** de ces comportements : effets sur les bénéficiaires des services rendus par le formé (le professionnel), effets sur l'entreprise (publique ou privée) de ce professionnel, effets sur la société plus généralement.

Un modèle pour penser et évaluer La pédagogie de l'enseignement supérieur : Alignement Tabulaire des Objectifs, Méthodes et Evaluations en Strates d'un curriculum de formation (D. Leclercq, 2010)



C'est, on s'en doute, la strate (ou profondeur) 1 (la satisfaction) qui fait le plus souvent l'objet d'évaluations en formations continues des adultes et la 2 (les acquis) dans les écoles, sous forme de contrôles et d'examens. La strate la moins souvent évaluée est évidemment la 4 non seulement parce qu'il est difficile de distinguer les effets de la formation de nombreuses autres composantes. D'intéressantes études ont néanmoins eu lieu et nous renvoyons à

- Leclercq et Van der Vleuten (1998) pour la description des comparaisons faites (méta-analyses) entre deux types de méthodes formations des médecins : la traditionnelle d'une part et celle(s) basée(s) sur l'Apprentissage Par Problèmes (APP).
- Leclercq (2005a) pour la description de mesures effectuées aux strates 2, 3 et 4 sur l'impact de formations d'adolescents diabétiques lors de camps d'été.

Dimension 4 : composantes intra-cellule

Dans chaque cellule ainsi constituée, des listes d'objectifs, de méthodes d'intervention, de méthodes d'évaluation aident à concevoir le curriculum plus en détails.

Pour les objectifs (peu importe le niveau), on peut définir des **référentiels** (de métier ou d'enseignement) de compétences. En entrant dans plus de détails encore, recourir aux **taxonomies** d'objectifs Affectifs (Maslow, 1943), Sensori-moteurs (Harrow, 1972). Cognitifs (Bloom et al., 1956), d'Identité ou Image de soi ou Métacognitifs (Leclercq et Poumay, 2008), Décisionnels ou d'engagement (Krathwohl et al., 1964), que nous avons résumées sous l'expression ASCID (Leclercq, 2010).

Pour les Méthodes, on peut s'inspirer du Modèle des Événements d'Apprentissage-Enseignement (EAE) ou Learning Events Model (LEM) de Leclercq & Poumay (2004).

Pour les évaluations, on peut prendre appui sur la Taxonomie d'Objectifs instrumentés au Service de l'Évaluation ou TOISE (Leclercq, 2005b).

Dimension 5 : le contexte

Ce contexte (en pointillés sur le schéma) englobe le système de formation)

Le **contexte le plus proximal** d'un cours est le programme des études auxquelles il contribue, et ce programme a lui-même pour contexte le(s) métier(s) au(x)quel(s) ce curriculum prépare, ces métiers eux-mêmes ayant pour contexte les besoins de la société. Dans une telle perspective, quand un enseignant rédige les objectifs de son cours, il est normal qu'il commence par « *Dans la cadre des objectifs suivants (du programme d'études tenant lui-même compte des besoins de l'individu et de la société)* »...et qu'il recopie dès lors une part de ces objectifs venus de l'amont. Il en va de même pour les méthodes et les évaluations qui, elles aussi, peuvent être inspirées de l'amont. Ainsi, on peut décider que les objectifs de chaque cours devront être mesurés non pas isolément, mais au sein d'une « épreuve intégrative ». Loin de « noyer tout », cette référence au « socle commun » permet de mettre en évidence les **valeurs spécifiques**² propres à chaque cours sur les trois piliers (objectifs, méthode set évaluations), aux quatre niveaux des compétences et pour les quatre strates d'impact.

Ce modèle a été appliqué pour décrire des systèmes de formation traditionnels et innovants, surtout au niveau de l'enseignement supérieur et de la formation des adultes. Il peut contribuer à la gestion de la qualité.

² Classification classique d'Aristote par genre prochain et différence spécifique.

Références

- Bloom, B.S., Engelhart, M.D., Forst, E.J., Hill, W.H. & Krathwohl, D.R., (1956) Taxonomie des objectifs pédagogiques, Tome I, Domaine cognitif, Montréal : Education nouvelle, 1969 (original 1956, New York : Longman).
- Harrow, A. (1972). Taxonomie des objectifs pédagogiques, Tome I : Domaine psychomoteur, (trad. française : 1977, Montréal : Presses de l'Université du Québec).
- Kirkpatrick, D.A. (1959), Technics for evaluating training programs, Journal of the American Society for Training Directors (ASTD), n° 13 ; 3-9 ;21-26 ; & n° 14 ; 13-18 ; 28-32.
- Knight, P. (2000) Skills Plus : Employability in higher education.
<http://www.lancs.ac.uk/users/edres/research/skillsplus/index.htm> consulté le 7 juin 2007
- Krathwohl, D., Bloom, B. & Masia, B., (1964). Taxonomie des objectifs pédagogiques, Tome II : Domaine affectif, (trad. française : 1970, Montréal : Entreprises d'éducation nouvelle).
- Leclercq, D. (1987). L'ordinateur et le défi de l'apprentissage, Horizon, Philips Professional Systems, novembre 1987, 13, 29-32 et mars 1988, 14, 22-24.
- Leclercq, D. & Van der Vleuten, C. (1998), PBL - Problem Based Learning ou APP - Apprentissage Par Problèmes, in D. Leclercq (Ed.), Pour une pédagogie universitaire de qualité, Sprimont : Mardaga, pp. 187-205.
- Leclercq, D., & Poumay, M. (2005). *The 8 Learning Events Model and its principles. Release 2005-1*. Retrieved April 29, 2005, from Université de Liège, LabSET, Web site : <http://www.labset.net/media/prod/8LEM.pdf>
- Leclercq, D. (2005a), Evaluation et Régulation. Chapitre 1 de D. Leclercq. Edumétrie et Docimologie pour praticiens-chercheurs, Editions de l'Université de Liège.
- Leclercq, D. (2005b) Une Taxonomie d'Objectifs Instrumentés au Service de l'Evaluation (TOISE), in D. Leclercq, D., Edumétrie et Docimologie pour praticiens-chercheurs, chap 4. Editions de l'Université de Liège.
- Leclercq, D., & Poumay, M. (2005). Sept principes de la Triple Concordance (Objectifs - Méthodes - Evaluation) en Formation.
<http://www.labset.net/media/prod/7PTCOME.pdf>
- Leclercq et Poumay (2008). La métacognition. Chap 6 de D. Leclercq. Psychologie éducationnelle de l'adolescent et du jeune adulte. Editions de l'université de Liège.
- Leclercq, D. (2009) La connaissance partielle chez le patient : pourquoi et comment la mesurer. Revue d'Education Thérapeutique du Patient. 1 (2) pp. 201-212.
- Leclercq, D. (2010). ASCID : le noyau central de douze facteurs expliquant la conduite en éducation thérapeutique du patient. In Education Thérapeutique du Patient, 2 (2), 213-2121.
- Maslow, A. (1943) *A Theory of Human Motivation*, Psychological Review 50(4):370-396.

Evaluations en grands groupes

Caractéristiques de dispositifs d'évaluation (... formative) Ou comment varier nos pratiques d'évaluation pour les rendre plus efficaces ?

Pour des illustrations de ces concepts dans le domaine de l'eLearning, voir article de 26p librement disponible sur Orbi :



Evaluation formative - Définition, justifications théoriques et illustrations



POUMAY, MARIANNE

Document pédagogique (2003)

Ce document propose quelques mises au point sur l'évaluation formative puis une grille d'analyse des facettes de cette évaluation formative, issue des travaux de Dieudonné Leclercq. Cette grille ... [\[plus ▼\]](#)



Marianne Poumay & Dieudonné Leclercq

L'évaluation formative

« *Evaluation intervenant, en principe, au terme de chaque tâche d'apprentissage et ayant pour objet d'informer élève et maître du degré de maîtrise atteint et éventuellement, de découvrir où et en quoi un élève éprouve des difficultés d'apprentissage, en vue de lui faire découvrir des stratégies qui lui permettent de progresser (...), les « erreurs » étant à considérer comme des moments dans (...) l'apprentissage, et non comme des faiblesses répréhensibles ou des manifestations pathologiques. »*

G. De Landsheere (1979). Dictionnaire de l'évaluation et de la recherche en éducation, p. 113



Marianne Poumay & Dieudonné Leclercq

Deux mises au point

~~• Formative, ... **DONC** pas de notation ?~~

~~• Formative... **A PRIORI** ?~~

Documentée

+ Informative

+ Formative



Marianne Poumay & Dieudonné Leclercq

Facettes : Évaluation à...	Dimensions de l'évaluation dans le cours de :	
Référence	Formative	Normative
Résultat	Sommatif	Diagnostic
Visée	Sanctionnante	Formative
Centrage sur	Processus	Produit
Dimension	Unidimensionnelle	Multidimensionnelle
Cible	Individuelle	De groupe
Destinataires	Privés	Public
Opérateurs	Expert / Pairs	Autoévaluation
Périodicité	Ponctuelle	Continue / Répétée
Révisabilité de notes	Éventuelle	Améliorable
Source	Objective	Subjective
Visibilité des critères	Annoncés	cachés
Contrat	Imposé	Négocié
Proximité professionnelle	Ecologique	Peu contextualisée



Marianne Poumay & Dieudonné Leclercq

Eval. continue diagnostique... QRO

Cours de J.L.Hornick, ULg

Test WebCT - Microsoft Internet Explorer

Question 4 (1 Point)

Facteurs déterminant l'ingestion

Regardez cette figure. Pour des animaux au pâturage, sur quels facteurs la saison sèche va-t-elle agir au niveau de l'ingestion alimentaire (par rapport à la saison des pluies)

Enregistrer une réponse

Marianne Poumay & Dieudonné Leclercq

Temps restant: 86 min.				
● Non répondue				
★ Répondue				
1	2	3	4	5
★	★	★	●	●
6	7	8	9	10
●	●	●	●	●
11	12	13	14	15
●	●	●	●	●
16	17	18	19	20
●	●	●	●	●

Quiz continus, diagnostics

Cours de Cl. Houssier, ULg

Chimie pour les sciences de la vie - WebCT 3.1.3 - Microsoft Internet Explorer

Test WebCT - Microsoft Internet Explorer

Quiz formatif 2

Nom: Equipe Labset

Nombre de questions: 13

Terminer Aide

Question 1 (5 Points)

Dans la réaction $3 \text{Cu} + 2 \text{NO}_3^- + 8 \text{H}^+ \rightarrow 2 \text{NO} + 3 \text{Cu}^{2+} + 4 \text{H}_2\text{O}$, chaque atome d'azote :

1. gagne 2 électrons et joue le rôle d'oxydant.
 2. perd 2 électrons et joue le rôle d'oxydant.
 3. gagne 3 électrons et joue le rôle d'oxydant.
 4. perd 3 électrons et joue le rôle de réducteur.
 5. gagne 1 électron et joue le rôle de réducteur.

Enregistrer une réponse

Question 2 (5 Points)

Une limonade contient une quantité inconnue d'acide citrique, un acide triprotique de formule $\text{H}_3\text{C}_6\text{H}_5\text{O}_7$; voir Table 8.2 p. 8.9). Quelle est la

Enregistrer une réponse

Enregistrer une réponse

Marianne Poumay & Dieudonné Leclercq

Temps restant: 86 min.				
● Non répondue				
★ Répondue				
1	2	3	4	5
●	●	●	●	●
6	7	8	9	10
●	●	●	●	●
11	12	13		
●	●	●		

Diagnosticité et accent sur le processus

Cours de Cl Houssier, ULg

Chimie pour les sciences de la vie - WebCT 3.1.3 - Microsoft Internet Explorer

Chimie pour les sciences de la vie
Accueil • Quiz & tests • Points • Essais

5 (5 Points)

Quel sera le coefficient stoechiométrique à faire intervenir dans la relation de calcul de concentration lors du titrage de l'acide phosphorique par l'hydroxyde de sodium si le saut de pH repéré par l'indicateur se situe aux deux-tiers de la neutralisation totale?

1. 5
2. 4
3. 3
100.0% ▶ 4. 2

Votre réponse est correcte. La réaction représentant cette neutralisation s'écrit : $\text{H}_3\text{PO}_4 + 2 \text{NaOH} \rightarrow \text{Na}_2\text{HPO}_4 + 2 \text{H}_2\text{O}$

Deux protons sur les trois que possède l'acide phosphorique seront neutralisés et la relation du titrage s'écrira : $V_{\text{H}_3\text{PO}_4} \times 2 \times M_{\text{H}_3\text{PO}_4} = V_{\text{NaOH}} \times M_{\text{NaOH}}$

5. 1

Note: 5 / 5

Marianne Poumay & Dieudonné Leclercq

Autoévaluation, accès répété

Cours de Ch. Hanzen, ULg

Applications de l'échographie en reproduction bovine
Accueil • Notes de cours • j. Mode échographique TM

MENU DES OPÉRATIONS: Précédent Suivant Contenus du cours Suivre Rafraîchir

Table des matières

- 1. Propédeutique du TG de la vache
- 2. Principes d'échographie : Introduction
 - 2.1. I. Principes physiques de fonctionnement : Table des matières
 - 2.1.1. a. Propagation des ondes
 - 2.1.2. b. Impédance et phases d'atténuation
 - 2.1.3. c. Atténuation, "pénétration" et absorption
 - 2.1.4. d. Réflexion
 - 2.1.5. e. Transmission et réfraction
 - 2.1.6. f. Exemples
 - 2.1.7. g. Diffusion et rétrodiffusion
 - 2.1.8. h. Effets "nocifs" et thérapeutiques
 - 2.1.9. i. Informations transportées vers l'écho
 - 2.1.10. j. Résumé
 - 2.2. II. Principes de fonctionnement de l'échographie : Table des matières
 - 2.2.1. Principes de fonctionnement de l'échographie
 - 2.2.2. a. Émission - réception
 - 2.2.3. b. Notion de champs

Marianne Poumay & Dieudonné Leclercq

Evaluation continue contextualisée

- + proche du réel professionnel
- + concordant avec objectifs
- (... et toujours aussi facile à corriger !)

(Ben Eggleston, Philosophy, KU)

In the following dialogue, which of the following statements is incompatible with cultural relativism?

- (A) Robert: "Child labor is found in many cultures. When children are forced to work in factories as early as the age of eight, they suffer serious consequences in terms of lost educational opportunities, limited time for peer interactions, and risks of injury or death."
- (B) Larry: "Some countries rely heavily on child labor, and would suffer devastating economic consequences if they were forced to give it up."
- (C) Robert: "Despite these consequences, the harms to children are too great to ignore. It is wrong of those cultures to force children to work."
- (D) Larry: "Perhaps those cultures could be persuaded, with carefully targeted economic incentives, to voluntarily discontinue their practices of child labor."



Marianne Poumay & Dieudonné Leclercq

	Facettes : Évaluation à...	Dimensions de l'évaluation dans le cours de : @@@ (préciser éventuellement l'activité)	
Quoi ?	Référence	Critérielle	Normative
	Résultat	Sommatif	Diagnostic
	Visée	Sanctionnante	Formative
Quoi ?	Centrage sur	Processus	Produit
	Dimension	Unidimensionnelle	Pluridimensionnelle
	Cible	Individuelle	De groupe
Qui ?	Destinataires	Privé	Public
	Opérateurs	Expert / Pairs	Autoévaluation
	Périodicité	Ponctuelle	Continue / Répétée
Comment ?	Révisabilité des notes	Définitive	Améliorable
	Source	Objective	Subjective
	Contrat	Imposé	Négocié
	Visibilité des critères	Annoncés	Cachés
	Proximité professionnelle	Contextualisée	Peu contextualisée



Marianne Poumay & Dieudonné Leclercq

Souhaitable!

Facettes : Évaluation à...	Dimensions de l'évaluation dans le cours de : @@@ (préciser éventuellement l'activité)	
Référence	Critérielle ★	Normative
Résultat	Sommatif	Diagnostic ★
Visée	Sanctionnante	Formative ★
Centrage sur	Processus	Produit
Dimension	Unidimensionnelle	Pluridimensionnelle ★
Cible	Individuelle	De groupe
Destinataires	Privé	Public
Opérateurs	Expert / Pairs	Autoévaluation
Périodicité	Ponctuelle	Continue/Répétée ★
Révisabilité des notes	Définitive	Améliorable ★
Source	Objective	Subjective
Contrat	Imposé	Négocié ★
Visibilité des critères	Annoncés ★	cachés
Proximité professionnelle	Contextualisée ★	Peu contextualisée



Marianne Poumay & Dieudonné Leclercq

Facettes : Évaluation à...	Dimensions de l'évaluation dans le cours de : Physique (Y. Lion) - COMPARAISON TRAVAUX/QCM	
Référence	Critérielle ★	Normative
Résultat	Sommatif	Diagnostic ★
Visée	Sanctionnante	Formative ★
Centrage sur	Processus	Produit
Dimension	Unidimensionnelle	Pluridimensionnelle ★
Cible	Individuelle	De groupe
Destinataires	Privé	Public
Opérateurs	Prof / Expert / Pairs	Autoévaluation
Périodicité	Ponctuelle	Continue/Répétée ★
Révisabilité des notes	Définitive	Améliorable ★
Source	Objective	Subjective
Procédure	Standardisée	Adaptative
Implication	Interne (ou participante)	Externe
Contrat	Imposé	Non imposé/Négocié ★
Visibilité des critères	Annoncés ★	cachés
Proximité professionnelle	Contextualisée ★	Peu contextualisée



Représentation synoptique des facettes et dimensions de dispositifs d'évaluation

	Facettes	Dimensions	Dimensions	Justif
Pourquoi ? (fonctions)	Référence	Critérielle	Normative	
	Résultat	Sommatif	Diagnostique	
	Visée	Sanctionnante	Formative	
Quoi ? (objets)	Centrage sur	Processus	Produits	
	Dimension	Uni-dimens	Pluri-dimens.	
Qui ? (sujets)	Cible	Individuelle	de groupe	
	Destinataires	Privée	Publique	
	Opérateurs	Allo-opératrice	Auto-opératrice	
Comment ? (méthodes, conditions)	Périodicité	Ponctuelle	Continue / Répét	
	Révisabil. des notes	Définitive	Améliorable	
	Source	Objective	Subjective	
	Procédure	Standardisée	Adaptative	
	Implication	Interne	Externe	
	Contrat	Imposé	Négocié	
	Repères/ échelle	Mobiles	Fixes	
	Critères	Annoncés	Cachés	
	Contexte	Ecologique	Scolaire	
Validités	Théorique »			
	Informative			
	Conséquentielle			
	Prédictive			
	Replicabilité			
	Acceptable			
	Déontologique			

Leclercq, D. (2006) L'évolution des QCM. In G. Figari et L. Mottier-Lopez. *Recherches sur l'évaluation en Education*. Paris : L'Harmattan, 139-146.

L'évolution des QCM

Dieudonné Leclercq

Université de Liège D.Leclercq@ulg.ac.be

La présente contribution a un double but. Tout d'abord faire perdre éventuellement au lecteur son innocence³ quant aux QCM : il n'existe pas une et une seule forme de QCM. Le second objectif est de montrer en quoi l'évolution des QCM est une succession de modifications visant à rencontrer des critères de qualité d'un système d'évaluation résumables sous le sigle **ETIC PRAD** (Leclercq, 2005) que nous énumérons d'emblée. Nous marquerons ces 8 mots-clés du signe * et nous mettrons leur première lettre en majuscule.

La validité **Ecologique*** (Brunswick, 1943) de l'évaluation pédagogique, ou « validité apparente » (en anglais *face validity*), est d'autant plus grande que la situation correspond à la situation de la vie réelle qu'elle est censée représenter ou prédire.

