

Colloque de l'ADMEE
(Association pour le Développement des Méthodes d'Evaluation en Education)
2009 Louvain-La -Neuve

Communication : **C107**

**Tests fractionnés
de la capacité à réagir à
des cas complexes en situation scolaire**

*Leclercq, Dieudonné,
Tooth, Alexandra,
Delcomminette, Séverine*
Université de Liège

Correspondance à d.leclercq@ulg.ac.be

Téléchargeable gratuitement à partir de <http://orbi.ulg.ac.be>

RESUME :

A l'Université de Liège, dans le cours de Psychologie éducationnelle aux futurs enseignants du secondaire supérieur, les étudiants sont entraînés et évalués entre autres à « Réagir à des cas complexes ». Et ce, selon une méthode d'Analyse de Cas Fractionnée (ACF) où l'évalué doit formuler des hypothèses d'explications successives à mesure qu'il reçoit des informations plus précises, pour finir par proposer des pistes d'action. Plusieurs réponses sont considérées comme correctes mais l'étudiant doit les justifier par la théorie. Les entraînements à cette compétence se font en présentiel et en e-learning. Des questionnaires d'avis (suggestions sur la méthode en général et sur les cas en eux-mêmes) ont permis d'ajuster la méthode et de tester son acceptabilité (pour les étudiants). Par dialogue avec les étudiants et recueil de leurs avis, nous avons testé l'hypothèse selon laquelle ces méthodes de formation et d'évaluation sont perçues comme ayant beaucoup de sens aux yeux des étudiants par rapport à leur formation professionnelle. Nous avons aussi étudié (1) le lien (ampleur de l'effet) entre les entraînements en ligne (utilisés par 181 étudiants sur 299) et les résultats à l'examen final ; (2) les validités écologique (isomorphisme par rapport à la situation professionnelle), théorique (triple concordance entre objectifs, méthodes et évaluations) et diagnostique (feedbacks lors des entraînements) perçues par les étudiants.

TEXTE :

La Psychologie Educationnelle, cours commun à toutes les disciplines de l'Agrégation de l'Enseignement Secondaire Supérieur (AESS), organisée à l'Université de Liège (ULg)¹, existe depuis plus de 20 ans. L'analyse par les encadrants de son dispositif a souligné certains manques dans la Triple Concordance (TC) entre Objectifs, Méthodes et Evaluations (Leclercq ; 1995, 2008). Les étudiants étaient également demandeurs de plus d'opérationnalité des cours généraux abstraits qui sont dispensés à l'agrégation à l'Université de Liège.

En effet, notre cours restait très théorique et ses applications présentées relevaient principalement de la vie quotidienne. « Ce cours n'était pas donc assez orienté vers l'enseignement des disciplines variées qui sont rassemblées au sein de ce public. Aussi avons-nous, comme Bédard (2000, p.23), infléchi sur le fonctionnement global du cours en nous appuyant sur des « dispositifs pédagogiques, dont l'apprentissage par problèmes, l'approche par projets ou encore la méthode des cas [...] afin de pallier ces lacunes » de lien(s) avec le terrain professionnel. Nos inflexions au niveau des méthodes recourent à la **méthode des cas**, une des façons de réaliser un enseignement « situé » (Anderson et al., 1996). » (Delcomminette, 2008, p.10)

¹ Aujourd'hui ce cours est intégré dans les masters à finalité didactique et le CAPAES.

1. Une formation en termes de compétences pour plus de cohérence

« En 2001, a été signé le décret définissant la formation initiale des agrégés de l'enseignement secondaire supérieur (Moniteur Belge. 22-02-2001 modification du M.B. 31-01-02) formulant au chapitre 2 les compétences que l'enseignant devait maîtriser. Parmi les 13 compétences énoncées, nous avons tenté d'identifier les liens avec notre cours. Nous avons ensuite précisé les buts du cours davantage en termes de compétences, selon la définition de Tardif (2006, p.22) : *un savoir-agir complexe prenant appui sur la mobilisation et la combinaison efficaces d'une variété de ressources internes et externes à l'intérieur d'une famille de situations.* » (Delcomminette, 2008, p.14)

Nous espérons ainsi nous rapprocher de la visée professionnalisante : en partant du cadre décretaal (référentiel d'emploi), nous centrons nos buts sur l'intégration des savoirs (compétences spécifiques) mis à leur disposition dans l'exercice du métier, donnant ainsi un sens professionnel (lien théorie-pratique) au cours. Les objectifs de formation et les évaluations par activité ont été visualisés comme suit (figure 1). Les connaissances théoriques seront les ressources internes que l'étudiant devra combiner pour répondre aux situations d'analyse, de justification et de production.

