

- **Transférer de la théorique à la pratique via un Testing Informatisé par CAs en Formation d'Enseignants (TICAFE)**
- Séverine DELCOMMINETTE & Dieudonné LECLERCQ
- Université de Liège – Centre Interfacultaire de Formation des Enseignants (CIFEN) et Laboratoire de Soutien à l'Enseignement Télématique (LabSET-IFRES). Boulevard de Colonster B9, 4000 LIEGE, Belgique
- severine.delcomminette@ulg.ac.be & d.leclercq@ulg.ac.be
- +32/4.366.20.72
- <http://www.ulg.ac.be/cifen>

A. Contexte

En Belgique francophone, les enseignants du secondaire supérieur bénéficient de la formation initiale suivante. D'abord, ils acquièrent un diplôme universitaire de maîtrise (master) nécessitant 4 à 5 années d'études (de 60 crédits chacune) selon les spécialités (mathématiques, physique, chimie, biologie, langues germaniques, langues romanes, éducation physique, sciences économiques, ...), plus d'une vingtaine constituant des « titres requis » pour enseigner dans le secondaire. Ensuite, les étudiants doivent acquérir un diplôme de 30 crédits (une demi-année), appelé « Agrégation de l'Enseignement Secondaire Supérieur (en abrégé AESS). A l'Université de Liège (ULg), cette formation est coordonnée par le Centre Interfacultaire de Formation des Enseignants (CIFEN). La loi prévoit que la formation doit, entre autres, porter sur la psycho-pédagogie et le CIFEN – ULg répond à cette exigence notamment par un cours intitulé « Psychologie éducationnelle de l'adolescent et du jeune adulte ». Ce cours est donné depuis vingt ans par Dieudonné Leclercq. L'arrivée récente d'une assistante, Séverine Delcomminette, a permis à cette équipe de s'attaquer à un problème ancien mais central de ce cours : la Triple Concordance entre Objectifs, Méthodes de formation et Evaluation.

B. La Triple Concordance

Le concept de triple concordance a été introduit par Ralph Tyler (1949). Successivement professeur à l'Ohio State University, à l'université de Chicago et finalement à Stanford University, Tyler a montré dès les années 40 que le testing pédagogique avait une autre fonction que de situer les résultats de chacun dans la distribution des résultats d'une population d'élèves (approche normative). Il montra en quoi le testing devait permettre de situer les résultats par rapport aux objectifs (approche qui sera appelée « critérielle » par Robert Glaser en 1963). En 1949, Tyler publie le livre « Basic principles of curriculum and instruction », et l'un de ces « basic principles » est la Triple Concordance (Triple Consistency).

Malgré le nombre d'années depuis lequel il est répété, ce principe est probablement le plus bafoué dans la pratique pédagogique. Ce n'est pas un hasard : même s'il est facile à comprendre, il est difficile à mettre en œuvre pour chacun d'entre nous.

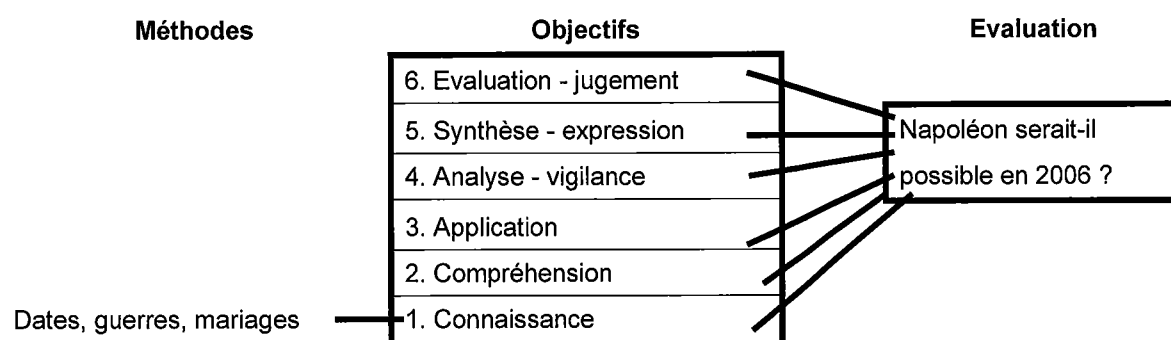
Un exemple suffira pour l'illustrer par l'absurde, c'est-à-dire par la non-concordance.