La validité **Théorique*** (Cronbach & Meehl, 1955) se décompose en validité de contenu (ou de « couverture » : tout ce qu'il faut tester l'est-il ?) et en validité de construct (le Système d'Évaluation est-il fondé sur un modèle crédible, scientifiquement fondé, par exemple des processus mentaux ?).

La validité **Informative*** (si possible diagnostique) est la multiplicité des informations résultant de l'évaluation, leur distinctivité (capacité de porter sur une capacité et non sur la voisine), leur précision (sensibilité), leur valeur explicative.

La validité **Conséquentielle*** (Green, 1998) s'apprécie aux suites que l'évaluation a sur les représentations, les actes (ex : réviser ou non la matière, changer ou non de méthode d'étude) des apprenants, des formateurs ou du système.

La validité **Prédictive*** ou concurrente des mesures obtenues est leur capacité de prédire efficacement (c'est-à-dire avec précision et exactitude) d'autres mesures souvent ultérieures, par exemple la réussite scolaire ou professionnelle, le rendement à une autre épreuve, etc.

La **Replacabilité*** ou stabilité (**fidélité**) d'une mesure est sa stabilité dans le temps ou entre corrections. Une formule (Ebel, 1969) précise le nombre de questions d'une épreuve et, pour les QCM, le nombre de solutions proposées (distracteurs) nécessaires pour obtenir un niveau de fidélité donné (0,8 par exemple). Des formules répondent à la question de la façon inverse : quel doit être le coefficient d'allongement n du test pour atteindre une fidélité donnée (par exemple 0,80 ou 0,90) d'un test qui existe déjà et dont on connaît la fidélité actuelle ?

L'**Acceptabilité*** ou praticabilité d'une évaluation, pour le professeur, concerne l'adhésion aux principes et l'applicabilité des méthodes (durée, matériel et lieux requis, concentration, précautions antifraude, moments possibles, etc.).

Pour l'étudiant, l'**Acceptabilité*** concerne l'adhésion et/ou la familiarité. Ainsi, il a été démontré (Leclercq, 1986) que plus l'étudiant est familier avec les procédures de *testing*, avec les barèmes de cotation, plus il est « aguerrri aux tests » (en anglais *test wiseness*) et plus ses chances de réussite sont élevées, tout spécialement avec les QCM.

La validité **Déontologique*** (ou éthique) prend diverses formes. L'équité est probablement la plus connue. Depuis longtemps, la docimologie négative (Piéron, 1963) a montré que les

³ Dans le sens que B. Bloom (1972) donnait à cette expression: ne plus pouvoir dire « je ne savais pas », parade des incompetents pour justifier tous leurs errements. Désormais, la personne formée devra plaider coupable.

corrections de copies par des juges sont l'objet non seulement de non concordance interjuges, de non constance intrajuge, mais d'autres effets regrettables (de contraste, de sévérité du correcteur, de halo, d'effet Posthumus, etc.) qui sont largement évités par le recours aux QCM. Par ailleurs, les droits des étudiants étant de plus en plus (et à bon droit) reconnus, les systèmes d'évaluation garantissent de plus en plus la transparence de l'évaluation en termes de recalculabilité de la note à partir de la copie brute, de contrôlabilité du processus, ce que les QCM permettent. Nous décrivons ci-après seulement quelques moments charnières de l'histoire des QCM.

1. La gloire de la consigne classique dès sa naissance

Pressés de sélectionner les officiers parmi les appelés à la guerre de 1914-1918, les Etats-Unis font confiance aux *Army tests* conçus par Otis. Il s'agit de tests constitués de Questions à Choix Multiple fonctionnant avec la consigne Classique (QCMC), i.e. : « Une seule des solutions proposées est correcte et vous avez droit à une seule réponse ». Le fait que les USA se sont retrouvés parmi les vainqueurs n'a pas peu fait pour assurer une crédibilité à ce mode de *testing* (validité Prédictive*). Au cours des années suivantes, les modalités de *testing* systématique ont encore accru l'exigence d'efficacité (rapport coût/efficacité) tant appréciée par les Américains du Nord (validité d'Acceptabilité*). Après la guerre 1940-1945, aux USA toujours, l'exigence grandissante de non-discrimination raciale (Civil Rights) dans la notation a fait apprécier ce que les américains ont appelé les « objective tests » (validité Déontologique*), alors que ces tests n'ont d'objectif que la correction. L'ajout de critères d'analyses a posteriori des réponses par les indices de discrimination ou corrélations point bisériales (Davis, 1946) ont donné aux épreuves par QCM une validité Théorique* (de construct) via la psychométrie. Ces quatre types de validité expliquent, à notre avis, le plus grand attachement des Américains aux QCM que ne l'ont été et le sont les Français par exemple.

Pourtant, dès 1963, dans son livre « La Docimologie », Piéron avait montré les discordances importantes pouvant exister entre les notes de différents juges d'une même copie « rédigée », et même l'instabilité de la note d'un même juge à une même copie. Au courant de ce problème, les autorités françaises ont cependant maintenu la notation subjective, sur la base du raisonnement selon lequel le correcteur ne connaissant pas l'identité de l'auteur de la copie, les injustices se répartissent au hasard selon un bon vieux principe (français lui aussi) d'égalité. Il se pourrait que la pratique de plus en plus courante de recours en justice (mode venue elle aussi des USA) des étudiants contre la note obtenue amène à reconsidérer la situation.

2. Une attaque théorique sur le hasard et la correction for guessing classique

Tversky (1964) définit la puissance d'un test par « 1 - la probabilité d'atteindre la performance parfaite par chance ». Or on sait qu'à chaque QCMC qui comporte k solutions, l'étudiant a $1/k$ chances de fournir la solution correcte par chance. Plusieurs parades ont été développées pour pallier ce défaut. Dès 1920, Mc Call recourt à la *correction for guessing* classique qui consiste à fixer comme suit les tarifs : le Tarif en cas de Réponse Correcte (TC) vaut + 1 point, le Tarif en cas d'Omission (TOM) vaut 0 et le Tarif en cas de réponse Incorrecte (TI) vaut $-1/(k-1)$ où k est le nombre de solutions proposées. Tout aussi tôt, West (1923) critique cette procédure. Nous prétendons, aujourd'hui encore, que cette procédure est inadéquate tout d'abord parce qu'elle est basée sur un modèle théorique dépassé de l'activité mentale d'un étudiant en train de répondre à une QCMC : le premier des trois modèles décrits par Bruce Choppin (1975).

Dans ce modèle 1, quand l'étudiant « sait », il choisit la réponse correcte et quand il ne « sait pas », il choisit au hasard parmi les réponses proposées. D'où la *correction for guessing* classique.

Le modèle 2 commence comme le premier, mais au lieu de répondre au hasard quand il « ne sait pas », l'étudiant commence par éliminer les solutions qu'il sait être fausses et choisit au hasard parmi celles qui restent. Ce modèle 2, dont le 1 n'est qu'une variante extrême, reconnaît la notion de connaissance partielle défendue par De Finetti (1965). Il a donné lieu à des consignes du type QCRM (Questions de Choix à Réponses Multiples) consistant à inviter l'étudiant à éliminer les solutions incorrectes, donnant lieu, dans une QCMC à des scores allant de $-(k-1)$ à $+(k-1)$, rendant la mesure plus subtile, plus diagnostique*.

Le modèle 3 de Choppin va jusqu'au bout du concept de connaissance partielle et dit que quand un individu est placé devant une question (à Choix multiple ou non), il commence par ranger les solutions possibles par ordre de plausibilité décroissante et, si la consigne l'oblige à ne fournir qu'une d'entre elles, alors il choisit celle dont la probabilité (subjective) est la plus élevée (à ses yeux) Ce modèle débouche sur le recours aux degrés de certitude car, comme le dit De Finetti (1965), « Seule la probabilité subjective peut donner une signification objective à toute méthode de mesure et de *scoring*. » (p. 111).

Des modèles de Choppin, on aura compris les faiblesses de la *correction for guessing* classique. (1) Elle manque de validité Déontologique* car elle est injuste : elle pénalise aveuglément les personnes à qui on a interdit d'exprimer leur degré de doute. De plus (2), elle manque de validité Informative* pour les enseignants puisqu'elle ne leur apprend rien de plus. Enfin (3), et pour les mêmes raisons, elle manque de validité Conséquentielle* pour les étudiants car, à part « omettre plus souvent », elle n'a pas d'effet sur leur comportement. Cross et Frary (1977) ont en effet démontré (voir détails dans Leclercq, 1986) pourquoi cette procédure dissuade peu de « deviner ».

Les tenants des Degrés de Certitude (voir plus loin) soutiennent que cette procédure répond aux trois manques signalés ci-dessus.

3. Une rafale de critiques théoriques sur les processus mentaux mesurés et non mesurés

3.1. Les QCMC ne mesurent pas l'évocation de mémoire

Il est évident que les QCMC ne peuvent prétendre mesurer la capacité d'évoquer des connaissances, mais bien celle de les « reconnaître », ce qui n'est pas la même chose. Depuis longtemps, en effet, on sait (Luh, 1922) que la performance de reconnaissance a un taux de réussite plus élevé que la performance d'évocation. Ces observations ont été maintes fois confirmées dans des contextes aussi différents que l'apprentissage de langues étrangères (Bahrick, 1984) ou de la médecine (Schurwitz, 1998). Ajouter la solution « Aucune » (ou « Autre ») aux solutions possibles améliore la validité Théorique*.

3.2. Les QCMC invitent au raisonnement à rebours

Même avec la solution « Aucune » ou « Autre », les QCMC induisent un processus mental ne correspondant pas à celui que les étudiants doivent pratiquer dans la vie courante. Avec une QCMC, l'étudiant a tendance à d'abord considérer (et éliminer) les solutions proposées, puis seulement choisir la solution au lieu d'en évoquer une personnelle. C'est le modèle 2 de l'activité mentale décrit par Choppin. Or, ce que l'on veut mesurer, c'est sa capacité à évoquer la solution, puis seulement à la confronter à des solutions possibles. C'est le principe des QCL (Leclercq, 2005) ou Questions à Choix Grandes : l'étudiant reçoit une liste de plusieurs centaines de solutions rangées par ordre alphabétique (comme l'index d'un livre) parmi lesquelles il doit choisir. Chaque solution possible est affectée d'un numéro d'ordre (par exemple de 1 à 700) et c'est par ce numéro en trois chiffres (lisible par le lecteur optique de

marques) que l'étudiant répond. On garde ainsi les avantages de l'automatisation de la correction, en donnant une plus grande validité Théorique* (de construct) L'automatisation de la correction permettant de poser beaucoup de questions (plus d'une par minute, par exemple 100 en une heure) contribue d'une autre façon encore à la validité Théorique* mais dans son aspect « validité de contenu ».

Les QCL seront cependant de plus en plus abandonnées avec le recours aux réponses par clavier. Il suffit, dans ce cas-là de taper le début du mot et le système propose la suite en choix large. Schurwartz (1998) a appelé cela *Long Menu Questions*. Ceci constitue une amélioration de la validité d'Acceptabilité* - applicabilité de la technique.

3.3. Les QCMC ou QCMR renforcent le curriculum caché de l'école

Le curriculum caché est ce que personne n'enseigne mais que tout le monde apprend à l'école. On y apprend, notamment, que quand une question est posée, il faut y répondre ; or certaines questions, parce qu'elles sont absurdes ou excessivement intrusives, ne doivent ou ne peuvent recevoir de réponse ! On y apprend que quand l'autorité pose une question, elle est forcément pertinente, bien posée, etc. On y apprend que toute question à une réponse et que si on ne la connaît pas, on ne peut pas la retrouver par le raisonnement. Bref, le curriculum habituel (il y a heureusement de plus en plus d'exceptions) n'exerce pas à la vigilance cognitive, à la détection des anomalies, des incohérences, etc. notamment par ses modalités de *testing*, les QCMC en étant la plus représentative. Grave lacune dans la validité Théorique* de cette technique ! Pour toutes ces raisons, nous avons développé (Leclercq, 1986) les QCM à Solutions Générales Implicites ou QCM SGI. Ces solutions sont au nombre de quatre : Aucune, Toutes, Manque de données dans l'énoncé, Absurdité dans l'énoncé. Elles sont Générales par ce qu'elles sont valables (et identiques) pour toutes les questions d'un test par QCM SGI. Elles sont implicites parce qu'elles ne sont présentées qu'une seule fois (au début du test) et ne sont pas répétées dans chaque question : l'évalué doit y penser tout seul. Du coup, cette procédure a aussi un impact sur la validité Informative* (ou diagnostique) car elle permet de distinguer deux niveaux de la taxonomie de Bloom : la compréhension (sans piège) et l'analyse (avec piège). Gilles (1999) a montré que les QCM SGI dont la réponse correcte est une SGI avaient une validité Prédicative* supérieure à celles dont la réponse correcte est une solution « visible » pour la réussite d'étudiants en médecine.

4. La rencontre entre QCM et DC (Degrés de Certitude)

Le recours aux Degrés de Certitude est indépendant des QCM. On peut très bien poser une question ouverte (du genre « En quelle année a eu lieu la révolution française ? » et demander à l'étudiant d'accompagner sa réponse d'un degré de certitude⁴. Shuford (1966), Van Naerssen (1965) et De Finetti (1965) ont montré que la consigne ne devait pas être verbale (« peu sûr », « moyennement sûr », « très sûr ») mais probabiliste (en pourcentages de chances). Nous avons en outre montré (Leclercq, 1982, 1993) qu'une précision plus grande que 20% était illusoire, d'où notre consigne en 6 degrés : 0%, 20%, 40%, 60%, 80%, 100%.

Avec les auteurs précités, nous pensons que ce procédé a une plus grande validité Ecologique* que le testing habituel qui empêche les étudiants d'exprimer leur doute. Choppin (1975) a décrit ce problème dans ses modèles 1, 2 et 3. Il dénonce la vision manichéenne (tout ou rien) de phrases telles que « Répondez uniquement si vous savez ; omettez si vous ne savez pas », alors que nous sommes très souvent (et en particulier lors de situations d'apprentissage) dans

⁴ Dans le cadre de l'opération MOHICAN (Leclercq, 2003), qui a posé des QCM (+ Autre et Toutes) dans dix matières à 4000 étudiants entrant dans les universités de la Communauté française de Belgique, nous avons posé cette question sur la date de la révolution Française. Il est intéressant de connaître non seulement le taux de réponse correcte, mais aussi la certitude moyenne (ou Confiance) accompagnant les réponses correctes ainsi que la certitude moyenne (Imprudence) accompagnant chacune des réponses incorrectes.

des états de connaissance partielle. (De Finetti, 1965). Le degré de doute explique les comportements de vérification (dans le dictionnaire par exemple) comme nous l'avons montré expérimentalement (Leclercq & Gilles, 1993, 45).

Avec les QCM, les Degrés de Certitude résolvent en outre (mais c'est un heureux effet secondaire, pas le but principal) le problème du *guessing*, ce qui contribue à la validité d'Acceptabilité* (par les enseignants) des QCM.

Enfin, les Degrés de Certitude montrent leur importante contribution à la validité Informative* des QCM quand les solutions erronées sont plus choisies et surtout avec une certitude plus élevée que la (ou les) solution(s) correcte(s), ce qui est anormal. Cette situation est révélatrice de conceptions erronées (*misconceptions*).

Nous arrêterons ici cette dialectique entre les améliorations apportées aux QCM et les critiques qui continuent à leur être faites, les deux contribuant à améliorer divers aspects de la validité des mesures. L'histoire des QCM n'est pas finie. Nous invitons ceux qui s'en sentent le désir... et le courage d'en écrire quelques pages.

Références

- Barhrick, H.P.(1984). Semantic memory content in permastore : 50 years of memory for Spanish learned in school. *Journal of Experimental Psychology : General*, 120, 1-29.
- Bloom, B.S. (1972). L'innocence en pédagogie, *Education - Tribune Libre*, 135, 14-20.
- Brunswick, E. (1943). Organismic achievement and Environment Probability, *Psychological Review*, 50, 255-272.
- Choppin, B.H. (1975). Guessing the answer on objective tests, *British Journal of Educational Psychology*, 45, 206-213.
- Cronbach, L. & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- 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.
- Davis, F.B. (1946). *Analyse des items*. Louvain : Nauwelaerts (1966)
- De Finetti, B. (1965). Methods for discriminating levels of partial knowledge concerning a test item, *British Journal of Mathematical and Statistical Psychology*, 18, 87-123.
- Ebel, R.L. (1969). Expected reliability as a function of choices per item, *Educational and Psychological Measurement*, 29, 565-570.
- Gardner, H. (1996). *Les Intelligences Multiples* (traduit de *Multiple Intelligences*, 1993). Paris : Retz.
- Gilles, J.L. (1999). Apports des mesures métacognitives lors d'un test de compréhension d'un article scientifique, in C. Depover & B. Noël (Eds), *Approches plurielles de l'évaluation des compétences et des processus cognitifs*, Actes de la 12^e Conférence de l'ADMEE Mons : UMH-FUCAM, 19-30.
- Green, D. R. (1998). Consequential aspects of the validity of achievement tests: A publisher's point of view. *Educational Measurement*, 17, 16-19, 34.
- Leclercq, D. (1986). *La conception des QCM*. Bruxelles : Labor.
- Leclercq, D. & Gilles, J.L.(1993). *Hypermedia : Teaching Through Assessment*, in D. Leclercq & J. Bruno, (1993), *Item Banking, Interactive Testing and Self Assessment*. NATO ASI Series F112. Heidelberg : Springer Verlag, 31-48.
- Leclercq, D. (Ed.) (2003). *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. (2005). *Edumétrie et docimologie pour praticiens chercheurs*. Editions de l'université de Liège.

- Luh, C.W. (1922). The conditions of retention. *Psychol. Monograph*, 31, 142, 401-410.
- Mc Call, W.A. (1920). A new kind of school examination. *Journal of Educational Research*, 1, 33-46.
- Messick, S. (1988, 3^e edition). Validity. In Linn R. (Ed), *Educational Measurement*. NY : Macmillan
- Pieron, H. (1963). *Examens et docimologie*, Paris : Presses Universitaires de France.
- Schurwirth, L.(1998). An approach to the assessment of medical problem solving : *Computerised Case-based Testing*, Ph. D., Rijksuniversiteit Maastricht : Datawyse Universitaire Pres.
- Shufford, E., Albert, A. & Massengill, N.E. (1966). Admissible probability measurement procedures, *Psychometrika*, 31, 125-145.
- Tversky, A. (1964). On the Optimal Number of Alternatives at a Choice Point', *Journal of Mathematical Psychology* 1(2): 386-391.
- Van Naerssen, R.F (1962). A scale for the measurement of subjective probability, *Acta Psychologica*, 20, 2, 159-166.
- West, P.V. (1923). A critical study of the right minus wrong method. *Journal of Educational Research*, 8, 1-9.

T.O.I.S.E.

Taxonomie d'Objectif
Instrumentée au Service de l'Evaluation

1

Recognition – Evocation

Evaluation de la connaissance de mémoire

D. Leclercq

Février 2011

d.leclercq@ulg.ac.be

Livres et articles téléchargeables gratuitement à <http://orbi.ulg.ac.be>

1

D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

La mesure des compétences a-t-elle remplacé la mesure
des ressources nécessaires pour ces compétences ?

NON

- Un interprète simultané ne peut interrompre la conférence parce que lui doit consulter le dictionnaire
- Un chirurgien ne peut arrêter l'opération pour demander via Google ce que sont ces organes qu'il découvre dans le ventre du patient
- Etc.

2

D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Gros plan sur 2 des qualités d'un dispositif d'évaluation des étudiants ou validités...

