



UNIVERSITÉ DE FRIBOURG SUISSE
UNIVERSITÄT FREIBURG SCHWEIZ

Evaluation des apprentissages

Promoteur : Bernadette Charlier

Mémoire de fin d'études
Présenté par François Mooser en vue de
l'obtention du

**Diplôme de formation continue
en Enseignement Supérieur et
Technologie de l'Education**

Année académique 2003/2004

Table des matières

Le présent mémoire est divisé en chapitres A., B., C., D., E. et F.
Chaque chapitre est subdivisé à son tour en sections numérotées.

	Page
A. Description du projet	1
B. Mise en œuvre du projet	2
1. Définition de l'évaluation	2
1.1 Formulation des questions	2
1.2 Observation des réponses (= informations) apportées par les étudiants en cours d'épreuve	4
1.3 Analyse-mesure	6
1.4 Jugement	7
2. Analyse d'une évaluation formative et d'une évaluation sommative	8
2.1 Evaluation formative	8
2.2 Evaluation sommative	9
3. Critères de qualité d'une épreuve pédagogique	12
4. Usage des questions à choix multiples (QCM) dans l'évaluation et degré de certitude	13
4.1 Degré de certitude et indice de réalisme	14
4.1.1 Evaluation des capacités cognitives de base : exemple d'un test d'entrée à l'université	15
5. Métacognition	22
6. Evaluation d'une thèse de doctorat	23
C. Conclusions	27
D. Liste des abréviations	29
E. Bibliographie	30
F. Webographie	31

Evaluation des apprentissages

A. Description du projet

Dans ma profession d'enseignant, je suis amené à évaluer les apprentissages des étudiants. Participant aux examens de 1^{er} et 2^{ème} propédeutique de médecine, je constate que l'évaluation des apprentissages occupe une place très importante dans l'enseignement. Les connaissances acquises par les étudiants peuvent être testées sous forme d'une évaluation formative au cours de leur processus d'apprentissage et sous forme d'une évaluation sommative certificative (examen propédeutique) à la fin de l'année académique.

L'utilisation de l'informatique (document « Excel », plate-forme d'enseignement à distance, CD-ROM) est fort intéressante et très utile dans la mesure où l'élève peut s'auto-évaluer par le « jeu » de questions-réponses et ainsi de garder à domicile une référence relative à la matière enseignée.

Mon objectif est d'être capable d'évaluer les apprentissages des étudiants de façon la plus judicieuse possible : avoir des outils nécessaires aidant à réaliser une évaluation qui soit la plus objective possible et ce à différents niveaux des apprentissages : les savoirs (connaissances acquises), les savoir-faire (convergent et divergent), les savoir-être et les savoir-devenir (Jean-Marie de Ketele). Comment préparer au mieux l'étudiant à affronter l'examen ? Au cours d'une évaluation (qu'elle soit formative et/ou sommative), comment stimuler, chez les étudiants, un processus de métacognition et développer une confiance en soi, une estime de soi dans la réponse à la question posée ? Comment peut-on mesurer la confiance qu'apporte un étudiant dans sa réponse ? Les nouvelles Technologies de l'Information et de la Communication (outils TICs) peuvent-elles jouer un rôle dans l'évaluation ? Finalement, comment mener à bien la métacognition chez l'enseignant tout au long de son parcours professionnel ?

Dans la suite du mémoire, je propose des réponses possibles aux questions posées en citant des applications concrètes et en les analysant.

Ma réflexion est basée sur les connaissances acquises au cours de ma formation continue en Enseignement Supérieur et Technologie de l'Education ainsi que sur des lectures personnelles.

B. Mise en œuvre du projet

1. Définition de l'évaluation

En référence à l'article « Les exigences de l'évaluation » (12)

EVALUER, c'est :

- recueillir, de manière systématique, valide et fidèle, des informations appropriées à ses objectifs d'enseignement (Observation)
- interpréter ces informations à l'aide de critères (Analyse-mesure)
- en vue d'établir des conclusions et/ou de prendre des décisions (Jugement)

⇒ trois phases dans une évaluation :

- Observation
- Analyse-mesure
- Jugement

Les deux dernières phases d'évaluation dépendent de la première et les 3 phases sont souvent imbriquées l'une dans l'autre. Mais avant la première phase, vient se greffer la formulation des questions.

1.1 Formulation des questions

La première étape consiste à formuler des questions et ce en tenant compte du principe de la cohérence. En effet, on évalue les performances des étudiants par rapport aux objectifs d'apprentissages définis dans le scénario pédagogique, au contenu du cours et aux critères définis sous la forme d'une grille d'évaluation. L'enseignant doit décider sur quoi, sur quel(s) niveau(x) du savoir (de Ketele, du plus bas au plus élaboré) il va évaluer :

- veut-il évaluer seulement les connaissances acquises à restituer : par cœur ; niveau « connaissances » dans la taxonomie de Bloom.

Exemples : a) géographie : quelle est la capitale de la Suisse ?

b) anatomie : quelles sont les branches de l'artère carotide externe ?

- veut-il évaluer les applications : savoir-faire convergent ; niveau « compréhension » et « application » dans la taxonomie de Bloom. Le niveau « compréhension » implique le fait d'avoir saisi le sens global, la portée de la matière enseignée. Et pour pouvoir appliquer, il faut bien sûr avoir compris.

Exemple 1 : anatomie : orientation d'un humérus (droit ou gauche) en se basant sur les 3 plans dans l'espace (horizontal, frontal et sagittal) (application formelle)

Exemple 2 : clinique : une personne est tombée et suite à sa chute, elle s'est fracturée le péroné. Le médecin fait alors une radiographie et opère le patient en posant une plaque sur la partie fracturée de l'os et ce en vue de le consolider (diagnostic évident, application instrumentale)

- veut-il évaluer les capacités de raisonnement, d'analyse, les compétences : savoir-faire divergent, c'est-à-dire être capable de mobiliser ses ressources pour résoudre une famille de situations à problèmes ou de tâches ; niveau « analyse » dans la taxonomie de Bloom.

Exemples : a) résolution d'un problème de physique : on fait appel à l'analyse découlant des connaissances acquises, de la compréhension et des applications. On dégage des éléments-clés du problème, on propose des hypothèses, on essaie de schématiser l'énoncé du problème dans le but de le résoudre.

L'énoncé en soi ne dit pas ce que l'étudiant doit faire, à lui de trouver un cheminement, un scénario qui lui permette d'apporter une solution au problème (voir exemple d'un problème de physique sur la dynamique des fluides dans la **section 1.2 du présent chapitre**).

b) clinique : une personne est victime d'un accident de la circulation. Quelques minutes plus tard, elle se plaint de douleurs violentes qui l'empêchent de bouger ses jambes et elle dit ne pas avoir perdu connaissance. Dans cette situation, le diagnostic est moins évident que dans celle du péroné fracturé. Le médecin va devoir analyser avec d'autres collègues ce cas et ce afin de pouvoir proposer le meilleur traitement possible, pour autant qu'un traitement existe.

- veut-il évaluer telle attitude à adopter face à telle situation : savoir-être ; niveau « évaluation » dans la taxonomie de Bloom. Et pour évaluer une situation, il faut être capable de l'analyser.
Exemple 1 : imaginer le comportement à adopter de la part du futur médecin face à son patient atteint de la maladie de Parkinson. Le médecin doit porter un regard critique sur la situation et ce en l'évaluant : par exemple, avant d'annoncer la maladie au patient, doit-il en parler d'abord à sa famille ?
Exemple 2 : porter un jugement critique à l'aide d'arguments pertinents sur un thème particulier comme par exemple celui de l'euthanasie et ce en évitant de se laisser emporter par ses émotions
- veut-il évaluer la créativité : savoir-devenir. Niveau « synthèse » (dans le sens de produire, de créer quelque chose d'original) dans la taxonomie de Bloom. Et pour pouvoir créer, il faut avoir apporté au préalable un jugement critique (évaluation).
Exemple : la mise en œuvre d'un projet d'ingénierie avec la conception d'appareils de mesure par exemple. Imaginer un appareil permettant de mesurer l'activité des neurones du cerveau.

Pour éviter tout malentendu le jour de l'examen, l'enseignant devrait expliciter dans ses objectifs d'apprentissage ses attentes vis-à-vis de ses étudiants.

En d'autres mots, sur quoi va porter l'évaluation, que veut-il réellement évaluer ? Qu'attend-il de ses étudiants ?

La formulation des questions devient donc très importante. C'est alors qu'apparaît toute la problématique de la communication entre pairs (enseignant-enseignant : examinateur et co-examinateur) et entre enseignant-étudiant. Le message sous forme de question à transmettre porte un double codage : celui de l'émetteur (l'enseignant qui élabore des questions d'examen) et celui du récepteur (ses pairs et les étudiants). En effet, l'émetteur émet un message (qui est meublé des marques de subjectivité, de la culture, de l'histoire de l'émetteur) décodé (par l'émetteur) au récepteur (qui reçoit le message à travers son propre décodage) ⇒ l'objectivité est située entre l'émetteur et le récepteur. Le message porte ainsi un double codage : celui de l'émetteur et celui du récepteur.

Il devient donc impératif que les questions puissent être comprises sans ambiguïté aussi bien par les étudiants que par les examinateurs. En d'autres mots, elles doivent être pertinentes.

De plus, l'aspect temporel joue un rôle crucial. En effet les questions doivent être adaptées au temps imparti de l'épreuve. Il appartient alors à l'enseignant de calculer le temps que mettra l'étudiant pour répondre à la question. Une des possibilités pour le faire serait que l'enseignant réponde lui-même à ses propres questions en se mettant dans la peau d'un étudiant et ce en doublant le temps mis à disposition pour y répondre, ou sous forme de prétest auprès d'un groupe d'étudiants.

1.2 Observation des réponses (= informations) apportées par les étudiants en cours d'épreuve

Une fois les questions formulées sans ambiguïté possible, apparaît la problématique quant aux réponses attendues de la part des étudiants.

Qui dit questions pertinentes, dit réponses recueillies pertinentes. Cela signifie qu'elles (questions et réponses) doivent être conformes à l'objectif que l'évaluateur s'est fixé pour l'évaluation. Autrement dit, les informations données par les étudiants qu'il recueille doivent lui permettre de vérifier ce qu'il veut vérifier. Par exemple, si un enseignant veut évaluer la qualité du dialogue maître-élève en classe, il ne doit pas chercher à connaître le niveau de connaissances des élèves, mais bien l'avis des élèves sur la qualité du dialogue. Les informations à recueillir ne sont pas des connaissances, mais des avis, c'est-à-dire des représentations. Par contre, s'il veut vérifier si les élèves ont acquis les connaissances nécessaires dans une branche, il ne lui sert à rien de demander aux élèves si les cours ont été intéressants (1)

Les réponses à apporter doivent être conformes aux attentes de l'enseignant certes, mais ce dernier doit aussi accepter d'autres voies possibles pour parvenir à la même bonne réponse. Cela vaut particulièrement dès que l'on évalue les niveaux du savoir à partir des savoir-faire divergents.