Imaginons un professeur d'histoire qui se donne comme objectifs d'amener ses étudiants à faire preuve d'esprit critique, ou plus précisément, en termes de la taxonomie de Benjamin Bloom (1956) d'analyse et de synthèse (être vigilant, détecter les éléments -même implicitement- en relation,

concevoir un raisonnement et l'exprimer avec ses propres mots). Lors de l'examen, ce même professeur pose la question à ses étudiants : « Napoléon serait-il encore possible à notre époque, l'an 2006, en Europe ? ». Répondre à la question exige en outre de la connaissance et de la compréhension (de notre époque) et de l'application (d'une époque à l'autre) pertinente (d'où jugement). Belle question ! Et belle concordance entre Objectifs et Evaluation qui met en jeu tous les niveaux de processus mentaux de la taxonomie des objectifs cognitifs de Bloom.

Le hic est que le cours, lui, dans ses contenus et ses Méthodes, n'est pas du tout en concordance avec ces deux autres composantes du cours. En effet, tout au long de l'année, le professeur n'a fait qu'énumérer dates, batailles et mariages royaux, que les étudiants seraient capables, dans une certaine mesure de réévoquer (niveau 1 de la taxonomie de Bloom : connaissance de rappel).

Nous référant à une représentation du « sucre électrique » proposée par Leclercq (2006), nous schématisons la situation comme suit :



Le manque de concordance est criant et peut être exprimé de toutes sortes de façons :

- soit le cours (méthodes et contenus) ne prépare pas les étudiants à l'examen,
- soit la question d'examen est d'un niveau d'exigence bien plus élevé que celui qui a été exercé au cours (la mémorisation – restitution),
- soit le professeur s'est fixé des objectifs ambitieux sans mettre en œuvre des méthodes permettant d'y entraîner les étudiants,
- soit ...

C. La Triple Concordance dans le cours « Psychologie Educationnelle » jusqu'en 2005

Les Objectifs de notre cours sont de rendre les étudiants capables, en situation réelle, d'interpréter et de concevoir des situations scolaires d'apprentissage et de communication interpersonnelle sous les angles ASCID, c'est-à-dire de l'Affectivité (A), des habiletés Sensori-motrices (S), des processus Cognitifs (C), de l'Image de soi ou métacognition (I), de la Décision (D), d'où l'acronyme ASCID.

Le contenu du cours est exposé dans un livre et réparti en 8 chapitres, un pour chacune des 5 rubriques ci-dessus (mais C en comporte trois) et un sur l'adolescence et le métier d'enseignant.

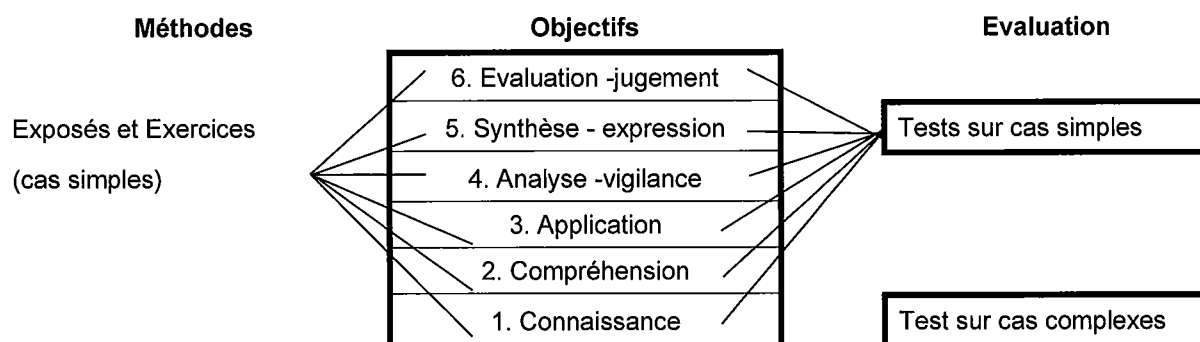
La méthode est très influencée par la contrainte horaire. Ce cours doit être donné en 15h, soit 7 à 8 réunions de 2 heures. Pour cette raison, le titulaire a, jusqu'à présent, recouru à l'exposé magistral (illustré par des diapositives), à la simulation d'examens (un test de 10 QCM en fin de séance avec correction immédiate sur questionnaire spectral¹), à la séquence LQRT (Lisez le chapitre X à domicile

¹ Cf. Leclercq, Rinaldi & Arnould (2003)

et lors de la séance, vous posez vos Questions, le professeur y Répond et un Test clôture l'échange)². Une seule séance est consacrée à l'échange de « cas » (l'adolescence et le métier de professeur), essentiellement apportés par les étudiants et explicités par tous (et davantage par le professeur).