- **E**cologique (authenticité – proximité au réel)
- **T**héorique (ex : Taxonomie de Bloom)
- **I**nformative (diagnostique et/ou détaillée, compréhensible)
- **C**onséquentielle (sur les Es, les Profs, l'Univ)
- **P**rédictive (ex : pour les réussites ultérieures)
- **R**eplicable (fiable - *reliable* : peu d'erreur de mesure)
- **A**ceptable (pour les Es, pour le Prof, en temps, en coûts...)
- **D**éontologique (standardisée, exempte de biais)

D.
Lecl
ercq

1. Critères de qualité (validité) d'un dispositif d'évaluation

E = **E**cologique (degré d'authenticité / artificialité, proximité à la réalité, sa complexité)
Exemple

Examen du Permis de conduire Poids lourds



Vraie cabine



Cabine simulée



Jouet

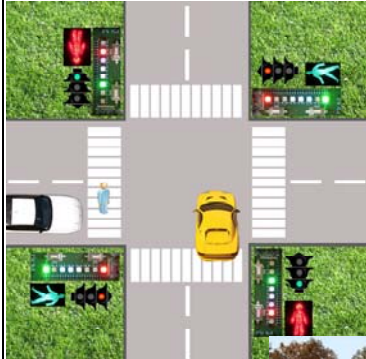
4

D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Critères de qualité (validité) d'un dispositif d'évaluation

E = Ecologique (degré d'authenticité / artificialité, proximité à la réalité, sa complexité)

Examen du Permis de conduire Poids lourds



Plan



Photo



Schéma avec perspective

5

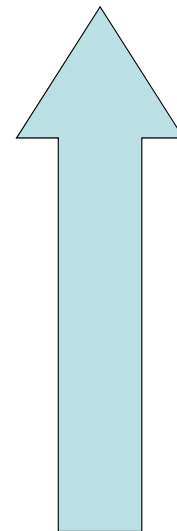
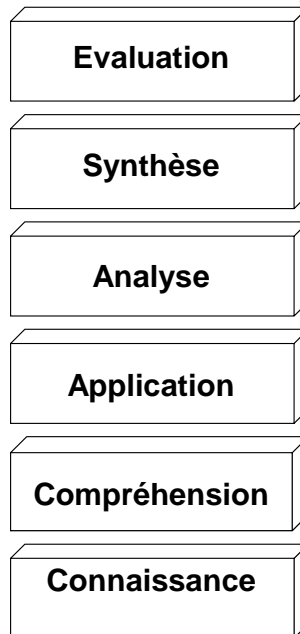
D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

T = théorique

La taxonomie de BLOOM sur les objectifs (processus) cognitifs (1956)



Benjamin Bloom
Université de Chicago

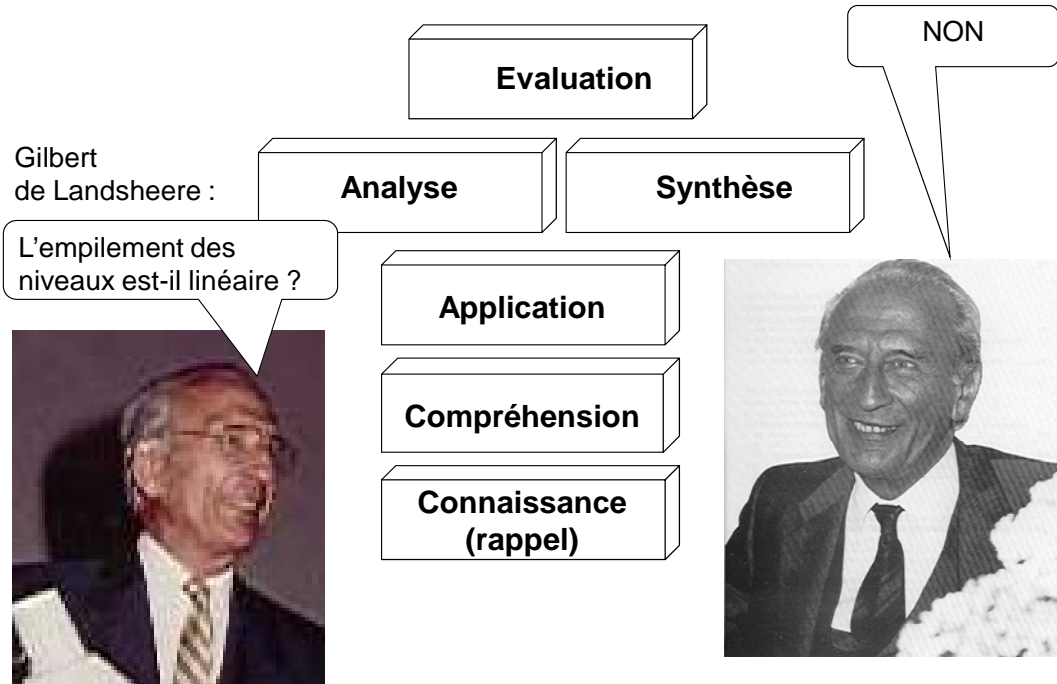


6 processus mentaux
rangés en ordre
croissant d'ouverture
(de complexité)

6

D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

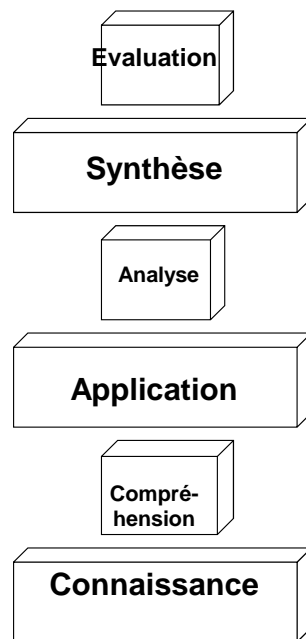
Taxonomie des processus cognitifs de Bloom et al. (1956)



D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

La taxonomie de BLOOM sur les objectifs (processus) cognitifs (1956)

Tous les niveaux sont-ils évalués aussi souvent dans l'enseignement ?




Leclercq :
Non : 3 sont moins utilisées



D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

La taxonomie de BLOOM sur les objectifs (processus) cognitifs (1956)

Le Rappel doit-il être distingué de la Recognition ?



Schurwitz (2000) :

Evaluation

Synthèse

Analyse

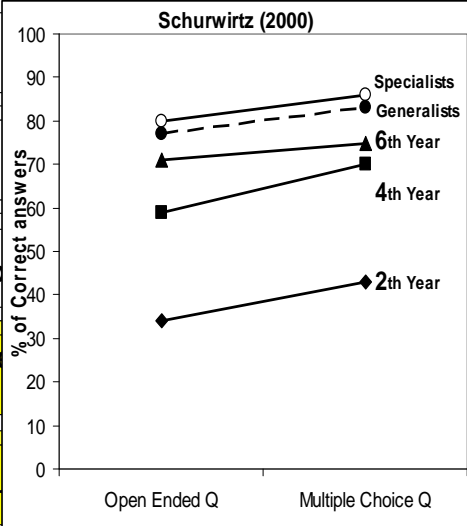
Application

Compréhension

Connaissance (de rappel)

Connaissance (de Reconnaissance)

Schurwitz (2000)



Group	Open Ended Q	Multiple Choice Q
Specialists	~80	~85
Generalists	~78	~82
6th Year	~72	~75
4th Year	~60	~70
2th Year	~35	~45

D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Des situations où...

...on peut se contenter de **reconnaître**.

- Les mots dans un exposé oral en langue étrangère

We offer you self-banking facilities.

...il faut pouvoir **évoquer, produire**.

- Interprétation simultanée (vers SA langue)

Nous vous offrons des facilités de

- Un accent (parisien, marseillais, américain, anglais, belge, suisse, etc.)

S'il est bien fait, ce camembert ? Ma petite dame, il sera rentré avant vous !.....

- Un acteur doit PRODUIRE ces accents :

- Une différence de ton (plagiat) dans un texte écrit.
- Le rédacteur du mémoire doit maîtriser SON ton.
- Le lecteur doit reconnaître Un personnage de BD
- Un dessinateur doit PRODUIRE ces personnages

10

D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Quels outils ?

Evaluation

Synthèse


Analyse

Application

Compréhension

Connaissance (Rappel)

Reconnaissance



Quel est le nom du petit du sanglier ?

QROC

QCM A + Aucune »

QCM A = Simple
1 des k solutions est la RC

.....

1. Haze 2. Laie 3. Veau
4. Goret 5. Poney 6. Aucune

1. Haze 2. Laie 3. Veau
4. Goret 5. Marcassin 6. Poney

11

D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Gros plan sur 2 des qualités d'un dispositif d'évaluation des étudiants ou validités...

- **E**cologique (authenticité – proximité au réel)
- **T**héorique (ex : Taxonomie de Bloom)
- **I**nformative (diagnostique et/ou détaillée, compréhensible)
- **C**onséquentielle (sur les Es, les Profs, l'Univ)
- **P**rédictive (ex : pour les réussites ultérieures)
- **R**eplicable (fiable – ~~reliable~~ : peu d'erreur de mesure)
- **A**ceptable (pour le Prof, en temps, en délais ex : correction de 30 réponses (et certitudes) pour 400 étudiants = 24000 données...à lire ou encoder, et à noter !)
- **D**éontologique (standardisée, exempte de biais)

12

D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Comment rendre le dispositif

• **Acceptable** (pour le Prof, en temps, en délais ex : correction de 30 réponses (et certitudes) pour 400 étudiants = 24000 données...à lire ou encoder, et à noter !)



Une formule : les QCL = Questions à Choix Large

Principes :

1. Encodage sur formulom (**F**ormulaire de **l**ecture **o**ptique de **m**arques) par les étudiants d'un nombre entre 001 et 999.
2. Liste de mots rangés alphabétiquement qui reçoivent un numéro.
3. Liste des mots à disposition
4. Annoncer les concepts et auteurs à connaître de mémoire

13

D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

A CHOIX LARGE AVEC DEGRE

11 0 1 2 3 4 5 6 7 8 9

C 0 20 40 60 80 100

12 0 1 2 3 4 5 6 7 8 9

C 0 20 40 60 80 100

14

D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Cours PSYCHO 058 Matière à mémoriser



Chapitre 1

Imitation
Acquiescement, Identification, Intériorisation, Imitation, Teach
Façonnage, Modelage social, Opérant, Agressivité, Shaping, Vicariant
Créativité Outcome expectancy Self efficacy
Attente d'efficacité, Attente du résultat, Auto-efficacité, Outcome expectancy
Barrières, Belief, Coût, Curabilité, Détectabilité, Effort, Freins, Gravité, Honte, Modelage social, Susceptibilité, Temps
Autoempathie, Autofixation, Difficulté, Développement
Prestigieux
Préconceptions
SITCOMS, Subliminales

Chapitre 2

Code, Galaxie Gutenberg, Mac Luhan
MACON
Action, Instrumentalisme, Intériorisation, Imitation, Opération mentale
Diffuseur, Diffusion, Ecriture, Graphisme, Gratuité et obligatoire, Imprimerie, Imprimeurs, Internet, Lecteurs, Ordinateurs, Opérateurs, Scripteurs
Bayeux
Conventions
Flesh, Lisibilité

p. 15 Confirmation, Internalisation, Progression Sollicitation, Spontanéité
p. 16 Behavioriste
p. 18 Cognitiviste
p. 19 Liaisons, Lien
p. 20 Relations
p. 21 Iceberg
p. 22 Exercisation

Chapitre 4

p. 1 Hypermédias
p. 4 Filet, Hypertexte, Métaphore, Toile, Web
p. 5 Backtracks, Bookmarks, Désorientation, Guided Tours, Hyperespace, Marques, se perdre, Visites guidées
p. 7 Brouter, Browsing, Butinage, Picorage
p. 8 Virtuelle
p. 9 Brouter, Frame, Marge
p. 10 Fictif
p. 11 Maestros
p. 12 ELIZA Rogerienne
p. 13 Turing
p. 14 DéLINE
p. 15 Analytique, Approfondissements, Globaliste Holiste, Sérialiste, Superficielle
p. 16 Contraintes, Croisée, Interaction, Intra-individuelle
p. 17 Note
p. 19 Gains absolus et relatifs, MAX, MIN, Gain POSSIBLE. POST-TEST. PRE. PRETEST. Stvle

15

D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

021: ABSOLU	075: Butinage	129: De Finetti	183: Façonnage	237: Instrumentalisme
022: Absolue(s)	076: Cartésiennes	130: de-Landsheere	184: Feedback	238: Intelligence Artificielle
023: acquiescance	077: Cas programmés	131: Défense	185: FOK	239: Interaction
024: Acquiescement	078: Cercle	132: Défi	186: fermés	240: Intérieurs
025: Action	079: Cerveau	133: Degré(s)-de-certitu	187: Fichiers	241: Intériorisation
026: Adhésion	080: Classement	134: Degré de détails	188: Fictif	242: Internalisation
027: Advance Organizer	081: coconstruction	135: DELIN	189: Fidélité	243: Internet
028: Algorithmes	082: Code(s)	136: Descartes	190: Filet	244: Intra-individuelle
029: Alvéolaires	083: Collaboration	137: Descendante	191: Flash-backs	245: intrapersonnels
030: analogie	084: Cognitiviste	138: Désorientation	192: Flesh	246: Intuition
031: Analyses	085: Collaboration	139: Détectabilité	193: Flexibilité	247: Inutilisable(s)
032: Analytique	086: Communisme	140: Développement	194: Fluidité	248: Isolable(s)
033: Analytiquement	087: compagnonage	141: Devoir	195: Frame	249: Jugement(s)
034: Angle de vue	088: Compensation	142: Diagnostique	196: Franco	250: Lacunaire(s)
035: Animateur	089: Compétition	143: Dialogues	197: Freinet	251: Langage
036: Anticipation	090: Complexe(s)	144: Difficulté	198: Freins	252: Lecteurs
037: Approfondissement	091: Compréhension	145: Diffuseur	199: Gagné	253: Liaisons
038: Apprentissage	092: Compression	146: Diffusion	200: Gains	254: Lien
039: Aptitude	093: Conceptions	147: Dimensions	201: Galaxie	255: limitatives
040: Argument(s)	094: Concepts	148: Discriminance	202: Gauss	256: Linéaire
041: Ascendante	095: Conceptuelle	149: Dissonance-cognit	203: Gestion-de-soi	257: Lisibilité
042: a-sens	096: Concordance	150: Distribution spectr	204: Globaliste	258: Locus-of-control
043: Assistée	097: Cône de Dale	151: Divergence	205: Graphique	259: LOGO
044: Asynchrone	098: Conduite	152: Divergente	206: Graphisme	260: Mac Luhan
045: Attente d'efficacité	099: Confiance	153: Douleur	207: Gratuite	261: Maestros
046: Attente de résultat	100: Confiance	154: Dressage	208: Gravité	262: Mailles
047: Attribution(s)-caus	101: Confirmation	155: Dumb	209: Guemica	263: Majorante
048: Ausubel	102: Conformisme	156: Dyades	210: Guessing	264: mal-définis
049: Auto-correctifs	103: Confrontation	157: Dynamiques	211: Guided-Tours	265: MARCONI

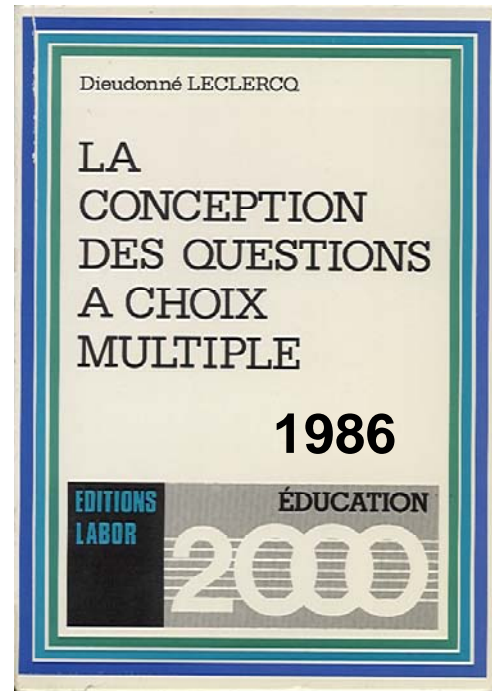


6

D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Le présent PPT
reprend le chapitre 1
de ce livre (épuisé)
de 1986

Pour copies :
<http://orbi.ulg.ac.be>



17

D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

1. Définition d'une QCM

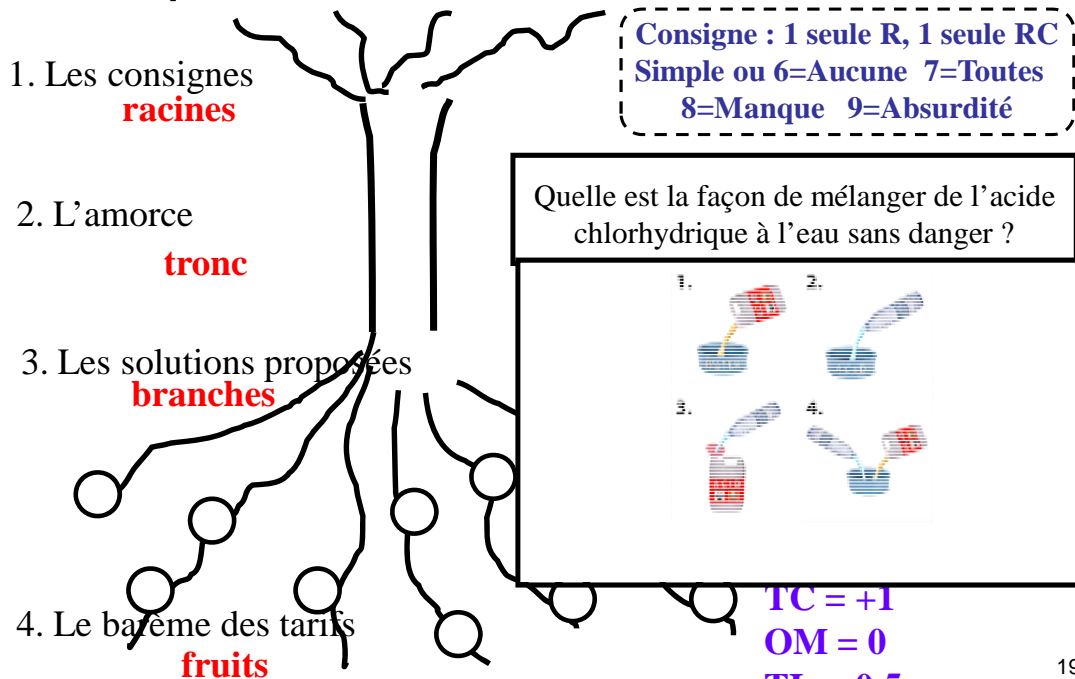
«Une question à laquelle l'étudiant répond en opérant une sélection (au moins) parmi plusieurs solutions proposées, chacune étant jugée (par le constructeur de l'épreuve et par un consensus entre spécialistes) correcte ou incorrecte indépendamment de l'étudiant qui doit y répondre»

(D. Leclercq, 1986)

18

D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Composantes d'une QCM **Arbre inversé**



D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

19

Désavantages des QCM

Validités (faiblesses)

- L'impuissance des QCM à mesurer certains types de performances et d'une présentation orale → **Ecologique**
- L'incapacité des QCM à entraîner à la formulation des réponses
- La tentation de poser des questions de détail } **Théorique**
- La possibilité offerte aux étudiants de fournir des réponses correctes en répondant au hasard → **Informative**
- La présentation de solutions erronées aux étudiants → **Conséquentielle**
- La facilité de copier les réponses du voisin → **Replicabilité**
- Le travail de mise au point d'une QCM → **Acceptabilité**
- Les étudiants n'y sont pas préparés → **Déontologique**

20

D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Avantages des QCM A

Exercice : Appliquer ETIC PRAD
à cette liste

- La clarté de la question
- La brièveté de la réponse
- La simplicité de la correction
- La capacité de couvrir une matière
- La précision des exigences
- L'objectivité de la correction
- La concision du code permettant les analyses et l'auto-régulation

21

D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

3. Les solutions générales implicites

- Les QCM comportent une et une seule réponse correcte.
- Certaines solutions font appel à votre vigilance :
les solutions générales implicites.
- Elles sont proposées sous les numéros de code suivant :
6-7-8-9 et ne sont pas dactylographiées dans les QCM.