Prenons un exemple d'un problème concret de la vie quotidienne énoncé lors d'un examen de physique sur la dynamique des fluides (7) :

Un avion vole à la vitesse de 200 m/s. Ses ailes ont une forme telle que la vitesse d'écoulement de l'air est plus grande au-dessus qu'au-dessous. On admet qu'elle vaut 230 m/s au-dessus et 200 m/s au-dessous. La surface utile des ailes vaut 220 m². Là où vole l'avion, la masse volumique de l'air vaut 0,6 kg/m³. Calculer la force qui soutient l'avion.

Le professeur s'attend parfois à un seul cheminement possible dans le raisonnement de l'étudiant pour arriver à la solution du problème (« calculer la force qui soutient l'avion »). Or, lors de l'épreuve, pour parvenir à la même conclusion, au même résultat, l'étudiant apporte un cheminement dans son raisonnement (pour analyser le problème posé) tout autre auquel le professeur s'attendait. Il y a donc un risque que le professeur n'accepte pas le raisonnement de l'étudiant alors qu'il s'avère être parfaitement correct. D'où le rôle d'une deuxième personne (co-examineur) participant à l'évaluation qui peut intervenir en faveur de l'étudiant le cas échéant. Au travers de cet exemple, cela montre bien combien il est primordial d'entrevoir toutes les possibilités pour arriver à la bonne réponse et ce déjà lors de la première phase d'une évaluation. Cela évitera des désagréments, des surprises lors de l'interprétation des réponses données par les étudiants. L'exemple décrit pose particulièrement problème dans un examen oral où le professeur doit prendre une décision, apporter son jugement dans un temps très restreint.

Donc, en conclusion, plus on va vers des niveaux élevés dans les savoirs, plus il devient difficile de construire un examen et l'enseignant aura besoin d'engager tout un processus de métacognition dans lequel il va se questionner sur toutes les possibilités de cheminement qu'il entrevoit pour arriver à la réponse correcte.

Un autre problème surgit : la validité.

La validité

Les questions que l'enseignant pose portent bien sur ce qu'il veut évaluer et donc les scores des apprenants doivent représenter ce que l'enseignant veut évaluer.

Reprenons l'exemple du problème de physique. Dans un examen, il est tout à fait envisageable de poser 2-3 problèmes analogues à l'exemple décrit, pourvu que l'aspect temporel soit respecté et que l'examen portant sur la dynamique des fluides couvre toute la matière enseignée au cours.

Dans un tel problème, plusieurs niveaux du savoir entrent en ligne de compte et peuvent être énumérés comme grille de critères d'évaluation :

- a) les connaissances (addition, soustraction, multiplication et division),
- b) la compréhension et les applications : appliquer les opérateurs mathématiques, et ce après avoir saisi leur signification, pour résoudre une équation que l'étudiant va poser dans l'analyse de la situation,
- c) finalement l'analyse du problème posé (cheminement dans le raisonnement).

Voici un exemple de questionnement de la part du concepteur de l'épreuve d'évaluation :

va-t-on alors apporter un même jugement pour un étudiant (**premier cas de figure**) maîtrisant les connaissances et les applications avec un cheminement incorrect dans son raisonnement, son analyse du problème, que pour celui (**deuxième cas de figure**) qui aura entamé un cheminement correct dans son raisonnement, son analyse du problème, mais en ne maîtrisant pas les connaissances et les applications ?

Dans le cas où l'enseignant désirerait accorder davantage de valeur, d'importance au raisonnement, à la capacité d'analyse plutôt qu'au simple fait d'appliquer les connaissances acquises (ex. : les opérateurs mathématiques) pour résoudre telle équation, le **deuxième cas de figure**, correspondant aux attentes de l'enseignant certes, mais ne maîtrisant pas les deux premiers niveaux du savoir, obtiendra-t-il toutefois une meilleure note que le **premier** ? Tout dépend alors de la valeur quantitative, de l'importance accordée dans la mesure de tel niveau du savoir. Et cette décision se prend dès la conception de l'épreuve et devra être précisée dans la formulation de la question : par exemple, combien de points accorde l'enseignant pour chacun des 3 niveaux du savoir : les connaissances, les applications et la capacité d'analyse du problème (raisonnement) ?

Et le fait de trouver la réponse correcte (« force qui soutient l'avion ») découle des 3 niveaux du savoir qui doivent être alors parfaitement maîtrisés.

En outre, vient encore se greffer le cas où l'étudiant maîtrise partiellement un niveau du savoir. Toute cette problématique quant à savoir quelle valeur quantitative (nombre de points) l'enseignant doit-il accorder à tel élément de réponse apporté par l'étudiant montre combien il devient important de bien analyser chacun des niveaux du savoir avec une échelle d'intervalle et donc de définir la manière dont les points sont attribués aux performances des étudiants. L'aspect quantitatif avec différents scénarios envisageables dans les réponses données sera explicité dans la **section 1.3 du présent chapitre**.

Cet exemple montre combien la complexité de préparer un examen augmente avec le nombre de niveaux du savoir que l'on veut évaluer.

Finalement, il est important qu'il n'y ait pas de pénalité chez l'étudiant, d'où la notion de fidélité d'une évaluation.

La fidélité

En effet, les résultats doivent être similaires selon que l'étudiant sera évalué par tel ou tel professeur ou quel que soit le jour d'examen. Donc il s'agit d'établir une concordance dans les critères d'évaluation en vue d'interpréter de la même façon les informations (=réponses) données par les étudiants. Ainsi le jugement deviendra identique selon le professeur et le calendrier.

1.3 Analyse-mesure

En plus du découpage temporel du module, des objectifs d'apprentissages et des activités proposées qu'elles soient sous forme de travaux pratiques, de séminaires ou d'apprentissages par problèmes (APP), il devient important, dans le scénario pédagogique, de bien expliciter les critères d'évaluation. Ainsi l'enseignant pourra éviter tout malentendu et plus le professeur s'est explicité les critères, meilleure sera son interprétation et cela évitera tout malentendu avec ses pairs et ses étudiants.

Reprenons l'exemple du problème de physique précité...

Parfois, le professeur s'attend à un seul cheminement possible dans le raisonnement de l'étudiant pour arriver à la solution du problème (« calculer la force qui soutient l'avion »). Il doit donc tenir compte de plusieurs cheminements possibles dans le raisonnement pour résoudre le problème posé. En d'autres mots, pour résoudre le problème posé, si un étudiant raisonne en suivant un cheminement A que le professeur n'a pas envisagé, son score devrait être le même que pour l'étudiant ayant raisonné en suivant un cheminement B que le professeur a envisagé lors de la conception de l'épreuve.

Un problème réside dans le fait de quantifier l'évaluation de la solution apportée par l'étudiant → échelle d'intervalle (attribuer un nombre). Les coefficients (nombres) attribués dépendent des objectifs que se fixe l'enseignant lors de la conception de l'épreuve, c'est-à-dire de bien définir sur quoi porte l'évaluation et quelle importance il va attribuer à tel niveau du savoir.

Exemple : si le problème est totalement résolu, l'étudiant obtiendra un maximum de **15 points**.

A présent, il s'agit d'attribuer un nombre de points pour chacun des niveaux du savoir. L'enseignant tiendra compte de l'ordre croissant d'importance des niveaux du savoir.

Par exemple, en reprenant la grille des critères d'évaluation proposée dans la **section 1.2 du présent chapitre**, il peut procéder comme suit :

a) Connaissances (savoirs)	:	2/15
b) Compréhension et applications (savoir-faire convergent)	:	4/15
c) Analyse (savoir-faire divergent)	:	9/15
Total	:	15/15

Voici alors les différents scénarios que peut envisager l'enseignant lors de la conception de l'épreuve : les différentes possibilités sont présentées dans l'ordre croissant de réussite de l'épreuve (du plus mauvais au meilleur) et ce en fonction de la hiérarchisation des savoirs décrits ci-dessus :

1. non maîtrise de a), b) et c)
2. maîtrise de a), non maîtrise de b) et c)
3. maîtrise de b), non maîtrise de a) et c)
4. maîtrise de a) et b), non maîtrise de c)
5. maîtrise de c), non maîtrise de a) et b)
6. maîtrise de a) et c), non maîtrise de b)
7. maîtrise de b) et c), non maîtrise de a)
8. maîtrise de a), b) et c)

Comme le montre la complexité des scénarios décrits ci-dessus, il devient alors nécessaire de rendre l'examen le plus objectif et le plus crédible possible et ce pour enlever tout risque de subjectivité dans l'interprétation des résultats, d'où l'importance de la quantification des niveaux du savoir.

A présent il s'agit d'établir une échelle d'évaluation. Dans le canton de Fribourg, nous avons une échelle allant de 1 à 6. Le critère de réussite est d'obtenir la note 4 (règle des 2/3). Voici un tableau décrivant les points, les notes et reprenant les 8 scénarios précités → passage du score, sous forme de points attribués (échelle d'intervalle), à la note (échelle ordinale) :

Points	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Note	1	1.3	1.7	2	2.3	2.7	3	3.3	3.7	4	4.3	4.7	5	5.3	5.7	6
Scénarios	1.		2.		3.		4.			5.		6.		7.		8.

Dans cet exemple, l'enseignant met l'accent sur la capacité d'analyse. Ainsi l'étudiant peut réussir l'épreuve en maîtrisant parfaitement le savoir-faire divergent. Chacun des scénarios proposés peut être affiné dans le cas d'une maîtrise partielle d'un niveau du savoir. Dans ce cas, on peut prendre en compte des demi-points (0.5, 1.5, etc.). Prenons un exemple : si l'étudiant ne maîtrise que partiellement ses connaissances et rien d'autres, il obtiendra un nombre de points compris entre 0.5 et 1.5, selon le degré de maîtrise de ses connaissances. Au contraire, si l'étudiant maîtrise, en plus de ses connaissances, partiellement les applications et/ou le raisonnement et la compréhension, il obtiendra un nombre de points compris entre 2.5 et 14.

Comme le fait de trouver la réponse correcte à la question (« force qui soutient l'avion ») découle des 3 niveaux du savoir qui doivent être alors parfaitement maîtrisés, si un étudiant trouve par hasard la réponse correcte, le professeur va devoir alors interpréter la façon à laquelle l'étudiant y est parvenu. L'étudiant va alors récolter des points en fonction de la maîtrise des niveaux du savoir. En aucun cas l'étudiant ne récoltera davantage de points par le simple fait d'avoir trouvé « par hasard » la bonne réponse : en effet, dans ce cas, le professeur peut légitimement soupçonner l'étudiant de fraude (regarder chez son voisin par exemple).

Cette quantification contribue à renforcer la validité et la fidélité d'un examen.

