

Pénalité et métacognition dans l'évaluation par les QCM

Eric Vachey

▶ **To cite this version:**

Eric Vachey. Pénalité et métacognition dans l'évaluation par les QCM. Médecine humaine et pathologie. 1997. dumas-00824874

HAL Id: dumas-00824874

https://dumas.ccsd.cnrs.fr/dumas-00824874

Submitted on 22 May 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE Victor SEGALEN BORDEAUX 2

Centre de Recherches Appliquées en Méthodes Educatives

C.R.A.M.E.

Année 1997

MEMOIRE

pour le

Diplôme Universitaire de

PEDAGOGIE DES SCIENCES DE LA SANTE

Eric VACHEY

PENALITE ET METACOGNITION DANS L'EVALUATION PAR Q.C.M.

SEMINAIRES SUIVIS DANS LE CADRE DU D.U. DE PEDAGOGIE

1° : Objectifs de formation dans les sciences de la santé

2° : Communication, enseignement interactif et groupe

3° : La recherche documentaire - Méthodologie et technologies

4° : Docimologie-QCM et cas clinique QCM

5° : Programmation sur Hypercard

6° : Evaluation et justification des décisions médicales dans un contexte de contraintes médico-économiques

REMERCIEMENTS

à André QUINTON, Professeur d'Hépato-Gastro-Entérologie, Directeur du CRAME, pour son dévouement, son aide précieuse, son engagement dans la pédagogie médicale

à Georges DORIGNAC, Professeur de Pédodontie, Directeur de l'UFR Odontologie, pour sa constante bienveillance et son esprit d'ouverture, notemment vis à vis des différents axes pédagogiques développés à Bordeaux II

aux membres du jury et aux enseignants d'origine pludisciplinaire qui animent avec beaucoup d'enthousiasme et de compétence les différents séminaires de pédagogie

SOMMAIRE

INTRODUCTION p 7

PREMIERE PARTIE : REGLES DE DOCIMOLOGIE

1. PROCEDURES LIEES A L'APPRENTISSAGE

1.1. LA METACOGNITION p 10

1.1.1. Définition
1.1.2. La prise de conscience cognitive : l'explicitation, l'anticipation, la décentration
1.1.3. Le rôle de la métacognition dans le fonctionnement cognitif
1.1.4. Le traitement de l'autonomie en sciences cognitives

1.2. L'AUTOAPPRENTISSAGE p 13

1.2.1. Objectifs
1.2.2. Méthode : Apprentissage par résolutions de problèmes et Enseignement Assisté par Ordinateur (EAO)

1.3. L'EVALUATION p 17

1.3.1. Définition
1.3.2. Objectifs
1.3.3. Rôles de l'évaluation : pronostique, jaugeage, diagnostique
1.3.4. Réponse ouverte courte ou fermée

1.4. L'AUTO-EVALUATION p 21

2. PROCEDURES LIEES AU CALCUL DE LA PENALITE

2.1. PENALITE PAR SCORES p 23

2.1.1. Pénalité et hasard
2.1.2. Méthode de la "correction for guessing"
2.1.3. Choix du type d'épreuve

2.2. PENALITE PAR DEGRES DE CERTITUDE p 27

2.2.1. Méthode des "Admissible Probability Measuring Procedures"
2.2.2. Méthode des degrés de certitude

DEUXIEME PARTIE : EXPERIMENTATION

1. BUT p 30

2. MATERIEL

3. METHODE p 31

3.1. Protocole
3.2. Pénalités expérimentées
3.3. Statistiques : dénombrement descriptif, analyse

4. RESULTATS, DISCUSSION p 33

4.1. Statistiques descriptives : notes de QROC et EAO, distribution de la validation des items des QCM, degrés de certitude, corrélation certitude / taux de réussite, pénalité
4.2. Statistiques analytiques : valeur prédictive de la pénalité/note QROC, valeur prédictive du degré de certitude, optimisation de la pénalité conditionnelle

CONCLUSION p 43

APPENDICE : APPROCHE MATHEMATIQUE

1. ESTIMATION DU DEGRE DE FIABILITE de la QCM p 47

1.1. Définition
1.2. Procédure

2. INDICE DE REUSSITE de la QCM

2.1. Définition
2.2. Pocédure

3. INDICE DE DISCRIMINATION de la QCM p 48

3.1. Définition
3.2. Procédure

4. OPTIMISATION DE LA PENALITE PAR SCORES

4.1. Consigne
4.2. Combinaison des réponses possibles
4.3. Règles de pénalité selon DAMIN

4.3.1. Une assertion exacte attendue
4.3.2. Deux assertions exactes attendues
4.3.3. Trois assertions exactes attendues
4.3.4. Quatres assertions exactes attendues
4.3.5. Synthèse

BIBLIOGRAPHIE p 55

INTRODUCTION

"A l'heure actuelle on charge de faire passer des examens ou concours telle ou telle personne sans que jamais on ne soit préoccupé d'attirer l'attention sur la manière de procéder, sur les erreurs possibles, sur les modalités de notation, etc. Chacun fait son éducation à la diable, apprend à examiner en examinant, et les erreurs commises le seront au bénéfice ou dépends des candidats, dans tous les cas aux dépends de notre organisation sociale, dans la mesure où elle se fonde sur les résultats des examens et des concours "

PIERON (22) Examens et docimologie, 1963 ...

Bien que le souci des examens ne doive pas nous faire oublier qu'ils ne sont qu'un moyen pour contrôler l'efficacité de notre enseignement et non un but, comme le souligne PIERON (22), il faut reconnaître que l'évaluation finale est un problème pédagogique incontournable. Elle engage avant et après l'épreuve, une "reflexion étudiante" et une "reflexion enseignante".

Dans la **première partie** de ce mémoire, nous développons cette "reflexion étudiante" qui se construit dans la phase d'apprentissage avec la nécessité de la prise de conscience de la qualité de son savoir.

L'étudiant doit être capable de juger par lui même ce qui mérite un effort, adopter les stratégies d'étude efficaces, observer son propre travail, l'auto-évaluer en fonction de ses propres critères et ceux de l'enseignant.
Implicitement apparait le concept d'autoapprentissage, mais aussi de l'appréciation crtique de son savoir. Il s'agit d'un savoir du savoir, d'une "métacognition", plus simplement selon QUINTON (23) de la "lucidité" sur la qualité de son savoir.

Si l'évaluation finale aboutit à un constat d'échec, elle suscite des interrogations. LAVEAULT (14) remarque que les explications avancées sont souvent ...
"Je ne savais pas que cette question serait posée"
"Je croyais que je savais la réponse à la question"
"Je ne savais pas qu'il fallait préparer cette question".

Ces justifications à priori puériles sont pourtant utiles dans la "reflexion enseignante", sa communication avec l'étudiant et finalement nous le verrons, dans l'établissement de ses règles de docimologie.

Il apparaît que l'examen nous informe autant sur les processus d'apprentissage que sur les différents aspects du programme

enseigné, à savoir ce qui paraît important, ce qui est considéré comme acquis et sur la certitude du savoir.
Ces informations concernent les raisons d'échec ou de réussite, les corrections à apporter et le type d'aide qui convient le mieux à l'étudiant

Il semble que les avantages à engager l'étudiant dans des processus métacognitifs soit nombreux.
Ils favorisent le développement de leurs capacités d'anticipation et d'autoévaluation formative, en discernant mieux leurs lacunes, ainsi que le développement de l'autodidactique, en percevant mieux les objectifs des enseignants, en améliorant l'autoapprentissage.

Les enseignants ont tendance à imputer le retard de leurs acquis universitaires à leur manque d'assiduité aux cours magistraux, oubliant qu'ils sont aussi causés par une mauvaise connaissance de nos objectifs, un manque d'autonomie et de certitude sur leur connaissances .

Il est difficile aux enseignants d'ignorer ces paramètres alors qu'ils sont contraints à l'évaluation.
La "reflexion enseignante" concerne aussi la règle de notation de l'épreuve, souhaitée la plus équitable et la plus cohérente possible.

C'est la raison pour laquelle nous nous proposons dans la **seconde partie** de ce mémoire d'expérimenter un nouveau modèle docimologique intégrant la notion de "métacognition".

Cette expérimentation se déroule en trois étapes invitant à une reflexion étudiante dans les deux premières, enseignante dans la troisième.

La première étape précède l'examen et a pour thème l'auto-apprentissage par la résolution de problèmes préparés, exposés et critiqués par les étudiants selon les recommandations de FORTIN (10).

La seconde concerne l'examen intégrant la métacognition par une contrainte à l'autoévaluation des degrés de certitudes accompagnant les propositions de réponses, lors d'une épreuve sous forme de questions à choix multiples de réponses.

Enfin la troisième, postérieure à cette épreuve, a pour but d'établir des règles de pénalité favorisant une juste évaluation de la connaissance, pour corriger les bénéfices acquis par des bonnes réponses faites au hasard.

PREMIERE PARTIE

REGLES DE DOCIMOLOGIE

1. PROCEDURES LIEES A L'APPRENTISSAGE

1.1. LA METACOGNITION

"Connais-toi toi-même" (temple de Delphes)

Nous avons tous entendu les commentaires des étudiants qui abordent l'examen de fin d'année; parfois anxieux, "je ne sais rien", "j'ai oublié de mettre ...", "si j'avais eu deux jours de plus ..."; parfois confiants "je sais tout par coeur ...".
Ces mêmes comportements se retrouvent à la sortie de l'épreuve, "j'ai oublié de mettre ...", "je ne sais plus ce que j'ai mis ..." ou au contraire, "cette fois-ci j'ai tout mis ..." etc ...

Devons nous ignorer la métacognition, qui exprime ce doute ou cette certitude, qui accompagne leur savoir ?

1.1.1. Définition

WOLFS (30) nous rappelle que le terme "métacognition", apparu dans la littérature scientifique à la fin des années 70, signifie étymologiquement "connaissance de la connaissance".

Elle intéresse toute personne qui apprend et qui réfléchit aux stratégies et aux méthodes de travail qu'elle utilise pour apprendre. D'ailleurs pour MAUBANT(20), l'apprentissage devrait être le lieu de la maîtrise de ses propres stratégies de connaissance .

C'est la surveillance de ses propres mécanismes cognitifs qui définit, selon LAVEAULT (14), la métacognition .

WOLFS (30) propose cette définition "domaine qui regroupe les connaissances introspectives conscientes qu'un individu a de ses propres états et processus cognitifs, les capacités que cet individu a de délibirément contrôler et planifier ses propres processus cognitifs en vue de la réalisation d'un but ou d'un objectif déterminé"

Selon CHARTER (3) ce concept de métacognition est flou (comme celui d'intelligence), il a d'ailleurs suscité il y a une vingtaine d'année de nombreux travaux de recherche aux résultats décevants, le problème reste donc ouvert.
Il souligne que dès que nous prenons notre propre activité cognitive comme objet d'observation nous exerçons une activité qui peut être dite "méta" cognitive .
Certains sont partisans de limiter le concept de métacognition aux connaissances et stratégies conscientes, d'autres y incluent les procédures inconscientes.

En définitive si ce terme de métacognition est généralement utilisé pour désigner la connaissance et le contrôle qu'un système cognitif peut avoir de lui même et de son propre fonctionnement, il n'y a pas vraiment de consensus sur ce que recouvre exactement le concept de métacognition.

Les processus cognitifs s'appliqueraient aux objets, au sens large, tandis que les processus métacognitifs s'appliqueraient aux processus cognitifs en adoptant un point de vue "exterieur" par rapport à ceux ci.

Nous pouvons établir une distinction entre les "habiletés métacognitives" dont dispose un étudiant pour orienter, contrôler ou réguler ses activités cognitives et ses "connaissances métacognitives" désignants, sa connaissance de ses propres processus cogntifs et des facteurs qui les influencent .

1.1.2. La prise de conscience cognitive

Pour WOLFS (30), la "prise de conscience" par le sujet de ses propres processus cognitifs fait partie des habiletés métacognitives avec l'évaluation de leurs efficacités ou "jugement métacognitif" et la "prise de décision".

Comme l'indique GAGEY (11) cette prise de conscience ne renvoit pas au vécu d'une donnée enfin acceptée et reconnue ou au revécu d'un événement tel qu'il fut, mais plutôt à "une autre façon de se dire".
WOLFS (30) distingue trois aspects particuliers de prise de conscience : l'explicitation, l'anticipation et la décentration.