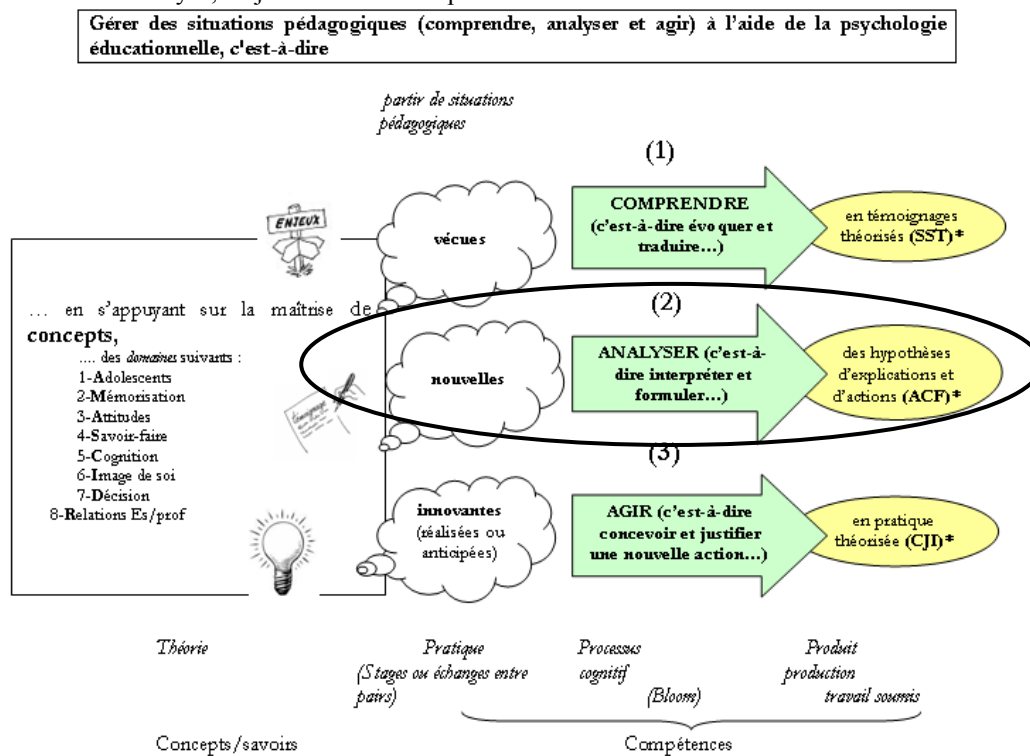


Figure 1 : schéma des compétences à destination des étudiants

1.1 Méthodes utilisées pour entraîner les étudiants

« Puisque nous visons l'acquisition de compétences, nous ne pouvons plus nous limiter à l'exposé des différentes théories. Aussi avons-nous, très modestement, diversifié nos méthodes en nous basant sur le Modèle des Événements d'Apprentissage (Leclercq et Poumay, 2005) sur nos préoccupations du lien théorie-pratique, sur les conceptions des élèves (Giordan et al, 1996), ... [...] Nous avons essayé de rendre le cours plus actif en développant des **micro-activités de sensibilisation**. Il est difficile de pratiquer de longues activités qui permettent la participation de tous (300 étudiants dans un amphithéâtre, 6 séances de 2h30). L'une de ces micro-activités, l'**exploitation de témoignages de pairs** des années précédentes, permet l'exposé et l'analyse de situations concrètes et réalistes ; elle permet aussi de faire émerger les représentations initiales et de contextualiser les contenus qui vont être ensuite exposés pour enrichir la discussion d'analyse de la situation. » (Delcomminette, 2008, pp. 15-16) Nous avons donc consacré le début de chaque séance de cours (15 min sur les 150 minutes de chaque séance), à la présentation d'un cas et à son analyse par les étudiants sur base de leurs représentations initiales puis, en 10 minutes à la fin de la séance, nous avons demandé de réinterpréter le cas (et les actions) en fonction de la théorie abordée lors de cette séance. Certains témoignages ont été remis en scène sous forme de cas photos. Voici la première diapositive (sur 5) d'un de ces cas².