Un autre exemple relatant le travail de doctorat en Sciences Naturelles et faisant intervenir tous les niveaux du savoir sera décrit à la section 6. du présent chapitre.

1.4 Jugement

La question que l'on se pose est la finalité d'une évaluation ; pourquoi, dans quel but évalue-t-on ? Douze finalités d'une évaluation ont été décrites (**10**) : conscientiser, motiver, activer, orienter, réguler, placer, consolider, sélectionner, classer, comparer, certifier et connaître.

En regard à la description qui précède, on pense alors tout de suite à une évaluation certificative sommative. Après avoir décrit le processus de préparation d'une évaluation de la part de l'enseignant principalement en regard à son scénario pédagogique (principe de cohérence, voir plus haut), vient alors la question de se demander comment préparer au mieux l'étudiant à affronter l'examen ? Est-ce au travers de tests intermédiaires non certificatifs réalisés durant le processus d'apprentissage ? En d'autres mots, quel est le rôle d'une évaluation formative ?

Voici à présent dans un premier temps une description analytique d'une évaluation formative tirée de mon enseignement à l'unité d'anatomie. Dans un second temps, je vais décrire une évaluation sommative réalisée dans le même institut et qui reprend les deux premiers niveaux du savoir (savoirs = connaissances et savoir-faire convergents = compréhension/application). Il s'agit d'un extrait de mon activité à distance réalisée dans le cadre du module « analyse et évaluation des apprentissages et des dispositifs ».

2. Analyse d'une évaluation formative et d'une évaluation sommative

Pour chacun des modes d'évaluation, je désire introduire une définition d'après G. De Landsheere (2)

2.1 Evaluation formative

Evaluation formative : « c'est l'évaluation 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 proposer ou de lui faire découvrir des stratégies qui lui permettent de progresser. » (2)

Tout d'abord voici le contexte dans lequel l'évaluation est réalisée :

Description de mon activité d'enseignement de l'anatomie humaine

Les séances de travaux pratiques s'articulent autour de 3 grands thèmes :

- a) Dissection des membres supérieurs et inférieurs (1^{ère} année) : avec un accent particulier sur l'appareil locomoteur, - myologie et cinétique des mouvements, arthrologie, vaisseaux et nerfs (voies périphériques), ostéologie, topographie - : 15 cours de 3 heures chacun (total 45 heures). Parallèlement à la dissection, 6 cours sont consacrés à l'enseignement de l'ostéologie.
- b) Dissection du corps humain (2^{ème} année) : objectifs identiques à ceux de la 1^{ère} année, mais avec un accent particulier sur le système nerveux en général, les organes et les vaisseaux (sanguins et lymphatiques) : 27 cours de 3 heures chacun répartis sur 4 blocs. (total 81 heures)
- c) Cours du système nerveux central (cerveau et moelle épinière préparés) (2^{ème} année) : 11 cours de 1 ou 2 heures chacun (total 15 heures).

Au début de la 1^{ère} année, les étudiants reçoivent une série d'objectifs d'apprentissages à atteindre en vue de l'évaluation formative (matière d'attestation) pour la fin des travaux pratiques de 1^{ère} année et de l'évaluation sommative certificative intervenant à la fin des 1^{ère} et 2^{ème} année de médecine (examens propédeutiques I et II).

Moment et rôle + qualité du feedback

Concernant le point b) décrit précédemment : l'évaluation formative intervient au terme de chaque bloc sous forme d'un test oral d'évaluation (posttest) des connaissances que les étudiants auront acquises au cours du bloc, voire des blocs précédents. Ce test oral d'évaluation dure en moyenne 5 minutes par étudiant. Lors de ces tests, il n'y a pas de notes contrairement aux examens. Mais, si l'étudiant a échoué, il doit refaire le test la semaine suivante. Et si l'apprenant échoue pour la seconde fois, il peut toutefois se présenter à l'examen et le fait qu'il n'ait pas réussi la seconde fois ne joue aucun rôle.

Concernant les points a) et c) décrits précédemment : elle intervient à la fin de tous les cours pratiques sous forme d'un test oral où on interroge l'étudiant sur la totalité des cours pratiques. Les questions que l'on pose au cours de chaque test oral sont des questions qui peuvent être posées lors de l'examen (pour la liste, voir rubrique « **Qualité des questions posées** » de la présente section). De plus, concernant le point a), un test oral d'ostéologie uniquement est réalisé vers le milieu de la période des cours pratiques.

Concernant les points a) et c), on aurait tendance à classer ces tests dans le mode d'évaluation sommative étant donné qu'ils interviennent à la fin de tous les cours pratiques. Mais, comme il

s'agit de tests portant sur les cours pratiques et non sur la totalité des cours (cours pratiques + cours magistraux) et que l'élève est encore en train d'apprendre (dans ce cas, le bilan d'apprentissage est prématuré), je les classe donc dans le mode d'évaluation formative.

Le rôle de ce type d'évaluation est de contrôler l'état des connaissances de l'apprenant, de l'informer sur l'état de ses connaissances et ce en posant des questions similaires à celles posées à l'examen propédeutique. Au travers de cette information, ce mode d'évaluation permet ainsi de repérer les difficultés de l'apprenant (diagnostiquer) et d'améliorer mon enseignement si nécessaire (la façon de transmettre le message pour que l'étudiant soit réceptif à ce que je raconte). Il est aisé à ce niveau de pronostiquer, c'est-à-dire de voir si l'étudiant est capable de réussir à l'examen, étant donné qu'il s'agit ici d'évaluer l'étudiant en posant des questions semblables à celles des examens. En fonction du diagnostic et du pronostic, l'enseignant a alors la possibilité de discuter avec l'apprenant des problèmes rencontrés et de lui donner des conseils dans ses études. Le diagnostic et le pronostic donnent un bon feedback surtout pour l'étudiant dans la mesure où ce dernier pourra se poser des questions sur ses réelles motivations, sa capacité à étudier l'anatomie et éventuellement se réorienter si nécessaire ou le rassurer pour la suite de son processus d'apprentissage au cas où il réussirait le test (acquérir une plus grande confiance en soi). A ce sujet, le département de médecine dispose d'un conseiller aux études.

Ce qui est intéressant dans ce mode d'évaluation, c'est qu'il peut aussi permettre à l'apprenant d'évaluer lui-même ce qu'il sait ou ce qu'il ne sait pas et, de ce fait, peut amener à une auto-évaluation. Au sein de notre unité, les étudiants ont la possibilité d'acquérir un CD-ROM d'embryologie humaine (produit maison) et avec ce CD, ils peuvent s'auto-évaluer au travers d'une série de questions à choix multiples (QCM, trois réponses à choisir). Ils reçoivent ainsi un bon feedback et, comme dans le cas des travaux pratiques, pourront alors se concentrer sur des domaines qu'ils ne maîtrisent pas bien : auto-régulation (se gèrent eux-même sans l'intervention de l'enseignant) dans leur processus d'apprentissage.

L'évaluation formative va amener l'apprenant à s'auto-évaluer (ce qu'il a appris, trouvé difficile, ce qu'il pense devoir réviser davantage, ce qu'il a bien compris, ...etc.) et à faire le bilan de ses apprentissages. Ainsi il va pouvoir développer des compétences métacognitives. Ce mode d'évaluation donne un bon feedback aussi bien pour le professeur (amélioration de son enseignement) que l'apprenant (sait ce à quoi il s'attend à l'examen et travaillera davantage la matière qu'il n'a pas encore maîtrisée). Ce mode d'évaluation a donc pour objectif principal la régulation du processus d'apprentissage chez l'apprenant et de l'enseignement chez le formateur et ce par un processus de rétroaction, de feedback. En effet, en fournissant à l'apprenant des informations sur ses progrès par rapport à un objectif déterminé, l'évaluation formative permet d'intervenir dans le processus d'apprentissage, c'est-à-dire de permettre à l'apprenant de transformer ou de varier ses méthodes et ses stratégies d'apprentissage. Le feedback devient alors bénéfique tant pour l'apprenant que l'enseignant lui-même. Un exemple concret concerne l'orientation d'un humerus (gauche ou droit ?) : si l'étudiant ne sait pas argumenter correctement, et ce car il a des problèmes de vision en 3 dimensions par exemple, il aura alors beaucoup de peine à l'orienter. Le rôle de l'enseignant sera alors de détecter ce problème (se représenter les structures en 3 dimensions, ce qui est essentiel en anatomie) chez l'étudiant et réexpliquer de manière plus approfondie, par exemple en donnant des outils nécessaires au raisonnement à faire. L'enseignant va alors montrer à l'étudiant sur le squelette ce qu'est la position anatomique (debout, bras tendu avec les paumes des mains orientées en avant) et, à partir de cette position, ce qu'on entend par plan horizontal - parallèle au sol et séparant l'os en une partie proximale (supérieure) et une partie distale (inférieure) - ; frontal - perpendiculaire au sol et séparant l'os en une partie antérieure et une partie postérieure - ; sagittal - perpendiculaire au sol et séparant l'os en une partie médiale et une partie latérale -.

2.2 Evaluation sommative

Evaluation sommative : « alors qu'une évaluation formative est normalement effectuée au terme de chaque tâche d'apprentissage, notamment pour intervenir immédiatement là où une difficulté se

manifeste, l'évaluation sommative revêt le caractère d'un bilan. Elle intervient donc après un ensemble de tâches d'apprentissage constituant un tout, correspondant, par exemple, à un chapitre de cours, à l'ensemble du cours d'un trimestre, etc. » **(2)**

Moment et rôle + qualité du feedback

L'évaluation sommative intervient à la fin de la 1^{ère} année et 2^{ème} année de médecine sous forme d'examens propédeutiques. (évaluation sommative certificative)

Le **rôle** principal de l'évaluation sommative est de fournir un bilan qui situe l'apprenant par rapport à une échelle de notes (de 1 à 6, avec 4 comme note minimale pour réussir l'examen d'anatomie) et à des objectifs établis, ainsi de prendre des décisions de type institutionnel. Il s'agit donc d'une certification. Le terme bilan me plaît beaucoup, car il s'agit bien de la somme des connaissances accumulées au cours de l'année écoulée. Toutefois, contrairement à un examen final dans le domaine de l'ingénierie par exemple où l'étudiant est interrogé sur ses connaissances accumulées au fil de son cursus dans le but de contrôler la maîtrise des situations réelles et donc extérieures à l'école (appliquer ses connaissances pour résoudre un problème pratique ou développer un nouveau produit dans le cadre de son travail de diplôme), et ce en faisant appel au raisonnement, les examens propédeutiques (1^{er} et 2^{ème}) de médecine portent sur les savoirs scolaires, ne préparent pas directement à la vie active, sont conçus comme outils indispensables (servent de « tremplin ») pour des formations ultérieures (la clinique).