1.1.2.1. L'explicitation

Il s'agit de "l'explicitation" par l'étudiant de ses propres processus et produits cognitifs, passés ou présents.
Un des avantages de cette technique est qu'elle modifie son comportement .
Cette prise de conscience des représentations intuitives amène l'étudiant à les remettre en question et à en construire de nouvelles intègrant les nouveaux acquis.
Pour DENEF (7), la métacognition peut aider à exprimer la connaissance erronée et prendre conscience du doute qui accompagne la connaissance : "Savoir qu'on ne sait pas, c'est aussi un savoir".
Or dans un système d'évaluation fondé sur des objectifs pédagogiques, il est souhaitable qu'un étudiant puisse terminer un examen en étant capable d'estimer son score. S'il est conscient des objectifs à atteindre et qu'il s'évalue correctement, il devrait pouvoir identifier ses lacunes de façon autonome.

Pour ROMAINVILLE (1990) cité par WOLFS (30), il existerait une relation positive entre la métacognition et la réussite en première année d'études universitaires.
Les étudiants qui expriment des connaissances métacognitives plus élaborées (quantitativement et qualitativement) auraient tendance à se retrouver parmi les étudiants qui réussissent .

Cette explicitation serait sollicitée bien avant le cursus universitaire.
QUINTON (24) a montré qu'un des facteurs de réussite au concours de PCEM1 est la mention obtenue lors du baccalauréat.

Par exemple en 1996, pour les redoublants du P1 de Bordeaux 2 issus d'un bac D, le taux de réussite est de 57,7% pour les "mention bien - très bien" et seulement de 5,2% pour les "sans mention", 10 fois moins ...

1.1.2.2. L'anticipation

TOCHON (27) la qualifie d'habileté métacognitive. Il s'agit d'un processus d'explicitation de l'avenir.
Il consiste à se représenter les résultats attendus d'une action cognitive ou les stratégies à mettre en oeuvre pour y parvenir, comme par exemple l'anticipation par l'étudiant des buts, objectifs ou résultats à atteindre, ou des questions qui pourraient lui être posées lors de l'examen.

1.1.2.3. La décentration

La "décentration" consiste à s'expliciter ses propres processus cognitifs dans le temps présent.
C'est ce qui passe quand en apprenant un cours, on se le remémorre en imaginant mentalement être en train de l'expliquer à un tiers.

1.1.3. Le rôle de la métacognition dans le fonctionnement cognitif

1.1.3.1. Les différents degrés de métacognition

Le rôle de la métacognition dépend de l'état de lucidité du sujet et peut être quantifié selon WOLFS (30), en quatre degrés .

Dans le degré 0, l'étudiant récite par coeur sans réfléchir. La relation entre performance et métacognition est négative, il s'agit d'automatismes, de connaissances implicites ou encore d'une attention soutenue à une activité cognitive.

Dans le degré 1, l'étudiant relit plusieurs fois sa question ainsi que sa réponse initiale. Il est susceptible d'amener une solution de rechange.
La relation entre performance et métacognition est positive et intervient en cas de difficulté lors de procédures d'autoévaluation, d'anticipation, explicitation ou celui du cadrage métacognitif d'un problème .

Dans le degré 2, la métacognition devient active. L'étudiant se met à la place de l'enseignant, se fixe des buts, refait le cours et adopte une stratégie d'autoquestionnement stimulant sa compréhension.
Il identifie la nature des obstacles rencontrés par ses défauts de compréhension l'amenant à réviser ses stratégies cognitives.

Enfin dans le Degré 3, il y a un enrichissement un perfectionnement métacognitif nécessitant une interaction avec l'enseignant

1.1.3.2. Evaluation et métacognition

Selon TOCHON (27) les fonctions institutionnelles de l'évaluation auraient intérêt à viser d'avantage la réflexion et la métacognition et un peu moins les tests finaux dont les exagérations ont été dénoncées, en caricaturant l'enseignement proche de la robotisation.

On pourrait en effet, à force de demander aux étudiants de cocher des QCM avec des stratégies d'apprentissage appropriées, les démunir de leur droit à réfléchir.

Cependant les tests intermédiaires sous forme de QCM évaluent bien la métacognition; l'étudiant peut alors dire "je croyais en savoir plus..." ou "je suis content, à présent je sais où j'en suis..."

Il semble que l'évaluation pertinente dans le développement d'objectifs transdisciplinaires doit intègrer la métacognition individuelle, la prise en charge par l'individu de son processus de connaissance.
Entraîner l'étudiant à l'autoévaluation, l'autocritique en lui demandant de réfléchir à la confiance qu'il accorde à ce qu'il écrit lors d'une évaluation le responsabilise et contribue à notre mission d'enseignement.

Cette démarche renvoit aussi à la morale, "je fais bien ou mal...", "je me sent concerné ou pas...", "je suis honnête ou je triche...", "je dissimule mes erreurs ou pas...", etc ...

1.1.4. Le traitement de l'autonomie en sciences cognitives

Ce traitement fait souvent défaut selon TOCHON (27). Or puisque la fonction d'apprentissage est interne, l'évaluation pédagogique devrait être complétée, sur le plan de ces compétences internes et relever aussi bien de la cognition que de la métacognition .

A ce jour, le processus de contrôle de la métacognition sur la fonction d'apprentissage est loin de faire l'unanimité.

1.2. L'AUTOAPPRENTISSAGE

C'est cette forme de pédagogie qui a retenu notre attention en 1994, alors que nous devions dispenser des enseignements dirigés sur des thèmes non traités par ailleurs, à des étudiants inscrits en P2 d'odontologie.

1.2.1. Objectifs

Le principe de la pédagogie par autoapprentissage consiste, comme son nom l'indique, à construire soi-même son savoir.

Selon GRENIER (12), cette procédure contribue à donner un maximum de certitudes à ce savoir, car il est issu non pas de l'avis de

l'enseignant "expert" mais de la consultation d'articles scientifiques .

L'autoapprentissage nous aide à faire face à l'augmentation dramatique des connaissances, dans un contexte actuel de diminution des volumes horaires et du nombre d'enseignants.

Il incite enfin l'étudiant à apprendre à apprendre, autant pour la la vie professionnelle que pour la formation universitaire en lui donnant d'avantage d'autonomie.
Or il ne faudrait pas comme le souligne MAUBANT (20), que l'université devienne le lieu de la passivité de l'étudiant apprenant, retranché dans une logique "transmission-reception", sans épanouissement individuel.

A l'heure des nouvelles technologies, notemment du multimédia, cette autonomie se traduit selon JEANNEL (13) par la possibilité de trier les informations offertes. C'est ce qui différencie, selon lui, la personne dite "plus", de celle dite "solo" qui reste dépendante de procédures extérieures même si elle sait les utiliser.
Ainsi ARMSTRONG (1) précise que l'objectif de l'autoapprentissage n'est pas seulement d'assimiler des savoirs, des procédures mais aussi de développer une attitude réflexive pour maîtriser sa propre démarche d'appropriation.

A l'issue du processus d'autoformation, nous n'évaluons pas un étudiant avec ses seules ressources conceptuelles à un moment précis, mais un individu qui, selon JEANNEL (13), parce qu'il a des ressources multiples à sa disposition, est capable en un temps déterminé d'adopter la meilleure stratégie pour répondre au problème qui lui est posé .

Ce constat est conforté par les travaux de LAVEAULT (14) qui démontre que les étudiants qui réussissent le mieux à un examen sont ceux qui ont su anticiper seuls correctement son contenu.
Il semble qu'une bonne surveillance par l'étudiant de ses propres processus d'apprentissage exerce une influence sur sa prise de conscience et son fonctionnement cognitif pour résoudre un problème.
Toutefois on peut supposer que plus un étudiant est laissé en charge de ses apprentissages, plus il est important que le "tuteur" (10) lui fasse part des objectifs d'apprentissage. Cette étape est nécessaire au développement d'une plus grande autonomie dans les études.

1.2.2. Méthode

Nous nous limiterons à la description de deux méthodes dont nous présenterons les résultats dans la seconde partie. Tout d'abord la méthode dite de l'apprentissage par résolutions de problèmes puis celle de l'Enseignement Assisté par Ordinateur.

1.2.2.1. L'Apprentissage par résolutions de problèmes

Il s'agit d'un auto-apprentissage réalisé en petits groupes, suivant les recommandations de FORTIN (10).
Chaque groupe doit analyser un thème, chercher les sources d'informations requises, exposer son savoir, le confronter avec celui des autres, éventuellement se remettre en question ce qui suppose aussi une aptitude à l'autoévaluation (19).

Malheureusement comme l'indique TOCHON (27), les méthodes d'évaluation traditionnelles sont inadaptées à ces fonctions d'apprentissage, métacognitives et personnelles.

1.2.2.2. L'Enseignement Assisté par Ordinateur (EAO)

Il a bien sûr comme support l'ordinateur, il nous dégage de tout discours répétitif, mais surtout responsabilise l'étudiant face à son apprentissage : "Plus le maître enseigne , moins l'élève apprend " (CONFUCIUS : 551-479 Av J.C.).

Il existe différents types d'EAO.
Les EAO tutoriaux sont utilisés en linéaires ou en arborescence s'apparentant à des cours dirigés, les EAO hypermédia donnent accès à une information sous forme de texte et d'image, les EAO de modélisation reproduisent des cinétiques notamment biologiques, les EIAO "tutoriaux intelligents" confrontent le modèle de l'expert et celui construit au cours de l'épreuve par l'étudiant.

Nous utilisons un EAO de simulation de cas cliniques.

L'étudiant doit établir un diagnostic précis à partir d'un cas clinique tiré au sort, défini par une arborescence cohérente de données disposant de noeuds de hasard .

Aucun étudiant ne réfléchit sur le même cas au même moment dans la salle d'informatique, et n'aura deux cas identiques à résoudre.
La durée de l'entraînement n'est pas limitée avant l'évaluation, puis un autre logiciel réunit trie et analyse les réponses des étudiants .

Voici des copies d'écran, à titre d'exemple du logiciel que nous utilisons dans la seconde partie. Nous l'avons écrit en BASIC puis compilé avec des fichiers d'autoexecution (28).

Vous avez fini L'INTERROGATOIRE vous abordez L'EXAMEN
que devez vous observer ? (deux lettres)

===GLOSSAIRE===
AIR.....A
DOULEUR.....B
ELECTRIQUE...C
EXAMEN.....D
FROID.....E
INTERRO.....F
LESION.....G
MOTIF CONSUL.H
NATURE LESI..I
PALPA APIC...J
PERCUSSION...K
PRESSION.....L
PROVOQUEE....M
RADIO.....N
SIEGE LESION.O
SPONTANEE....P
SURVEILLANCE.Q

A partir des informations que vous avez recueillies
vous allez nous dire si la dent est VIVANTE
(maintenez la touche ^ enfoncee et appuyez sur la touche +)
ou si elle est MORTE (tapez uniquement sur -)

DENT No : 4 5
LESION : II
D SPONTANEE : NON
D PROVOQUEE : OUI
AIR : -
FROID : +
ELECTRIQUE : +
PRESSION : -
PALPATION APICALE : -
PERCUSSION : -

S'il vous plait M. * appuyez sur une touche

quand vous aurez lu votre evaluation,puis appelez un etudiant

qui n'est pas a la methodologie

NOTES : 3 1 3 2 1 2 2 2 2 3 0 2.5= 23.50/24

= 19.58/20
= 1.96/2
DATE : 01-01-1980
HEURE DEBUT : 00:01:09
HEURE FIN : 00:05:19

1.3. L'EVALUATION

1.3.1. Définition

Nous devons à PIERON (22) le terme de docimologie, il l'a proposé à partir de mots grecs relatifs aux examens : épreuve, examiner, examinateur, apte à examiner.
De LANDSHEERE (6) la définit comme "une science qui a pour objet l'étude systématique des examens, en particulier des systèmes de notation, et du comportement des examinateurs et des examinés".
Il évoque donc une psychologie de l'évaluation

1.3.2. Objectifs

Parmi ses nombreux objectifs, nous retiendrons l'étude des effets favorables ou défavorables des différentes formes d'examen, des réactions émotionnelles des étudiants, des processus mis en jeu et des effets obtenus par les différents types de notation des épreuves.

1.3.3. Rôles de l'évaluation

L'évaluation peut intervenir à trois moments différents par rapport à l'acquisition du savoir ce qui lui confère un triple rôle .

1.3.3.1. Evaluation formative : rôle pronostique

Il s'agit selon de LANDSHEERE (6) d'une évaluation formative spécifique à l'étudiant destinée à établir un état d'avancement et reconnaître ses difficultés afin de les surmonter.
Elle mesure les qualités cognitives et les connaissances nécessaires de l'étudiant avant d'aborder un nouveau thème .
Cette évaluation appelle donc une action corrective en retour sur son activité cognitive ("feed back" des anglo saxons).
Elle peut donner lieu selon CHARTIER (3) à des connaissances métacognitives à condition de bien analyser les causes de réussite ou d'échec .