L'évaluation se déroule sur une demi-journée. Deux épreuves écrites se succèdent : la première à livres fermés, par QCL³, la seconde à livres ouverts, par QCM avec Solutions Générales Implicites (solutions Aucune, Toutes, Manque, Absurdité)⁴ ; toutes deux intègrent les Degrés de Certitude⁵. Dans l'épreuve à livres ouverts, les questions portent sur des principes généraux, tels que les auteurs les exposent, mais nous n'exploitons jamais de cas très concrets issus de la réalité du terrain.

La Triple Concordance semble assurée, si on omet la mention « en situation réelle » (cas complexes) figurant dans les objectifs. Or les étudiants interviewés font part de leur difficulté à faire les liens entre les notions théoriques apprises, les nombreux exemples du cours et LEUR pratique (cas complexes). La Triple concordance avec les cas simples, mais la non-concordance avec les cas complexes est illustrée dans le schéma suivant :



Une solution, pour rétablir la Triple Concordance serait la règle 1 suggérée par Leclercq et Poumay (2005): « Si votre service ne peut atteindre la promesse, réduisez la promesse ».

D. Une recherche d'amélioration de la Triple Concordance

Depuis 2004-2005, nous avons cependant décidé d'essayer de mettre en œuvre la deuxième règle⁶ « améliorez le service ». Or celui-ci se décompose en « Méthodes » et « Evaluation », nécessitant toutes deux des changements, et ce dans deux directions : plus d'interactivité et plus de personnalisation. Par interactivité, nous entendons la sollicitation et la prise en compte des divers processus mentaux et non seulement leur résultat final (LA réponse à la QCM). Par personnalisation, nous entendons ici le choix d'exemples portant sur des situations ayant le plus possible de sens pour les étudiants, tant par leur complexité que par le domaine (le leur).

² Cf. Leclercq (1998,165-185)

³ Questions à Choix Large : Les (nombreux) trous de textes doivent être complétés par un mot d'une liste de plusieurs centaines (500 par exemple) de termes rangés par ordre alphabétique sur 8 colonnes d'une feuille A4 recto-verso. Cette technique est intermédiaire entre les QCM (quelques solutions proposées) et les Questions à Réponse Ouverte Courte (aucune solution proposée) : l'étudiant doit d'abord évoquer la réponse, puis chercher son code numérique (3 chiffres) dans la liste alphabétique et reporter ce code sur le formulaire destiné à la lecture optique. Voir Leclercq (2005, chap. 4, 8-10)

⁴ Cf. Leclercq (1986,127-144)

⁵ Cf. Leclercq (1982)

⁶ de Leclercq et Poumay (2005)

Un tel programme est ambitieux et nous nous sommes donnés quatre ans pour y parvenir :

Année 1 : imaginer (et tester) un système interactif d'évaluation (par ordinateur) permettant d'évaluer des processus mentaux subtils à partir de cas complexes se déroulant dans l'enseignement secondaire supérieur. (2005).

Année 2 : recueillir, pour tous les points de la matière, des exemples « situés », c'est-à-dire dans les divers domaines (une vingtaine) des étudiants. (2006).

Année 3 : concevoir (et appliquer) une méthode exploitant les exemples de l'année 2 et revoir le système d'évaluation sous cet angle également.

Année 4 : appliquer les modifications tant dans le volet « Méthodes » que dans le volet « Evaluation », pour aboutir, enfin, à la Triple Concordance sans avoir diminué les ambitions des objectifs.

Le présent article rend compte d'une partie de la recherche-action menée lors de l'année 1 : le testing sur des cas complexes rédigés par nous. Notre intervention ne s'est pas limitée à l'évaluation ; aussi avons-nous commencé les cours oraux par des questions issues du milieu professionnel, mais cet aspect ne sera pas développé ici .

Pour le testing interactif par cas, nous nous sommes principalement inspirés de 2 projets existants :

- a) la recherche internationale PePCAA (Pedagogical Psychology Computer Assisted Assessment) à laquelle nous avons participé,
- b) la synthèse théorique de l'approche par cas (Case Based Learning Methods) réalisée par Poumay en 2001.

On constatera que la structure (ci-après) de nos tests interactifs comporte deux blocs :

- le bloc « Introduction » (assez long) explicitant les enjeux, exposant le cas ainsi que le thème global des questions s'y rapportant et posant des questions métacognitives PRE.

- le bloc des Questions, comportant lui aussi une part importante de métacognition.

Rappelons ici la définition de « métacognition » proposée par Leclercq et Poumay (2005, 7, p. 3) :

« jugements, analyses et/ou régulations observables effectués par l'apprenant sur ses propres performances (processus ou produits d'apprentissages), ceci dans des situations de PRE⁷, PER⁸ ou POST⁹ performance ».