² Ce cas a été réalisé avec l'aide de Perrine Fontaine, Déborah Malengrez et Carla Englebert, que nous remercions.



1.2 L'évaluation

Il importe de distinguer trois situations d'évaluation :

(1) Lorsqu'on évalue à partir d'un problème bien défini, dans lequel l'étudiant dispose d'assez de données pour répondre, l'enseignant peut avoir recours à des Questions à Choix Multiple (QCM) classiques (ou avec les solutions « Aucune » ou « Toutes ») ou à des Questions à Réponse Ouverte (QRO).

(2) Pour évaluer à partir d'un problème mal défini, pour lequel l'étudiant n'a pas assez de données à sa disposition, nous recommandons de recourir à des QCM à Solutions Générales Implicites (SGI) telles que « Aucune », « Toutes », « Manque » et « Absurdité » (Leclercq, 1986, pp. 127-144).

(3) Pour évaluer à partir d'un problème mal défini, ne présentant pas assez de données, et pour lequel on demande à l'étudiant de faire des hypothèses, par exemple en médecine, Charlin et al. (2002) ont créé le Test de Concordance de Script (TCS). C'est de cette forme de test que nous nous sommes inspirés pour créer notre nouvelle évaluation.

2. L'analyse de cas fractionné inspirée du test de concordance de script

L'Analyse de Cas Fractionnée (ACF) a des points communs avec le Test de Concordance de Script ou TCS (Charlin et al., 2002) développé en médecine.

2.1 Le test de concordance de script

« Le TCS est conçu pour évaluer les sciences cliniques. Il n'est pas adapté pour évaluer les sciences fondamentales. » (Charlin et al., 2002, p.136).

Le TCS présente les caractéristiques suivantes :

- A. Une série de questions (par exemple de 3 à 6) sont liées à un même cas (vignette clinique d'un patient). Chacune de ces questions comporte trois volets : (1) Une hypothèse de départ (ex : diabète de type 1), (2) une information supplémentaire (ex : pression artérielle) et (3) une échelle de Likert dans laquelle l'étudiant doit choisir un des 5 échelons à propos de la probabilité de l'hypothèse de départ : Quasi exclue, Nettement moins probable, Inchangée, Nettement plus probable, Quasi certaine.
- B. Un test est composé de plusieurs cas. Gagnon et al.(2008) ont estimé que la meilleure combinaison était entre 15 et 25 cas comportant chacun 2 à 4 questions, soit au total une soixantaine de réponses).
- C. Pour chaque question, une série d'experts (ex : 10) ont répondu et la dispersion de leurs réponses servira de référence : si l'étudiant choisit une réponse qui a été choisie par beaucoup d'experts, sa note sera plus élevée que si cette solution a été choisie par peu d'experts. Ce principe tient compte du fait que, contrairement aux *problèmes bien définis*, « la capacité à résoudre des problèmes mal définis caractérise davantage les médecins expérimentés et compétents » et qu'en médecine, « les médecins expérimentés divergent considérablement dans leur cheminement vers la solution de problèmes mal définis... dès que l'on essaye d'introduire dans des examens des questions correspondant à des problèmes complexes, les experts ne parviennent pas à se mettre d'accord sur la bonne réponse à exiger... » (Charlin et al., 2002, p.136).

2.2 L'analyse de cas fractionnée

En vue d'évaluer la compétence des étudiants à « Analyser (c'est-à-dire formuler des hypothèses d'explications des **situations pédagogiques nouvelles** et proposer des pistes d'action) », nous avons développé la méthode d'évaluation par **Analyse de Cas Fractionnés (ACF) à livre fermé** dont plusieurs principes sont communs avec ceux du TCS, mais qui en diffère par plusieurs aspects fondamentaux. Notamment, dans un grand amphithéâtre, lors des phases successives du test, des volets de questions sont distribués puis repris avant que soient fournies les informations suivantes. On comprendra pourquoi ci-après.