1.3.3.2. Evaluation normative : rôle de jaugeage

Ce jaugeage est réalisé par rapport aux autres apprenants. Selon de LANDSHEERE (6), son but est de contrôler les acquisitions et les progrès.

1.3.3.3. Evaluation sommative : rôle diagnostique

De LANDSHEERE (6) la qualifie de sommative car elle revêt le caractère d'un bilan d'ensemble après l'enseignement.
C'est l'épreuve qui sanctionne, mais dans ce genre d'évaluation, la question qui se pose est de savoir pourquoi un apprentissage parfait ne s'est pas produit.

1.3.4 Choix entre la question à réponse ouverte ou fermée

Dans une question à réponse ouverte, la plus tradionnelle,l'étudiant répond avec son propre vocabulaire. En revanche si la réponse est fermée il choisit parmi plusieurs réponses proposées

1.3.4.1. Réponse rédactionnelle ouverte courte

Avantages

Ce sont les questions les plus naturelles, elles conviennent en particulier à des épreuves de contrôle réalisées rapidement. L'évaluation des capacités supérieures de créativité, de jugement et d'esprit critique semble de leur ressort .
En construisant entièrement sa réponse ouverte, l'étudiant prouve aussi bien sa connaissance de la matière que sa capacité à l'exprimer verbalement.
Cette traduction verbale de la pensée n'est qu'une forme d'expression parmi d'autres, mais y renoncer serait certainement une erreur puisque ce type de communication est dominant.

Inconvénients

Le caractère unique d'une réponse formulée en toute liberté se prête mal à l'évaluation par comparaison avec les réponses des autres individus, d'où un possible désaccord entre correcteurs, si souvent dénoncé en docimologie.

Ce sont les raisons pour lesquelles nous avons adopté le principe de la réponse fermée dans nos évaluations normatives.

1.3.4.2. Réponse fermée

Définition des questions à choix multiple (QCM)

Pour LECLERCQ (18) il s'agit d'une "question à laquelle l'étudiant répond en opérant une sélection parmi plusieurs solutions proposées, chacune étant jugée correcte ou incorrecte indépendamment de l'étudiant qui doit y répondre" .

Une QCM est donc une question à réponse fermée pour laquelle le candidat exprime son choix en cochant une ou plusieurs cases, dont le nombre sera limité à quatre dans notre expérimentation.
Ce mode de questionnement peut paraître très direct, cependant les travaux de LECLERCQ (18) ont prouvé qu'il peut être mis au service de sollicitations plus indirectes de la réponse.

Les compléments, ou propositions de réponses peuvent être multiples, Dans la consigne ces compléments peuvent être reliés suivant différentes lois entre eux . Dans l'ouvrage de LECLERCQ (18) nous en avons répertorié 19 ...

Ces réponses peuvent être saisies sur ordinateur comme c'est le cas au CRAME de Bordeaux 2.

Inconvénients

Inconvénients majeurs

Les principaux inconvénients des QCM semblent liés à leur mauvaise formulation et utilisation .

D'autre part, les QCM n'entraînent pas à la formulation de réponses. Ils sont donc impuissants à mesurer l'expression spontanée, l'aptitude à rédiger ou résumer sa pensée.

On peut craindre aussi que la présentation de solutions erronées, soit à l'origine d'erreurs fixées dans la mémoire. Ce danger paraît réel dans certains domaines.

Inconvénients secondaires

A côté de celà d'autres désavantages plus secondaires ne doivent pas être oubliés comme par exemple notre travail de mise au point beaucoup plus important pour une QCM que pour une épreuve productive.
En principe pour concevoir une QCM, l'enseignant utilise les erreurs connues des étudiants, dans le cas contraire il découvre concrètement le prix à payer pour les obtenir...
Nous avons aussi pu nous rendre compte que la construction de QCM appelle une réelle compétence sur le fond et une riche expérience pédagogique. Autrement dit, l'aide collégiale et les essais expérimentaux sont indispensables.

Par ailleurs, étant donné que toute présentation orale est exclue, la QCM doit comporter une amorce dense où comme le souligne QUINTON (23) chaque mot compte, sans futilité, inexactitude, ambiguité, chauvinisme ,complexité ni suggestivité, avec des solutions proposées ne différant parfois que par un élément ou deux.
L'étudiant doit pourtant analyser ces données à son aise.
Nous ne devons pas céder non plus à la facilité, en posant des questions de détail à l'origine d'une parcellisation des connaissances, faisant souvent appel à la mémoire beaucoup plus qu'à la logique. Il y a là un risque réel, surtout dans la perspective d'un examen, qui encouragerait les étudiants à des stratégies de préparation en accord avec cette parcellisation.

On reproche aussi aux QCM de faciliter la tricherie lors des épreuves en copiant les réponses du voisin . L'ordinateur peut remédier à celà par des éditions triées.

Mais surtout subsiste la possibilité aux étudiants de fournir des réponses correctes en répondant au hasard .
Dans le cas des QCM à choix unique, il est difficile de distinguer ces choix heureux par ignorance des choix corrects par compétence. A l'opposé, les choix incorrects sont eux aussi, difficiles à interpréter (aléa ou erreur?).

Pour compenser cet inconvénient on augmente classiquement le nombre de solutions proposées à au moins quatre, afin de réduire la part du hasard au maximum à 25% . Mais ce principe n'est pas toujours applicable, c'est le cas notemment des choix multiples sans patron de réponse, où de toute façon la part du hasard reste faible.
Une autre façon de procéder que nous expérimentons dans ce mémoire est de pénaliser les erreurs .
Pour dissuader les choix aveugles, LECLERCQ (18) utilise la consigne "ne devinez pas aveuglément, car des points vous seront retirés en cas d'erreur". Ce retrait de points est connu sous le nom de "correction for guessing" que de LANDSHEERE (6) traduit par "correction pour divination"

Avantages

La maîtrise de la majorité des défauts évoqués engendrent des qualités, donc des avantages. De LANDSHEERE (6) distingue les avantages quantitatifs des avantages qualitatifs .

Avantages quantitatifs

Selon PIERON (22) le français avait la réputation en 1938 "d'un animal qui sait faire une dissertation"; précisément la structure des QCM s'accomode mal de toute phraséologie de réponse.
Dans l'énoncé, les termes et la syntaxe d'une QCM sont choisis avec soin afin de rendre la question aussi précise et claire que possible.

Une même attention sera portée quand au choix et au nombre des solutions proposées .

Par ailleurs, notre évaluation n'est pas affectée par une difficulté qu'aurait l'étudiant à exprimer correctement sa réponse.

La correction des QCM est simple, rapide et juste.

L'impossibilité pour l'étudiant de masquer une absence de savoir est un autre avantage quantitatif, les QCM sont effet capables de couvrir toute une matière. Il en résulte une meilleure évaluation des connaissances réelles de l'étudiant.

Avantages qualitatifs

D'après LECLERCQ (18), une des principales qualités des QCM est d'éliminer les effets liés au seul correcteur, l'évaluation est plus fidèle et moins injuste.
Les effets de sévérité ou de clémence ne sont pas à craindre. Pour prédire sa note, dans une épreuve de production, il vaut mieux parfois connaître son examinateur que soi même.
Il n'y a pas non plus d'effet de tendance centrale ou centripète, en effet la majorité des correcteurs évitent les extrêmes en concentrant leurs notes au milieu de l'échelle, donnant à leur distribution une loi gaussienne

La relation entre le professeur et l'étudiant ne vient pas non plus perturber l'évaluation.
L'effet de halo est impossible, la notation n'est pas influencée par des caractéristiques étrangères à la performance de l'étudiant; son aspect physique, sa présentation , sa prononciation, son accent.
Enfin il n'y a pas d'effet de stéréotype de logique ou d'inertie, le correcteur n'a pas tendance à attribuer à un étudiant des notes qu'il a acquises auparavant.
L'influence de la correction de plusieurs copies n'existe pas d'avantage par des effets de proximité, de séquence ou de contraste. Une copie suivante à une copie brillante risque d'être désavantagée ou inversement par relativisation.

Pour de LANDSHEERE (6), un autre avantage qualitatif important est de pouvoir évaluer l'étudiant "dans ses derniers retranchements". Dans les épreuves de production, il peut masquer une absence de savoir en développant ce qu'il sait et en esquivant ce qu'il ne sait pas.
Les questions à choix multiples bien construites permettent des explorations beaucoup plus subtiles des capacités supérieures qu'il n'y paraît. Il lui semble souhaitable d'isoler au moins partiellement les connaissances des capacités d'expression verbale, ce que permettent les QCM.

Enfin les QCM favorisent l'autorégulation. Si elles sont assorties de leurs grilles de correction, les étudiants trouvent les moyens d'une autoévaluation rigoureuse et rapide.

Comme nous allons le voir à présent, ceci peut avoir de nombreuses conséquences favorables.

1.4. L'AUTO-EVALUATION

1.4.1. Définition

L'autoévaluation consiste à mesurer sa connaissance .

Elle désigne selon WOLFS (30), l'évaluation par l'individu de ses propres comportements cognitifs et des facteurs qui les influencent.

Avant de passer un examen c'est l'estimation par l'étudiant de son degré de préparation à répondre aux questions.
Pendant l'examen, c'est le degré de certitude exprimé à propos d'une réponse fournie.
Après l'examen ce sont les raisons données au choix des réponses, ou l'explicitation des facteurs de réussite ou d'échec.

En définitive, la notion d'auto évaluation se rapproche de celle de "sentiment d'autoefficacité ,... résultat d'un processus transactionnel entre l'estimation que fait la personne, des exigences et des conditions d'une tâche particulière, des ressources qu'elle possède ou qu'elle croit posséder et surtout de

sa capacité à les utiliser adéquatement dans cette situation précise"

Finalement, elle aboutit à l'évaluation de son degré de préparation, en sachant déterminer le seuil de performance ou des conditions de réussite exigées par l'enseignant et selon d'IVERNOIS (8), au jugement correct des aspects interessants ou importants de la matière. L'autoévaluation nécessite donc l'anticipation correcte des tâches d'un examen .

1.4.2. Autoévaluation et anticipation

Selon BARRET (2), tout étudiant confronté à une épreuve d'évaluation, se prépare à celle ci en mettant en place une stratégie anticipatrice qui s'établit, à partir ...

- d'une estimation, plus ou moins juste, des attentes de l'enseignant concernant l'acquisition ou l'utilisation de connaissances et de procédures de raisonnement.
Cette estimation dépend de la clarté des objectifs et des critères d'évaluation .

- d'un niveau personnel d'exigence de réussite, c'est à dire du degré d'aspiration propre à l'étudiant concernant la note qu'il désire obtenir.
Dans certain cas cette stratégie amène même au jeu subtil des "impasses".

Le degré de réussite ultérieure ne dépend pas seulement, au regard de ces quelques éléments, de la quantité de préparation mais aussi de sa qualité .

Cependant, il ne faut pas conclure trop hâtivement en négligeant la quantité de travail à fournir.
QUINTON (24) montre par exemple que pour le P1, le taux de succès des étudiants est inversement proportionnel au temps de loisir qu'ils s'accordent. Ainsi parmi les primants reçus en 1996, 17% ne s'accordaient aucun loisir et 3% (6 fois moins) s'accordaient tout le week-end ...

1.4.3. Autoévaluation et comportement

Selon BARRET (2), l'une des premières préoccupations de l'étudiant consiste en la recherche de cohérence entre les questions posées et les apprentissages effectués préalablement.

En l'absence d'objectif clairement formulé par l'enseignant, la compréhension des demandes détermine la représentation des objectifs à atteindre et les procédures nécessaires pour y parvenir, en même temps que les connaissances auxquelles il faut faire appel.

1.4.4. Auto évaluation et confiance en soi

Ainsi il apparaît que l'étudiant est amené à produire un discours interne, dit métacognitif, dont l'objectif est d'évaluer, au fur et à mesure de sa production, la pertinence des connaissances utilisées, la validité des procédures mises en oeuvre et la justesse de la représentation.

Notre objectif selon BARRET (2) est que l'étudiant porte un jugement adéquat sur ses prestations. Cela passe donc par une véritable formation à la métacognition, en portant un regard critique sur ses productions dans le but de les améliorer.