6 ou NUL ou REJET
7 ou TOUTES
8 ou MANQUE
9 ou ABSURDITE

22

D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

3. Les solutions générales implicites

NUL ou REJET :	Aucune des solutions proposées n'est correcte.
TOUTES :	TOUTES les solutions proposées sont correctes simultanément.
MANQUE :	Il est impossible de répondre parce que de l'information (au moins une donnée) NECESSAIRE MANQUE dans l'énoncé de la question (donc pas dans le cours ni dans les connaissances actuelles du problème).
ABSURDITE :	Une ABSURDITE dans l'énoncé rend toute la question sans objet parce qu'il y a par exemple une CONTRE-VERITE dans l'énoncé (il ne s'agit pas ici de dénoncer une absurdité dans les solutions ni dans le livre !).

23

D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Les Solutions Générales Implicites (SGI) pour lutter contre le **curriculum caché** *

*Définition : « *Ce que personne n'enseigne, mais que tout le monde apprend...* » à l'école, à l'armée, dans l'entreprise

Proposition d'exercisation systématique de la **vigilance cognitive** :
ajouter dans la consigne des QCM et même des

Questions à Réponse Ouverte (QRO) les **S**olutions possibles :

6. **Aucune**
7. **Toutes**
8. **Manque** de données dans l'énoncé
9. **Absurdité / impossibilité** dans l'énoncé

Générales car elles concernent TOUTES les questions d'un test

Implicites parce qu'elles ne seront plus rappelées dans les questions₂₄

D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Exemples de QCM SGI

Une seule réponse permise !

SGI en vigueur :

6. **Aucune** car **Autre**
7. **Toutes** en même temps
8. **Manque** de données dans l'énoncé
9. **Absurdité / impossibilité** ds l'énoncé

Q1. « Vienne la nuit, sonne l'heure,
les jours s'en vont, je demeure. »
(Apollinaire).

« **Vienne** » est à quel mode et temps ?

1. indicatif imparfait
2. Impératif
3. conditionnel

Q2. Que signifie « le cours » ?

1. l'entourage du roi
2. Le trajet d'un ruisseau
3. L'enseignement d'un professeur
4. Une forme de chasse à cheval

Q3. Quel âge avait Rimbaud quand
Verlaine avait 5 ans ?

1. Un an
2. Trois ans
3. Dix ans

Q4. Quel est l'antonyme d'*attirance* ?

1. Répulsion
2. Aversion
3. répugnance

25

D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Avantages des QCM SGI

Exercice : Appliquer ETIC PRAD

26

D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

3. Nombre de solutions, nombre de questions et fidélité

Lorsque la réponse consiste à choisir *une* des solutions présentées, il est clair que plus la question présente de solutions, plus il faut de temps pour la lire et donc pour y répondre.

Donc, dans une unité de temps donnée (une heure, par exemple) on pourra obtenir beaucoup plus de réponses à des questions vrai-faux qu'à des QCM à cinq solutions proposées. Si la vitesse de réponse est *double* pour une question vrai-faux que pour une QCM à cinq solutions, alors le recours à l'un ou l'autre de ces deux types de questions a le même impact sur la fidélité des scores au test, comme le montrent les exemples ci-dessous.

Ainsi, imaginons qu'en une heure, on puisse répondre à 60 questions vrai-faux, tandis qu'on ne peut répondre qu'à 30 QCM à 5 solutions proposées. Si on consulte le tableau ci-devant, on constate que dans les deux cas, la fidélité prévue par la formule d'Ebel vaut 0,55.

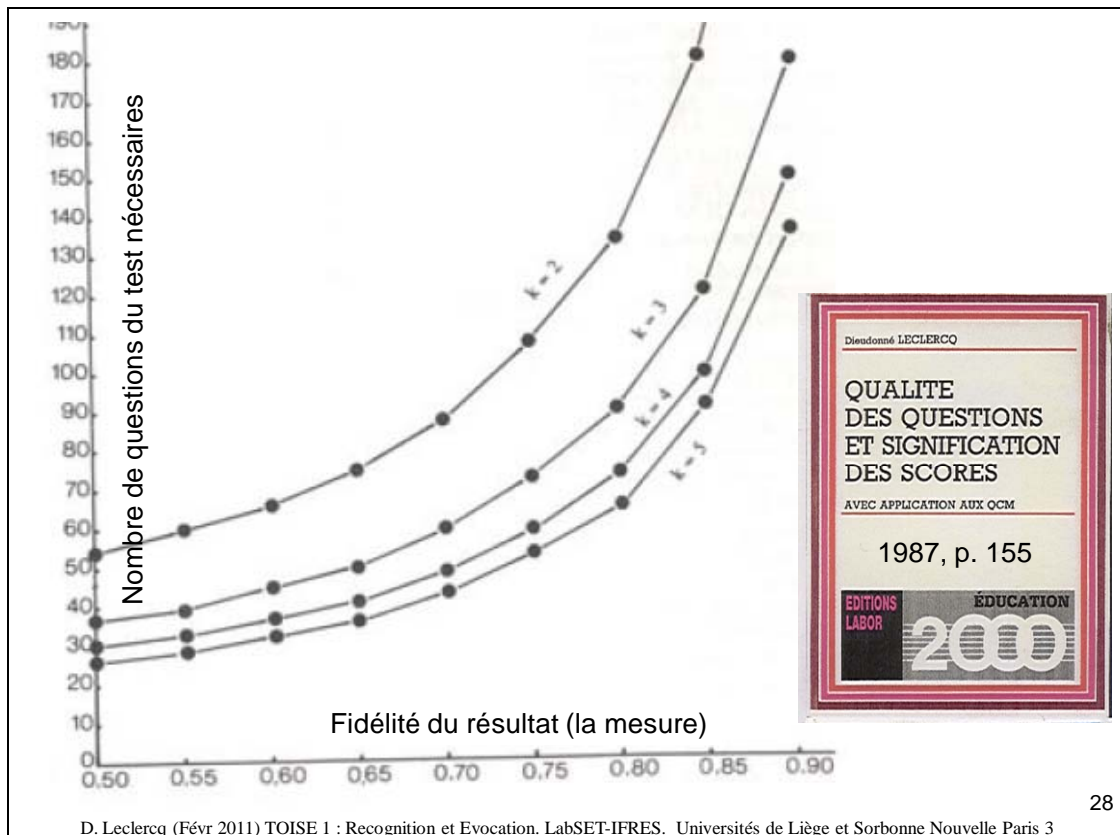
De même, au bout de trois heures, on pourrait répondre à 180 questions vrai-faux, à 90 QCM à 5 solutions proposées et, dans les deux cas, la fidélité prévue par la formule d'Ebel vaut 0,85.

Le graphique ci-dessous permet de lire, pour k (le nombre de solutions proposées) valant 2, 3, 4 ou 5, les nombres de QCM nécessaires pour obtenir (selon la formule d'Ebel), un coefficient de fidélité donné (compris entre 0,50 et 0,90).

Ce graphique peut aussi être lu autrement : étant donné k et un nombre de QCM, on peut toujours selon la formule d'Ebel, la fidélité prévisible ?

D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

27



D. Leclercq (Févr 2011) TOISE 1 : Recognition et Evocation. LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

28

T.O.I.S.E.

Taxonomie d'Objectif
Instrumentée au Service de l'Evaluation

2

de la Compréhension à la Synthèse

D. Leclercq

Janvier 2011

d.leclercq@ulg.ac.be

**Livres et articles téléchargeables
gratuitement à <http://orbi.ulg.ac.be>**

1

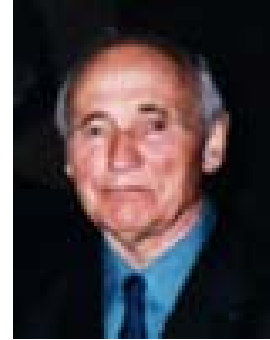
D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Evaluer la COMPREHENSION

2

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Jan Smedslund (1997)



- « Une personne E (comme Elève) **comprend** correctement ce qui est signifié par une expression X produite par une personne P (comme Professeur) si les deux personnes s'accordent sur ce qui, pour P :
- est équivalent à X , c-à-d signifie la même chose que X
- est impliqué par X , c-à-d découle de X
- est contredit par X , ou nié par X
- est non pertinent par rapport à X , c-à-d n'a rien à voir avec X
- Un désaccord sur un seul de ces critères indique une mauvaise compréhension (misunderstanding).
- Cela suppose que E est logique.
- **On ne peut en effet distinguer un manque de compréhension d'un manque de logique. »**

3

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

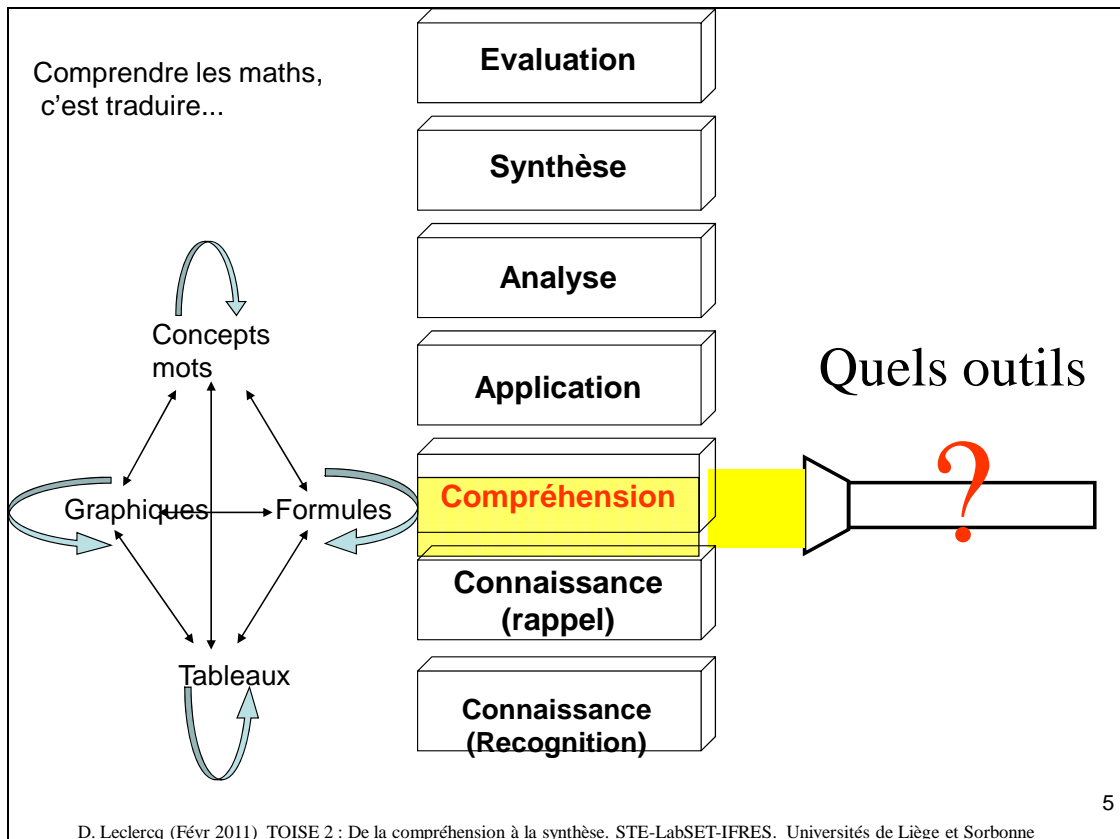
Une métaphore
pour tester la

**Comprehension
en math :**
des traductions
comme un
échangeur
d'autoroutes



4

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3



Quels outils ?

Quel est le nom du petit du sanglier ?

Rappel

Avec dictionnaire

.....

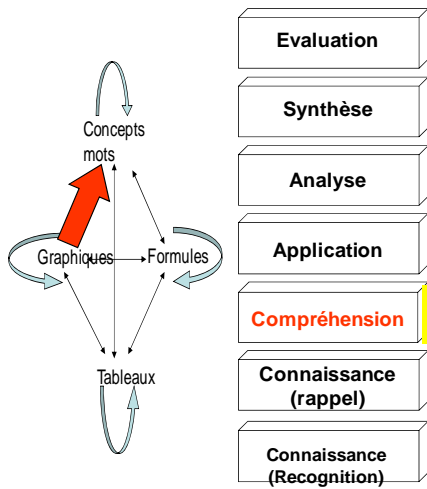
1. Haze 2. Laie 3. Veau
4. Goret 5. poney 6. Aucune

1. Haze 2. Laie 3. Veau
4. Goret 5. Marcassin 6. poney

6

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

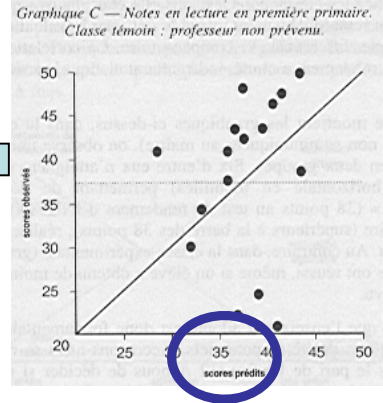
Comprendre = traduire dans tous les sens



- Evaluation
- Synthèse
- Analyse
- Application
- Compréhension**
- Connaissance (rappel)
- Connaissance (Recognition)

1 seule réponse, 1 seule solution correcte soit Simple soit 6. Toutes 7. Aucune 8. Manque 9 Absurdité

- Les points entourés en bleu sont des enfants avec :
1. Une bonne prédiction et un mauvais score observé final
 2. Une bonne prédiction et un bon score observé final
 3. Une mauvaise prédiction et un bon score observé final
 4. Une mauvaise prédiction et un mauvais score observé final

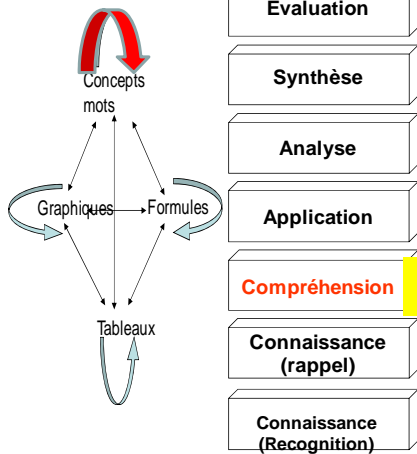


RC = 1

7

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Comprendre = traduire dans tous les sens



- Evaluation
- Synthèse
- Analyse
- Application
- Compréhension**
- Connaissance (rappel)
- Connaissance (Recognition)

Une ou plusieurs Réponses Ouvertes COURTES

« C'est un joli nom, « camarade »,
C'est joli nom, tu sais,
Qui marie cerise et grenade
Aux cent fleurs du mois de mai. »

Quel(s) mot(s) dans cette chanson en fait (font) une chanson politique et pourquoi ?

.....

RC =

Cerise = allusion à la chanson « le temps des cerises, écrite par J-B Clément, emblématique de la Commune de Paris

Grenade : la ville où fut assassiné Federico Garcia Lorca par les franquistes

Mai : le premier mai est le jour des travailleurs

8

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Comprendre = traduire dans tous les sens

Une ou plusieurs Réponses Ouvertes COURTES

« C'est un joli nom, « camarade »,
C'est joli nom, tu sais,
Qui marie cerise et grenade
Aux cent fleurs du mois de mai. »

Quel(s) mot(s) (en plus de « camarade ») Jean Ferrat a-t-il utilisé pour évoquer le Communisme ?

.....

RC =

Cerise = allusion à la chanson « le temps des cerises, écrite par J-B Clément, emblématique de la Commune de Paris

Grenade : la ville où fut assassiné Federico Garcia Lorca par les franquistes

Mai : le premier mai est le jour des travailleurs

9

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Comprendre = traduire dans tous les sens

1 seule réponse, 1 seule solution correcte
dont 6. Aucune 7. Toutes 8. Manque 9 Absurdité!

« C'est un joli nom, « camarade »,
C'est joli nom, tu sais,
Qui marie cerise et grenade
Aux cent fleurs du mois de mai. »

Quel(s) mot(s) (en plus de « camarade ») Jean Ferrat a-t-il utilisé pour évoquer le Communisme ?

1. Cerise
2. Grenade
3. Mois de mai

RC = Toutes car

1. Cerise = allusion à la chanson « le temps des cerises, écrite par J-B Clément, emblématique de la Commune de Paris (19871)

2. Grenade : la ville où fut fusillé Garcia Lorca par les franquistes

3. Mai : le premier mai est le jour des travailleurs

10

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Comprendre = traduire dans tous les sens

1 seule réponse, 1 seule solution correcte
dont 6. Aucune 7. Toutes 8. Manque 9 Absurdité

Dans leurs conclusions de la page 51, les auteurs suggèrent que :

1. Il faut introduire un quota restrictif
2. Il faut limiter les possibilités de changement de carrière
3. L'incidence du triplement est élevée
4. Les efforts déployés pour les étudiants ayant échoué est un investissement perdu

RC = Aucune

11

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Evaluer **l'application**

Quels outils

12

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne

Comprendre = traduire dans tous les sens

1 seule réponse, 1 seule solution correcte
dont 6. Aucune 7. Toutes 8. Manque 9 Absurdité

Concepts mots

Graphiques Formules

Tableaux

Evaluation

Synthèse

Analyse

Application

QCM T A M A

Compréhension

Connaissance (rappel)

Connaissance (Recognition)

Vous, les belles, vous êtes vous laiss....aller ?

Compléter par

é

ée

és

ées

ez

RC = ées.

Application =
Un cas où

- la question est explicite,
- La règle n'est ni nommée ni rappelée
- Toutes les données sont disponibles

13

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Evaluer l'analyse

Evaluation

Synthèse

Analyse

Application

Compréhension

Connaissance (rappel)

Connaissance (Recognition)

Quels outils

14

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne

Comprendre = traduire dans tous les sens

1 seule réponse, 1 seule solution correcte dont 6. Aucune 7. Toutes 8. Manque 9 Absurdité

Evaluation
 Synthèse
Analyse (QCM S T A M A)
 Application
 Compréhension
 Connaissance (rappel)
 Connaissance (Recognition)

Vous êtes vous laiss....aller ?

Compléter par

- é
- ée
- és
- ées
- ez

RC = **Manque** (le contextes) car les 4 premières sont possibles, mais pas **Toutes** car ez est exclu.