2.2.1 Schéma d'analyse fractionnée sur base d'un exemple « Question(s) de perspective(s) »




- « Phase 1 : Interpréter une situation exposée, ses causes, ses enjeux possibles (Q1)

1. ACF2-Q1 (Points : 9,0)

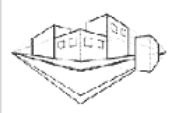
Cas 2 : Le professeur d'arts plastiques enseigne à ses élèves la perspective. Il en donne une définition.

« En dessin, en géométrie ou en architecture, on parle de perspective lorsqu'on projette un objet sur un plan telle que sa représentation sur ce plan coïncide avec la perception visuelle qu'en aurait un observateur en un point donné ».

Il donne un exemple de chacun des trois types de perspective suivants : horizontale, plongeante et plafonnante.

Horizontale (vue de face ou de côté)	Plongeante (vue de dessus)	Plafonnante (vue de dessous)
		

Il leur propose alors l'exercice suivant : « Voici une perspective horizontale. Transformez-la en perspective plongeante. »



Après 10 minutes, l'enseignant constate que ses étudiants sont incapables de réaliser l'exercice.

Q1. Pourquoi les étudiants sont-ils incapables de faire ce qui leur est demandé ?

Faites 3 hypothèses de relations causales. Précisez à chaque fois l'explication (théorique de la psychologie éducationnelle, et s'il y a lieu son auteur) de chaque relation causale.

En évaluation à visée certificative (sur papier) en présentiel, les feuilles sont reprises et les correcteurs donneront, par après, 1 point par explication plausible, 1 point pour son degré de précision et 1 point par référence théorique pertinente, soit 9 points au total pour chaque question (puisque 3 hypothèses au total par question pour cette phase). NB : Lors de cette phase, 4 cas (donc 4 questions) sont présentés.

- Phase 2 : Sur base d'une information supplémentaire, décider des interprétations les plus probables. (Q2)

2. ACF2-Q2 (Points : 2,0)

Information supplémentaire

Le professeur demande à ses étudiants : « Pourquoi ne réalisez-vous pas l'exercice demandé ? »
Arthur, le délégué de la classe répond : « Nous ne savons pas comment faire ».

Q2. L'incapacité des étudiants à résoudre l'exercice peut être expliqué par le fait que...

- 1. Le professeur n'a pas appliqué le principe de la triple concordance (Leclercq) et les étudiants ne sont pas motivés par la matière (Viau)
- 2. Le professeur n'a pas appliqué le principe de la triple concordance (Leclercq) et les étudiants n'ont pas le schéma conceptuel adéquat pour réaliser l'exercice (Rumelhart)
- 3. Le professeur a principalement fourni des connaissances déclaratives (Anderson) mais pas les connaissances procédurales (Anderson) pour réaliser (scripts = Schank) la transformation demandée
- 4. Le professeur a principalement fourni des connaissances déclaratives (Anderson) mais pas les connaissances procédurales (Anderson) et les étudiants ne sont pas motivés par la matière (Viau)
- 5. Le professeur n'a pas créé un contexte d'apprentissage motivant (Viau) et les étudiants n'ont pas le schéma conceptuel adéquat pour réaliser l'exercice. (Rumelhart)

- Réponse formulée par les experts à la Q2

3. ACF2-Q3 (Points : 4.0)

L'hypothèse des experts est :

Le professeur a principalement fourni des connaissances déclaratives (Anderson) mais pas les connaissances procédurales (Anderson) pour réaliser (scripts = Schank) la transformation demandée.

Q3. A partir de ces 2 hypothèses des experts,

I) **Que feriez-vous (donnez un exemple précis et concret) en tant que professeur d'arts plastiques dans cette classe ?**

II) **Pourquoi feriez-vous cela (Justifiez votre réponse par un élément théorique) ?**

- Phase 3 : Imaginer des solutions justifiées et argumentées sur base de la psychologie éducationnelle (Q3)

2.2.2 Mise en œuvre³

L'entraînement des étudiants a pris deux formes :

- (1) en version papier **en amphithéâtre** (donc comme à l'examen)
- (2) en e-learning (**sur WebCT**) via 3 cas. Les étudiants peuvent s'auto-corriger en comparant leurs réponses avec celles des experts.