Ce mode d'évaluation permet aussi de repérer les difficultés de l'apprenant (diagnostiquer) et d'améliorer mon enseignement (la façon de transmettre le message pour que l'étudiant soit réceptif à ce que je raconte). Il arrive souvent que le candidat apprenne de la matière à l'examen et, de ce fait, se rende compte des points qu'il ne maîtrise pas. Malheureusement, il est souvent trop tard à ce niveau, car il n'est parfois plus possible de refaire l'examen (après 2 tentatives soldées par deux échecs). Quant à l'enseignant, si par exemple plusieurs étudiants échouent à la même question (par exemple orienter un os), alors il pourra s'interroger sur la qualité de son enseignement. Le feedback joue aussi ici le rôle de régulateur surtout pour l'enseignant qui en sera le principal bénéficiaire, pour l'étudiant également s'il a la possibilité de refaire l'examen. Concernant le pronostic, il est en effet plus judicieux de l'établir au cours de l'évaluation formative, car il donne un meilleur feedback aussi bien pour l'étudiant que l'enseignant (voir **section 2.1 du présent chapitre**).

Dans une telle évaluation, on peut s'interroger sur la qualité des questions posées. Elles impliquent les 3 niveaux du savoir : connaissances, application et compréhension/analyse.

Qualité des questions posées

Dans cette rubrique, je traite des examens oraux d'anatomie.

Remarques : concernant les examens oraux de première année, les thèmes se rapportent aux évaluations réalisées jusqu'en automne 2003 inclus. Depuis lors, il y a une réforme du programme des études de médecine à Fribourg qui a débuté avec le curriculum de 1^{ère} année et ce en vue de la réforme de Bologne. Les thèmes des examens oraux de 2^{ème} année de médecine restent encore valables pour l'année 2004.

En premier lieu, voici une brève description de thèmes intervenant pour tous les étudiants à l'examen. Il s'agit, pour prendre le terme tiré du rapport Romainville **(8)**, de pratiques standardisées au sein de notre unité d'anatomie, c'est-à-dire que tous les thèmes interviennent pour chaque candidat et que pour chaque thème, il y a une question générale. Les questions posées oralement par l'examineur s'articulent autour de la question générale. A ce niveau on pourrait discuter sur la standardisation des questions posées par l'examineur car elles peuvent différer d'un candidat à l'autre et ce pour un même thème. Toutefois, je les considérerais comme standardisées dans la mesure où s'il y a effectivement des différences entre les questions posées

au candidat, cela pourrait être dû à l'ajustement des questions en cours de route pour l'aider dans son processus de raisonnement par exemple.

L'examen oral de 1^{ère} année porte sur 3 thèmes ou domaines : (1) embryologie humaine ; (2) anatomie systématique et topographique des membres ; (3) anatomie fonctionnelle des membres. (durée : 12 minutes) L'étudiant tire deux cartes : la première avec une question générale relative à l'embryologie humaine et la seconde avec une question générale qui attrait à l'anatomie systématique et topographique des membres ainsi qu'à l'anatomie fonctionnelle des membres. (voir annexes, pages 7 et 8)

L'examen oral de 2^{ème} année porte sur 5 thèmes ou domaines : (1) embryologie humaine ; (2) tête, cou et organes de la tête ; (3) thorax, systèmes cardio-vasculaire et respiratoire et colonne vertébrale ; (4) abdomen et bassin, systèmes digestif, urinaire et génitaux ; (5) système nerveux central. (durée : 20 minutes). L'étudiant tire 5 cartes (1 carte par thème) avec chacune une question générale relative au thème correspondant. (voir annexes, pages 7 et 8)

Voici une liste d'exemples de questions posées à l'examen oral et pratique d'anatomie :

Exemples de questions :

- a) Muscles (Myologie): origine, insertion, innervation et fonction
- b) Vaisseaux : origine (artères), confluence (veines), trajectoire (topographie)
- c) Nerfs : origine, structure (organe, muscle, articulation, ...etc.) innervée
- d) Articulations (Arthrologie) : types d'articulation (familles), mouvements possibles (degrés de liberté, axes), description détaillée (capsule, cavité, ligaments, innervation, vascularisation)
- e) Os (Ostéologie) : type (long, court, plat), orientation selon 3 plans (horizontal, frontal et sagittal, gauche ou droit ?), description détaillée, origines et insertions des muscles
- f) Organes : description et situation (topographie), vascularisation, innervation

Au cours des examens, il est important que les questions soient compréhensibles, concises et qu'ils correspondent aux objectifs fixés en début d'année. Les questions formulées correspondent aux objectifs et se rapportent donc à la matière enseignée au cours de l'année ou des deux ans. Les questions posées sont en nombre suffisant et restent fidèles à la question générale de la carte avec la possibilité toutefois d'ajuster les questions en cours de route et ce en fonction des réponses apportées. De plus, concernant l'examen pratique, les régions à disséquer correspondent à celles qui l'ont été au cours de l'année. Les questions que l'on pose aux étudiants sont là également pour aider l'étudiant dans son processus de raisonnement. Par exemple, si un étudiant décrit par cœur la fonction d'un muscle en se trompant (décrit en fait la fonction d'un autre muscle), la question posée a trait alors à l'origine et l'insertion de ce muscle et ce afin d'en déduire sa fonction. A l'examen, l'étudiant qui fait appel à son raisonnement (en appliquant ses connaissances acquises pour répondre à la question posée) aura beaucoup plus de chance de le réussir qu'un autre récitant le cours par cœur et ne sachant pas appliquer ses connaissances acquises. En se basant sur la taxonomie de Bloom, les questions demandent des connaissances acquises et de la compréhension/application de la matière enseignée. Concernant le niveau compréhension/application, on demande par exemple à un étudiant d'orienter un os (droit ou gauche ?) et, pour se faire, l'étudiant doit argumenter en se servant des connaissances acquises (dans l'exemple : selon les plans horizontal, frontal et sagittal, structures spécifiques de l'os permettant de l'orienter dans ces 3 plans). Le niveau « analyse » n'est pas pris en compte ici car ce dernier concernerait plus la clinique, où il s'agit par exemple de poser un diagnostic à partir de symptômes d'une maladie, que l'anatomie. En effet, en clinique, l'étudiant va ausculter (« analyser ») le patient et choisir ensuite le traitement adéquat en mobilisant son savoir et savoir-faire. Au niveau de l'anatomie, les types de questions posées font appel à des compétences spécifiques (savoirs et savoir-faire disciplinaires où on demande par exemple à l'étudiant de montrer un humerus et de le décrire) et démultiplicatrices (savoir-faire, techniques généraux où, à l'examen pratique, on demande à l'étudiant de dégager un muscle en laissant en place vaisseaux et nerfs). Et ces compétences peuvent être décrites par des outils tels ceux de la taxonomie de

Bloom (voir ci-dessus) ou de Gagné. Un élément de la taxonomie de Gagné me semble applicable à notre évaluation sommative : c'est l'apprentissage d'informations verbales qui est très important en anatomie. En effet, les étudiants doivent apprendre les termes anatomiques en latin selon la « nomina anatomica » et ce pour pouvoir converser à l'examen. Le latin est reconnu sur le plan international et il permet par exemple à un Chinois de pouvoir converser avec un Suisse. Le latin évite ainsi toute confusion, ambiguïté dans les termes anatomiques. On peut juger de l'importance de cette langue avec l'exemple suivant : le terme « nerf spinal » (en français) est synonyme de « nerf accessoire » (N. accessorius en latin), tandis que le terme « nervus spinalis » (en latin) ou « Spinalnerv » (en allemand) se traduit en français par « nerf rachidien » qui n'a rien à voir avec le « nerf accessoire ». Le latin apporte ainsi davantage de validité de l'évaluation sommative. Les connaissances latines peuvent parfois influencer la note de l'examen dans la mesure où on demande, au début de la première année d'études, à l'étudiant d'apprendre systématiquement les termes en latin. Ainsi il pourra utiliser cette langue pour décrire les structures anatomiques dans un vocabulaire approprié. Toutefois, beaucoup d'étudiants acquièrent le latin au cours de leur processus d'apprentissage ; ils n'ont pas besoin de prérequis, donc pas de prétest (connaître le latin n'est pas une condition à remplir avant de commencer les études de médecine). Dans le cadre de l'examen pratique en anatomie, un autre élément de la taxonomie de Gagné vient s'ajouter au précédent (voir ci-dessus) dans l'évaluation sommative : il s'agit d'habiletés motrices. En effet, l'étudiant doit avoir acquis, dans le cadre des travaux pratiques une certaine dextérité pour disséquer (aiguiser le scalpel, façon de tenir une pincette et un scalpel, ne pas sectionner des structures, ...etc.). En outre la qualité de la préparation fait partie de la note de l'examen pratique.

3. Critères de qualité d'une épreuve pédagogique

Au vu des descriptions d'une évaluation formative et sommative dans mon milieu professionnel, on retrouve dans une évaluation formative certaines des 12 finalités d'une évaluation comme par exemple la régulation aussi bien pour les étudiants que l'enseignant lui-même. De plus, lors d'une évaluation, il faut à tout prix éviter toute forme de discrimination entre étudiants, avoir conscience de l'aspect temporel, etc.

Jean-Luc Gilles (3) a établi une liste de critères d'une épreuve pédagogique « **SMART** », Système Méthodologique d'Aide à la Réalisation de Tests, qui apportent une synthèse à ma réflexion menée aux **sections 1. et 2. du présent chapitre** :

- Validité** : les scores des apprenants doivent représenter ce que l'enseignant veut mesurer, permettre des inférences (liens avec d'autres matières enseignées) solides (validité de *construct* ou théorique), et couvrir les aspects importants des apprentissages (validité de contenu).
- Fidélité** : la note (résultat d'une épreuve corrigée) classée dans une catégorie donnée doit l'être de la même façon si elle est traitée par d'autres correcteurs (concordance interjuges) ou/et à un autre moment (consistance intrajuge)
- Sensibilité** : la mesure doit être précise, refléter des phénomènes subtils
- Diagnosticité** : les résultats doivent permettre le diagnostic précis des difficultés d'apprentissage (et idéalement de leurs causes), des processus maîtrisés, de ceux qui ne le sont pas
- Equité** : tous les étudiants doivent être traités de façon juste, en principe de la même façon (standardisation)
- Praticabilité** : la réalisation des évaluations doit être faisable dans des délais raisonnables et à l'aide des ressources en personnel et en matériel disponibles et proportionnées à l'importance relative des enjeux
- Communicabilité** : les informations non confidentielles relatives au déroulement du processus doivent être communiquées et comprises par les partenaires (enseignants, étudiants) engagés dans la réalisation des épreuves

Authenticité : les questions doivent être en rapport avec des situations qui ont du sens pour les étudiants, pertinentes par rapport au monde réel et propre à favoriser les transferts dans des contextes et des pratiques qu'on peut rencontrer dans la vie

Jusqu'ici, je me suis focalisé sur le processus de préparation d'une évaluation de la part de l'enseignant principalement en regard à son scénario pédagogique (principe de cohérence, voir **section 1. du présent chapitre**), sur des moyens pédagogiques comme l'évaluation formative permettant de préparer au mieux l'étudiant à affronter l'examen (**section 2. du présent chapitre**), sur les critères de qualité d'une évaluation (**section 3. du présent chapitre**). Le processus de métacognition entrepris de la part de l'étudiant lui permet d'ajuster, de réguler son processus d'apprentissage.