Analyse = Un cas présenté comme un cas d'application (donc la règle n'est ni nommée ni rappelée) mais où

- soit la question n'est **pas** explicite (**Absurdité**),
- soit toutes les données nécessaires ne sont **pas** disponibles (**Manque**)

15

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Quino

Le curriculum caché

Evaluation
 Synthèse
Analyse
 Application
 Compréhension
 Connaissance (rappel)
 Connaissance (Recognition)

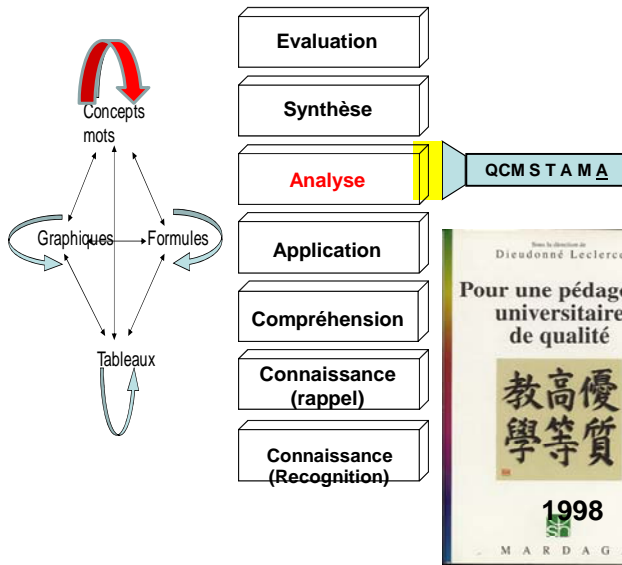
Enjeu : Mesurer la détection d'erreurs ou la vigilance cognitive et faire contreponds au **Curriculum caché**

- ☛ Quand on vous pose une question, vous devez répondre
- ☛ Il y a une réponse à toute question
- ☛ Une question posée par l'autorité est bien posée

16

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Comprendre = traduire
dans tous les sens



1 seule réponse, 1 seule solution correcte
dont 6. Aucune 7. Toutes 8. Manque 9 Absurdité

A la page 16 de son livre ci-
contre, lequel de ses 4
principes de la page 53
Leclercq applique-t-il ?

Le 1 Le 2 Le 3 Le 4

RC = 9. Absurdité : ce
n'est pas Leclercq qui
a écrit la page 16, car
il est l'éditeur du livre
(il n'a pas écrit tous les
chapitres).

17

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Comment distinguer la compréhension de l'analyse ?

18

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Comprendre = traduire dans tous les sens

1 seule réponse, 1 seule solution correcte
dont 6. Aucune 7. Toutes 8. Manque 9 Absurdité

Evaluation
 Synthèse
 Analyse (highlighted in yellow)
 Application
 Compréhension
 Connaissance (rappel)
 Connaissance (Recognition)

QCM S T A M A

PRIM : A la page 16 de son livre ci-contre, lequel de ses 4 principes de la page 53 Leclercq applique-t-il ?
Le 1 Le 2 Le 3 Le 4

BIS : il y a une absurdité dans l'énoncé. Voyez-vous laquelle ?

19

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Une vision plus subtile des compétences
(observations cours D. Leclercq PEA, 1998)

	Avaient mal répondu (RI)		Avaient bien répondu (RC)	
Lors de la PRIM	35%		65%	
Lors de la BIS	Ont encore mal répondu (RI)	Ont bien répondu (RC)	Ont mal répondu (RI)	Ont encore bien répondu (RC)
	20%	15%	15%	50%
Ni compréhension ni vigilance	Compréhension mais manque de vigilance		Réussite injustifiée	Compréhension ET vigilance

20

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Evaluer

**La Synthèse
Formulation
personnelle
(créativité)**

Quels outils

Evaluation

Synthèse

Analyse

Application

Compréhension

Connaissance (rappel)

Connaissance (Recognition)

21

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne

Exemple of question

Les 4 critères de Torrance's se rapportent aux QR ouvertes

- fluidité
- flexibilité
- originalité
- élaboration

Mesurer la créativité :

Evaluation

Synthèse

Analyse

Application

Compréhension

Connaissance (rappel)

Connaissance (Recognition)

QRO

Pour des performances complexes (Rapports, dissertations, projets, ...) qui comportent plusieurs dimensions

22

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne

Production longue (QROL) Grille de critères de critères (poids cachés)

L'omission est aussi codée 0.

	Exactitude	Certitude
C01 : Exactitude des 3 exemples donnés		
C02 : Degré de détail		
C03 : Originalité		
C04 : Structuration de la réponse		
C05 : Pertinence des liens établis		
C06 : Explications claires		
C07 : Reformulation par ses mots		
C08 : Présence du vocabulaire technique		
C09 : Orthographe d'usage		
C10 : Orthographe grammaticale		

Evaluation

Synthèse

Analyse

Application

Compréhension

Connaissance
(rappel)

Connaissance
(Recognition)

23

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Guy Verhofstadt
Président en exercice de l'Union européenne, Premier ministre de Belgique

Seattle, Göteborg, Gênes... Des dizaines de milliers de personnes qui descendent dans la rue pour proclamer leurs opinions. Une bouffée d'air dans cette ère post-idéologique. S'il n'y avait eu cette violence gratuite, on aurait été tenté d'applaudir. Le mérite du mouvement antimondialiste, c'est de faire contrepoids à une époque où la politique est devenue si stérile, ennuyeuse et technique. Cette contestation est bénéfique pour la démocratie.

Mais au fond, vous, les antimondialistes, que voulez-vous nous dire au juste? A l'instar du «black block», voulez-vous agir avec violence contre toute forme de propriété privée? Ou voulez-vous promouvoir le mouvement «slow food», un club mondain qui publie de prestigieux dépliants faisant l'éloge de la consommation d'une alimentation correcte dans les meilleurs établissements?

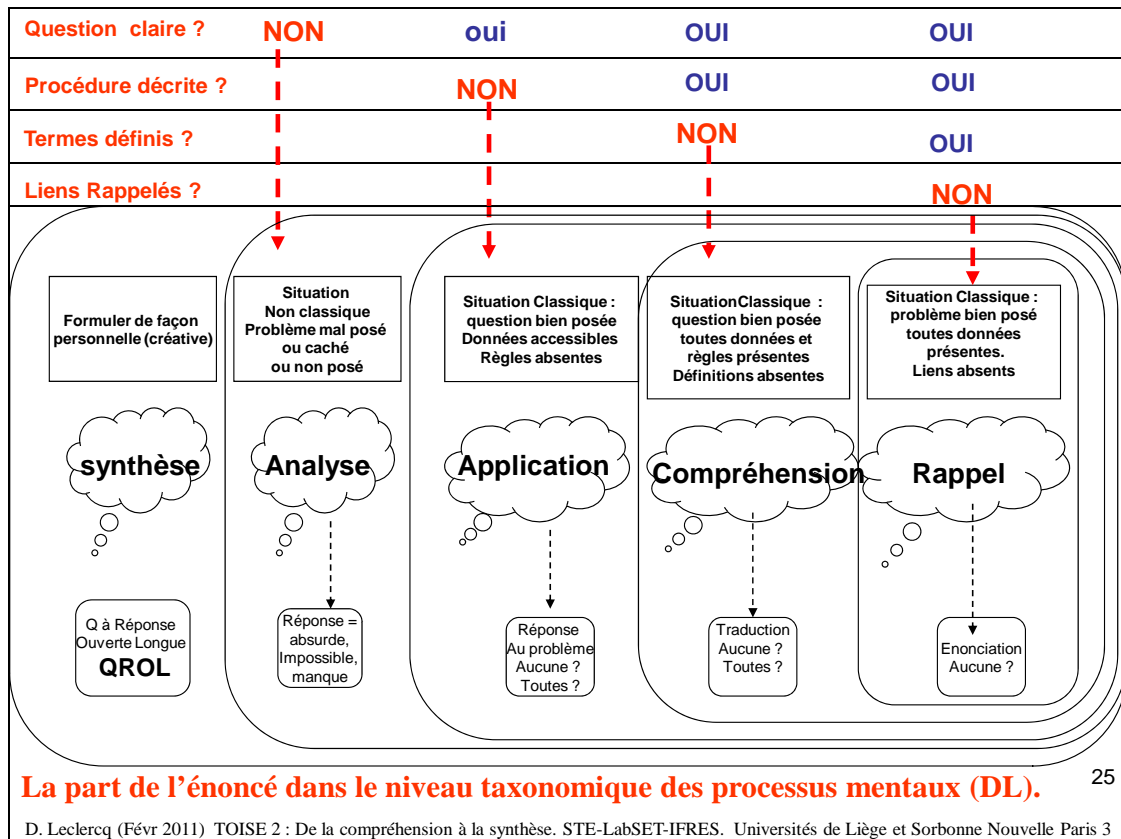
D'où sortent tout à coup ces accusations à l'adresse de la mondialisation?

Jusque récemment, les intellectuels progressistes chantaient eux aussi les louanges du marché mondial censé porter le bien-être et la prospérité dans des pays qui n'avaient connu que la pauvreté et le déclin. Et à juste titre. L'expérience nous apprend que chaque pour cent supplémentaire d'ouverture de l'économie d'un pays génère une augmentation de 1 % du revenu par habitant de ce pays. Ce qui explique la richesse des Singapouriens, une richesse qui contraste radicalement avec la pauvreté qui règne dans une économie fermée comme celle du Myanmar. Bref, jusqu'à Seattle, la mondialisation n'était pas considérée comme un péché, mais comme un bienfait pour l'humanité. En opposition radicale avec

Trouver un titre !

24

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3



2	Colonne 1 : le domaine	Besoins physiologiques satisfaits (plaisir) ou non (souffrance). Ex : nourriture (alimentation - faim), boisson (boire - soif), absence de douleur (indolore - douloureux), chaleur (bien-être - froid - brûlure), etc.
3	Affectif	Sensoriel : perception visuelle, auditive (reconnaître la différence entre le "r" et le "l" en espagnol entendu oral, tactile, olfactive, gustative, etc. (voir taxonomie de Harrow)
4	Sensori-Moteur	Moteur : vocalisation (ex : produire le r roulé de l'espagnol ou le "la" en musique) manipulation, gestes, postures, mouvements (voir tax
5		
6	Cognitif	
7	reconnaître - identifier	Pour un objet ou un phénomène ou une idée présenté, choisir sa bonne appellation parmi plusieurs présentées et l'inverse : reconnaître un objet dénommé parmi une série d'objets (ou situations, ou idées); juger comme "déjà rencontrée (ou non)" une situation, un objet
8	rappeler - évoquer	reproduire à l'identique. Pour un objet ou un phénomène ou une idée présenté, évoquer (produire, formuler, se appeler de) sa bonne appellation, et l'inverse : à partir d'un concept (un ou plusieurs mots), en évoquer la caractéristique demandée ex : la traduction anglaise de "cheval" ou "la formule chimique de l'acide sulfurique" ou "le nom du premier président des USA".
9	comprendre - traduire	Décider si un message (sous forme de mot, phrase, dessin, tableau de nombres, graphique, formule mathématique) produit par un émetteur donné (ce professeur, ce livre, cette émission TV, ce site web...) est équivalent (signifie la même chose) qu'un autre message.
10	appliquer - déduire	En situation simple : A partir d'un énoncé précisant sans ambiguïté le résultat à produire (ex : calculez l'aire) à partir de données immédiatement disponibles, (ex : de ce triangle dont la hauteur et la base sont précisés), produire ce résultat à partir d'un principe (ex : la formule du calcul de l'aire d'un triangle) qui n'est pas fourni. En situation plus complexe : les données sont disponibles moyennant une opération (ou 2 ou 3) intermédiaire(s).
11	analyser - détecter	Détecter : Dans une question mal définie (mais non annoncée comme telle), être capable de détecter un erreur factuelle (ex : une absurdité factuelle du genre "Quand Pavlov décrit les réflexes conditionnés en 1945) ou une erreur de catégorisation (ex : lors de sa migration vers le sud en hiver, le merle...) ou une erreur de principe (ex : de combien faut-il baisser la température d'un litre d'eau à 25° pour la faire bouillir ?). Réorganiser : attribuer à des composants leurs vraies nature, faussement présentées dans l'énoncé d'un problème (ex : requalifier comme victime celui qui est présenté comme accusé).
12	SYNTH-produire - CRÉER-reformuler	Formuler de façon originale (non apprise à l'identique) de façon courte (inventer un nom), plus longue (inventer un titre, une formule) ou plus longue encore (écrire un texte ou prononcer un discours ou construire un schéma)
13	évaluer - juger	En termes de critères internes (sinon, en termes de critères externes, nous sommes dans l'application), porter un jugement de qualité (c'est beau-laid, moral - non moral, peu coûteux - trop coûteux, correct - non correct), cette OPINION pouvant parfois être confrontée à l'opinion d'un expert (le professeur par exemple) ou de plusieurs experts (un jury par exemple).
14	auto-évaluer (DC) - avec réalisme	Juger de la qualité de correction d'une réponse, idéalement sur une échelle de probabilités ou de pourcentages de chances (sinon le sens des termes "pas du tout sûr", "peu sûr", moyennement sûr", "très sûr", totalement sûr", qui forment une échelle ORDINALE (et non métrique) peuvent donner lieu à des interprétations différents selon les personnes.

26

D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Leclercq (2010) Profils docimologiques d'instruments d'évaluation LabSET IFRES 2010	
Définitions de la ligne 1	
Série de VF ou VFG	Quand toutes les questions d'un test sont des Vrai-Faux, on parle de "Vrai-Faux Généralisé" (ou VFG).
Vrai-Faux	Attention à la tendance à l'acquiescement (voir Leclercq, 1986, p.) Une correction for guessing est à envisager (Leclercq, 1987 p.) ou des Degrés de Certitude. Les Degrés de Certitude ne peuvent aller que de 50% (ignorance totale) à 100%. Le score de passage est ici 75% de RC (et non 50%). Incompatible avec la Solution Générale "Absurdité" mais compatible avec "Manque".
Degré de Certitude	-Doivent être exprimés en valeurs métriques : soit des probabilités (de 0 à 1), soit des pourcentages de chances (de 0 à 100). -Peuvent être 0 10 20 30 40 50 etc. mais la précision est illusoire (Leclercq, 1982) La meilleure échelle : 0 20 40 60 80 100 -Susceptibles de déformations "stratégiques" (ne a sdire la vérité sue son de gré de doute) sous l'influence du système de scoring annoncé. (Leclercq, 1982). -Préférer un système de bonus (Leclercq et Poumay, 2008) pour les seuls étudiants réalistes pour leur Confiance et leur Prudence.
QCL	Questions à Choix Large : Toutes les questions d'un test ont les mêmes solutions proposées et celles-ci, très nombreuses (ex : 590), sont classées par ordre alphabétique (pour les retrouver facilement) et dotées d'un numéro (pour encoder facilement sur formulaire destiné à la lecture optique de marque (formulom).
QCM	Une QCM a, comme un arbre, 4 composants : son tronc (la question), ses branches (les solutions proposées), ses racines (sa consigne décrivant le type de question, par ex le nombre de solution correctes, l'existence ou non de SGI) et la façon dxe répondre, par exemple le nombre de réponses, avec ou sans degrés de certitude) et enfin les fruits (les points que l'on gagne ou perd dans les diverses situations).
QCM A	Il existe beaucoup de QCM, selon les consignes (voir ci-dessus). Les plus connues sont - les QCM A = 1 seule solution correcte, une seule réponse permise - les QCM B ou "the Best" = "choisissez la meilleure des solutions "
27	
D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3	

Solution Générale Explicite	
	= valable pour toutes les Questions du test et répétée à chaque question les plus classiques (Leclercq, 1986, p.) sont : Aucune, Toutes, Manque, Absurdité et même 6. Aucune, 7. Toutes, 8. Manque et 9. Absurdité afin que les codes numériques restent les mêmes quel que soit le nombre de solutions explicites
Solution Générale Implicite	
	= annoncé une seule fois et non rappelé à chaque question
QROC	-un mot, une locution de 2 mots, un nombre
QROM	'-une phrase maximum, une formule mathématique, une schéma (simple)
QROL	'-une demi page ou plus
Présentation orale	
28	
D. Leclercq (Févr 2011) TOISE 2 : De la compréhension à la synthèse. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3	



L'évaluation des performances complexes ou des compétences - Principes et illustrations -

Prof. Marianne Poumay, Dr. François Georges,
Université de Liège, IFRES

1

Plan



- Rappel des principes à la base de l'évaluation des compétences (Tardif, 2006), cadrages théoriques (Tardif 2006 ; Prigent, Bernard et Kozanitis 2009)
- Exemples de situations qui permettent d'évaluer des performances complexes ou des compétences :
 - ECOS
 - Portfolios
 - Cartographie conceptuelle

2

Principes à la base de l'évaluation des compétences (Tardif 2006, 104 -119)



1. Rendre compte d'une progression dans une trajectoire de développement ;
2. Mettre l'accent sur les compétences et non sur les ressources ;
3. Déterminer les ressources mobilisées et combinées ;
4. Déterminer les ressources mobilisables et combinables ;
5. Circonscrire les situations de déploiement des compétences ;
6. Documenter la trajectoire de développement à partir de preuves partagées;
7. Rendre compte de l'autonomisation dans le déploiement des compétences;
8. Recourir à des critères multiples dans chaque évaluation ;
9. Intégrer les différences individuelles dans l'appréciation.

Poumay, M. & Georges, F. (2011). *L'évaluation des performances complexes ou des compétences – principes et illustrations*. Liège: ULG – IFRES - LabSET.

3

Nécessité des situations authentiques



L'évaluation en situation authentique :

- simule le contexte d'une situation de travail complexe (nécessite la prise en compte de nombreux paramètres présents dans le domaine du travail : règles, éthiques, comportements humains, budget, etc.) ;
- est réaliste ;
- demande aux étudiants d'accomplir une réalisation plutôt que de simplement dire, répéter ou reproduire (favoriser une démarche personnelle, pas une réponse toute faite);
- favorise la consultation, la rétroaction et l'amélioration (favoriser les échanges avec l'enseignant pour s'améliorer)

Prégent, R., Bernard, H., & Kozanitis, A. (2009).

Enseigner à l'université dans une approche-programme. Montréal: Presses internationales Polytechniques.

Poumay, M. & Georges, F. (2011). *L'évaluation des performances complexes ou des compétences – principes et illustrations*. Liège: ULG – IFRES - LabSET.

4

Nécessite des situations authentiques



Tableau 5.2 Exemples de situations authentiques d'apprentissage et d'évaluation

Création d'une œuvre Exposition Expérimentation Tribunal école Clinique de droit Clinique dentaire, d'optométrie, etc. Création Concert Simulation d'un bureau d'experts Performance artistique	Stage d'observation APP (apprentissage par problème) Étude de cas Présentation (colloque, conférence, etc.) Projet Rédaction d'articles ou d'autres documents à caractère réel Soutenance Stage Etc.
--	--

Prégent, R., Bernard, H., & Kozanitis, A. (2009). *Enseigner à l'université dans une approche-programme*. Montréal: Presses internationales Polytechniques.

Poumay, M. & Georges, F. (2011). *L'évaluation des performances complexes ou des compétences – principes et illustrations*. Liège: ULG – IFRES - LabSET.

5

L'ECOS, présentée par Geneviève PHILIPPE (pharmacie, ULg, 2007)



L'évaluation clinique objective et structurée (ECOS)

Situations cliniques simulées

Facteurs dus à la subjectivité atténués

Contenu élaboré en tenant compte d'objectifs bien définis

- Mode d'évaluation développé par Harden dans les années 1970 pour tester la **compétence clinique** des étudiants médecins
- Circuit de stations d'examen où l'étudiant est confronté à des **tâches professionnelles fréquentes** (patients simulés)
- Très fréquemment utilisée dans les études médicales (ULg : pédiatrie, médecine générale, pharmacie)

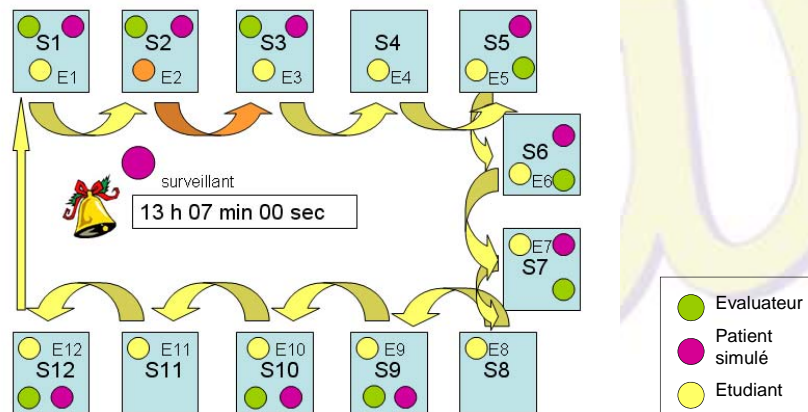
Poumay, M. & Georges, F. (2011). *L'évaluation des performances complexes ou des compétences – principes et illustrations*. Liège: ULG – IFRES - LabSET.