Nous avons aussi convié les étudiants à deux postdictions au terme de l'évaluation du cas :

- a) *« Connaissant le cas et le genre de questions, jugez-en la pertinence pour votre situation professionnelle. »* sur l'échelle *« 1. Pas du tout pertinent – 2. Faiblement pertinent – 3. Modérément pertinent – 4. Fortement pertinent – 5. Extrêmement pertinent. »*. Nous avons préféré la placer au terme de l'évaluation, afin que le jugement porte un regard plus global sur le cas qui a ainsi été approfondi par le testé et non seulement lu brièvement.
- b) *« Ayant analysé le cas et éventuellement consulté les aides, estimez la qualité de votre performance future (% de réponses correctes) à une série de questions sur ce sujet (situation d'apprentissage), dans ce même thème (sans aide). .../100 ».*

E. Le schéma du bloc « Introduction » sur un exemple

Ancrer une habitude : prononcer correctement

Dès le départ, nous

⁷ Avant

⁸ Pendant

⁹ Après



Règles du testing

ENJEU du cas qui suit :

Prendre de bonnes habitudes ne coûte pas plus cher que de prendre des mauvaises, dit-on. C'est à la fois vrai et faux. Vrai parce que la répétition d'actes (mentaux ou gestuels) bons ou mauvais va les « ancrer » tels qu'ils sont. C'est faux parce que pour que de mauvaises habitudes s'installent, cela ne coûte ni temps ni attention au professeur. Elles s'installent TOUTES SEULES. Pour que de bonnes s'installent, par contre, vous, en tant que professeur, devrez y consacrer du temps, de l'attention et de la compétence. Quelle est la vôtre sur ce point ?

Le cas : épisode 1

Monsieur Gates donne un cours d'informatique à des étudiants de 17 ans. Il constate qu'ils ne savent pas prononcer correctement beaucoup de mots en anglais figurant dans des textes qu'ils doivent lire. La plupart les prononcent à la « française » quand ils rencontrent ces mots pour la première fois.

Aussi en discute-t-il avec sa collègue Madame British professeur d'anglais : « Quels conseils me donneriez-vous ? »

Madame British répond : « Pensez d'abord à énoncer oralement correctement le mot AVANT de l'écrire au tableau. »

Monsieur Gates continue : « Et vous, au cours d'anglais, que pourriez-vous mettre en place comme activités linguistiques ? »

Madame British répond : « Pour ma part, je vais essayer de leur faire découvrir et prendre conscience des analogies par deux phrases mnémotechniques :

- les mots qui se terminent en « ind » (sauf wind) se prononcent souvent comme « find » (ex : « A kind of mind »).

- les mots qui se terminent en « ing » ou « ink » se prononcent comme « king » et « ring » (ex : « The king of rings »).

Les questions relatives à ce cas concerneront : les principes d'apprentissage – enseignement que les professeurs vont mettre en œuvre.

Question métacognitive : Connaissant le cas et la question, estimez la qualité de votre performance (% de réponses correctes) à une série de questions sur ce sujet (situation d'apprentissage), au départ de CE cas-CI. /100

souhaitons travailler l'attrait du cas par une illustration originale et un titre accrocheur.

Les règles de fonctionnement du test sont accessibles via le bouton « Règles du testing ».

Nous cherchons à impliquer le testé en tant que professionnel de l'éducation. Ces enjeux décrits dans la perspective des *advance organisers* (Ausubel, 1968) servent de fil rouge à la lecture du cas proposé.

Après ce premier cadrage macroscopique, nous décrivons la situation extraite d'un vécu au secondaire supérieur. Cette histoire peut être relatée en plusieurs actes.

1. Nous ciblons les facettes d'analyse à la lumière de théories.

2. Nous demandons aux étudiants de prédire leur maîtrise de la matière en regard de la situation présentée.

F. le schéma du bloc « Les Questions » sur le même exemple (une seule question ici)

Q1. Madame British a donné un conseil à Monsieur Gates (dire le mot à voix haute avant que les étudiants ne le voient).

a. Selon **votre opinion**, qu'est-ce qui justifie le conseil de Madame British à Monsieur Gates de faire entendre le mot aux étudiants avant qu'ils le voient par écrit ?

.....
.....

Quel est votre degré de certitude ? 0% 20% 40% 60% 80% 100%

Tapez le terme sur lequel vous voulez une information :

Cette question ouverte (a) d'opinion n'est jamais scorée. Les étudiants ont le droit d'être en désaccord avec les théories vues au cours pour autant qu'ils puissent développer et soutenir leur prise de position.

Le glossaire est accessible à tout moment.