Mais qu'en est-il de l'attitude, du comportement, des habiletés intellectuelles que devrait adopter un étudiant au cours d'une épreuve ? En d'autres mots, comment peut-on évaluer le savoir-être chez un étudiant ?

Au cours d'une évaluation, les étudiants doivent pouvoir entamer et développer une confiance en soi, une estime de soi dans la réponse à la question posée. Les outils TICs peuvent-ils jouer un rôle dans l'évaluation ?

La définition suivante de ce qu'est l'estime de soi permet d'introduire la notion de degré de certitude : « La plupart des professionnels définissent l'estime de soi comme l'évaluation globale de la valeur de soi en tant que personne. Il s'agit de l'évaluation qu'un individu fait de sa propre valeur, c'est-à-dire de son degré de satisfaction de lui-même » (5).

La suite de la réflexion va aborder l'introduction du degré de certitude à l'aide d'outils TICs au sein d'une évaluation.

4. Usage des Questions à Choix Multiples (QCM) dans l'évaluation et degré de certitude

Les QCM sont des questions à choix multiples pour lesquels l'étudiant doit donner une ou plusieurs réponses à une question formulée. Beaucoup de personnes reprochent aux QCM de ne mesurer que les connaissances (1^{er} niveau du savoir). En est-il vraiment ainsi ? Après avoir décrit deux exemples : un problème de physique et un examen oral d'anatomie, la question que l'on peut se poser au sujet du QCM est la suivante :

un test QCM peut-il prendre en compte les 3 niveaux du savoir issus de la taxonomie des objectifs cognitifs de Bloom : connaissances, compréhension/application et analyse ?

Voici dans un premier temps 3 exemples (17) :

QCM de connaissances (réponse correcte : 1.)

Le diamètre normal des cornes utérines d'une vache ou d'une génisse non gestante est :

1. De 2,5 à 3,5 cm
2. De 1 à 2 cm
3. De 3,5 à 5,5 cm

L'évalué doit retrouver la valeur correcte parmi les propositions. Notez que celles-ci ne se recouvrent pas, ce qui pourrait influencer le travail mental de l'évalué.

QCM de compréhension (réponse correcte : 1.)

Au cours du pro œstrus, l'utérus de la bête bovine s'œdématie, ce qui le rend à la palpation transrectale :

1. Plus ferme
2. Plus tonique
3. Plus flasque

L'évalué ne doit plus connaître mais comprendre comment l'oedème influence la palpation. Il doit néanmoins connaître les différences entre ferme, tonique et flasque comme caractéristiques d'un utérus lors d'une palpation transrectale.

QCM d'analyse (réponse correcte : 2.)

Lors de l'examen échographique d'une vache pie noire de 3 ans, vous mettez en évidence sur l'ovaire droit, une structure d'une taille supérieure à 3 cm, à paroi mince (inférieure à 5 mm) et de forme polygonale. Votre diagnostic le plus probable sera :

1. Un corps jaune cavitaire
2. Un kyste folliculaire ou lutéal
3. Un follicule
4. Une tumeur

La connaissance des 4 solutions est indispensable pour que l'étudiant puisse appliquer cette connaissance à l'énoncé de cette question. D'autres facteurs contenus dans l'énoncé vont lui permettre de donner un diagnostic.

Concernant le niveau « Analyse », il manque un paramètre de mesure permettant à l'étudiant de se questionner sur sa propre confiance apportée dans sa réponse ; d'où la problématique de **l'estime de soi, de la confiance en soi** apportée par l'étudiant dans sa réponse et des degrés de certitude choisis.

4.1 Degré de certitude et indice de réalisme

Avant de commencer la réflexion, voici une citation qui introduit la notion du degré de certitude :

«Savoir à quel point on sait» est précieux pour définir ses propres stratégies d'apprentissage, pour décider de ses actions, pour apprendre à mieux se connaître... (Jans & Leclerq, 1997, cité dans **15**)

« Avec les degrés de certitude on ne mesure plus la connaissance de faits mais la connaissance que l'étudiant a de son savoir » (**15**).

Le fait d'être sûr de ce que l'on fait est très important dans la vie quotidienne, particulièrement en médecine : par exemple connaître la dose d'un anesthésique et en être certain permet de passer à l'acte d'injecter. Avant l'injection, la personne qui doute de sa connaissance ira vérifier alors que celui qui est certain et qui se trompe (ignorance ignorée) commettra une erreur légère (en sous dosant) ou dangereuse (en sur dosant) l'anesthésique (**16**).

Le site web « **SMART** » (**14**) montre que l'on peut parfaitement concilier les nouvelles technologies de l'information et de la communication avec l'utilisation des QCM.

« **SMART** » implique la possibilité d'utiliser les degrés de certitude dans des QCM, et ce à l'aide d'un site dérivant de « **SMART** » : le site web « **TIMI** » (**18**), un système de Testing Interactif

Multimedia via Internet. Ce site web permet de réaliser un test QCM on-line avec degré de certitude.

4.1.1 Evaluation des capacités cognitives de base : exemple d'un test d'entrée à l'université

Dans un travail réalisé en triade dans le cadre du module « évaluation des apprentissages », nous avons voulu introduire la notion de degré de certitude et implicitement celle d'indice de réalisme pour un test d'aptitudes d'entrée à l'université, test similaire à une partie du *numerus clausus* des études de médecine à Fribourg (**13**). Ce type d'épreuve permet de tester les capacités cognitives de base chez un futur étudiant. Il fait appel à la capacité de compréhension, d'analyse et de synthèse. L'étudiant doit lire le texte, le comprendre, l'analyser et ce afin de répondre aux questions y relatives. Ce type de test va développer aussi chez l'étudiant un esprit de synthèse, c'est-à-dire la capacité à ressortir des éléments importants du texte et ce afin de faciliter les réponses à donner aux questions.

Dans ce travail, nous avons utilisé le logiciel « Excel » comme plate-forme pour les QCM avec degrés de certitude. Voici un extrait de ce travail qui a été adapté au présent mémoire et expliquant la façon de quantifier le degré de certitude et l'indice de réalisme :

Le texte choisi illustre un domaine des sciences naturelles et de la médecine (**6**).

Il y a au total 9 questions type QCM et 1 réponse correcte possible par question. En outre, pour chaque question, nous avons introduit la possibilité pour l'étudiant de choisir un degré de certitude (3) qui lui semble le plus approprié par rapport à sa proposition de réponse.

Pour chaque question, nous comparons ce que l'évalué annonce qu'il connaît et la mesure de cette connaissance. On peut donc comparer la prédiction de l'évalué (la certitude qu'il a choisie) et son résultat. De là calculer s'il se connaît bien lui-même et lui donner son réalisme en plus de sa cote globale.

Pour réaliser ce travail, nous nous sommes inspirés :

- a) du test d'aptitudes pour les études de médecine (**13**) en ce qui concerne le texte et les types de questions
- b) du système méthodologique d'aide à la réalisation de tests « **SMART** » (**14, 18**) ainsi que des degrés de certitude (**3**) explicités par Jean-Luc Gilles en ce qui concerne l'évaluation des capacités cognitives de base d'un étudiant (compréhension/application, analyse, habiletés métacognitives et attitudes).

Tout ce qui est décrit ci-dessous se trouve dans le document « Excel » où l'étudiant doit cocher une des réponses possibles ainsi qu'un degré de certitude qui lui convient le mieux. A la fin du test, les résultats sont affichés sur écran.

Voici une description du fonctionnement de QCM avec degrés de certitude (**3, 16**) composée de 6 situations cognitives, allant de la méconnaissance ignorée à la connaissance parfaite, en fonction des degrés de certitude :

Degré de certitude	Score		Situations cognitives	
	RC	RI	RC <u>Compétence</u> :	RI <u>Ignorance</u> :
0 : 0-25%	+13	+4	ignorée	reconnue
1 : 25-50%	+16	+3		
2 : 50-70%	+17	+2	simple	simple
3 : 70-85%	+18	0		
4 : 85-95%	+19	-6	parfaite	ignorée
5 : 95-100%	+20	-20		

Pour chaque question, un maximum de **20 points** est possible (réponse correcte avec degré de certitude 5). Comment cela fonctionne-t-il ?

Voici un extrait de 2 questions parmi les 9 questions que nous avons formulées :

Question 3 :

Laquelle de ces propositions est juste :

- a) Le bon fonctionnement du foie et du pancréas dépend de l'équilibre moléculaire de l'organisme.
- b) L'équilibre moléculaire de l'organisme dépend du bon fonctionnement du pancréas et du foie.
- c) Le bon fonctionnement de l'organisme dépend du bon fonctionnement du foie et du pancréas qui eux-mêmes dépendent de l'équilibre moléculaire de l'organisme.
- d) Le bon fonctionnement du foie et le pancréas ne dépendent que partiellement de l'équilibre moléculaire de l'organisme.
- e) L'équilibre moléculaire de l'organisme n'apparaît pas dans le texte.

Question 4 :

Laquelle ou lesquelles des propositions suivantes est-elle ou sont-elles correcte(s) en vertu du texte ?