6

L'ECOS en pratique... (I)



- Comporte une succession de **stations d'examen** (10-20) à temps d'examen **chronométré** (7 min) à chaque station ;
- Passage d'une station à l'autre synchronisé → les étudiants ne communiquent pas entre eux :



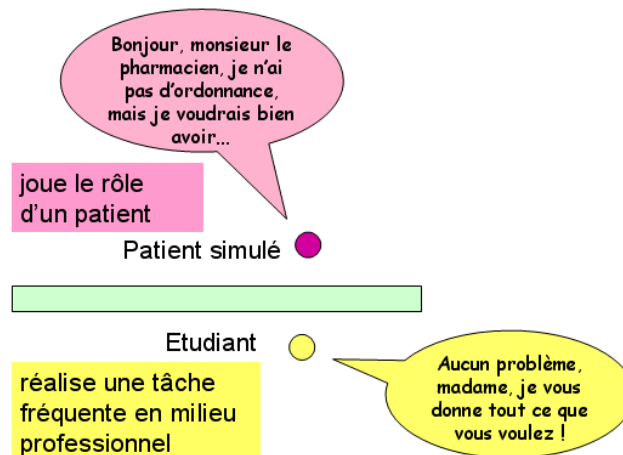
Poumay, M. & Georges, F. (2011). *L'évaluation des performances complexes ou des compétences – principes et illustrations*. Liège: ULG – IFRES - LabSET.

7

L'ECOS en pratique... (II)



- Stations interactives avec intervention de **patients simulés** ou stations non-interactives (tâche à effectuer par écrit)
- **Grilles d'évaluation** reprenant les réponses attendues et pondérées remplies par des évaluateurs en cours d'examen ;



Poumay, M. & Georges, F. (2011). *L'évaluation des performances complexes ou des compétences – principes et illustrations*. Liège: ULG – IFRES - LabSET.

8



Exemple de scénario

(pédiatrie, ULg, JP Bourguignon)

« Olivier, qui est âgé de 2 ans 1/2, a toujours été en bonne santé, à l'exception d'une otite vers 1 an. Il est malade depuis 24 heures : il présente une température qui est montée jusqu'à 40°C. Il a le nez qui coule et tousse un peu. Il a mal au ventre. Il n'a pas eu de vomissements ni de diarrhée. Il continue à boire normalement, mais son appétit a diminué. Au moment de la conversation téléphonique, il court partout dans l'appartement. Il n'a pas d'éruption sur la peau. Il pèse 13 kg.

Le grand-père lui a fait prendre 2 pipettes deTEMPRA ; la fièvre est tombée à 37°5. »

Poumay, M. & Georges, F. (2011). *L'évaluation des performances complexes ou des compétences – principes et illustrations*. Liège: ULG – IFRES - LabSET.

9

Département de Pédiatrie - Université de Liège
EVALUATION CLINIQUE OBJECTIVE ET STRUCTUREE (E.C.O.S.)
Grille de réponse

34

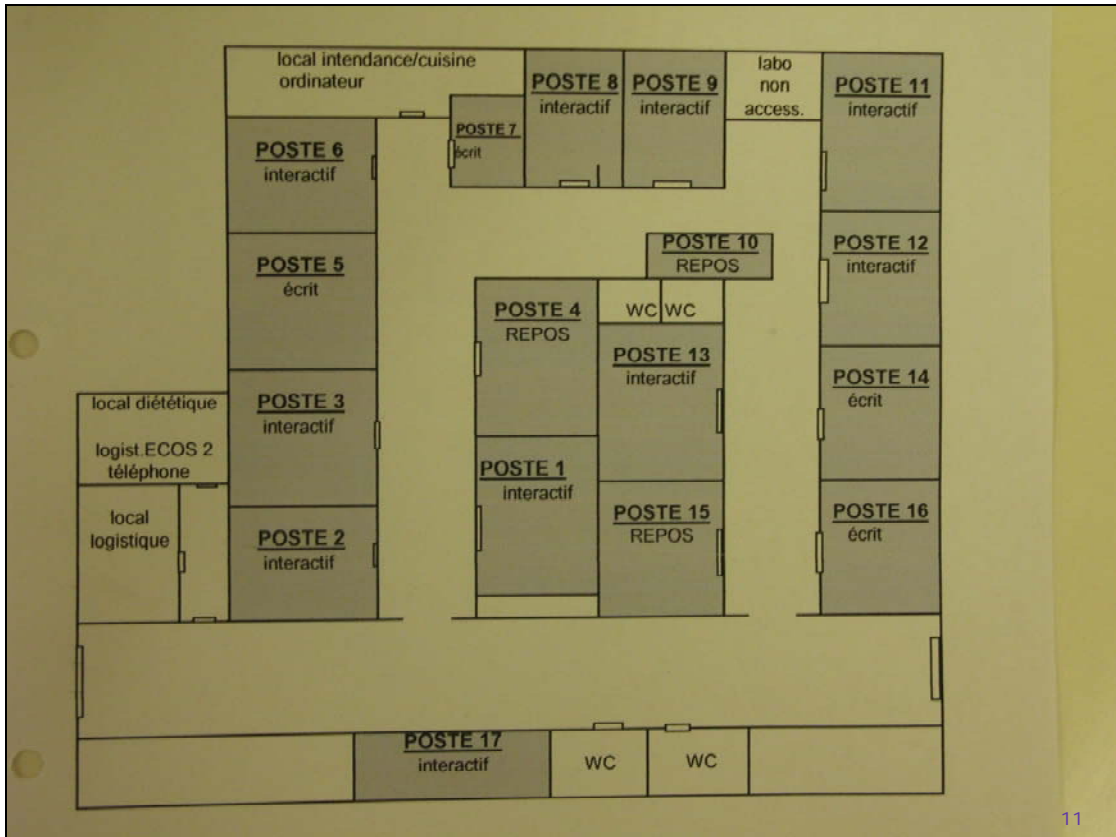
Nom de l'étudiant :
Nom de l'examineur : Docteur J.M. CHANTRAINE
Date : 24.02.96

Station n° : 10

Grille de réponse

(Bourguignon, S.D.)

ITEMS	Cocher	SCENARIO	Valeur	Cote
1. Quel âge a-t-il ?		2 ans 1/2	1	
2. Depuis quand a-t-il de la fièvre ?		24 heures	2	
3. Niveau de la fièvre		40°C	1.5	
4. A-t-il mal quelque part ?		Au ventre	0.5	
5. Quel fut l'effet de l'antipyrétique ?		↓ 37°5C	1	
6. A-t-il déjà convulsé ?		Non	1	
7. Dans quel état est-il ?		Bon état	1	
8. Y-a-t-il des malades dans la famille ?		Non	1	
9. A-t-il toujours été en bonne santé ?		1 otite il y a 1 an 1/2	0.5	
10. Rhinorrhée ?		Nez coule	0.5	
11. Toux ?		Un peu	1	
12. Vomissements ?		Non	0.5	
13. Diarrhée ?		Non	1	
14. Appétit ?		Diminué	1	
15. Boit-il ?		Oui	1	
16. Eruption ?		Non	1.5	
17.				
- L'étudiant suggère de voir l'enfant en consultation.			2	
- L'étudiant suggère de continuer le traitement symptomatique			1	
- L'étudiant demande au grand-père de le rappeler si l'évolution n'est pas bonne/ou de l'amener en fin de journée			1	
			20	/20



11



12

Station 13 :

Une dame vient vous demander des produits à base de plantes.

Répondez à chacune de ses assertions et conseillez-la au mieux.

13



14

Portfolios



Poumay, M. & Georges, F. (2011). *L'évaluation des performances complexes ou des compétences – principes et illustrations*. Liège: ULG – IFRES – LabSET.

15

Le portfolio - **définition**



- « Le portfolio est un échantillon de preuves, sélectionnées par l'étudiant pour rendre compte de ses apprentissages » (Tardif, 2006)
- C'est un dossier (classeur, blog, site...) qui permet de collectionner, puis de sélectionner des traces (écrits, photos, références, schémas, cartes conceptuelles, ...) qui témoignent d'un apprentissage ou même d'un développement

Poumay, M. & Georges, F. (2011). *L'évaluation des performances complexes ou des compétences – principes et illustrations*. Liège: ULG – IFRES – LabSET.

16

Apport du portfolio pour l'étudiant du supérieur



- Démultiplie l'apprentissage car prise de conscience à la fois du « processus » et du « progrès » par l'étudiant.
- Permet le développement de compétences dites « génériques » (auto-évaluation, pensée critique, communication, analyse, synthèse et intégration des connaissances).
- Fournit aussi à l'étudiant une « trace » de ses apprentissages (bien plus qu'un relevé de notes), qu'il peut inscrire sur un CV.
- Dans les programmes professionnalisants, il peut servir de mécanisme « accélérateur de carrière » en permettant à l'étudiant de relier plus facilement ses apprentissages à sa pratique professionnelle (et ainsi obtenir une promotion verticale ou horizontale).

Poumay, M. & Georges, F. (2011). *L'évaluation des performances complexes ou des compétences – principes et illustrations*. Liège: ULG – IFRES - LabSET.

17

Contenus des portfolios



- Déterminés par l'apprenant, l'enseignant, ... ou les deux.
- Constitués de (à titre indicatif) :
 - productions (preuves d'apprentissage) réalisées en cours d'apprentissage;
 - commentaires sur ses productions (analyse des preuves);
 - commentaires par pairs, tuteurs ou maîtres de stage sur ses productions;
 - bilan personnel de son action;
 - plan d'action, perspectives de poursuite de l'apprentissage.

Poumay, M. & Georges, F. (2011). *L'évaluation des performances complexes ou des compétences – principes et illustrations*. Liège: ULG – IFRES - LabSET.

18

réalisation de préparations (y compris précision et propreté) et application des principes de calcul		
Insuffisant	Manque de précision ou de soin dans la réalisation de préparations, ne respecte pas les consignes de travail, nécessite une aide régulière et une surveillance constante	<input type="checkbox"/>
Basique	Est capable de réaliser avec précision et propreté les préparations simples, mais nécessite une aide et une surveillance régulières	<input type="checkbox"/>
Bon	Est capable de réaliser seul, avec précision et propreté, la plupart des préparations, mais nécessite une aide ponctuelle pour la réalisation de préparations originales ou complexes	<input type="checkbox"/>
Excellent	Est capable de réaliser toute préparation (y compris les préparations originales ou complexes) et ce, sans aide, avec précision et propreté	<input type="checkbox"/>

Preuves les plus probantes :		
Date de la preuve	Titre / résumé	Validation du maître de stage ?

Source :
Geneviève Philippe,
ULg

Commentaire / justification :

19

Détail de la prescription manuscrite :

Titre : Jean Packer 1/06/2008

Texte principal : Détection d'une erreur au sein d'une prescription. Le médecin a probablement confondu la scopolamine et la butylscopolamine. La dose prescrite est 10mg 5x/j, correspondant à 50 x la DMS.

Prescription : Bromure de Scopolamine 10 mg, Navalgine 150 mg, Of 20 gélules. S. 1 gélule jusqu'à 5 fois par jour, 6 douleurs.

Annotations : '10mg' (en rouge), '10mg' (en bleu), 'OK à remplacer par butylscopolamine' (en bleu), '1/06/2008' (date), 'En accord avec mon maître de stage, j'ai contacté le médecin par lui signaler son erreur. Ensemble, nous avons décidé de réaliser la préparation en remplaçant la scopolamine par de la butylscopolamine (= butylhyosine)' (en bleu).

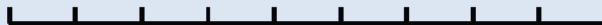
Source :
Geneviève Philippe,
ULg

Exemple de preuve n°2 : compétence « résolution de problèmes et sens des responsabilités »

Sélectionner des données pertinentes, les confronter entre elles

Très peu maîtrisé

Bonne maîtrise

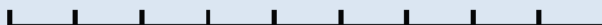


Illustrations :

Ecouter avec empathie

Très peu maîtrisé

Bonne maîtrise



Illustrations :

Source : grille de stage en psychologie clinique
Jean-Jacques Detraux, Ulg, 2009
Compétence 1 : « travailler la demande »

Une carte conceptuelle - Définition



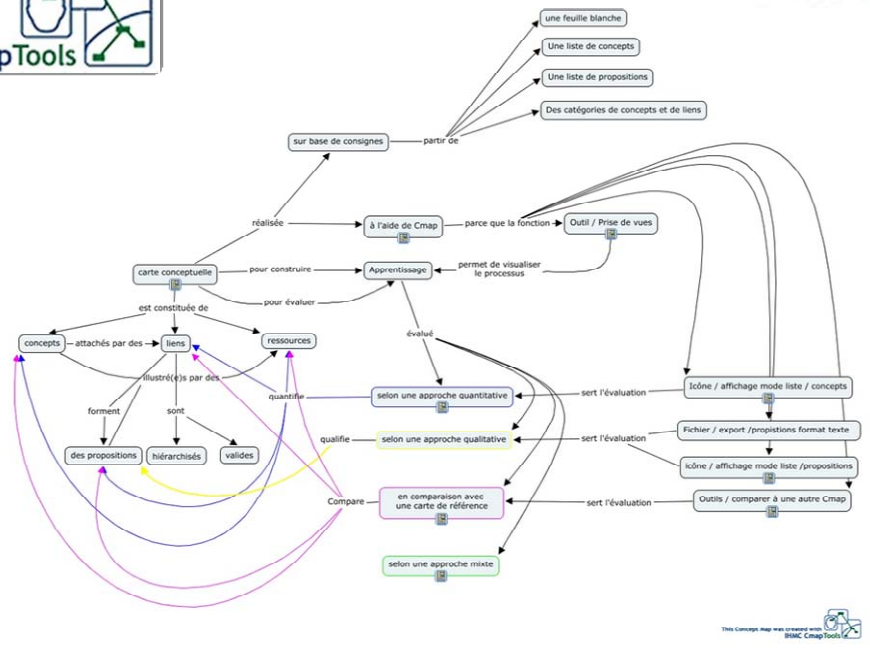
« Une carte conceptuelle est

une **représentation graphique d'un domaine de la connaissance** tel que **perçu par un ou plusieurs individus**.

Cette perception – **évolutive** –

établit des liens entre des concepts – interconnexions sémantiques – selon des règles plus ou moins formelles. »

Laflamme, A. (2006). *Les cartes conceptuelles*, Montréal : CEFES.



Cartes conceptuelles - **Avantages**

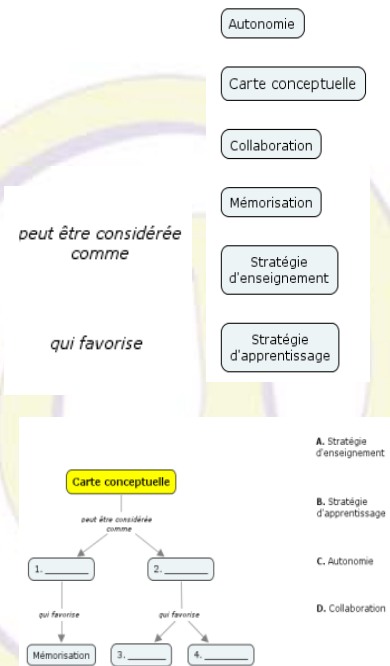
Produite par l'apprenant,

- « elle permet à l'**apprenant** de **verbaliser ses connaissances** d'un domaine particulier, et donc de **découvrir ses lacunes** sur certains aspects (pensée réflexive)
- À l'**enseignant** de **découvrir les conceptions de l'apprenant**, ou plus généralement d'**évaluer ses connaissances**, que ce soit de manière formative ou sommative. »

Delorme, F. (2005). *Évaluation et modélisation automatiques des connaissances des apprenants à l'aide de cartes conceptuelles*. Thèse de doctorat. p. 50-51 cité dans Belflamme, 2006.

Des cartes pour 4 types de tâches

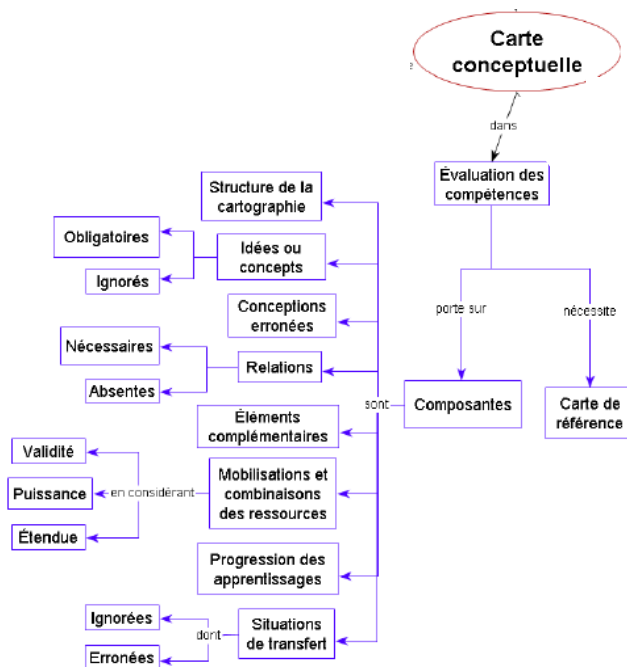
1. Elaborer une carte conceptuelle depuis une feuille blanche
2. Elaborer une carte conceptuelle depuis une liste de concepts préétablis à hiérarchiser et à exemplifier
3. Elaborer une carte conceptuelle depuis une liste de concepts et de mots de liaisons préétablis à hiérarchiser et à exemplifier
4. Compléter une carte conceptuelle depuis une liste de concepts déjà hiérarchisée (cases vides)



Poumay, M. & Georges, F. (2011). *L'évaluation des performances complexes ou des compétences – principes et illustrations*. Liège: ULG – IFRES - LabSET.

25

Critères d'évaluation des cartes

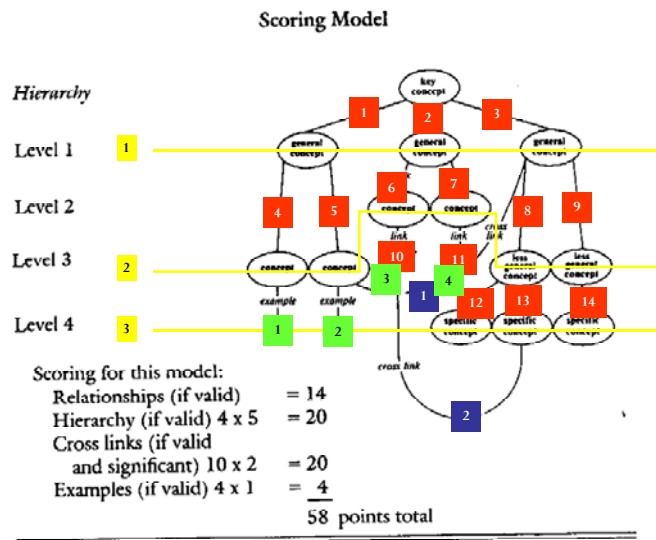


Tardif, J. (2008). *Des usages multiples de la cartographie conceptuelle dans la formation*. Conférence organisée à l'IFRES le 20 novembre 2008.