Q1. Madame British a donné un conseil à Monsieur Gates (dire le mot à voix haute avant que les étudiants ne le voient).

b. Selon **Skinner**, cette stratégie consiste à utiliser un principe de progressivité. Lequel ?

1. l'intériorisation progressive
2. le modelage-façonnage
3. la spontanéisation
4. Aucun des trois ci-dessus
5. Toutes
6. Manque de données
7. Absurdité

Votre degré de certitude : 0% 20% 40% 60% 80% 100%

Justifiez votre réponse (si vous le souhaitez) :

.....

Tapez le terme sur lequel vous voulez une information :

.....

Une ou plusieurs question(s) fermée(s) font référence à des auteurs et sont ainsi « décidables »¹⁰.

Nous invitons à la vigilance cognitive en proposant, en plus des solutions propres à la question, les Solutions Générales Implicites (Aucune, Toutes, Manque de données, Absurdité).

Quelques principes :

1. Ces sous-questions (a) et (b) sont toujours suivies d'une ou deux question(s) métacognitive(s) : le degré de certitude relatif aux réponses fournies tant ouverte que fermée et une justification possible pour la QCM.
2. Idéalement (toutes les plateformes de test ne le permettent pas), chaque question est suivie d'un feedback relatif à la question (b) et rapportant l'avis d'un auteur précis sur la question.
3. Il est quelques fois nécessaire d'apporter des informations supplémentaires rapportées dans un acte complémentaire afin de relancer le débat ou de l'élargir.
4. De nouvelles questions surgissent selon la séquence précédemment décrite. Et ainsi de suite, au besoin de nouveaux actes seront joués et l'analyse continuera. Le précédent schéma de questions peut-être répété un certain nombre de fois jusqu'à l'épuisement éventuel des ressources sur le sujet.
5. Bien que le testing se fasse à **livre ouvert**, un glossaire est disponible en ligne. Il définit différents concepts-clés, auteurs, etc. (exemple ci-contre pour le terme « Pavlov »).



Pavlov (1932) : a décrit le fonctionnement du mécanisme d'association, c'est-à-dire d'extension à un objet des propriétés attractives ou répulsives, une personne, un événement ou des circonstances aux autres objets et aux lieux associés à ces objets.

G. Le dispositif d'action - recherche

Nous avons rédigé 8 cas selon deux versions : en version A, les étudiants devaient émettre d'opinion (avis) personnel (4 cas) découpé en 5 questions principales ; en version B, ils n'avaient pas cette possibilité (4 cas) et se découpaient en moyenne en 10 questions. Répondre à un cas prenait entre 20 min et 60 min (en moyenne 40 min).

Nous avons annoncé que des évaluations formatives interactives seraient accessibles entre le cours (octobre-décembre 2004) et l'examen (juin 2005). Pour les six facultés impliquées, 250 étudiants environ étaient concernés. Ils sont en grande partie des « fractionneurs », étalant l'épreuve sur deux ans. 198 étudiants se sont portés volontaires. S. Delcomminette les a accueillis par groupes de 15

¹⁰ Selon l'expression de Noizet et Caverni (1978, p. 155) cité par Leclercq (1986, p. 31)

dans une même salle au cours de 18 soirées en février et mars 2005. Ils se répartissent dans les différentes facultés de la façon suivante :

Faculté	Nombre	Proportions
Economie, Gestion et Sciences Sociales	33	16,67 %
Médecine	21	10,61 %
Philosophie et Lettres	93	46,97 %
Sciences	46	23,23 %
Sciences Appliquées	5	2,53 %
Total	198	100%

En fin de chaque cas, des questions ont été posées sur l'utilité perçue du mode de testing, de la consultation du glossaire, de la métacognition et des justifications et/ou opinions. Nous avons étudié les relations entre plusieurs de ces variables, les opinions et leurs relations avec les scores.

Ces données statistiques et « cliniques » permettent de concevoir des améliorations du dispositif sur plusieurs de ses composantes. Ces modifications sont en cours.

Pour répondre à la question de recherche « *Quelle est la réaction des étudiants à l'innovation ?, leur satisfaction (Kirkpatrick, 1959) par rapport à l'évaluation par cas est-elle positive ? Ont-ils des suggestions ?* », au terme de chaque cas la question suivante était posée : « *Avez-vous des suggestions pour améliorer ce système d'évaluation ?* ».

Afin de mesurer l'effet de satisfaction à moyen terme (non plus à chaud à la fin du test) mais quelques mois plus tard lors de l'examen en mai 2005, nous avons rédigé un questionnaire d'avis anonyme.