- I. Les acides aminés servent de base à la synthèse d'autres protéines telles l'albumine, les enzymes, ...etc.
- II. Lors d'un repas riche en protéines et pauvre en glucide, il y a libération de l'insuline par les cellules du pancréas et donc stockage du glucose en excès au niveau du foie, des muscles et du tissu adipeux
- III. Lors d'un repas riche en protéines et pauvre en glucide, il y a risque d'hypoglycémie et par conséquent il y a libération du glucagon par les cellules du pancréas

- a) Seule la proposition I est correcte
- b) Seules les propositions I et II sont correctes
- c) Seules les propositions I et III sont correctes

- d) Seules les propositions II et III sont correctes
- e) Les trois propositions sont correctes

Extraits du document « Excel » :

Répondez aux questions sur cette feuille de calcul, puis cliquez sur "Résultats" pour connaître le résultat du test

Pour chaque question : choisir une réponse, un degré de certitude, entrer le chiffre "1" et appuyer sur "enter"

Question 3

Certitude	0	1	2	3	4	5
Choix réponse	0-25%	25-50%	50-70%	70-85%	85-95%	95-100%
a)						
b)			1			
c)						
d)						
e)						

Question 4

Certitude	0	1	2	3	4	5
Choix réponse	0-25%	25-50%	50-70%	70-85%	85-95%	95-100%
a)					1	
b)						
c)						
d)						
e)						

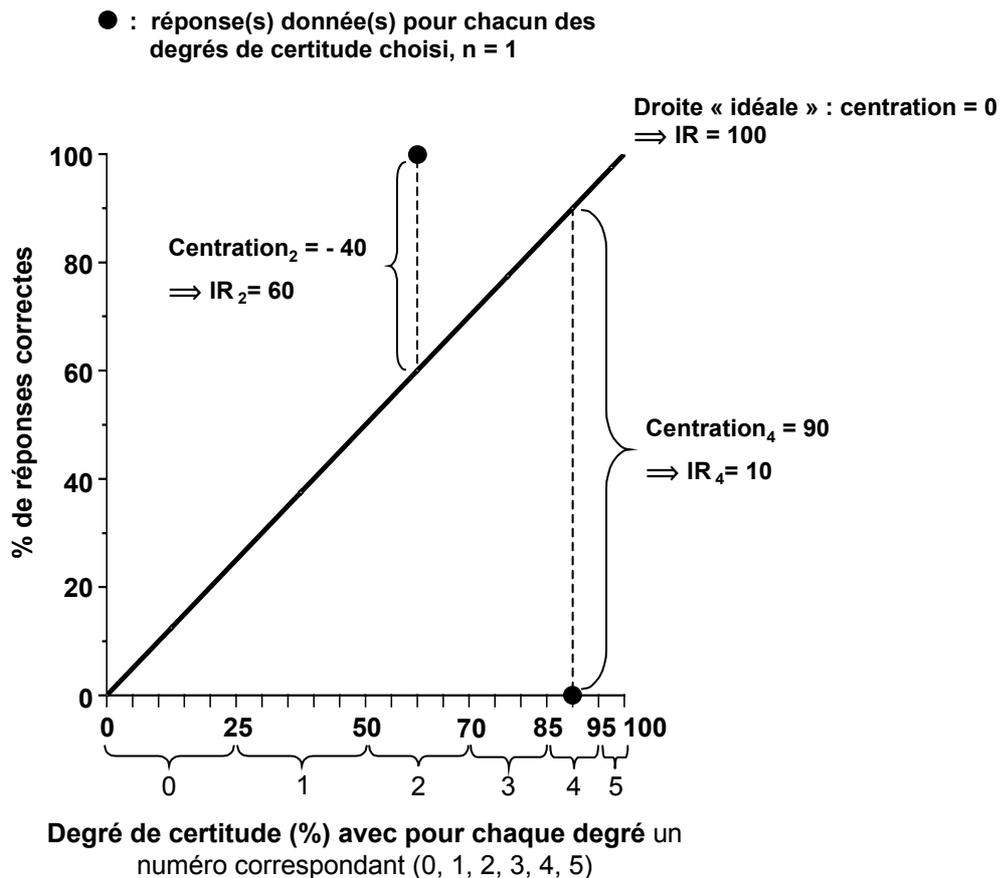
Comme le montrent les extraits du document « Excel », si l'étudiant répond correctement à la question 3 (réponse correcte : b) et qu'il coche le degré de certitude « 2 », il va alors récolter 17/20. Si le même étudiant répond faux à la question 4 (réponse correcte : c) et qu'il coche le degré de certitude « 4 », il aura un total de : 17 (1^{ère} question) – 6 (2^{ème} question) = 11/40, 40 car il a répondu à deux questions.

Pour chaque étudiant, on peut déterminer :

- **score total** : **11/40**
- **score moyen** : **5.5/20**
- **certitude moyenne** : $[60\% (1^{\text{ère}} \text{ question : } (50+70\%)/2) + 90\% (2^{\text{ème}} \text{ question : } (85+95\%)/2)]/2$ (= nombre total de réponses données) = **75%**
- **% de réponses correctes** : **50%** (correcte à la 1^{ère} question, faux à la 2^{ème})
- **centration** : $75\% (\text{certitude moyenne}) - 50\% (\text{réponses correctes}) = \mathbf{25}$
(tendance à se sur-estimer)
- **indice de réalisme (IR)** : $[(100 - 40) \times 1 + (100 - 90) \times 1] / 2 = \mathbf{35}$

$$\underbrace{\hspace{10em}}_{IR_2} \qquad \underbrace{\hspace{10em}}_{IR_4}$$

Graphique inspiré de 15



Plus la réponse correcte est proche de cette droite idéale, plus le réalisme de l'étudiant sera bon.

$IR = \text{somme } [IR_x \text{ (pour 1 degré de certitude}_x \text{ donné)} \times n \text{ réponses données pour ce degré de certitude}_x] / \text{nombre total de réponses données au test (2 dans l'exemple précité)}.$

$IR_x \text{ (pour 1 degré de certitude}_x \text{ donné)} = 100 - \text{valeur absolue de la centration}_x \text{ (pour ce degré de certitude}_x)$

$\text{Centration}_x \text{ (pour 1 degré de certitude}_x \text{ donné)} = \% \text{ certitude}_x - \% \text{ réponses correctes (obtenues sur } n \text{ réponses données pour ce degré de certitude}_x)$

$x : 0 (0-25\%), 1 (25-50\%), 2 (50-70\%), 3 (70-85\%), 4 (85-95\%), 5 (95-100\%)$

Si la centration est négative, l'étudiant a tendance à se sous-estimer
Si la centration est positive, l'étudiant a tendance à se sur-estimer

L'indice de réalisme tient compte de plusieurs paramètres : % de réponses correctes, la centration, la certitude et donc implicitement le score moyen et total. Il pourrait donc englober, quantifier les habiletés métacognitives ainsi que les attitudes des étudiants.

Jean-Luc Gilles (1996, **15**) a établi des bornes à cet **indice de réalisme** variant de **0 à 100** :

Niveau	Valeur de réalisme
Idéal	96-100
Excellent	91- 95
Très bon	86- 90
Bon	81- 85
Moyen	76- 80
Faible	71- 75
Insuffisant	66- 70
Très insuffisant	0- 65

Critères de réussite d'un tel test d'entrée à l'Université :

Dans un tel test QCM, les niveaux suivants du savoir sont évalués :

- a) Compréhension/application \Rightarrow savoir-faire convergent
- b) Analyse \Rightarrow savoir-faire divergent
- c) Habiletés intellectuelles, métacognitives \Rightarrow savoir-être

Il y a au total 9 questions, chacune d'entre elles comptant 1 point. Si on prend une échelle de notation allant de 1 à 6, on obtient le tableau suivant :

Points	0	1	2	3	4	5	6	7	8	9
Note	1	1.6	2.1	2.7	3.2	3.8	4.3	4.9	5.4	6
Niveau	Très insuffisant			Insuffisant		Faible-Suffisant	Moyen	Bon	Très bon	Idéal-Excellent

L'étudiant réussit le test QCM avec degré de certitude si :

- il répond correctement à 5 questions sur 9 (règle des 2/3 dans une échelle de notation de 1 à 6 et les notes sont arrondies au demi) : compréhension/application et analyse (niveaux savoir-faire convergent et savoir-faire divergent)

et

- il obtient un minimum de 71 dans l'indice de réalisme : habiletés métacognitives et attitudes (niveau savoir-être)

Les QCM avec **degrés de certitudes** et le calcul des **indices de réalisme** permettent, outre le fait de vérifier si une réponse est correcte ou non et ce en appliquant des éléments du texte en vue de répondre aux questions (savoir-faire convergent), de prendre en compte également d'autres paramètres tels :

- le savoir-faire divergent : l'étudiant développe des compétences en mobilisant ses ressources (compréhension, application, analyse, synthèse, savoir-être) en vue de répondre aux questions relatives au texte (résolution d'une situation à problèmes = texte). A ce stade, le degré de certitude devient intéressant dans la mesure où l'étudiant peut se questionner sur sa propre compréhension du texte : a-t-il suffisamment bien compris pour pouvoir apporter sa réponse ? Est-il alors sûr de sa réponse ? Jusqu'à quel point est-il satisfait de sa réponse ? Il peut ainsi engager tout un processus de métacognition à partir de laquelle découlent
- les habiletés métacognitives : être sûr de sa réponse (estime de soi, confiance en soi), pouvoir se questionner par rapport à ce qu'on a compris du texte, estime de soi-même, et
- les attitudes, les habiletés intellectuelles : prendre l'habitude de relire systématiquement le texte dans le cas où l'étudiant ne serait pas sûr de sa réponse, développer la curiosité, une indépendance dans la mesure où l'étudiant se retrouve seul face à un texte, la logique, du bon sens. Être capable d'identifier les éléments-clés, les lignes directrices d'un texte par exemple.

En d'autres mots de développer un savoir-être chez l'étudiant.

Avec le degré de certitude et l'indice de réalisme, on retrouve le critère de diagnosticité **(3)** dans la mesure où les résultats doivent permettre le diagnostic précis des difficultés d'apprentissage (et idéalement de leurs causes), des processus maîtrisés. On pourrait rechercher des manières d'appliquer ces deux paramètres à d'autres types de questions comme par exemple les questions à réponse ouverte.

Le fait de prendre en compte l'indice de réalisme pour la réussite d'un test est très intéressant : en effet, un étudiant peut répondre correctement à toutes les questions et échouer au test pour la simple et bonne raison qu'il aurait choisi des degrés de certitudes faibles (0-25% par exemple) pour chacune des réponses apportées. Les degrés de certitude permettent d'éviter l'effet hasardeux dans les réponses données. Ce type de test permet de sensibiliser l'étudiant à la notion de « être sûr de soi, de sa réponse, de ce que l'on fait », ce qui est particulièrement crucial dans toutes les professions, en chirurgie par exemple !

En outre, il a été démontré **(4)** que le test d'aptitudes pour les études de médecine à Fribourg **(13)** a un bon pouvoir prédictif aussi bien pour le 1^{er} propédeutique que pour le 2^{ème} propédeutique. En effet, les personnes obtenant de meilleurs résultats au test :

- ont une plus grande probabilité de réussir leurs examens plutôt que celles ayant de moins bons résultats au test et,
- font des études d'une plus grande qualité, exprimées par de meilleures notes à l'examen **(4)**.

Il existe certes une corrélation entre les résultats du test et ceux des examens de 1^{er} et 2^{ème} propédeutique, mais on ne peut pas justifier le numerus clausus comme critère d'admission aux études de médecine.

Contrairement au *numerus clausus* consistant en QCM sans degré de certitude, l'exemple illustré dans le présent mémoire propose des degrés de certitude et ne fait que renforcer tout un processus de métacognition chez l'étudiant.