Pour l'examen, afin de réduire la charge d'encodage due à la correction par nos soins des réponses ouvertes (hypothèses « phase 1 » et actions « phase 3 » formulées), nous avons,⁴ créé deux formuLOMs *ad hoc* (figure 2). Pour la phase 2, il s'agit d'un formuLOM QCM classique. Tous ont été lus au SMART par lecture optique de marque.

HYPOTHESES

IDEES D'ACTION

Figure 2 : FormuLOMs spécifiques à l'Analyse de Cas Fractionnée

³ Les cas ont été créés avec l'aide de Charlotte Alexandre, Gaëlle Bourtembourg, Isabelle Sente et Alexandra Tooth que nous remercions.

⁴ avec l'aide du SMART (Service Méthodologique d'Aide à la Réalisation de Tests à l'Université de Liège)

3. Questions de recherche

La mise en place de ce dispositif innovant a suscité chez les encadrants une série de questions de recherche desquelles nous rendons compte ci-dessous.

3.1. *Opinion des étudiants sur la validité écologique et la validité théorique*

Nous nous sommes interrogés sur la validité écologique parue par les étudiants (l'isomorphisme de cette évaluation avec leur situation professionnelle). et sur la validité théorique : (l'ACF est-elle en triple concordance avec les autres éléments du dispositif ?, les étudiants y ont-ils été entraînés ?).

Des réponses des participants à la question finale « Vos suggestions sur l'ACF », posée lors des entraînements en *e-learning*, pour chacun des cas, ressortent trois types de commentaires positifs. (1) l'interaction entre la théorie et la pratique : chaque cas était contextualisé et pouvait être authentique, mais surtout il était lié à une ou plusieurs théories. (2) Les théories présentées dans le cours interagissaient entre elles. (3) l'isomorphisme entre la méthodologie du cours et les théories enseignées (respect du principe de triple concordance) car cela encourage, selon eux, à transcrire en actes une théorie et, dès lors, ne plus la considérer comme un savoir mort. Par ailleurs, ils ont souligné, à juste titre, une faiblesse de cette évaluation ACF : une hypothèse basée sur une argumentation solide ne peut permettre au correcteur de la réfuter (d'où les notes élevées en moyenne, chez ces étudiants brillants, au terme de leur cursus universitaire).

3.2. *Opinion des encadrants sur la lourdeur de conception et la lourdeur de correction*

Du point de vue des encadrants, nous insistons sur deux aspects : la conception et la correction. En ce qui concerne la tâche de conception des différents cas contextualisés, elle est assez lourde dû au fait principalement que nous ne disposons pas d'un modèle théorique qui nous aiderait à générer les cas. En effet, la création et la rédaction des cas demandent beaucoup de confrontations entre experts afin de les rendre le plus précis et compréhensible possible. Pour la correction, la tâche est également assez lourde car il n'y a pas une automatisation suffisante (chaque cas appelant des hypothèses, des théories et des solutions pratiques différentes). Toutefois, l'*e-learning* (ici s'appuyant sur WebCT avec l'aide du LabSET-ULg) est plus gérable aussi bien parce que la présentation fractionnée des cas est plus aisée à mettre en place que parce que la correction est réalisée par les étudiants eux-mêmes en confrontant leur réponse à des réponses-types (les cas en ligne ont une visée formative).

3.3. *La distribution des notes*

L'auto-estimation de la qualité de leur examen par les étudiants a été confrontée aux notes attribuées par les correcteurs.