Poumay, M. & Georges, F. (2011). *L'évaluation des performances complexes ou des compétences – principes et illustrations*. Liège: ULG – IFRES - LabSET.

26

Méthode d'évaluation des cartes



Comptage
(méthode
Novak &
Gowin, 1984)

Poumay, M. & Georges, F. (2011). *L'évaluation des performances complexes ou des compétences – principes et illustrations*. Liège: ULG – IFRES - LabSET.

27

Exemple d'usage ULg



Utilisation des cartes au sein de la formation Formasup (LabSET)

Production d'une carte qui présente la triple concordance de son cours
(ex. carte de Ellen Harry - HEC)

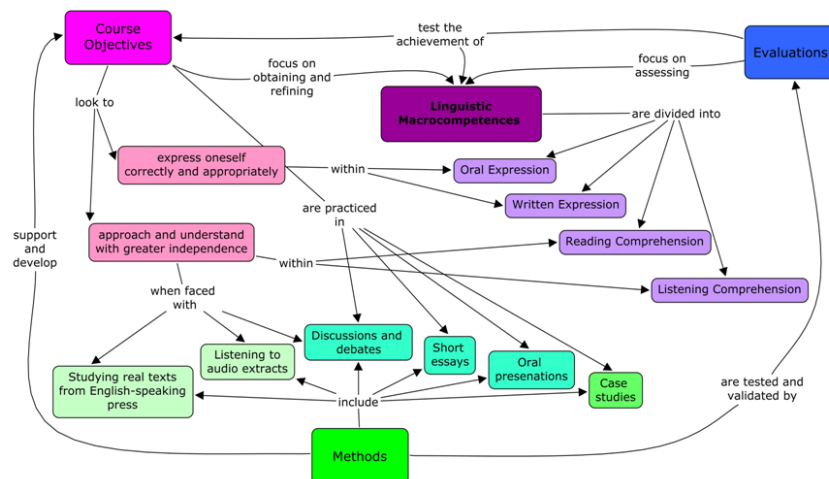


Figure 1. Triple alignment in English Level 2

Poumay, M. & Georges, F. (2011). *L'évaluation des performances complexes ou des compétences – principes et illustrations*. Liège: ULG – IFRES - LabSET.

28

T.O.I.S.E.

Taxonomie d'Objectif
Instrumentée au Service de l'Evaluation

3

L'évaluation- jugement -certitudes

D. Leclercq

Février 2011

d.leclercq@ulg.ac.be

Livres et articles téléchargeables gratuitement à <http://orbi.ulg.ac.be>

1

D. Leclercq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

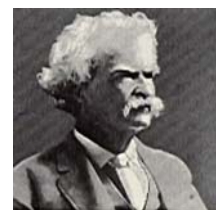
1. Ignorance, connaissance et doute

L'ennui dans ce monde est
que les imbéciles
sont sûrs de tout
et les sages pleins de doute.



Bertrand RUSSEL

Ce n'est pas ce que nous ignorons qui nous nuit,
c'est ce que nous savons **et qui est faux.**



Mark TWAIN

Autocognition avec réalisme

2

D. Leclercq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Methods of discriminating levels of partial knowledge concerning a test item

Brit. Journ. of Math. & Statist. Psychol., 1965, 18, 87-123.

- 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 toute méthode de
notation (p. 111)



Bruno De Finetti

....et de mesure
des gains d'apprentissage. 3

D. Leclercq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

ETIC PRAD

Validité Théorique



Bruce Choppin

3 modèles de l'activité mentale d'une personne répondant à un QCM

Modèle 1 :

- soit sait
- soit au hasard

Modèle 2 :

- soit sait
- soit au hasard éclairé

Modèle 3:

- mise en ordre des
solutions par ordre de
plausibilité

Scoring - notation

Correction for
guessing classique

Correction for
guessing spéciale

Degrés de Certitude

4

D. Leclercq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

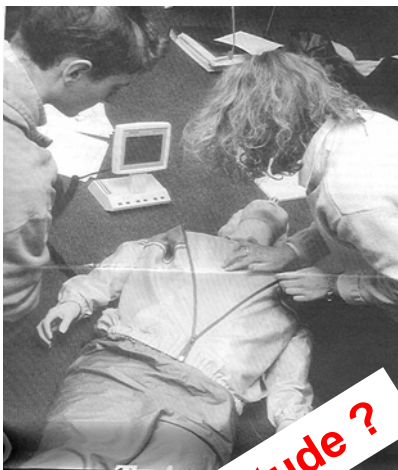
Pilotes et mécaniciens d'avion



5

D. Leclercq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Urgentistes



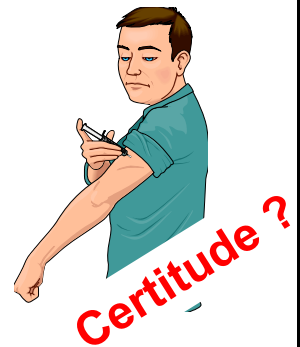
Certitude ?

Infirmières



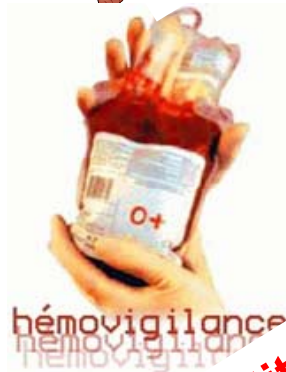
Certitude ?

Education du patient

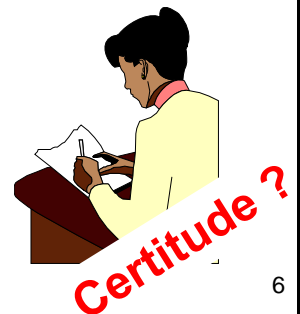


Certitude ?

Étudiant(e)s



Certitude ?



Certitude ?

6

D. Leclercq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

1. Une consigne probabiliste

Admissible Probability
Measurement Procedures



Emir Shuford (Dallas)

- 1-not sure at all
- 2-weakly sure
- 3-average
- 4-partly sure
- 5-fairly sure
- 6-strongly sure

Une échelle de probabilité :

0 20 40 60 80 100

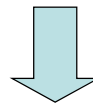
7

D. Leclercq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

3. Spectre de la qualité des performances en 13 positions

Réponses +
Certitudes

0 20 40 60 80 100



Réponses Incorrectes

Réponses Correctes

-100 -80 -60 -40 -20 0 OM 0 20 40 60 80 100

Spectral de la qualité des réponses

Hémi-Spectre gauche (des RI)

Hémi-Spectre droit (des RC)

8

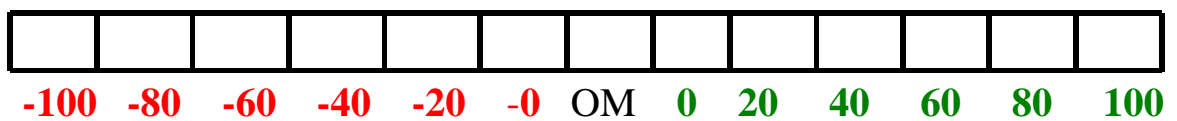
D. Leclercq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Essayons avec le *Leclercq Guessing Game*

0 20 40 60 80 100

Lettre _____
DC _____

Lettre _____
DC _____



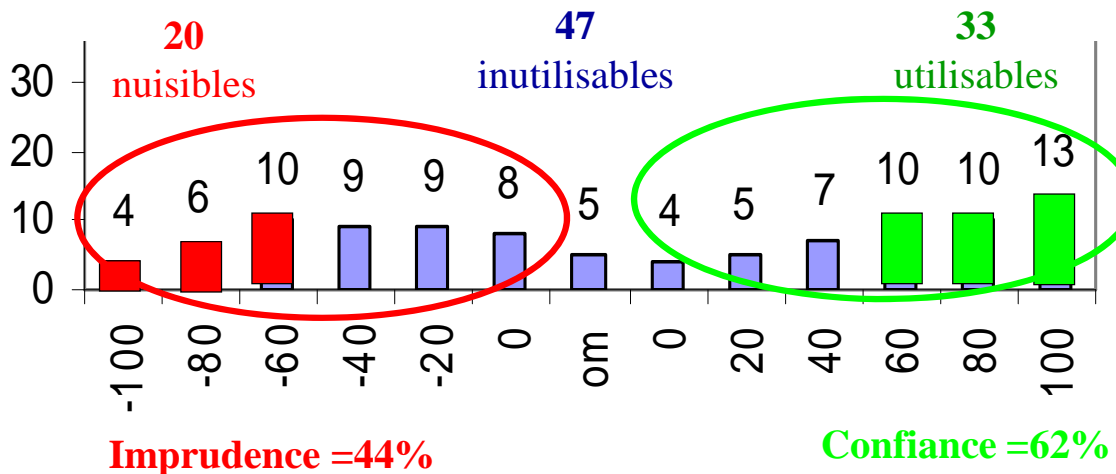
Réponses Incorrectes

Réponses Correctes

9

D. Leclercq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

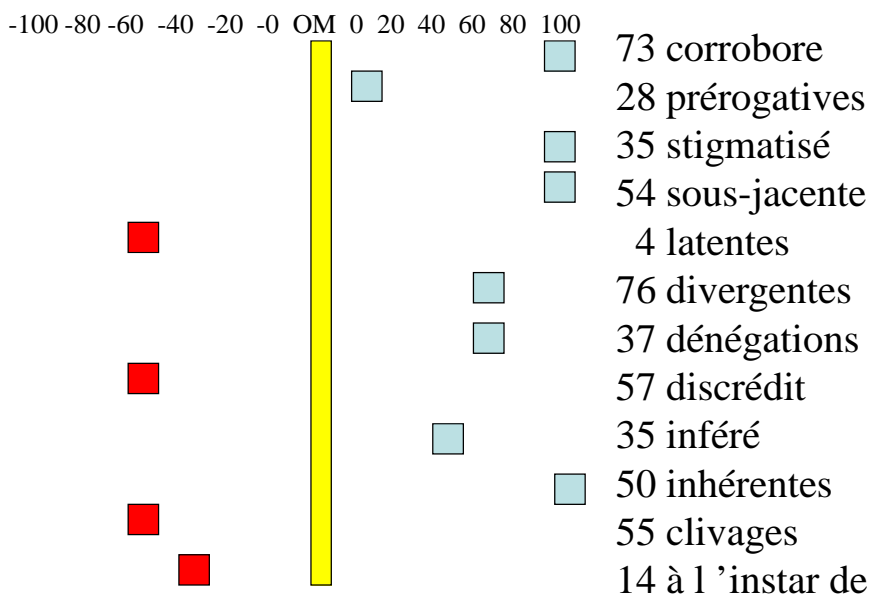
Vocabulaire : Spectre des performances pour le test (4000 étudiants)



10

D. Leclercq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

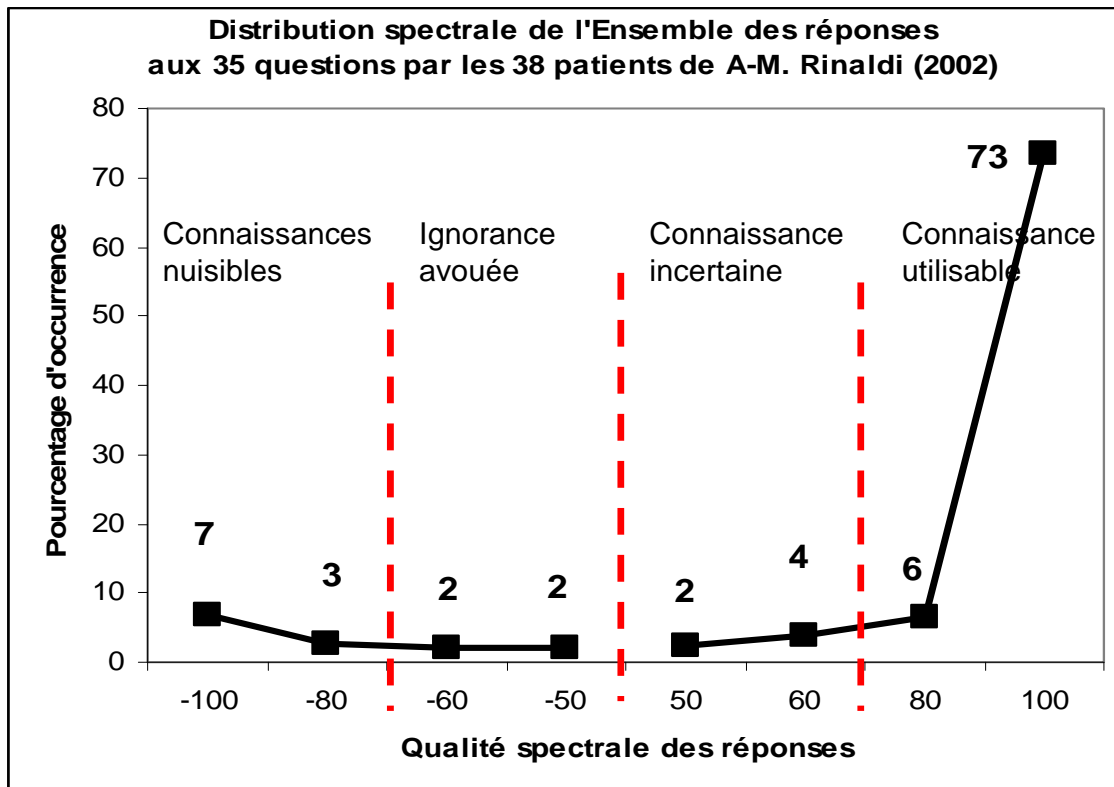
Feedback aux étudiants : répartition spectrale des réponses



11

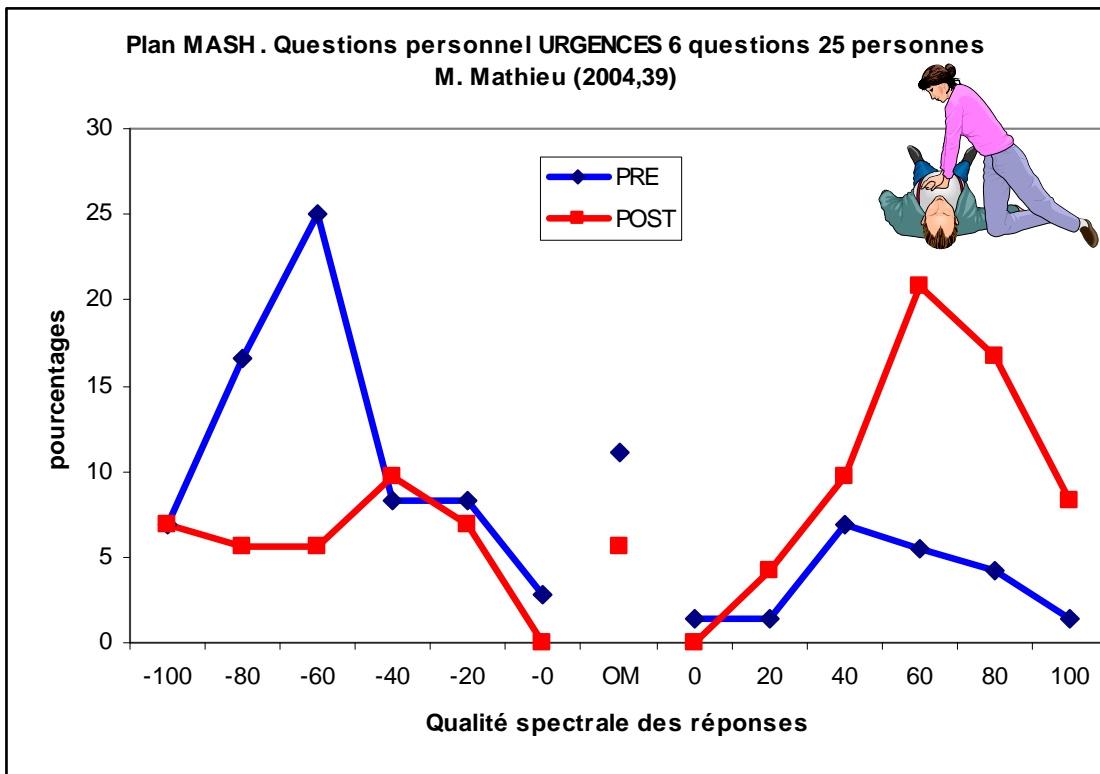
D. Leclercq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Distribution spectrale de l'Ensemble des réponses aux 35 questions par les 38 patients de A-M. Rinaldi (2002)



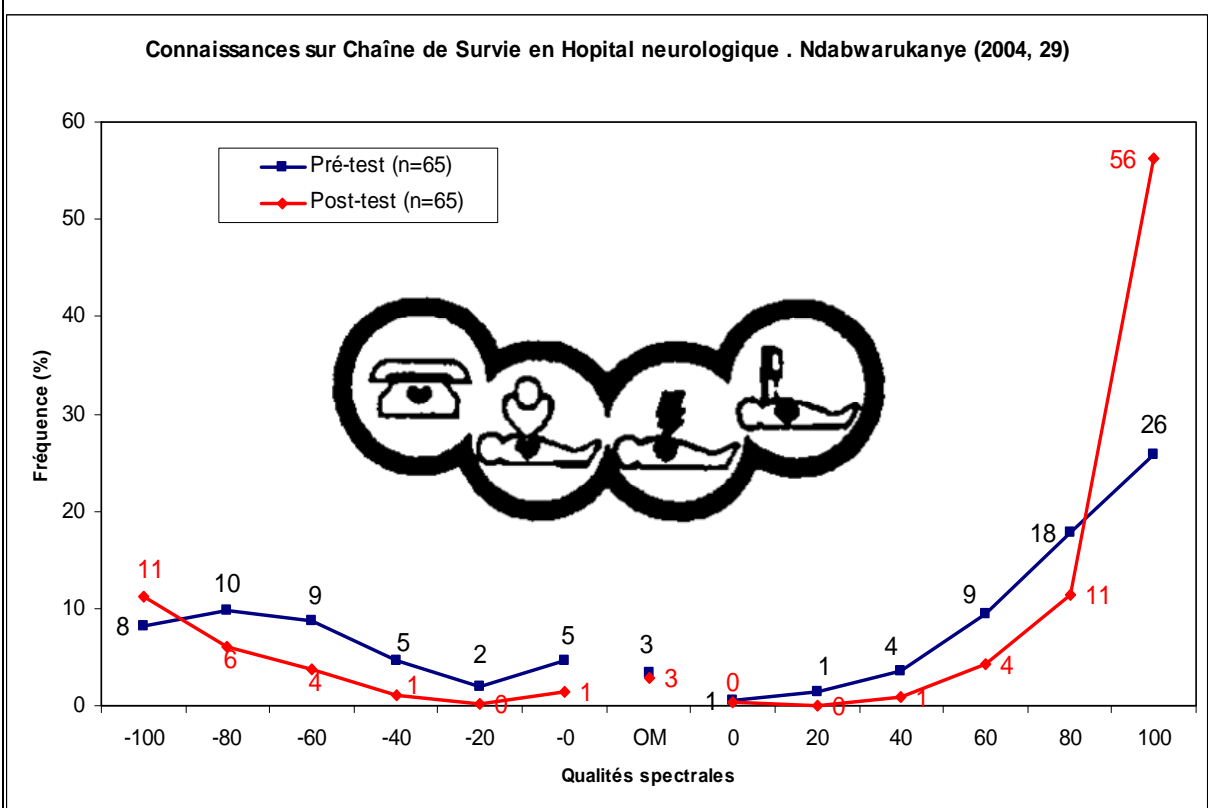
12

D. Leclercq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3



13

D. Leclercq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3



14

D. Leclercq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Quelles matières ré-enseigner ?

-100	-80	-60	-50	50	60	80	100	
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