Dans notre étude, nous demandons à l'étudiant d'accompagner ses réponses à des QCM, d'un degré de certitude indiquant la confiance qu'il leurs accorde.
Ceci met en jeu selon selon CHARTIER (3), la notion d'estime de soi et permet de quantifier sa propre métacognition sans attribuer réussites et échecs au hasard ou à une incapacité générale.

Cette quantification peut être prise en compte dans le calcul de la note finale, c'est le problème posé à l'enseignant, du choix de la pénalité que nous développons à présent.

2. PROCEDURES LIEES AU CALCUL DE LA PENALITE DES QCM

2.1. La pénalité par scores

Le problème qui se pose est le choix de la règle à adopter pour donner ou enlever des points, en fonction des réponses apportées aux QCM.

2.1.1. Pénalité et hasard

Selon PIERON (22), il est nécessaire qu'une part de hasard soit présente dans les examens et concours, puisque c'est le cas dans les épreuves de la vie.

Cependant il paraît souhaitable d'en limiter les effets en docimologie.
Dans le cas par exemple d'une épreuve comportant n QCM indépendants à complément simple à N réponses, la probabilité de répondre correctement à une QCM est de 1/N, celle de réussir la totalité de l'épreuve est de $1/N^n$

Les travaux mathématiques de SHUFORD (26), ARMSTRONG (1) et de FINETI (5), ont montré les relations existantes entre hasard et résultats lors des évaluations par QCM.

Il en résulte qu'une stratégie appropriée permet d'optimiser les réponses.

PIERON propose pour les QCM, 5 choix de réponses avec une seule juste, en fixant comme pénalité le 0 de l'échelle de notation à partir de 20% de bonnes réponses.

Selon DAMIN (4) si les règles des pénalités sont bien établies, alors toute stratégie, autre que la reflexion, ne sera jamais payante voire même pénalisante. Dans ce cas, la pénalité "idéale" doit donner en moyenne pour ceux qui répondent au hasard, donc sans processus cognitif, une note négative.
L'écart-type associé à cette moyenne est un indicateur de l'aléa qui accompagne chaque réponse.

Cette pénalité varie en fonction de l'objectif de l'évaluation. Elle peut être fortement négative si l'on préfère la non réponse à une réponse fausse, ou faiblement négative de façon à encourager les réponses notemment chez les anxieux et ne pas pénaliser une erreur matérielle.

Le comportement des étudiants est lié à la connaissance des consignes de pénalité qui ne doivent être ni dissuassives, ni favorables à des réponses hâtives ou hasardeuses.

2.1.2. Méthode de la "correction for guessing"

Comme le rappelle de LANDSHEERE (6), dans une question fermée à deux choix de réponses, l'une correcte et l'autre fausse, la probabilité de réussite de l'élève qui répond au hasard est de 50%
Les constructeurs de tests le savent depuis longtemps (22), c'est la raison pour laquelle on a le plus souvent recours aux questions à cinq choix dont l'un est correct et les autres qu'il nomme distracteurs, seulement vraissemblables.

Dans la méthode de la "correction for guessing", à chaque réponse incorrecte correpond la pénalité p...

$$p = 1/(k-1)$$

où k désigne le nombre de solutions proposées à chaque QCM.

On détermine donc un retrait de 1 point pour 2 solutions proposées (Vrai/Faux)

1/2 point pour 3 solutions proposées
1/3 point pour 4 solutions proposées
1/4 point pour 5 solutions proposées
etc ...

Dans ce modèle, la notion de degrés de certitude n'existe pas.

Par exemple dans la QCM suivante présentée avec sa grille de correction, (les * correspondant aux bonnes réponses), le score maximal étant de 1 point et le nombre de solutions proposées de 4, la pénalité pour chaque réponse fausse sera de 1/(4-1) de point .

N°3 : Le ternissement d'un amalgame provient ...

[ ] A: d'un polissage prématuré

* [ ] B: d'une corrosion de surface

* [ ] C: d'un taux élevé de mercure résiduel

[ ] D: des habitudes alimentaires du patient

Si l'étudiant coche la réponse A et B il a 2 erreurs, "A" est une réponse fausse et il n'a pas répondu à C réponse juste attendue. A ces deux erreurs correspond donc une pénalité de 2 fois 1/3, le score de la QCM est de 1/3.

2.1.3. Choix du type d'épreuve

Selon LAVEAULT (14), l'examen mesure aussi bien l'apprentissage de l'étudiant que la qualité de sa préparation à une série d'objectifs d'apprentissage jugés important par l'enseignant.

Il distingue deux types d'épreuves, celles à critères relatifs et celles à critères absolus.

2.1.3.1. Critères relatifs et absolus

L'épreuve à critère relatif est fondée sur la référence à la normale nommée par TOCHON (27) "référence normative", c'est à dire selon la courbe des résultats de tous les étudiants ayant passé le même test, elle risque donc de valider un groupe de niveau faible.

L'épreuve à critères absolus, que nous avons choisi dans notre expérimentation, dépend de critères qu'il nomme "références critérielles".
La norme est celle exigée par l'enseignant, elle ne correspond plus à une mesure de tendance centrale de mode, médiane ,moyenne ou seuil issu de la loi de distribution des notes du groupe.
En revanche l'étudiant doit absolument avoir validé certaines questions pour avoir la moyenne.

En réalité l'analyse de TOCHON (27) démontre qu'en dernier ressort, l'enseignant utilise souvent une courbe normative en adaptant son seuil d'admissibilité, la référence critérielle devient alors "normative sérielle".

Il existe une technique docimologique dite du "calcul du niveau acceptable de performance" qui est adaptée aux épreuves à critères absolus. Celle ci semble plus rigoureuse et mieux adaptée aux objectifs pédagogiques fixés.

2.1.3.2. Calcul du niveau acceptable de performance

Définition

Il correspond au seuil qui permet de décider, selon des critères absolus, du passage ou non d'un étudiant "qui en sait juste assez".

Son calcul n'a de valeur que si le nombre de QCM est supérieur à la trentaine, il repose sur la décision collective de plusieurs enseignants ayant d'abord porté un jugement indépendant.

Son emploi impose un jugement à priori de la difficulté relative de chaque question et implique un jugement sur l'épreuve tout entière.

L'examen doit être objectif. Ceci est vérifié si des corrections, indépendantes ou par le même correcteur à intervalle de temps, donnent le même résultat.

Procédure

Les enseignants décident de la réponse exacte pour chaque QCM. Puis ils s'entendent sur la (ou les) réponse(s) que l'étudiant doit absolument éliminer sans bénéficier du hasard.
Enfin ils calculent l'indice d'acceptabilité pour chaque QCM, qui correspond en fait au score attribué pour chaque bonne réponse.

Le calcul du niveau acceptable pour l'épreuve est égal à la somme des indices d'acceptabilité de chaque QCM.

Par conséquent à chaque QCM il faut étudier tous les choix offerts et décider celui (ou ceux) qui doit être absolument rejeté.
Si nous appelons N le nombre de réponses de la QCM et L le nombre de réponses fausses (leurres), l'indice I d'acceptabilité pour la QCM a pour expression

$$I = 1 / (N-L)$$

Donc pour un QCM à 4 choix , on obtient le résultat suivant :

Nbre de leurres	Indice
0	0,25
1	0,33
2	0,5
3	1

Par exemple dans les 2 QCM qui suivent, N=4.

N°6 : Les causes techniques d'un échec du traitement canalaire sont liées ...
* [ ] A: au déplacement du foramen apical
[ ] B: à l'utilisation d'une instrumentation mécanisée
* [ ] C: à la présence d'une perforation
* [ ] D: un défaut d'assèchement du tiers apical du canal

Pour la QCM N°6 L=1, l'indice I=0,33 ,en cochant A, C ou D, le score est à chaque fois de 0,33.

N°7 : Les causes biologiques d'un echec endodontique dépendent de

* [ ] A: du non respect des limites apicales lors de la préparation
[ ] B: de l'importance de la lésion apicale
[ ] C: d'une nécrose pulpaire préalable
[ ] D: d'un défaut de microcirculation desmodontale

Pour la QCM N°7 L=3, l'indice I=1 ,en cochant A le sore est de 1

Le calcul du niveau acceptable de performance équivaut à l'addition de tous les indices et repose sur l'analyse détaillée de chacune des questions.

La validité de son estimation dépend pour chaque QCM du jugement indépendant de plusieurs enseignants qui prennent bien en considération les objectifs et le niveau de l'examen .

Si les différences qui apparaissent entre les enseignants sont grandes, il importe de clarifier les objectifs.

Selon LAVEAULT (14) cette clarification devrait faire l'objet de réunions de consensus en présence des étudiants afin d'évaluer et corriger le degré decompréhension au niveau du groupe.

L'enseignant dispose de plusieurs outils statistiques pour l'aider dans cette tâche

2.2. Pénalité par degrés de certitude

Lors d'une évaluation formatrice, il n'est pas choquant que les chances de réussite d'un étudiant qui ne connaît pas son cours soient minimes et qu'elle soient un peu augmentées s'il est lucide de son niveau de connaissance.
Il faut pour cela accompagner la QCM d'une procédure métacognitive, on peut lui demander d'évaluer son niveau de certitude par une échelle.

Degré de certitude -> 0: [ ] 1: [ ] 2: [ ]

Selon LECLERCQ (15,16) et FABRE (9), il est inconcevable de traiter de la même façon ...
l'ignorance avouée avec une réponse fausse,
la connaissance insuffisante avec une réponse en partie correcte sans certitude,
la connaissance satisfaisante avec une réponse correcte et assez sûre,
la connaissance totale par une réponse correcte avec certitude,
et la méconnaissance sur la base d'idées fausses par une réponse incorrecte et sûre.

L'ignorance par l'enseignant de cet état de doute est certainement préjudiciable à son évaluation et à la possibilité de remédiation.

2.2.1. Méthode des "Admissible Probability Measuring Procedures" (16,17)

Dans cette méthode mise au point par LECLERCQ (16,17), la consigne est probabiliste et le barème de l'évaluation avantage l'expression de son doute ou de sa certitude sans biais.

Comme on peut le voir dans la règle de pénalité qu'il propose, ce sont ceux qui s'autoévaluent bien qui gagnent le plus de points.

Si vous considérez que votre réponse a une probabilité d'être correcte comprise entre	Ecrivez	Dans une notation sur 20, vous obtiendrez les points suivants cas de réponse Correcte	Incorrecte
0% et 25%	0	+13	+4
25% et 50%	1	+16	+3
50% et 70%	2	+17	+2
70% et 85%	3	+18	0
85% et 95%	4	+19	-6
95% et 100%	5	+20	-20

2.2.2. Méthode simplifiée

Pour réduire le rôle de la chance une règle plus simple de pénalité utilisée par LECLERCQ (15) et DENEF (7) consiste à inviter l'étudiant à un processus métacognitif en lui demandant de préciser le degré de certitude qui accompagne sa réponse à chaque QCM .

Cette évaluation de la certitude peut se faire selon l'échelle suivante ...

2 : pour une réponse absolument certaine,
1 : pour une réponse probable sans certitude
0 : dans le cas d'ignorance totale .

Au lieu de 6 choix, il n'y en a plus que 3.

Un système adapté de notation sur lequel nous reviendrons, récompense le sujet qui évalue correctement la certitude de ses connaissances et pénalise celui qui se trompe à cet égard.

GRENIER (12) propose une répartition de la certitude sur cinq degrés. Ceci lui permet de l'exploiter sous la forme de courbes ROC (Receiver Operating Characteristic curve) et de connaître la valeur discréminante du degré de certitude séparant les bons étudiants des mauvais.

DEUXIEME PARTIE

EXPERIMENTATION

"Que le juste milieu ne soit pas suivi, je sais pourquoi : l'homme intelligent va au-delà, l'imbécile en deçà..." (CONFUCIUS)

1. BUT

Inspiré des reflexions de TOCHON (27) et des travaux de LECLERCQ (15), le travail que nous présentons se propose d'optimiser le calcul de la pénalité des QCM en intégrant notamment la notion de métacognition ou de "lucidité" sur la qualité de son savoir. Cette métacognition concerne on l'a vu, la connaissance et le contrôle de soi, des processus et la prise de décision.

Ces QCM sont destinés à pronostiquer la compétence acquise lors d'enseignements dirigés utilisant le principe de l'auto apprentissage par résolution de problème.