5. Métacognition

Le fait de pouvoir mesurer la métacognition me paraît fort intéressant dans la mesure où elle intervient avant, pendant et après l'apprentissage. En effet, avant l'apprentissage, l'élève peut être amené à essayer d'anticiper les difficultés qu'il pourrait rencontrer, à évaluer son attitude vis-à-vis de la nouvelle matière, à se questionner si cette nouvelle matière va lui apporter quelque chose pour son parcours professionnel ou s'il se sent apte à suivre telle formation. Et l'enseignant peut vérifier l'état des connaissances prérequis pour suivre tel module au travers d'un prétest sous forme de QCM avec degrés de certitude, ou s'il préfère mesurer les capacités cognitives de base, il va concevoir un test d'entrée à l'université comme l'exemple décrit dans la **section 4.1 du présent chapitre**. En cours d'apprentissage, l'évaluation formative est une bonne opportunité pour se questionner sur l'état de ses apprentissages réalisés (voir **section 2.1 du présent chapitre**). Après l'apprentissage, l'étudiant va faire le bilan de ses connaissances acquises au travers d'une évaluation sommative (voir **section 2.2 du présent chapitre**), va pouvoir se questionner sur ce que sa formation lui aura réellement apporté. Elle permet ainsi une auto-évaluation et une auto-régulation (qu'est-ce qu'il modifiera pour s'améliorer et pourquoi ?) du processus d'apprentissage dans la mesure où l'étudiant peut en tout temps se réajuster au besoin (continuer dans la même voie, changer de discipline, etc.)

Outre le fait d'amener l'élève à s'auto-évaluer, à s'auto-réguler, à juger de par les degrés de certitude (jusqu'à quel point l'étudiant est-il satisfait de sa réponse ?), une question peut favoriser un comportement métacognitif chez l'étudiant également si par exemple **(11)** :

- elle insiste sur la justification (pourquoi a-t-il procédé ainsi ?)
- elle provoque des prises de conscience (qu'a-t-il appris de sa formation ?)
- elle amène l'élève à verbaliser sa réflexion (que pense-t-il de sa formation ?)
- elle permet l'analyse des stratégies utilisées (quelle est l'utilité des stratégies utilisées ?)

Ainsi, en guise de synthèse en regard de ce qui précède et comme le cite Jacques Tardif **(9)**, *« la métacognition est un concept plus englobant que l'auto-régulation et elle se réfère à la fois à la conscience qu'une personne a de ses stratégies et de ses caractéristiques (tant affectives que cognitives) de même qu'à leur contrôle. Quant à l'auto-régulation, il s'agit d'une conséquence observable de la métacognition chez une personne qui gère les stratégies auxquelles elle recourt dans une tâche particulière ainsi que les caractéristiques personnelles qui marquent son approche dans la tâche en question »*.

L'évaluation favorise la métacognition chez l'étudiant et l'enseignant. A son tour, la métacognition sert à développer une motivation à apprendre et à construire un concept de soi comme apprenant ainsi qu'à développer davantage d'autonomie dans la gestion des tâches et dans les apprentissages. Elle sert aussi à apprendre des stratégies de résolution de problèmes favorisant la réussite en se questionnant sur ses propres capacités à aborder telle nouvelle situation. La métacognition favoriserait le développement de cette forme de savoir-être chez l'étudiant, comme cela a déjà été relevé dans la **section 4.1 du présent chapitre** en ce qui concerne l'estime de soi, la confiance en soi.

La question qui se pose alors est la suivante : y aurait-il donc la possibilité d'évaluer cette forme de savoir-être et, implicitement le savoir-devenir ?

Pour tenter d'apporter quelques éléments de réponse, je prends l'exemple d'une thèse de doctorat. En effet, une thèse de doctorat est à mes yeux un excellent exemple car il englobe non seulement le savoir-être et le savoir-devenir, mais également tous les autres niveaux du savoir.

En d'autres mots, comment peut-on évaluer une thèse de doctorat ?

6. Evaluation d'une thèse de doctorat

Comme décrit précédemment, le savoir-être peut être évalué parallèlement aux connaissances et de façon quantitative à l'aide des degrés de certitudes et ce en utilisant des TICs. Mais il s'agit d'une certaine forme de savoir-être, l'estime de soi, la confiance en soi face à une question précise. Mais qu'en est-il des autres formes du savoir-être comme par exemple la capacité d'initiative, l'organisation du travail, le travail en équipe, l'autonomie ? Et implicitement, qu'en est-il de la créativité ?

Plus on va vers les niveaux élevés du savoir, plus se pose le problème de l'arbitraire dans l'évaluation. Afin de limiter au maximum cet effet, il devient primordial de définir une grille de critères d'évaluation (échelle ordinale).

L'exemple que j'ai choisi est une grille de critères d'évaluation pouvant s'appliquer à un travail de doctorat en Sciences naturelles. Pour se faire, je me suis inspiré de la grille présentée pour les Sciences pharmaceutiques de l'Université Catholique de Louvain (19).

Critères d'évaluation d'une Thèse de Doctorat en Sciences Naturelles

En premier lieu, je désire introduire une remarque émanant de l'Université Catholique de Louvain (19) s'avérant être très intéressante et qui, à mes yeux, doit être pris en compte dans l'évaluation d'une thèse de doctorat :

« Ces critères sont relatifs et doivent être envisagés de façon large. La nature des travaux de Thèse et les conditions de réalisation étant très différentes d'un candidat à l'autre, certains critères sont nécessairement contradictoires. Le but de l'usage de ces critères est de pouvoir objectiver le fait que le candidat soit dans l'ensemble d'une qualité suffisante (certaines appréciations défavorables pouvant être compensées par les appréciations favorables concernant des critères différents). En tout état de cause, le Jury reste souverain. (Conseil d'Ecole du 11 décembre 1997) »

Dans le présent mémoire, je propose une grille de critères d'évaluation avec une échelle de notation à 5 échelons (échelle ordinale) pour chacun des critères d'évaluation. On retrouve cette échelle de notation dans le règlement du doctorat de certaines facultés (comme celle des sciences économiques) de l'Université de Fribourg. De plus, comme on évalue ici du savoir-être et du savoir-devenir (en plus des autres niveaux du savoir), et, pour pallier les imprécisions de notation et pour essayer d'éviter les désaccords entre correcteurs, je propose pour chacun des critères d'évaluation d'ajouter à l'échelle une description aussi précise que possible des comportements à adopter et ce en vue d'augmenter la validité de ce type d'évaluation.

Echelle de notation :



0	1	2	3	4	5
Non sufficit	Rite	Cum laude	Magna cum laude	Summa cum laude	
Insuffisant	Suffisant	Bien	Très bien	Avec distinction	

Critères d'évaluation inspirés de l'Université Catholique de Louvain :

- **Originalité:** élaboration de nouveaux concepts et d'innovations (la thèse ouvre-t-elle des nouvelles voies de recherche?)
 - Non sufficit = néant (recherche déjà réalisée avant de débiter la thèse)
 - Sufficit = les méthodes utilisées sont standardisées et le sujet de recherche n'a pas encore été publié par un autre groupe.
 - Cum laude = certaines parties de la méthodologie sont innovantes et les résultats obtenus sont susceptibles d'ouvrir une perspective d'avenir dans le domaine relatif au sujet de recherche.
 - Magna cum laude = la méthodologie est entièrement innovante et apporte une perspective d'avenir. Possibilité de publier les méthodes dans un journal scientifique « peer reviewed ». Les résultats obtenus ouvrent de très bonnes perspectives de recherche dans le domaine relatif au sujet de thèse et peuvent servir aux autres groupes de recherche. Les résultats obtenus intéresseront une large part de la communauté scientifique et sont susceptibles d'être publiés dans un journal considéré parmi les meilleurs de la discipline dans laquelle la thèse a été faite.
 - Summa cum laude = sujet de recherche avant-gardiste susceptible d'être publié dans un journal prestigieux de type « Nature ». Donc le sujet va ouvrir de nouvelles perspectives de recherche comme par exemple le développement de thérapies visant à guérir telle maladie (la régénération axonale par exemple)

- **Rigueur:** maîtrise et précision expérimentale, justesse de l'analyse et adéquation des résultats obtenus
 - Non sufficit = néant, aucune analyse justifiant les résultats obtenus. Aucune maîtrise expérimentale. Résultats discutables.
 - Sufficit = analyse certes présente, mais manquant de clarté, de conviction, comme par exemple absence totale de données statistiques (dans le cas où ces dernières s'avéraient nécessaires à l'argumentation). Maîtrise expérimentale pas totalement parfaite comme par exemple argumentation peu convaincante dans la justification de telle méthode plutôt qu'une autre
 - Cum laude = bonne analyse, mais manquant parfois de précision comme par exemple dans le choix discutable de telle analyse statistique. Bonne maîtrise expérimentale, bonne argumentation dans la justification de telle méthode. La précision expérimentale peut cependant être quelque peu améliorée comme le fait de tenir compte de tous les paramètres (par exemple si un animal doit rester dans le noir absolu avant une expérience, prouver qu'il ne voit effectivement rien du

-
- Magna cum laude = tout en plaçant le négatif d'un film dans la pièce) très bonne analyse des résultats, bonne adéquation des résultats obtenus. Très bonne précision expérimentale avec toutefois quelques corrections mineures à apporter. Pertinence dans la recherche bibliographique
 - Summa cum laude = parfaite maîtrise dans l'interprétation, l'analyse des résultats et la précision expérimentale. La partie discussion est écrite de façon très réfléchie. Thèse pouvant servir d'exemple pour d'autres candidats au doctorat
- **Attitude du candidat:** capacité à travailler de façon autonome, capacité d'organisation, de planification des expériences, de collaboration éventuelle avec d'autres groupes de recherche, sens de la communication, esprit de curiosité, d'initiative
 - Non sufficit = dépendance totale vis-à-vis de son directeur de thèse. Incapable de travailler seul, de planifier des expériences, aucune envie d'aller chercher par lui-même
 - Sufficit = au début de la thèse, dépendance vis-à-vis de son directeur de thèse, puis progressivement développement d'une certaine indépendance. Capacité à aller chercher soi-même les informations nécessaires, faible capacité à planifier les expériences (a souvent besoin d'aide)
 - Cum laude = capacité à travailler de façon plus ou moins autonome (cependant, dès qu'il y a un problème, il a tendance à solliciter son directeur de thèse), à planifier des expériences. Mais un peu de peine à communiquer ses résultats avec ses pairs lors de congrès par exemple. Parfois manque de curiosité
 - Magna cum laude = capacité à travailler de façon autonome, effort dans la résolution par soi-même des problèmes rencontrés en cours d'expérience, capacité à communiquer ses résultats avec ses pairs, volonté de trouver la solution par soi-même, bonne recherche de littérature. Parfois manque d'initiative dans les prises de décision (changer de méthodologie, proposer un autre modèle animalier le cas échéant)
 - Summa cum laude = grand esprit d'initiative, d'entrepreneur. Capacité à monter un setup complet. Le candidat serait capable de diriger son propre laboratoire.
- **Publications scientifiques** (en rapport avec le sujet de la thèse)
 - Non sufficit = pas de publications, journal local
 - Sufficit = chapitre dans des "proceedings"
 - Cum laude = journal "peer reviewed" mais pas en premier auteur
 - Magna cum laude = un article dans un journal "peer reviewed" en premier auteur
 - Summa cum laude = deux articles ou plus dans un journal "peer reviewed" en premier auteur
-

- **Mises au point expérimentales:** contribution du candidat dans la mise au point de techniques nouvelles, voire de l'importation au laboratoire de techniques décrites
 - Non sufficit = néant, quelques modifications mineures
 - Sufficit = adaptation et modification moyennes
 - Cum laude = mise au point d'une technique (créativité, innovation)
 - Magna cum laude = mise au point de deux techniques
 - Summa cum laude = mise au point de plus de deux techniques

- **Activités de communication scientifique:** séminaires, congrès, symposiums
 - Non sufficit = néant
 - Sufficit = abstract à des sociétés nationales (poster),
présentation aux séminaires départementaux (lieu de travail)
 - Cum laude = communication orale lors d'une réunion scientifique nationale
 - Magna cum laude = abstract à des réunions internationales (poster)
 - Summa cum laude = communication orale lors d'une réunion internationale

A la lecture de ces critères d'évaluation d'une thèse de doctorat, on se rend compte que tous les niveaux du savoir sont impliqués et imbriqués, qui vont des connaissances acquises tout au long du processus d'apprentissage à la notion de créativité, d'innovation dont doit faire preuve un candidat au doctorat. Au fond, le doctorant ne doit-il pas être l'artisan principal d'un travail original de recherche de portée internationale ?