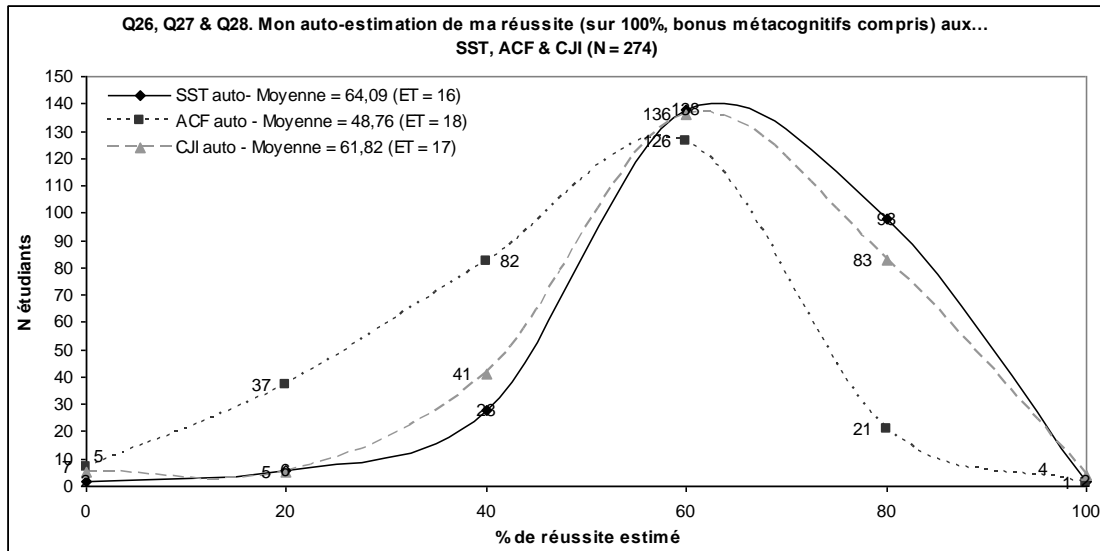
3.3.1. *Distributions des auto-estimations*

« A l'observation de la dispersion des **auto-estimations** (graphique 1), la performance auto-estimée la plus favorablement est celle des Souvenirs Sollicités par la Théorie (SST) – travail d'analyse de leur propre vécu que les étudiants rédigent à domicile et rendent un mois avant l'examen.

Ce taux de réussite nous paraît avoir au moins deux explications possibles :

- (1) les étudiants ont le choix concernant le point matière qu'ils illustrent
- (2) le syllabus en propose beaucoup d'exemples.

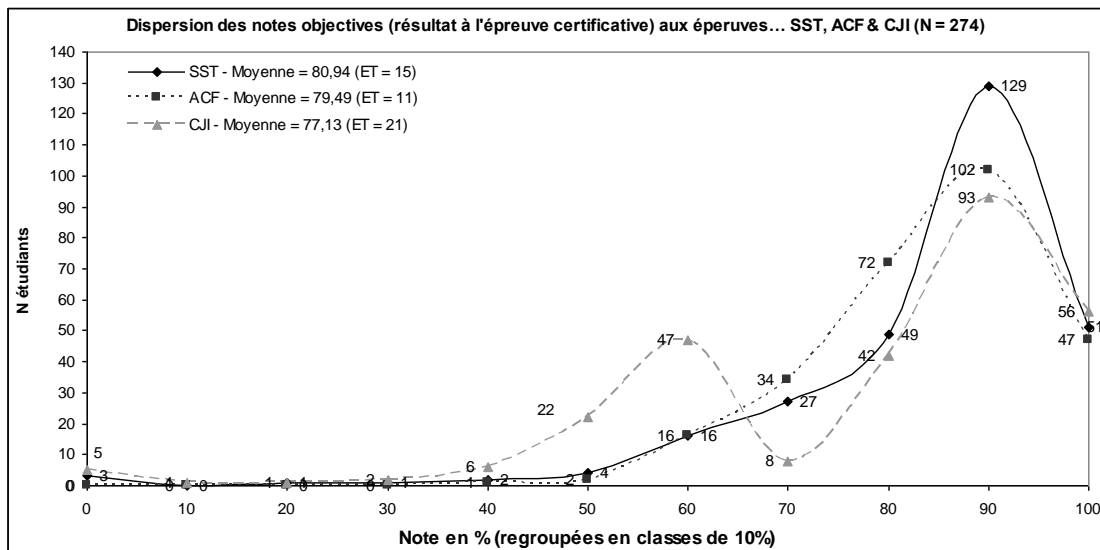
C'est l'Analyse de Cas Fractionnée (ACF) qui est la plus faiblement estimée. L'auto-estimation moyenne 48,76% est un échec si on la situe sur une échelle de décision certificative. Nous en sommes assez surpris car les étudiants pouvaient recourir à des entraînements similaires mis à leur disposition (3 en ligne et 1 en présentiel). Nous émettons une hypothèse explicative : cette auto-estimation a pris place juste après la communication des réponses correctes de l'étape 2 de l'ACF ; or beaucoup d'étudiants n'avaient pas les mêmes réponses que celles jugées correctes au départ. Le professeur en a accepté beaucoup d'autres, par la suite, sur base de justifications du raisonnement qui lui sont parvenues par mail.