La saisie des réponses à chaque QCM associée à l'autoévaluation du savoir nous permet selon NOEL (21) de mesurer pour chaque étudiant

sa performance, c'est à dire la qualité de son activité cognitive

sa lucidité, c'est à dire la qualité de son estimation de son niveau de compréhension ou de sa perfomance.

son pessimisme ou son optimisme, par la sous estimation ou surestimation, de son niveau de compréhension ou de la qualité de sa performance

Deux facteurs de confusion possibles sont analysés, celui lié à l'enseignant par la comparaison des moyennes des notes entre les groupes, et celui lié au groupe par la comparaison des degrés de certitude rapportés aux taux moyens de réussite par Q.C.M.

2. MATERIEL

Notre expérimentation date de 1994 et concerne 61 étudiants de P2 de l'U.F.R. d'Odontologie de Bordeaux 2 répartis en 3 groupes possédant chacun un enseignant différent.
La discipline concernée est l'Odontologie Conservatrice-Endodontie.

La saisie des données est effectuée sur un ordinateur AMSTRAD 1512 HD grâce au module de bases de données du logiciel WORKS (Ver.1)
Les statistiques sont réalisées à l'aide du tableur de WORKS pour la partie descriptive et du logiciel EPIDEMIO (Ver.3) pour la partie analytique

3. METHODE

Nous rédigeons des QCM à 4 compléments, sans patron de réponse, il peut y avoir de 0 à 4 propositions de réponse vraies.

A chacune des 60 QCM proposées, la procédure métacognitive qui accompagne la réponse, consiste à décider entre ces trois propositions :

0 : certain de ne pas savoir
1 : incertain
2 : certain de savoir

Ce choix fait appel au premier degré de métacognition en invitant l'étudiant à la relecture et à trouver une solution de rechange.

3.1. Protocole

La méthode pédagogique utilisée dans chaque groupe est celle de l'apprentissage par résolution de problème.

Les étudiants sont soumis ...

à une évaluation formative sous forme de QCM
puis à une évaluation sommative de fin d'année sous la forme de Questions Réponses Ouvertes Courtes corrigées par 4 enseignants
enfin, deux années plus tard, à une évaluation formative mais cette fois sous la forme d'un Enseignement Assisté par Ordinateur sur le thème du diagnostique avec des cas cliniques tirés au sort.

Respectant les recommandations de LAVEAULT (14) concernant l'intérêt de l'anticipation du contenu d'un examen, pour une meilleure utilisation des processus métacognitifs et l'acceptation d'un éventuel échec, les étudiants sont informés à l'avance de la nature de l'épreuve sous forme de QCM.

3.2. Pénalités expérimentées

En revanche la pénalité qui détermine le barème de l'évaluation n'est pas explicité dans la consigne. Ceci évite toute stratégie de la part de l'étudiant, et donne plus de liberté dans le choix de la pénalité.

La pénalité de chaque QCM sera déterminée à partir de 4 méthodes de calcul que nous définissons ainsi ...

Méthode Exclusive : "0" à la Q.C.M. si une réponse est fausse
Méthode Additive : somme des réponses justes
Méthode Internat : additions par paliers
Méthode Conditionnelle : méthode intégrant des degrés de certitude, pénalisant selon le modèle de DAMIN (4) développé en appendice, indécision et réponses fausses, surtout celles assorties d'une certitude.

Le détail du mode de calcul en fonction du nombre d'erreurs à la QCM est donné dans le tableau ci dessous, exprimant les gains en pourcentages

NBRE ITEMS JUSTES ->	0	1	2	3	4
EXCLUSIVE	0	0	0	0	100
ADDITIONNELLE	0	25	50	75	100
INTERNAT	0	0	40	50	100

% de points à la QCM fonction du nombre de choix corrects

CONDITIONNELLE

HASARD	0	25	50	75	100	id. à ADDITIONNELLE
PROBABLE	0	0	0	0	100	id. à EXCLUSIVE
CERTAIN	- 100	0	0	0	200	
	%	%	%	%	%	

Tableau de pénalités
Les valeurs inscrites correspondent au % de point que peut rapporter la question

3.3. Statistiques

3.3.1. Dénombrement descriptif

A chaque QCM, les 4 scores liés aux différentes pénalités sont calculés par ordinateur.

Ces QCM rassemblent 60 questions construites à partir des exposés portant sur l'ensemble du programme sous la forme déjà évoquée des 4 compléments sans patron de réponse. Les étudiants doivent donc répondre à un total de 60x4=240 propositions.
Il y a 2^4x60=960 combinaisons de réponses possibles par copie.

Puisqu'il n'y a qu'une combinaison juste par QCM la réussite due au hasard est de 1/16 ce qui n'est pas négligeable et nous conforte dans l'exploitation de la certitude avouée.

Il y a ainsi 60 certitudes à déterminer et donc pour chaque étudiant 240+60=300 propositions à saisir .

A chaque question correspond une fiche type de la base de donnée identifiée de 3 éléments; le nom, le groupe, le numéro de la question. Il y a donc 180 éléments d'identification par étudiant (60*3).

Au total nous avons 300+180=480 données par étudiant et puisqu'il y a 61 étudiants, 29280 données expérimentales .
La place mémoire allouée par l'ordinateur est de 160 kOctets, ce qui n'est pas excessif (480*61).

L'ordinateur applique les 4 méthodes de notations à partir des 18300 propositions de réponses (61x300) structurées autour de 10980 éléments d'identification(61x180).

3.3.2. Analyse

Les 4 pénalités sont comparées, par la méthode du coefficient de corrélation des rangs de SPEARMAN, avec les notes des Q.R.O.Q. de fin d'année et celles archivées lors du module d'Enseignement Assisté par Ordinateur.
Ce test est choisi par ce qu'il est non paramétrique, et qu'il ne nous impose pas d'hypothèse sur la distribution des notes.
Le lien statistique ainsi évoqué correspond à ce qu'on appelle la concordance en pédagogie.
Plus ce coefficient est élevé moins il y a d'écart entre les résultats des épreuves.

4. RESULTATS, DISCUSSION

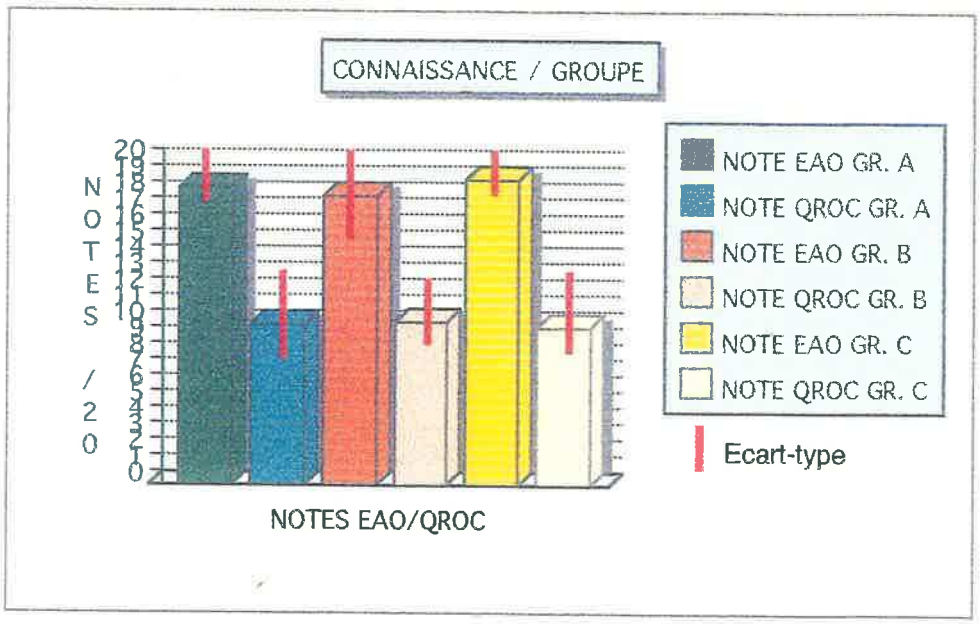
4.1. Statistiques descriptives

4.1.1. Comparaison des groupes

Etant donné que les trois groupes ont reçu un enseignement sous la forme de résolutions de problèmes coordonnées par des tuteurs différents, le problème est de savoir si acquis et certitudes sont comparables .

4.1.1.1. Notes de QROC et EAO

Comme on peut le voir sur les histogrammes qui suivent, les moyennes de notes et leurs écart-types sont identiques pour ce type d'épreuves.



Seul l'écart type de l'épreuve d'EAO du groupe B diffère des deux autres, ceci est provoqué par 2 notes moins bonnes un 15 et un 7,5 alors que toutes les autres sont centrées autour de 18/20.

Bien que l'épreuve d'EAO ne couvre pas l'intégralité du programme, on peut tout de même admettre que les acquis des trois groupes sont comparables.

Nous observons là un autre intérêt de l'APP. Le phénomène d'appropriation des connaissances par des étudiants de groupes différents ayant consulté la même bibliographie, semble identique et moins fortement fortement corrélé aux certitudes de l'enseignant (12).

Ce constat nous autorise à raisonner sur l'ensemble des étudiants et non par groupe.
Nous prendrons comme référence la note d'évaluation finale de CROQ

4.1.1.2. Distribution de la validation des items des QCM

Dans le tableau qui suit nous présentons les pourcentages correspondant au nombre d'items justes par question des différents groupes, c'est à dire au nombre de choix par QCM retenu et non retenu de façon juste.

Nombre d'items justes	% de QCM Gr A	% de QCM Gr B	% de QCM Gr C	% de QCM Total
0	2	2	1	1
1	10	10	9	10
2	27	26*	23	25
3	34	34	37	35
4	27	28	30	29
	100%	100%	100%	100 %

Tableau des pourcentages du nombre de réponses justes par QCM pour chaque groupe et pour l'ensemble des étudiants
(*) 26% des QCM du groupe B ont 2 réponses justes sur 4 pour l'ensemble des étudiants

La encore, la similitude des taux de réussites entre les groupes est évidente. Nous travaillerons donc sur les effectifs globaux.

Notons tout de suite la marginalité du nombre de questions n'ayant occasionnées aucune bonne réponse parmi les 4 propositions offertes (1%).

Les acquis sont satisfaisants, on voit qu'au total (colonne de droite) 60% (25+35) des questions engendrent 2 à 3 items corrects sur 4.
Mais il n'y a que 29 % des questions qui voient leurs quatres propositions de réponse validées.

Ce constat nous laisse présager le caractère trop sélectif d'une correction du type "exclusive" et nous invite à connaître le profil associé de la perception du savoir.

4.1.1.3. Degrés de certitude

Certitude	% de QCM Gr A	% de QCM Gr B	% de QCM Gr C	% de QCM Total
0	16	13	13	14
1	47	43,5	47	46
2	37	43,5	40	40
	100%	100%	100%	100%

Tableau des pourcentages des degrés de certitude par QCM pour chaque groupe et pour l'ensemble des étudiants

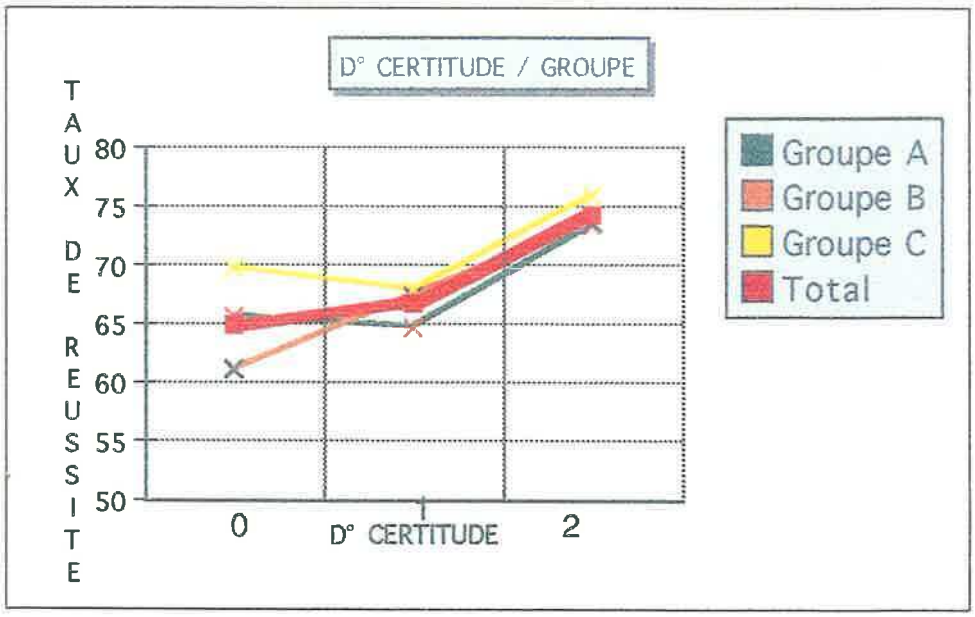
La encore il n'y a pas de différence statistique en ce qui concerne les certitudes des différents groupes. Cette homogénéité évoque une corrélation entre les acquis et la certitude qui les accompagne.
Il faut cependant noter le taux élevé de QCM (46% , colonne de droite) à l'origine de réponses en état de doute (degré 1).
Cette remarque nous semble importante et nous invite à retravailler les ressources bibliographiques (29) des étudiants.