C. Conclusions

Si j'analyse ma réflexion sur l'évaluation des apprentissages, je suis parti d'une définition de ce qu'est une évaluation et à partir de laquelle j'ai développé ma réflexion. Tout au long de mon travail, j'ai essayé de mettre en parallèle la taxonomie de Bloom ainsi que celle de Gagné dans une moindre mesure (par exemple dans la description d'une évaluation sommative certificative) avec les niveaux du savoir décrits par Jean-Marie de Ketele.

L'enseignant doit respecter le principe de cohérence dans une évaluation, c'est-à-dire tenir compte des objectifs d'apprentissages définis, du contenu de son cours ainsi que des critères d'évaluation. Tout au long de mon mémoire, je me suis efforcé d'insister sur les critères d'évaluation et ce en me basant sur les niveaux du savoir : quelle importance accorde l'enseignant à tel niveau du savoir par rapport à un autre ? Par exemple, insiste-t-il davantage sur le raisonnement, l'analyse d'un problème posé plutôt que sur les connaissances acquises et ses applications ? Pour aider dans ma tâche d'une définition de l'évaluation, j'ai décrit un cas concret : un exemple de problème de physique de dynamique des fluides faisant intervenir les trois premiers niveaux du savoir. La réflexion menée dans la **section 1. du chapitre B.** s'est concentrée essentiellement sur cet exemple du problème de physique. Dans cette partie de mon travail, je me suis donc concentré sur le processus de préparation d'une évaluation (sommative certificative pour la majeure partie de ma réflexion) de la part de l'enseignant et ce principalement en regard de son scénario pédagogique.

En regard de ce qui précède, je me suis alors posé la question de savoir comment préparer au mieux l'étudiant à affronter l'examen. Pour tenter d'apporter un élément de réponse, je me suis alors focalisé sur le rôle d'une évaluation formative réalisée dans le cadre de mon travail à l'unité d'anatomie de l'université de Fribourg et j'ai ensuite créé un parallélisme avec une évaluation sommative certificative également réalisée sur mon lieu de travail. Tout au long de ma réflexion dans les **sections 1. et 2. du chapitre B.**, j'ai introduit le terme de « métacognition ». L'évaluation formative n'induit-elle pas un questionnement de la part de l'étudiant et de l'enseignant également ? Et cette forme de métacognition n'aide-t-elle pas à se préparer aux examens finaux de fin d'année ? Dans une évaluation, qu'elle soit formative et/ou sommative, il est important d'accorder une priorité à sa qualité.

Après avoir décrit les critères de qualité d'une évaluation en guise de synthèse des **sections 1. et 2. du chapitre B.** ainsi qu'un exemple de QCM prenant en compte les connaissances, la compréhension et l'analyse, mais où le questionnement que l'étudiant se fait sur la justesse de sa réponse fait défaut, je me suis alors posé la question de savoir comment évaluer la confiance en soi dans la réponse apportée par l'étudiant à une question posée : « être sûr de sa réponse », « estime de soi », « être certain de la justesse de sa réponse », « être réaliste » ; en d'autres mots, comment peut-on évaluer cette forme de savoir-être chez un étudiant ? C'est alors que j'ai décrit l'usage des QCM avec degrés de certitude et ce en m'appuyant sur les outils TICs ainsi que sur un travail réalisé en triade dans le cadre du module « Evaluation des apprentissages ». Ce travail concerne l'analyse d'un test d'aptitudes d'entrée à l'université, et ce au travers d'un texte à lire, à comprendre et à analyser suivi d'une liste de 9 questions de type QCM avec degrés de certitude et calcul d'indice de réalisme. Au travers de ce travail, nous avons analysé le processus de métacognition que développe un étudiant dans un tel test. La notion de degré de certitude et implicitement celle d'indice de réalisme prend toute son importance dans la vie quotidienne et particulièrement en médecine où la vie du patient est entre les mains de médecins qui doivent être sûrs de leurs actes.

Un exemple (**16**) : connaître la dose d'un anesthésique et en être certain permet de passer à l'acte d'injecter. Avant l'injection, la personne qui doute de sa connaissance ira vérifier alors que celui qui est certain et qui se trompe commettra une erreur légère (en sous dosant) ou dangereuse (en sur dosant) l'anesthésique .

La métacognition découle de l'évaluation et favoriserait également le développement d'une autre forme de savoir-être (autre que l'estime de soi) chez l'étudiant, comme l'autonomie, les stratégies à appliquer dans la résolution de problèmes comme la planification d'une expérience ; ce qui devrait amener l'étudiant à développer un esprit de créativité, un savoir-devenir. Y a-t-il donc la possibilité d'évaluer cette forme de savoir-être et implicitement le savoir-devenir ? Pour tenter d'y répondre, je me suis appuyé sur un exemple de thèse de doctorat en établissant une grille de critères d'évaluation prenant en compte tous les niveaux du savoir : des connaissances au savoir-devenir.

A la lecture des critères d'évaluation d'une thèse de doctorat, on se rend compte que tous les niveaux du savoir sont impliqués et imbriqués, qui vont des connaissances acquises tout au long du processus d'apprentissage à la notion de créativité, d'innovation dont doit faire preuve un candidat au doctorat.

A la lumière de la réflexion menée dans le présent mémoire, on se rend compte combien il devient important, lors de l'élaboration de questions d'évaluation des apprentissages, de bien respecter le principe de cohérence – tenir compte des objectifs d'apprentissages définis dans un module, du contenu du cours ainsi que des critères d'évaluation –. Ainsi l'enseignant pourra éviter dès la conception de l'examen bien des malentendus et les étudiants pourront d'autant mieux s'y préparer. Et pour aider encore davantage dans la préparation, une évaluation formative permettra à l'étudiant de se questionner par exemple sur l'état des connaissances acquises et à l'enseignant de s'interroger sur la manière dont il donne son cours (transmission de la matière). Cette métacognition devient donc très importante tout au long du processus d'apprentissages chez l'étudiant et tout au long du processus d'enseignement chez le professeur.

D. Liste des abréviations

APP : Apprentissages Par Problèmes

QAA : Questions A Appariements

QCM : Questions à Choix Multiples

QRO : Questions à Réponses Ouvertes

SMART : Système Méthodologique d'Aide à la Réalisation de Tests

TIC : Technologie de l'Information et de la Communication

TIMI : Testing Interactif Multimedia via Internet

E. Bibliographie

1. De Ketele et al., 1997 « L'évaluation et l'observation scolaires : deux démarches complémentaires »
2. De Landsheere G., 1979 « Dictionnaire de l'évaluation et de la recherche en éducation »
3. Gilles Jean-Luc, 2002, « Qualité spectrale des tests standardisés universitaires », Thèse de doctorat
4. Hänsgen K.-D. et Spicher B. « Numerus clausus : le test d'aptitudes pour les études de médecine (AMS) permet-il de trouver les personnes les plus aptes ? In Bulletin des médecins suisses, vol. 83, Nr. 47, 2002, pp. 2562-2569
5. Harter S., 1998 « Comprendre l'estime de soi de l'enfant et de l'adolescent: considérations historiques, théoriques et méthodologiques » In Bolognini, M. & Prêteur, Y. (eds). Estime de soi. Perspectives développementales, Lausanne-Paris : Delachaux et Niestlé
6. Lefèvre A., 1994 « La chimie fine du foie et du pancréas », Sciences et Vie, No. 187, pp. 96-99
7. Monnard J.-A., 1981, Cours de physique « Mécanique »
8. Romainville M., rapport Romainville, décembre 2002, « L'évaluation des acquis des étudiants dans l'enseignement universitaire »
9. Tardif J., 1999, « Le transfert des apprentissages », Montréal, Editions Logiques
10. Tourneur Y. et Vasamillet C., 1982 « L'évaluation au service de la formation », pp. 91-92
11. « La métacognition ! » In Pédago +, volume 1, numéro 7, avril 2003
12. « Les exigences de l'évaluation », réseau 12-13 (Juin-septembre 1990), SPU – FUNDP
13. « Le test d'aptitudes pour les études de médecine », Hogrefe – Verlag für Psychologie, Göttingen-Bern-Toronto-Seattle, ISBN 3 8017 0954-X

F. Webographie

14. <http://www.smart.ulg.ac.be> Système Méthodologique d'Aide à la Réalisation de Tests (SMART) (Université de Liège)
15. <http://www.fmv.ulg.ac.be/oga/agenda/evaluation/dc.html> Système d'évaluation en OGA : les degrés de certitude (Faculté de Médecine Vétérinaire, Université de Liège)
16. <http://www.fmv.ulg.ac.be/oga/agenda/evaluation/1Doc/dc.html> Système d'évaluation en propédeutique spéciale : les degrés de certitude
17. <http://www.fmv.ulg.ac.be/oga/agenda/evaluation/qcm.html> Système d'évaluation en OGA (= Obstétrique des Grands Animaux) : la QCM (Faculté de Médecine Vétérinaire, Université de Liège)
18. <http://www.smart.ulg.ac.be/smart/ee/timi/index.html> Système Méthodologique d'Aide à la Réalisation de Tests (SMART) : Testing Interactif Multimedia via Internet (TIMI) (Université de Liège)
19. http://www.md.ucl.ac.be/pharma/cycle3/doctorat/instructions_jury_anc_proc.htm Ecole de Pharmacie et Département des Sciences Pharmaceutiques, Doctorat en Sciences Pharmaceutiques (Université Catholique de Louvain) : instructions et renseignements pour les membres des Jurys