H. L'utilité du glossaire durant le testing : volets subjectif et objectif

a) Volet subjectif (avis)

Au sein de chaque épreuve (cas) nous avons injecté deux questions invitant les étudiants à donner leur avis sur l'utilité de cet outil de consultation en ligne : « *L'apport des aides (liens vers un glossaire) vous a-t-il été utile ? OUI-NON. Expliquez, si vous le souhaitez, votre réponse à la question précédente (utilité des aides)* ».

L'outil glossaire est jugé utile par une majorité d'étudiants (59,8%), car (selon leurs dires) il permet d'effectuer un lien direct avec le cours, de se rassurer ou de se remémorer une définition, tout en gagnant du temps. Pour des proportions moindres d'étudiants (8,4%), le glossaire semble inutile, car il est redondant avec le syllabus ou n'apporte pas les informations espérées. Certains étudiants préfèrent se référer au syllabus plutôt qu'à l'outil en ligne. Enfin les étudiants n'ayant pas utilisé le glossaire sont essentiellement ceux qui ne l'ont pas vu, ont manqué de temps, ou n'en ont pas ressenti le besoin.

b) Volet objectif : Liaisons entre probabilités subjectives et recours au glossaire

C'est avec l'indice de confiance moyen que l'on observe une liaison (0,135) positive (mais non significative à P.05) : plus le nombre de mots consultés augmente, plus la confiance (certitude moyenne accompagnant les réponses correctes) est élevée.

Par rapport aux auto-estimations de pré et postdictions, les liens observés sont faibles.

Ultérieurement, il serait intéressant de concevoir un dispositif d'observation PRE-POST consultation permettant de vérifier l'existence d'une relation causale entre doute et consultation, puis l'efficacité de la consultation (comme l'ont fait Leclercq et Boskin, 1990).

I. Le scoring tenant compte ou non des justifications

Une de nos questions de recherche était « *Quelle est la différence des scores avant / après prise en compte, par les correcteurs, des justifications d'une part et des opinions personnelles d'autre part ?* ».

En regard de la différence des moyennes (d) sur 20 points, la progression des scores suite à la lecture (par les notateurs) des justifications et « Réponses opinions » est plus importante pour les cas invitant à donner un avis personnel ou « R opinion » (d = 5,4) que pour les cas sans prise de position (d = 3,4).

Les scores des cas sans « R opinion », après lecture de justifications, sont mieux corrélés (que ceux avec « R opinion ») avec les auto-estimations et davantage encore avec les estimations « post » (0,372). Les scores des cas avec opinion ont une corrélation plus élevée avec les postdictions qu'avec les prédictions, mais les coefficients de corrélation (que ce soit pré ou postdiction) diminuent après lecture des opinions comme l'illustre le tableau ci-dessous.

		PREdiction ¹¹		POSTdiction ¹²	
		Avant lecture des justifications	Après lecture des justifications	Avant lecture des justifications	Après lecture des justifications
Avec opinion »	« R	0,222	0,01	0,342	0,213
Sans opinion »	« R	0,243	0,313	0,274	0,372

Nous faisons l'hypothèse que les notateurs ont trouvé dans les justifications et opinions des sources de valorisation (d'attribution de points) supérieures aux attentes des étudiants.

J. L'avis des étudiants sur le nouveau système d'évaluation

Les positions sont partagées, avec une dominance d'étudiants satisfaits (50,7%), mais ils le seraient davantage si certains inconvénients (qu'ils signalent) disparaissaient. La rapidité, la convivialité de l'ambiance ainsi que l'impact sur les méthodes de travail sont les aspects positifs les plus fréquemment cités. Du côté négatif : la linéarité du test (liée à l'impossible retour en arrière) et l'ergonomie du travail sur ordinateur (taille des caractères, support parallèle, ...).

Les étudiants émettent des suggestions sur le dispositif et l'organisation, sur les aspects techniques et sur le contenu. La part de critiques positives est plus importante que celle des négatives, mais nous pouvons suspecter dans ces remarques positives une part de « désirabilité sociale ».

Se former de manière continue à distance ? Les étudiants se disent plutôt *pour* (77,5%), essentiellement pour des raisons de développement personnel, professionnel et des bénéfices offerts par la mise à disposition de ressources à accès direct. Ceux qui préfèrent ne pas le faire évoquent principalement des raisons personnelles et un manque d'habitude de l'outil informatique.

¹¹ Prédiction donnée au début de chaque cas après la lecture de l'énoncé et du thème global des questions.

¹² Postdiction fournie à la fin de l'analyse complète de chaque cas : estimation du % de réponses correctes dans une situation similaire sur le même thème.