Proposons nous à présent d'étudier globalement, l'évolution simultannée des taux de réussite et des degrés de certitude ainsi que le résultat des pénalités.

4.1.2. Etude globale

4.1.2.1. Lien : certitude, taux de réussite

Les droites ci dessous expriment ce lien. La certitude du savoir augmente effectivement avec sa qualité.



Lien entre degrés de certitude et pourcentage de réponses justes par QCM

L'idéal serait de trouver une certitude de 0 pour un taux de 0% , autrement dit aucun savoir à une question difficile, mais avec la certitude de ne pas savoir.
En réalité à ce degré 0 correspond un taux de réussite de 65% .
Par conséquent les étudiants se sousestiment.

Les réponses en situation de doute donnent un taux de succès à peine plus élevé : 1,75% de plus.

Mais celles en situation de certitude ne donnent pas le taux idéal de 100% mais de 74,25% , cette fois les étudiants se surestiment.

Heureusement la consultation par tris de la base de données nous apprend que les QCM associant 0 réponse juste et certitude d'avoir pourtant raison (degré 2), la pire des choses, ne correspond qu'à 1,1% des cas, ce pourcentage passe à 17% avec la certitude de ne pas savoir.

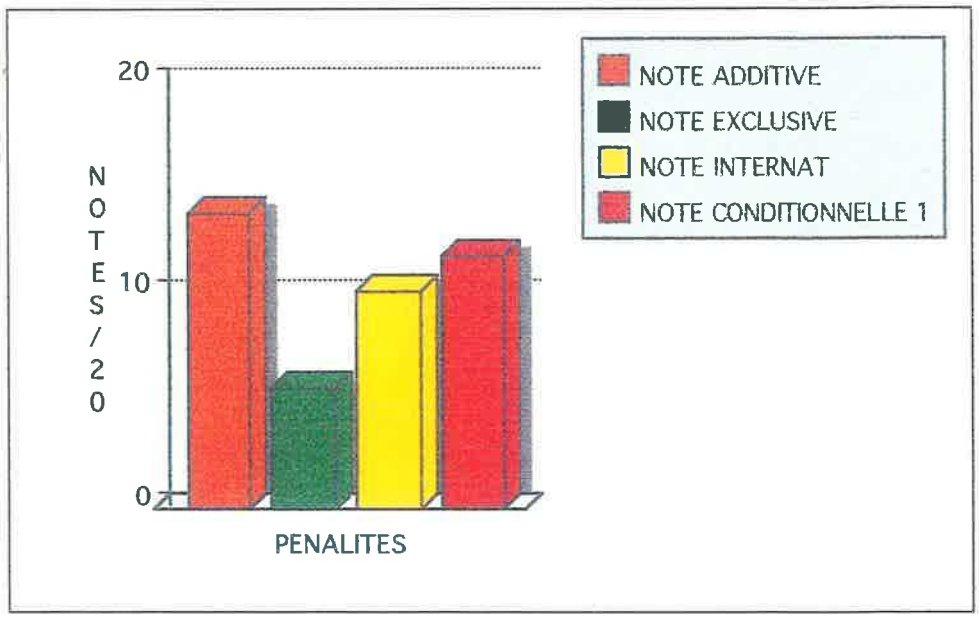
Cette autoévaluation du savoir est perfectible chez nos étudiants. Signalons à ce propos un article paru dans le Times du 5 février 1990 signé Charles KRAUTHAMMER cité par TOCHON (27), titré "Education : Doing bad and feeling good". Il fournit un exemple d'insuffisance métacognitive. Un test de mathématique est donné à des élèves de 13 ans, ceux ayant les scores les plus bas pensent en majorité (68%) être bons en mathématiques .

Pourtant LAVEAULT (14) prétend que le degré de concordance entre les sujets du groupe fort et faible est comparable tant en ce qui concerne l'autoévaluation par l'étudiant de sa performance à l'examen, que l'ordre de difficulté des items. En revanche les sujets ayant le moins bien réussi à l'examen sont aussi ceux qui sont le plus en désaccord sur le type de questions qui seront selectionnées par l'enseignant .

Cette piètre appréciation de leur niveau de connaissance peut être un handicap à leur progrès .

4.1.2.2. Pénalité

Comme on peut le voir sur l'histogramme qui suit, le choix de la pénalité influence de manière sensible le score final.



Note finale fonction de la pénalité

Les méthodes additive et conditionnelles donnent des notes deux fois plus élevées que la méthode exclusive qui correspond d'avantage à un objectif de sélection que d'évaluation.

Cependant la méthode additive paraît trop indulgente.

La méthode type internat est un compromis satisfaisant. Mais son but est de sélectionner, il n'y a pas de pénalisation de la connaissance erronnée (dans le sens "misconception").

Les distributions des notes par pénalité, suivent toujours une loi Normale.
Le volet analytique de notre travail statistique va nous aider à présent à faire un choix entre les différentes règles de pénalités.

4.2. Statistiques analytiques

4.2.1. Valeur prédictive de la pénalité/note QROC

La question qui se pose,est de savoir quelle pénalité choisir, parmi les quatre expérimentées.

Ce sera celle qui nous donnera la meilleure corrélation, au sens statistique du terme, pour chaque étudiant avec la note de QROC ou d'EAO.
Les évaluations par EAO ne couvrent pas tout le programme et sont très permissives

Nous représentons dans le tableau ci dessous le résultat du calcul des coefficients de corrélation de SPEARMAN entre les notes de QCM par pénalités et celles de l'épreuve QROC et d'EAO pour chaque étudiant

Pénalité	Conditionnelle valeur de r	Additive valeur de r	Exclusive valeur de r	Internat valeur de r
QROC	0,04	0,188	0,137	0,224
EAO	0,03	0,088	0,057	0,06

Corrélations entre la note de QROC et la note finale de QCM en fonction de la pénalité

Ces corrélations sont faibles et n'autorisent pas les régressions linéaires (risque α=0,01); cependant les meilleurs d'entre elles sont obtenues avec l'épreuve de QROC, ce qui nous conforte à l'utiliser comme épreuve d'évaluation sommative finale référencielle.

Le score le plus prédictif de l'épreuve de QROC est celui obtenu par la méthode internat .
Malheureusement dans ce type de pénalité, à moins de disposer d'une banque de questions très importante (16), nous ne nous s'affranchissons pas totalement du gain de points du au hasard ou à une stratégie.

Le coefficient de corrélation décevant de la pénalité conditionnelle, nous invite à optimiser ce type de pénalité.

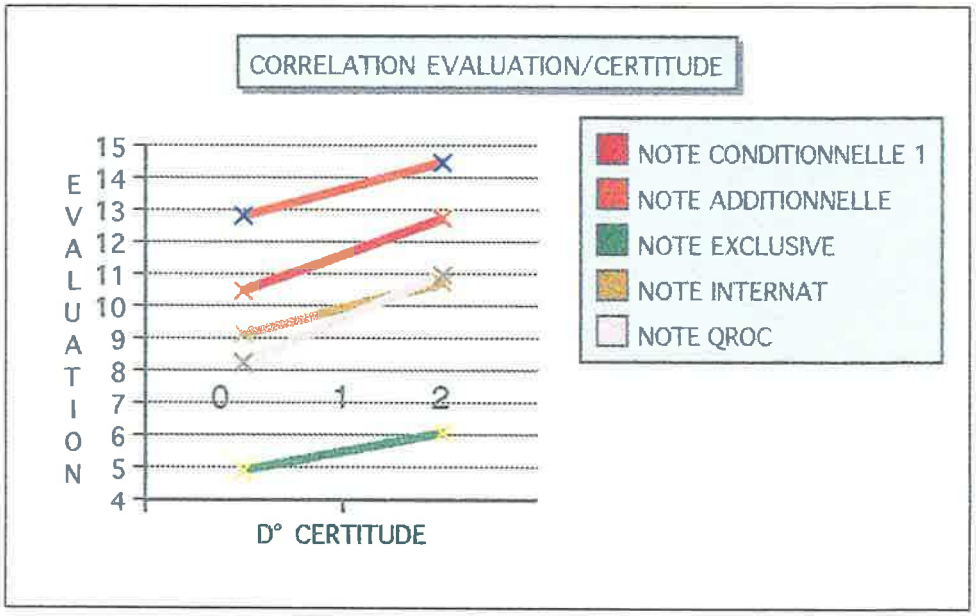
Etant donné qu'elle repose sur le degré de certitude de l'étudiant, il nous faut avant tout vérifier si cette certitude a une valeur prédictive.

4.2.2. Valeur prédictive du degré de certitude

Les regressions linéaires expriment le lien entre d'une part les différents scores des étudiants à l'évaluation formative par QCM pour les différentes pénalités ainsi que sommative par QROC et d'autre part leur degré moyen de certitude exprimé lors de l'évaluation formative par QCM .
Les résultats statistiques sont rassemblés dans le tableau qui suit :

Note conditionnelle 1	Note additionnelle	Note exclusive	Note internat	Note QROC
Cond=10,46+1,06*D°Cert	Add=12,81+0,84*D°Cert	Exc=4,92+0,56*D°Cert	Int=9,19+0,79*D°Cert	Qro=8,29+1,37*D°Cert
R=0,14	R= 0,25	R=0,09	R=0,18	0,14

Tableau des corrélations et régression linéaires entre évaluations et degrés de certitudes



Regressions linéaires des degrés de certitude moyen en fonction de la note finale par pénalité

Les coefficients de corrélation associés aux droites de régression ne prouvent l'existence d'un lien statistique que pour la pénalité additionnelle (α=0,05).

Le parallèlisme des "regressions" autorise un classement des différentes méthodes de pénalité, de la plus indulgente "additive", en haut du diagramme à la plus sélective "exclusive", en bas.

Ces droites sont ascendantes donc plus l'étudiant est sûr de lui, meilleure est sa connaissance.

Il apparaît que les questions qui suscitent plus de pessimisme ne sont pas particulièrement faciles mais plutot de difficulté moyenne, et que l'étudiant sûr de lui est meilleur que celui qui ne l'est pas.

Les degrés de certitude ont donc une réelle valeur prédictive sur le score de l'étudiant, cette corrélation existe même avec l'épreuve de QROC de fin d'année corrigée par différents enseignants.

Ainsi il nous apparaît que plus l'étudiant est sûr de lui (D° de certitude) en évaluation formative (QCM), plus il a de chance de réussir l'évaluation sommative (QROC).

On peut observer sur le graphisme que dans le cas de la pénalité conditionnelle, à un degré de certitude moyen de 0 (certain de ne pas savoir), correspond une note de 10,47/20 , c'est à dire à la moyenne ce qui n'est pas normal.

En revanche, la droite obtenue pour l'épreuve de QROC, dispose du meilleur taux d'accroissement ce qui est bon sur le plan pédagogique. Plus l'étudiant est sûr de lui meilleurs sont ses acquis. Son score score moyen est de 10/20, il est corréllé avec une certitude moyenne de 1 (incertain), la réponse en situation d'ignorance ne donne pas la moyenne et la certitude la moyenne.

Pour la méthode additionnelle la certitude de savoir (2) donne une bonne note .

Enfin dans la méthode exclusive la réponse en situation d'ignorance (0) est sanctionnée.

A partir de ces observations, nous de redéfinissons la pénalité conditionnelle en "conditionnelle 2".

4.2.3. Optimisation de la pénalité conditionnelle

Nous exploitons directement les observations graphiques précédentes pour adopter un compromis.

Pour une réponse en situation d'incertitude totale nous appliquons la pénalité exclusive, pour une réponse évaluée probable la méthode internat, enfin pour une réponse en situation de certitude, la pénalité additionnelle .

REUSSITE ->	0	1	2	3	4	
HASARD	0	0	0	0	100	EXCLUSIVE
PROBABLE	0	0	40	50	100	INTERNAT
CERTITUDE	0	25	50	75	100	ADDITIONNELLE

SENTIMENT	PENALITE
Hasard	Exclusive
Probable	Internat
Certitude	Additive

Définition de la pénalité "Conditionnelle 2"

Partant du constat que l'étudiant sûr de lui est de meilleur niveau que celui qui ne l'est pas, cette nouvelle pénalité avantage considérablement la confiance en soi.