15

D. Leclercq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Compétences isolées

Lire efficacement

Faire des hypothèses

Esprit critique

FOK

16

D. Leclercq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Quand on n'a que 3 minutes						Si vous deviez donner un synonyme de	
Q1 : Prurit	0	20	40	60	80	100
Q2 : Implicite	0	20	40	60	80	100
Q3: Intrinsèque à	0	20	40	60	80	100
Q4: (un) prescrit	0	20	40	60	80	100
Q5: (un) feed-back	0	20	40	60	80	100
Q6: Appréhender	0	20	40	60	80	100
Q7: Promouvoir	0	20	40	60	80	100
Q8: Drastique	0	20	40	60	80	100
Q9: Compétence	0	20	40	60	80	100
Q10: Recommandation	0	20	40	60	80	100
Q11: Carence	0	20	40	60	80	100
Q12: Volte-face	0	20	40	60	80	100
Q13: Prosaïquement	0	20	40	60	80	100
Q14: Enclin à	0	20	40	60	80	100
Q15: prospective	0	20	40	60	80	100
Q16: Congruence	0	20	40	60	80	100
Q17: Tangible	0	20	40	60	80	100
Q18: Mobilité (sociale)	0	20	40	60	80	100
Q19: Stipuler	0	20	40	60	80	100
Q20: (effets) pervers	0	20	40	60	80	100

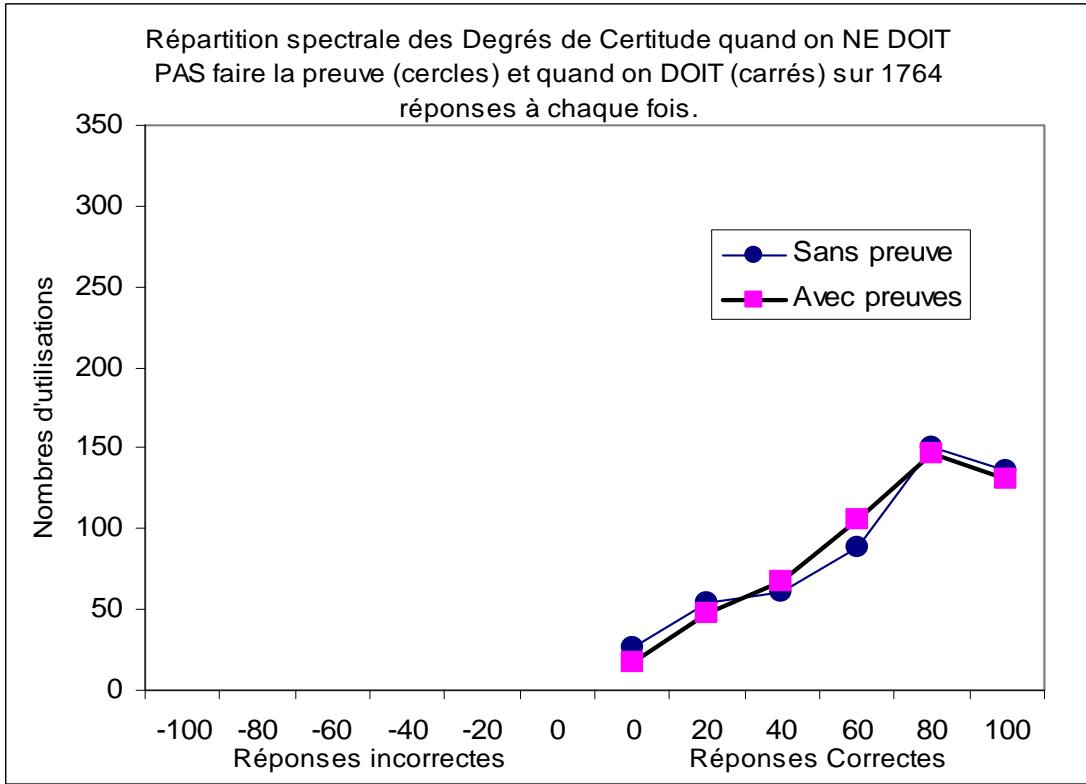
17

D. Lederqcq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSEF-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Et quand on doit faire la preuve ?						Donnez un synonyme	
Q1 : Prurit	0	20	40	60	80	100	<i>démangeaison</i>
Q2 : Implicite	0	20	40	60	80	100	<i>interne</i>
Q3: Intrinsèque à	0	20	40	60	80	100
Q4: (un) prescrit	0	20	40	60	80	100
Q5: (un) feed-back	0	20	40	60	80	100	<i>comprendre</i>
Q6: Appréhender	0	20	40	60	80	100
Q7: Promouvoir	0	20	40	60	80	100
Q8: Drastique	0	20	40	60	80	100	<i>exigeant</i>
Q9: Compétence	0	20	40	60	80	100
Q10: Recommandation	0	20	40	60	80	100
Q11: Carence	0	20	40	60	80	100
Q12: Volte-face	0	20	40	60	80	100
Q13: Prosaïquement	0	20	40	60	80	100
Q14: Enclin à	0	20	40	60	80	100
Q15: prospective	0	20	40	60	80	100
Q16: Congruence	0	20	40	60	80	100
Q17: Tangible	0	20	40	60	80	100
Q18: Mobilité (sociale)	0	20	40	60	80	100
Q19: Stipuler	0	20	40	60	80	100
Q20: (effets) pervers	0	20	40	60	80	100

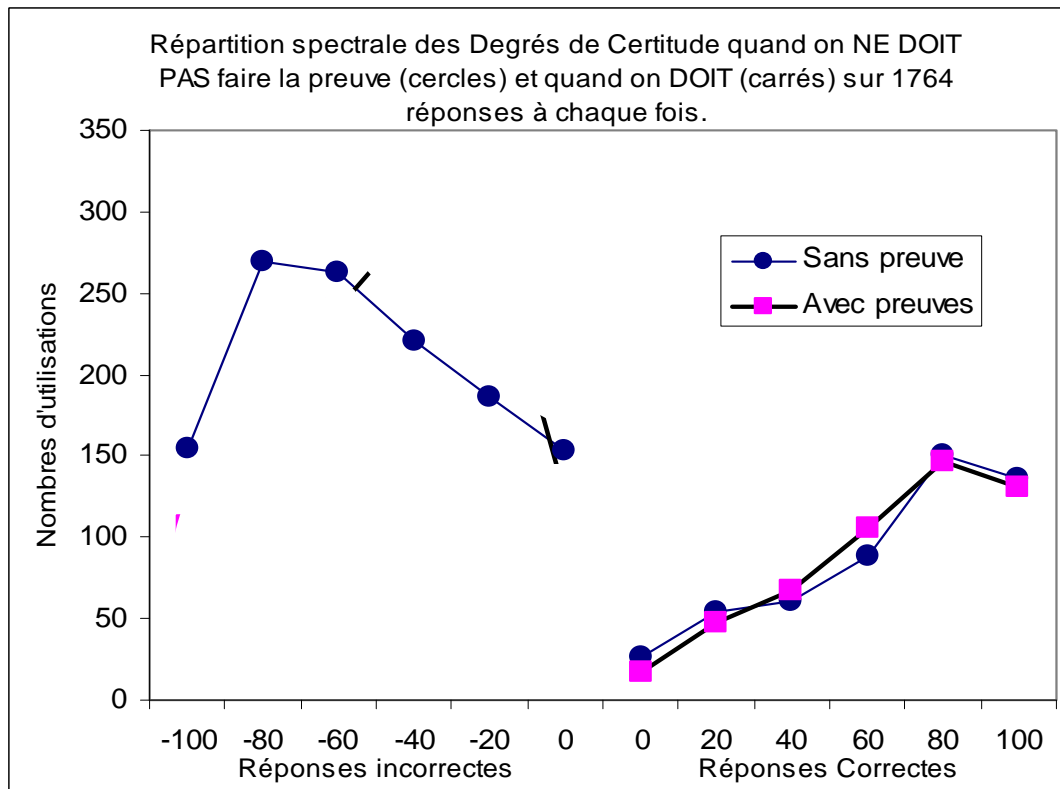
18

D. Lederqcq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSEF-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3



19

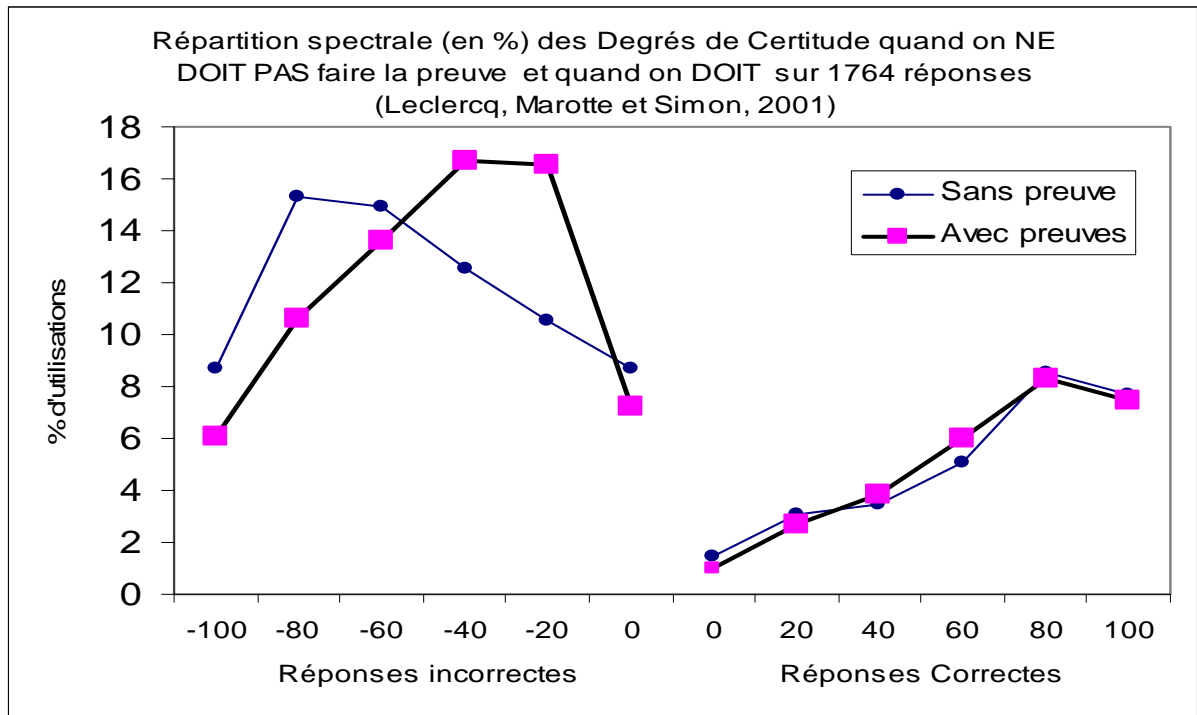
D. Leclercq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3



20

D. Leclercq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

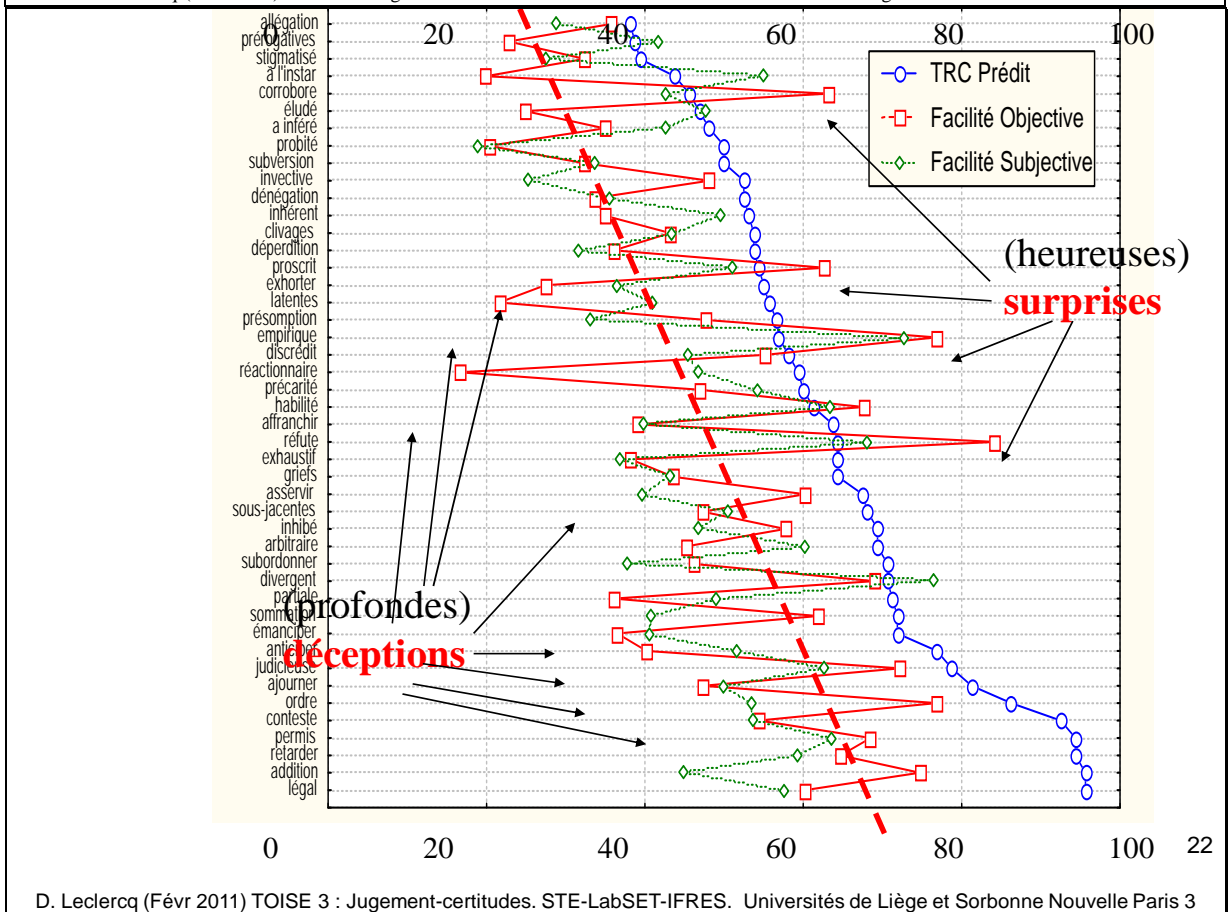
Quand on n'a que 3 minutes...



et quand on doit faire la preuve.

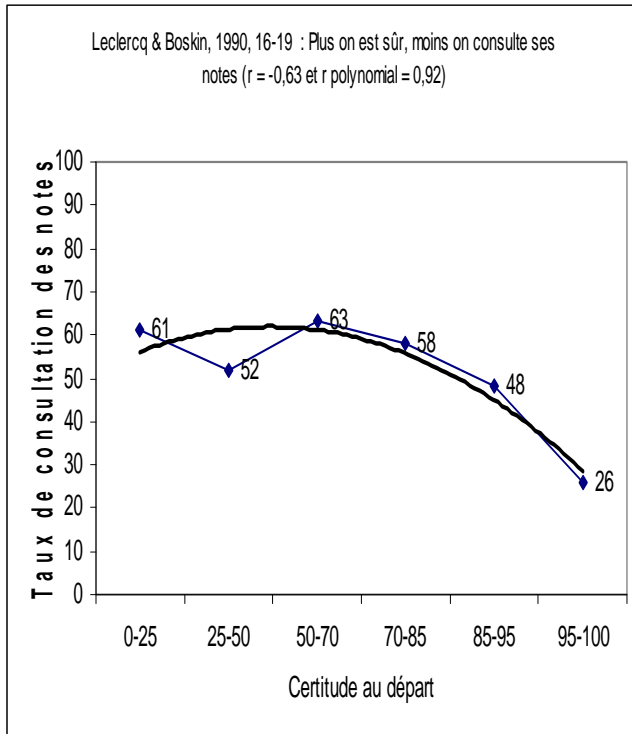
21

D. Leclercq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3



D. Leclercq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

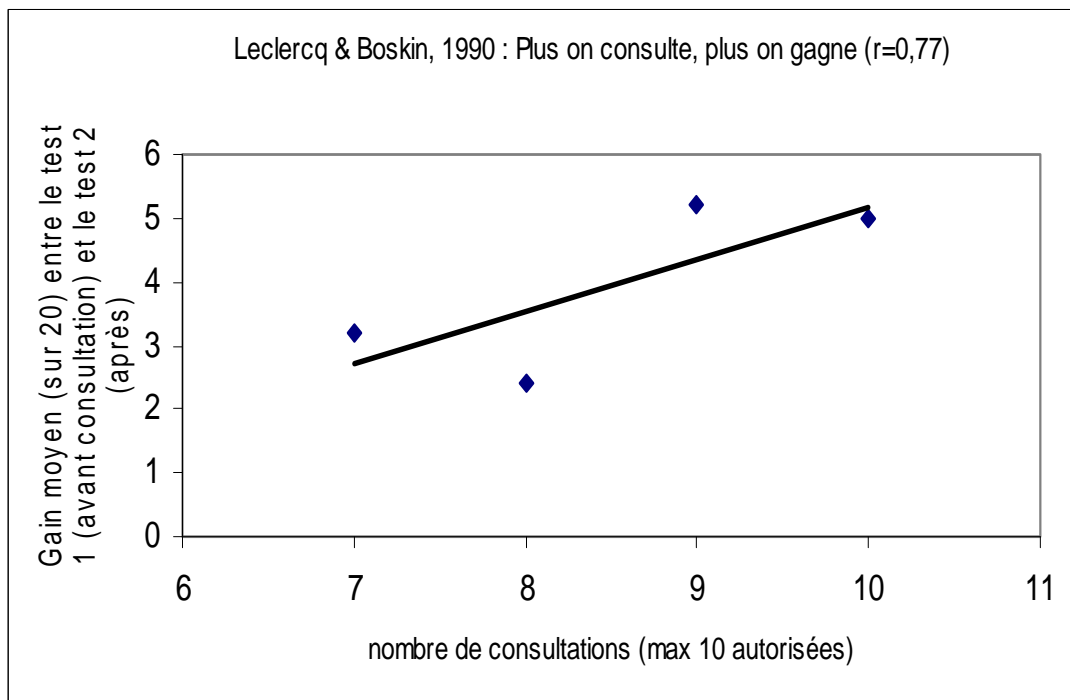
Plus la certitude est faible, plus la Consultation du dictionnaire est fréquente



23

D. Leclercq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Consultations et gains



24

D. Leclercq (Févr 2011) TOISE 3 : Jugement-certitudes. STE-LabSET-IFRES. Universités de Liège et Sorbonne Nouvelle Paris 3

Annexes

1. Leclercq, D., & Poumay, M. (2007-2008), Psychologie Educationnelle de l'Adolescent et du Jeune Adulte, chap 6, Editions de l'Université de Liège
2. Leclercq, D. (1986) « Evaluation et QCM ». chap 1 de La conception des QCM p.13-45. Bruxelles : Labor.
3. Leclercq, D. (1986) « La rédaction des QCM ». chap 3 de La conception des QCM, p.78-115, Bruxelles : Labor.
4. Gilles, Detroz & Bourguignon (2000). « Les QCM : utilisation pour l'enseignement en groupe avec boîtiers électroniques ». Revue Médicale de Liège, 55(12), p.1047-1050. (Téléchargeable à l'adresse : <http://hdl.handle.net/2268/5138>)
5. Leclercq, D. (2004) « Ce que MOHICAN a montré ou Les douze travaux de MOHICAN. » ULg
(Téléchargeable à l'adresse : http://www.labset.net/media/prod/ce_que_MOHICAN_a_montre.pdf)
6. Nicol D. (2007) Dix principes de bonnes pratiques d'évaluation.