K. Effet positif sur les résultats à livre fermé

Les étudiants volontaires qui se sont présentés au testing par cas ont-ils un meilleur score, que ceux qui n'ont pas passé cette épreuve dispensatoire, à l'épreuve à livre fermé passée par tous le même jour en mai 2005 ?

Bien que la population qui a présenté l'épreuve par cas ne soit pas considérée comme aléatoire, nous observons une différence exprimée en ampleur d'effet de 0,79, ce qui, selon l'échelle reprise par Kirk (1996, p. 750), est assimilable à un large effet. Cet effet n'est pas, en raison des caractéristiques probables des personnes qui se sont présentées en février / mars, uniquement expliqué par la passation du testing. Pourquoi se soumettre au testing par cas est-il corrélé à un meilleur résultat à livre fermé ? Entre autres peut-être parce que, pour se préparer à cette épreuve dispensatoire de mars,

Mesure	Moyenne à livre fermé
$\mu_{\text{présenté}}$	17,2
$\mu_{\text{pas présenté}}$	13,6
$\sigma^2_{\text{présenté}}$	13,4
$\sigma^2_{\text{pas présenté}}$	27,5
d (AE)	0,79

les étudiants ont étudié la matière à ce moment et l'ont réétudiée en mai pour l'examen à livre fermé. Cette analyse doit être nuancée : l'enjeu diffère entre l'épreuve dispensatoire et l'examen à livre fermé. Parallèlement à une distinction introduite par Mc Donnell en matière d'évaluation externe (1994 ; cité par Monseur & Demeuse, 2005, p. 4), nous qualifierions l'épreuve dispensatoire de février / mars 2005 de « low-stakes – épreuve à implication ou à enjeu faible », car un échec est « rattrapable » en mai ou en septembre. L'examen à livre fermé en mai 2005, par contre, est un examen « high-stakes – épreuve à implication forte ».

L. Les améliorations à apporter

Les perspectives d'améliorations qui suivent sont issues tant des avis des étudiants que de notre propre réflexion au terme de cette première année de recherche.

1. Dans un premier temps (expérimental) notre dispositif n'était destiné qu'aux futurs agrégés de l'enseignement secondaire supérieur. Ultérieurement, s'il fait ses preuves, il pourrait être **ouvert à tout enseignant**, et, avec la coopération des Hautes Ecoles Pédagogiques, aux agrégés de l'enseignement secondaire inférieur et aux instituteurs.
2. Un **manuel** d'utilisation devrait accompagner le dispositif (mis en ligne et téléchargeable en version papier).
3. Les contenus des cas devraient être « situés » dans des contextes disciplinaires. La banque de cas pourrait s'étoffer en impliquant les acteurs eux-mêmes du terrain dans le récit de faits vécus (année 2 de la recherche, en cours), ce qui permet d'enrichir et diversifier le **fond**.
4. Les **feedbacks** devraient comporter des ressources adéquates au diagnostic posé : des références précises afin de combler l'erreur ou la mauvaise compréhension observée.
5. La **forme** du testing pourrait être encore être améliorée afin de rendre le travail plus ergonomique : lisibilité des caractères, précision dans la rédaction des consignes et des questions, ...
6. L'**absence d'entraînement** des élèves au dispositif d'évaluation pourrait être palliée par un test à blanc mis à disposition « *on line* » et à distance.

7. Nos **cas** sont encore **restés trop simples** : chacun de nos cas ne porte que sur un seul chapitre. Ce défaut est difficile à éviter car il faudrait alors un énoncé long permettant d'analyser plusieurs facettes de la situation. Nous verrons à l'avenir comment pallier cet inconvénient.

8. Les apports de la **métacognition** et des justifications et/ou opinions permettent d'améliorer les scores, et les propriétés éducatives du testing. Cet aspect devra être amplifié.

9. L'outil **glossaire** est avantageux, mais il pourra être perfectionné en tenant compte des remarques des étudiants : accompagner chaque définition d'un exemple, veiller à procurer une information ciblée en regard des cas, ...

10. Des **extraits vidéo** permettraient de rendre les cas encore plus proches du réel (validité écologique de l'évaluation)¹³, et d'ouvrir la porte à des analyses faisant référence à diverses théories (sur les styles cognitifs par exemple). Ces cas seraient donc plus complexes, moins linéaires et apporteraient un surplus de sens ... qui est selon nous au centre du processus d'apprentissage. Sur ce sujet, citons Dupont et Ossandon (1994, p. 118, cités par Leclercq, 1998, p. 78) : « *A tous les niveaux d'enseignement, le sens évite le superficiel, l'artificiel et le cumulatif et privilégie la transversalité, la profondeur, la tierce place ; la logique linéaire a peu de signification par rapport à la logique dialectique ; l'ordre n'est rien face à la complexité.* ».