Etudions à présent le lien créé entre cette pénalité avec les autres pénalités ainsi que l'épreuve référencielle de QROC

Pénalité	QROC	Additive	Exclusive	Internat
Conditionnelle 2	0,225	0,759	0,631	0,738

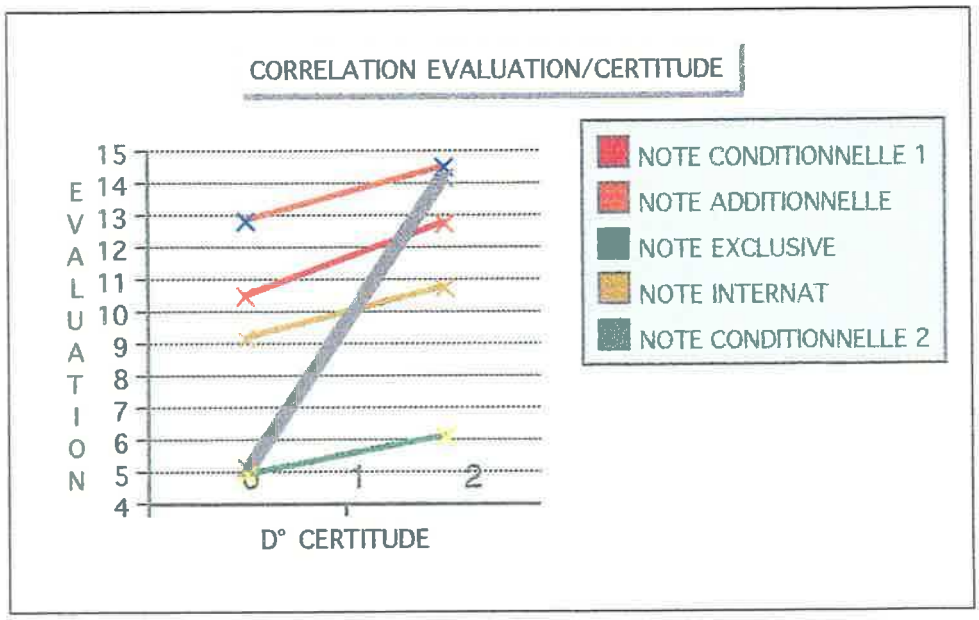
Tableau des corrélations entre la note obtenue par la nouvelle pénalité "Conditionnelle 2" et celles obtenues avec les autres pénalités et l'épreuve de QROC

La corrélation entre la nouvelle pénalité conditionnelle et l'épreuve de QROC est la meilleur qui soit (α=0,1). Le lien est aussi très fort avec les autres pénalités(α=0,01).

Le lien entre pénalité "conditionnelle 2" et degré de certitude est le meilleur qui soit avec un coefficient r=0,74

$$Cond2 = 5,16 + 4,5 * D° Cert$$

La nouvelle représentation graphique montrant l'incidence de la certitude sur cette nouvelle pénalité est évident



Regressions linéaires des degrés de certitude moyen en fonction de la note finale par la pénalité "Conditionnelle 2" et les autres

L'accroissement obtenu est le résultat de l'objectif fixé, privilégier la certitude .

La note moyenne en situation de doute absolu est de 5/20,
de doute relatif de 10/20,
et de certitude de 14/20 .

En revanche, on peut reprocher à cette pénalité de ne pas récompenser l'honnêteté .

CONCLUSION

"L'apprentissage le plus utile pour la vie est de désapprendre ce qui est faux"
(ANTHISTENE, 444-365 av J.C.)

Selon WOLFS (31), la majorité des auteurs s'accordent pour dire que l'entraînement métacognitif exercerait une influence positive sur les performances du sujet.
Il envisage le développement de la métacognition aussi bien pour aider à l'apprentissage, que pour aider au développement de la personne au sens large par une meilleure connaissance de soi, de ses possibilités et de son fonctionnement cognitif. Ce dernier s'articulerait en trois niveaux, nous mémorisons, puis nous contrôlons nos propres progrès, enfin nous réfléchissons sur nos certitudes.
Des facteurs d'ordre psychologiques interviennent sur cette certitude. NOEL (21) propose que le sujet pessimiste ou optimiste, non conscient de ses possibilités ou lacunes réelles, puisse bénéficier d'une action pédagogique destinée préalablement à éveiller cette conscience métacognitive.

Il nous est apparu un intérêt évident pour les questions à choix multiples de réponses dans le contrôle objectif et approfondi des connaissances. Elles obligent le candidat à se concentrer sur certains problèmes, sans devoir disserter comme le soulignait déjà PIERON (22).
Les critiques veulent que les QCM soient ce qu'elles ne sont pas ; "ne pas permettre de mesurer la tolérance à l'égard de l'avis d'autrui". Les QCM ont leurs faiblesses comme toutes les autres techniques d'examen. Mais ici au moins, le candidat sait exactement ce qu'il doit faire, alors que dans l'examen traditionnel, il doit parfois deviner ce que l'examinateur attend de lui.

Selon TOCHON (27) il faut que l'on distingue en évaluation pédagogique ce qui est de l'ordre de l'apprentissage (évaluation métacognitive) et ce qui est de l'ordre des stratégies d'enseignement et ne pas oublier qu'entre l'évaluation institutionnelle et l'apprentissage réside la conscience de l'apprenant.

En cours de formation, la généralisation des QCM nous semble souhaitable autant pour l'enseignant que pour l'étudiant surtout quand elles sont proposées sur ordinateur en libre accès, comme c'est le cas actuellement au CRAME de Bordeaux 2.
En revanche, en évaluation finale sommative non sélective, cela nous semble potentiellement inutile voir dangereux.

L'évaluation pédagogique vise l'apprentissage et ne doit pas l'empêcher. Elle doit intégrer la métacognition, cependant nous avons conscience qu'une grille ne saurait à elle seule provoquer un état où l'on se prend pour son propre référent.

Il est vrai que la pensée d'ANTHISTENE rappellée en exergue, est un peu provocatrice. Cependant nous pouvons tous mesurer dans nos

diverses spécialités médicales, les conséquences d'une certitude dans l'erreur.
Heureusement nous avons vu au travers de cette étude que cette situation reste très marginale, la certitude s'accompagne de la connaissance et le défaut de certitude d'un défaut conscient de connaissance.

Il en résulte un doute, nécessaire pour désapprendre ce qui est faux, d'où l'intérêt de connaître l'amplitude de ce doute .
Dans notre expérimentation, 40 % des étudiants d'odontologie sont en état de certitude en fin de P2.
Il faudrait donc désapprendre (ou apprendre) chez 60 % d'entre eux...

Pour cela nous n'oublions pas qu'au-delà des méthodes d'apprentissage par résolution de problèmes et de mise en équation des processus métacognitifs lors des évaluations, la rencontre ultime entre l'enseignant et l'étudiant reste nécessaire et apporte toutes les nuances que l'approche quantitative n'a pas pu percevoir.

APPENDICE : APPROCHE MATHEMATIQUE

1. ESTIMATION DU DEGRE DE FIABILITE DE LA QCM

1.1. Définition

Pour la mesure de la fiabilité, une méthode consiste à vérifier si les mêmes étudiants soumis au même examen dans des conditions identiques donneraient les mêmes réponses. C'est simple en théorie mais moins évidente en pratique.

La méthode proposée est celle dite des "demi-tests".
Elle permet de mesurer l'homogénéité, la constance interne du test.
Elle concerne exclusivement les QCM dichotomiques avec mécanisme de correction pour le hasard.

1.2. Procédure

Il faut administrer le test intégral à un même groupe, partager le test administré et corrigé en deux moitiés à l'aide d'une table de nombres au hasard (25), puis calculer le coefficient de corrélation de PEARSON (r) entre les deux moitiés du test (25), enfin estimer le coefficient de fiabilité r^{tt} de tout le test en appliquant la formule de SPEARMAN-BROWN

$$r^{tt} = 2r / (R+1)$$

Si l'écart type (S) qui accompagne la moyenne des notes du groupe d'étudiants est faible, le QCM n'a pas grand pouvoir de discrimination et son degré de fiabilité est bas.

Le calcul de l'erreur type (E) des notes est un moyen d'estimer les limites dans lesquelles peut se situer la vraie note de l'étudiant qui est forcément entachée d'erreurs non systématiques.

$$E = S \sqrt{(1-r^{tt})}$$

2. INDICE DE REUSSITE DE LA QCM

2.1. Définition

Cet indice permet de déterminer dans quelle mesure une question d'examen est facile ou difficile.
C'est le pourcentage (%) d'étudiants ayant répondu correctement à une question d'examen, il varie donc de 0 à 100 %

2.2. Procédure

A partir des paramètres suivants ...

F = nombre de réponses exactes dans le groupe fort (empiriquement les 28% de tête)
f = nombre de réponses exactes dans le groupe faible (même %)
N = nombre total d'étudiants appartenant à ces deux groupes

... on applique la formule :

Indice de réussite = 100 (F+f)/N

Une QCM possédant un indice de 50 à 60 % a de fortes chances d'être fiable.

Plus cet indice est corrélé avec une bonne qualité d'anticipation des questions d'examens, plus il est permi de croire que la transmission des objectifs a été effectuée coorectement.

3. INDICE DE DISCRIMINATION DE LA QCM

3.1. Définition

Il détermine dans quelle mesure une question est assez selective pour distinguer un groupe fort d'un groupe faible

3.2. Procédure

A partir des mêmes paramètres on applique la formule suivante :

Indice de discrimination = 2 (F-f) / N

Cet indice varie de -1 à +1

Une QCM possédant un indice de discrimination
supérieur à 0,35 est qualifiée d'excellente
bonne de 0,25 à 0,34
marginale de 0,15 à 0,24
et mauvaise s'il est inférieure à 0,15 .

Nous avons pu nous rendre compte que l'application de cet indice est facilement transposable à l'évaluation des objectifs sensori-moteurs des étudiants de P2 en odontologie.

Cet indice contribue à l'analyse des items des QCM.
Ceux possédant un mauvais coefficient de discrimination ont probablement pour les étudiants, peu de validité de contenu.

4. OPTIMISATION DE LA PENALITE PAR SCORE

4.1. Consigne

Les règles de pénalité sont souvent absente en début d'énoncé. Cette pratique contestée par LECLERCQ (15) est pourtant usuelle lors des évaluations. Comme l'avoue DAMIN (4), cela permet plus de souplesse lors de la correction des épreuves

4.2. Combinaison des réponses possibles

Nous nous intéresserons aux QCM à 4 compléments, sans patron de réponse.

Considérons une épreuve par QCM avec pour chaque question, 4 assertions exactes au plus, soit de 0 à 4.
Il y aura donc $2^4 = 16$ combinaisons par question.

A chaque question, les 16 combinaisons englobent des réponses partielles où l'étudiant donne une des réponses attendues, et des réponses érronées où il donne une des réponses attendues et des réponses fausses.

Prenons l'exemple de la QCM suivante .

N°33 : La digue est utile pour :
* [ ] A: améliorer la sécurité du patient
[ ] B: maintenir les cotons salivaires
* [ ] C: accroître le confort de travail du praticien
[ ] D: maintenir les dents mobiles lors d'un traitement endo

Si l'étudiant coche A et C nous avons la bonne réponse, s'il ne coche que C nous avons une réponse partielle, s'il coche C et D nous avons une réponse erronée, s'il ne coche que B la réponse est fausse .

Le détail de la distribution des 16 combinaisons en fonction du nombre de réponses exactes à chaque QCM se résume dans le tableau de synthèse suivant.
Il est adapté de celui de DAMIN (4), le sien concerne des QCM à 4 compléments dont deux au plus sont exactes.

Nous appelons ...

Bonne Réponse (BR) : toutes les réponses de la QCM sont cochées correctements
Réponse Partielle (RP) : une partie des bonnes réponses de la QCM n'est pas cochée
Réponses Erronnée (RE) : une partie des bonnes réponses de la QCM n'est oas cochée, d'autres propositions de réponses cochées sont erronées
Réponses Fausses (RF) : toutes les propositions de réponses cochées sont fausses

Nbre d'assertions attendues par le correcteur	Nbre d'assertions données par le candidat	Bonnes Réponses	Réponses Partielles	Réponses Erronées	Réponses Fausses
1	0 1 2 3 4	1		3 3 1	1 3 3 1
2	0 1 2 3 4	1	2	4 4 1	1 2 1
3	0 1 2 3 4	1	3 3	3 3 1	1 1
4	0 1 2 3 4	1	4 6 4		1

Tableau des combinaisons d'erreurs pour une QCM à 4 choix de réponses possibles

Prenons un autre exemple pour illustrer ce tableau, celui concernant 4 assertions attendues par le correcteur, il ne peut pas y avoir de réponses érronées (en cochant une mauvaise case).