M. Conclusion

Grâce au projet de développement, en termes de méthodes et d'évaluation, nous espérons avoir contribué à donner du sens à l'apprentissage de la Psychologie Educationnelle à l'Agrégation et à améliorer la concordance entre les objectifs du cours, les méthodes utilisées et les évaluations mises en place. Nous sommes cependant encore assez loin du compte, d'où la nécessité des trois années de recherches qui viennent.

Etre à l'écoute des avis des élèves restera important. Le dispositif ne pourra se poursuivre que si nous préservons des espaces de parole (tels des recueils d'avis), si nous aménageons des moments et phases de responsabilisation, en invitant les étudiants à créer des cas (projet de l'année 2) et si nous restons ouverts à la critique et à la négociation.

Bibliographie

- Ausubel, D. (1963). *The psychology of meaningful verbal learning: An introduction to school learning*. New York: Grune & Stratton.
- Ausubel, D. (1968). *Educational psychology: a cognitive view*. New York: Grune & Stratton.
- Bloom, B.S., Hastings, J.T., & Madaus, G. (1969). *Taxonomie des objectifs pédagogiques : tome 1 domaine cognitif*. Montréal : Education Nouvelle. (Version anglaise : 1956).
- Delcomminette, S. (2005). *Testing par cas en formation d'enseignants*. Mémoire de Licence en Sciences de l'Education, Université de Liège.
- Dupont, P., & Ossandon, M. (1994). *La pédagogie universitaire*. Paris : Presses Universitaires de France.
- Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. Some questions. *American Psychologist*, 18, 519-521.

¹³ Leclercq (2005, chap. 5, p. 36)

- Kirk, R.E. (1996). Practical significance: a concept whose time has come. *Educational and Psychological Measurement*, 56, 5, 746-759.
- Kirkpatrick, D.A. (1959). Technics for evaluating training programs, *Journal of the American Society for Training Directors (ASTD)*, 13, 3-9, 21-26, & 14, 13-18, 28-32.
- Leclercq, D. (1982). Confidence Marking, its use in testing, in Postlethwaite & Choppin (Eds), *Evaluation in education: An international review series*. (Vol.6-2, pp. 161-287). Oxford: Pergamon.
- Leclercq, D. (1986). *La conception des QCM*. Bruxelles : Labor.
- Leclercq, D., & Boskin, A. (1990). Note taking behavior studied with the help of hypermedia, in N. Estes, J. Heene & D. Leclercq (Eds.), *Proceedings of the 7th International conference on Technology and Education*, 2, 16-19.
- Leclercq, D. (Ed) (1998). *Pour une pédagogie universitaire de qualité*. Sprimont : Mardaga.
- Leclercq, D., Rinaldi, A.M., & Ernould, C. (2003). Un questionnaire spectral pour l'évaluation des connaissances chez le patient diabétique, in Gagnayre et al. (Eds). *L'évaluation de l'éducation thérapeutique du patient*. Paris : IPCEM.
- Leclercq, D. (2005). *Edumétrie et docimologie pour praticiens-chercheurs*. Liège : Editions de l'Université de Liège.
- Leclercq, D., & Poumay, M. (2005). Sept principes de la Triple Concordance (Objectifs – Méthodes – Evaluation) en Formation. Consulté sur <http://www.labset.net/media/prod/7PTCOME.pdf> le 11/04/06.
- Leclercq, D. (2006). Triple concordance O-M-E et Systèmes éducatifs. Consulté sur <http://www.labset.net/media/prod/TCOMESystEdu.pdf> le 11/04/06.
- Monseur, C., & Demeuse, M. (2005). Les évaluations externes permettent-elles une régulation efficace ? In M. Demeuse, A. Baye, M.-H. Straeten, J. Nicaise & A. Matoul. *Vers une école juste et efficace : 26 contributions sur les systèmes d'enseignement et de formation* (chap. 25). Bruxelles : De Boeck.
- Poumay, M. (2001). *Why and how using Case Based Learning Methods (CBLMs)*. LabSET report for Hewlett Packard in San-Francisco, University of Liège. Consulté sur <http://www.ulg.ac.be/labset/projets/hp/pgHOME.htm> vérifiée le 03/08/05.
- Noizet, G., & Cavezmi, J.P. (1978). *Psychologie de l'évaluation scolaire*. Paris : Presses Universitaires de France.
- Tyler, R.W. (1949). *Basic principles of curriculum and instruction*. Chicago : University of Chicago Press.