N° 21 Les facteurs de variation de temps de prise de l'eugénate sont :

* [ ] A: la granulométrie

* [ ] B: la quantité d'eau incorporée au mélange

* [ ] C: le rapport liquide/poudre

* [ ] D: la température de la plaque à malaxer

Si l'étudiant ne coche aucune case, la réponse est fausse : une seule combinaison possible.
Si l'étudiant ne coche qu'une case, la réponse est partielle : 4 réponses partielles possibles (4 cases)
Si l'étudiant coche 2 cases, la réponses est partielle : 6 réponses partielles possibles (AB, AC, etc ...); en terme de dénombrement nous sommes dans la situation des groupes non ordonnés avec tirage sans remise ayant pour expression :

$$C_4^2 = 4!/[2!*(4-2)!] = 6$$

Si l'étudiant coche 3 cases, la réponse est partielle : $C_4^3 = 4$ réponses possibles (ABC etc ...)

Enfin si l'étudiant coche 4 réponses, la réponse est bonne : 1 seule possibilité .

4.3. Règles de la pénalité selon DAMIN

Le problème que nous nous proposons de résoudre est de déterminer le nombre de points à enlever dans les différentes situations que nous venons de décrire où la réponse n'est pas bonne.

Dans le cas des réponses erronées la question ne rapporte aucun point et dans celui des réponses partielles ce sera au prorata du nombre de bonnes réponses.

Pour quantifier la pénalité "x", nous devons définir la loi de la variable aléatoire égale à la note obtenue .

Dans les QCM où il peut y avoir jusqu'à 4 assertions exactes quatre cas de figures peuvent se présenter avec à chaque une sele bonne réponse (BR).
Les valeurs seront exprimées en 1/100 ème de points par commodité.

4.3.1. Une assertion exacte attendue

Dans une épreuve notée sur 20 comportant 60 questions de valeur identique, chaque question vaut 20/60 = 0,33 point (exprimé 33 en 100ème de point).
A la lecture de la première ligne du tableau qui précède , les 16 combinaisons possibles se résument à 8 (1+3+3+1) réponses fausses (RF) possibles et 7 (3+3+1) réponses erronées (RE).
Les réponses erronées ne donnent pas de pénalité mais ne rapportent aucun point (0 en 100ème)

	BR	RE	RF
Note 1	33	0	- x
Prob	1/16	7/16	8/16

Dans ce cas la moyenne des points vaut : Moy1=(33-8x)/16

4.3.2. Deux assertions exactes attendues

Pour 2 assertions exactes, la réponse partielle doit apporter une note inférieure à la moyenne, elle est établie au maximum à 5/20, 60 questions donnant 5 points apportent 5/60 = 0,08 point chacune (8 en 100ème).
Toujours d'après la troisième ligne du tableau de synthèse pour une bonne réponse il y a 4 réponses fausses (RF), 9 réponses erronnées (RE) et 2 réponses partielles (RP) ce qui nous donne :

	BR	RP	RE	RF
Note 2	33	8	0	- x
Prob	1/16	2/16	9/16	4/16

La moyenne des points vaut : Moy2=(49-4x)/16

4.3.3. Trois assertions exactes attendues

Pour 3 assertions exactes, le gain des réponses partielles apporte au mieux le 1/3 du maximum de la note soit 20/3 = 7 points.
Donc pour 60 questions cela fait 7/60 = 0,11 (11 en 100ème) chacune .
D'après le tableau des distributions pour une bonne réponse il y a 2 réponses fausses (RF), 7 réponses erronnées (RE) et 6 réponses partielles (RP) ce qui nous donne :

	BR	RP	RE	RF
Note 3	33	11	0	- x
Prob	1/16	6/16	7/16	2/16

La moyenne des points vaut : Moy3=(99-2x)/16

4.3.4. Quatre assertions exactes attendues

Pour 4 assertions exactes , même raisonnement

	BR	RE	RF
Note 4	33	11	- x
Prob	1/16	14/16	1/16

La moyenne des points vaut : Moy4=(187-x)/16

4.3.5. Synthèse

La moyenne globale vaut :

$$Moy = a\ Moy1 + b\ Moy2 + c\ Moy3 + d\ Moy4$$

avec a+b+c+d=1

D'après nos statistiques, il faut attribuer les valeurs suivantes aux paramètres :

a=0,1 b=0,28 c=0,34 d=0,28

Ce qui nous donne pour moyenne pondérée :

Moy = (103,04-2,88x)/16

soit une pénalité de 35/100ème pour avoir une moyenne nulle .

En définitive,

le gain en cas de bonne réponse est de 33/100 ème,
celui en cas de réponse partielle de 8 à 11/100 ème,
il est nul en cas de réponse érronée,
la pénalité en cas de réponse fausse est de -35/100 ème.

Il en résulte que la pénalité selon DAMIN de la réponse fausse, consiste à enlever autant de points qu'on en accorde en cas de bonne réponse.

Une telle sanction est mieux adaptée dans un esprit de sélection (concours de l'ENAC en l'occurence) que d'évaluation.
Elle correspond à la pénalité conditionnelle initiale expérimentée dans les cas où l'étudiant donne une réponse fausse avec la certitude d'avoir raison.

Lors du calcul de la moyenne d'un groupe d'étudiants DAMIN exploite aussi l'écart type qui l'accompagne.
Plus celui ci est important, plus l'aléatoire augmente et plus le risque grandit.

BIBLIOGRAPHIE

(1) ARMSTRONG (A-M.).- The Development of Self-Regulation Skills Through the Modeling and Structuring of Computer Programming .- Educational Technology Research and Development, 1989, 37, 2, pp. 69-76

(2) BARRET (G.).- Les différentes étapes du processus métacognitif de l'autoévaluation .- Mesure et Evaluation en Education, 1990, 13, 3, pp. 27-33

(3) CHARTIER (D.), LAUTREY (J.).- Peut-on apprendre à connaître et à contrôler son propre fonctionnement cognitif ?.- L'orientation scolaire et professionnelle, 1992, 21, 1, pp. 27-46

(4) DAMIN (Gilbert) .- Quelles pénalités ? . in 3ième colloque international du groupe ESIEE Paris "QCM et questionnaires fermés : méthodes et pratiques". 5 au 7 Décembre 1994. Marne-la-Vallée. - ESIEE,pp 53-59

(5) DE FINETI (B.) .- Methods for discriminating levels of parial knowledge concerting a test item .- British Journal of Mathematical and Statistical Ppsychology, 1965, 18, 1, pp. 87-123

(6) DELANDSHEERE (G.) .- Evaluation continue et examens . Précis de docimologie .- Liège, Education 2000 Labor, 1992, 310 p.

(7) DENEF (J.F.).- Techniques modernes d'enseignement . in XIIèmes Journées du Collège National des Enseignants en Odontologie Conservatrice .28, 29 Mai 1993. Bruxelles

(8) D'IVERNOIS (J.F.) .- Objectifs de formation dans les sciences de la santé .- Séminaire du 28,29 Mars 1994, Centre de Pédagogie Médicale : Université Bordeaux 2 .

(9) FABRE (J-M.) .- Subjective Uncertainly and the Structure of Set of all Possible Events in : LECLERCQ (D.), BRUNO (J.).-Item Banking : Intractive testing and Self-Assessment .- NATO ASI Series, Series F: Computer and Systems Sciences, 1992, 112, pp. 99-114

(10) FORTIN (M.), MADARNAS (T.) .- La pédagogie par résolution de problèmes, Centre de Pédagogie Médicale, Unaformec-Aquitaine, 8 Mars 1995, Bordeaux, Université Bordeaux 2.

(11) GAGEY (J.) .- La scientificité de la clinique .- Psychanalyse à l'université, 1975, 1, 1, pp. 75-96

(12) GRENIER (B.) .- Evaluation de la décision médicale . Introduction à l'analyse médico-économique .- Paris : Masson, 1996, 412 p.

(13) JEANNEL (A.) .- La place du Multimédia dans la Formation .- Journée "Multimédia et Enseignement sur Mesure", Centre de Recherches Appliquées en Méthodes Educatives. 8 Fevrier 1996, Bordeaux, Université Victor Segalen Bordeaux 2.

(14) LAVEAULT (D.), FOURNIER (C.).- Evaluation par objectifs : une approche métacognitive .- Mesure et Evaluation en Education, 1990, 13, 1, pp. 57-74

(15) LECLERCQ (Dieudonné), GILLES (Jean-Luc).- GUESS, un logiciel pour entraîner à l'autoestimation de sa compétence cognitive . in 3ième colloque international du groupe ESIEE Paris "QCM et questionnaires fermés : méthodes et pratiques". 5 au 7 Décembre 1994. Marne-la-Vallée. -ESIEE, pp 137-157

(16) LECLERCQ (D.) .- Validity, Reliability, and Acuity of Self-Assessment in Educational Testing in : LECLERCQ (D.), BRUNO (J.).- Item Banking: Intractive testing and Self-Assessment .- NATO ASI Series, Series F: Computer and Systems Sciences, 1992, 112, pp. 114-131

(17) LECLERCQ (D.), BOXUS (E.), de BROGNIEZ (P.), WUIDAR (H.), LAMBERT (F.) .- The TASTE Approach: General Implicit Solutions in Multiple Choice Questions (MCQs), Open Books Exams and Interactive Testing in : LECLERCQ (D.), BRUNO (J.).- Item Banking: Intractive testing and Self-Assessment .- NATO ASI Series, Series F: Computer and Systems Sciences, 1992 ,112, pp. 210-32

(18) LECLERCQ (D) .- La conception des questions à choix multiple .- Bruxelles : Labor Ed., 1986, 153 p.

(19) MATTHEWS (M-D.).- Self-Assessment Programmes and Aspects of Audit .- Journal of the Royal College of Physicians of London, 1979, 13, 3, pp. 139-42

(20) MAUBANT (Ph.).- Pour une approche méta-cognitive de l'acte d'apprendre au service d'une pédagogie de l'alternance .- Ateliers Lyonnais de pédagogie - Comment mieux réussir l'alternance Ecole-Entreprise, 1992, 44, pp. 13-19

(21) NOEL (B.), MAYENCE (A-M.).- Construction et essai d'un test de jugement métacognitif .- Recherche en Education . théorie et Pratique, 1992, 8, pp. 13-20

(22) PIERON (H.).- Examens et docimologie .- Paris : Presses Universitaires de France, 1963, 190 p.

(23) QUINTON (A.).- Séminaire de docimologie : QCM . Centre de Recherches Appliquées en Méthodes Educatives : Université Bordeaux 2, Victor Segalen .

(24) QUINTON (A.).- Le concours de PCEM 1 en 1996 à l'Université de Bordeaux II : Candidats - Résultats - Analyse docimologique - Enquête sur le vécu du PCEM 1 et les modalités de travail - Tutorat.- Rapport édité par le Centre de Recherches appliquées en Méthodes Educatives : Université Bordeaux 2, Victor Segalen .

(25) SCHWARTZ (D.).- Méthodes statistiques à l'usage des médecins et des biologistes .- Paris : Flammarion Médecine-Sciences, 1983, 318 p.

(26) SHUFORD (E.), ALBERT (A.), MASSENGILL (E.).- Admissible probability measurement procedures .- Psychometrika, 1966, 31, pp. 125-45

(27) TOCHON (F-V.).- L'évaluation métacognitive : une évaluation d'apprentissage .- Mesure et Evaluation en Education, 1990, 13, 3, pp. 35-54

(28) VACHEY (E.) .- L'Entraînement informatisé au diagnostic Codé .- Journal d'odontologie Conservatrice, 1987, 5, pp.54-6

(29) VITRAC .- La recherche documentaire. Méthodologie et technologies , Séminaire du 20 et 21 juin 1994, Centre de Pédagogie Médicale : Université Bordeaux 2 .

(30) WOLFS (J-L.) .- Contribution à l'opérationnalisation du concept de métacognition .- Recherche en Education . théorie et Pratique, 1992, 10 ,pp. 3-13

(31) WOLFS (J-L.) .- Etude des relations entre performance cognitive et métacognition : Revue de la littérature .- Recherche en Education . théorie et Pratique, 1992, 10 ,pp. 15-23