Graphique 1 : Dispersion des auto-estimations des notes

3.3.2. Distributions des notes objectives

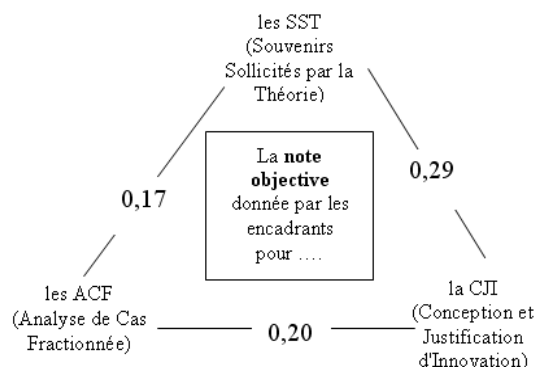
Pour faciliter la représentation et la lecture des **notes** dites « **objectives** », nous les avons regroupées par tranches de 10 %. Les distributions des notes pour l'Analyse de Cas Fractionnée (ACF) et pour la rédaction des Souvenirs Sollicités par la Théorie (SST) se présentent sous la forme de deux courbes en « j ». Les moyennes des ACF et des SST sont proches de 80% avec un écart-type de 11 et de 15. Par contre la distribution des résultats pour la Conception et la Justification d'une Innovation (CJI) se présente sous une forme bimodale, avec une moyenne un peu plus faible, mais surtout un écart-type plus important. Nous l'expliquons par le fait qu'il n'y avait qu'une seule innovation à rendre (contre 3 SST et 4 ACF) : elle répartit les étudiants en deux sous-distributions en fonction de sa réussite ou de son échec. » (Delcomminette, 2008, p.47)



Graphique 2 : Dispersion des notes objectives attribuées par les encadrants

3.3.3. Corrélations entre parties de l'épreuve de certification

Les distributions des résultats étant des courbes en « j », on comprendra que les corrélations calculées entre parties de l'épreuve sont relativement faibles. Cela conforte notre hypothèse selon laquelle ces trois types d'évaluation mesurent des compétences différentes.



Graphique 3 : Corrélations entre les notes objectives attribuées par les encadrants

3.4. Le lien (exprimé en ampleur de l'effet⁵) entre les entraînements en ligne et les résultats à l'examen final.

« Les étudiants, ayant passé le testing d'entraînement formatif à l'ACF en e-learning (181 étudiants), ont un meilleur score à l'examen écrit portant également sur l'ACF que ceux ne l'ayant pas effectué, ... (Ampleur de l'Effet = 0,68..., et les étudiants s'étant testés sur plusieurs cas d'entraînement en ligne ont obtenu un meilleur score à l'examen que ceux ne s'étant entraîné que sur un seul cas, (AE = 0,51) ce qui nous amène à l'hypothèse qu'un seul cas n'est pas suffisant pour s'entraîner. (Tooth, 2008, p.20).

4. Analyse « qualités » de la méthode ACF à l'aide de la grille ETIC PRAD

Nous intéressants aux qualités d'un dispositif d'évaluation, nous comparons ci-après la méthode TCS et l'ACF, selon la grille « ETIC PRAD » (Leclercq, 2006). La signification de chacune des lettres de cet acronyme sera rappelée ci-après.

4.1. La validité Ecologique (ou proximité à la situation de la vie réelle)

Le TCS porte sur une tranche réaliste (le raisonnement clinique et la révision d'un diagnostic en fonction d'une information) de l'acte médical (d'autres tranches, non prises en compte par le TCS sont l'anamnèse, la prescription, etc.).

L'ACF constitue plusieurs tranches (premières hypothèses et justifications théoriques, révision du diagnostic, projets d'action et justifications). La phase « question à poser (à qui et comment ?) pour obtenir de nouvelles informations manquante, hélas, pourrait être développée en e-learning.

4.2. Validité Théorique

« La théorie des scripts (Feltovich et al, 1984 ; Schmidt et al, 1990 ; Charlin et al, 2000) postule qu'en situation clinique, les médecins mobilisent des réseaux de connaissance préétablis, des scripts, qui sont utilisés pour comprendre cette situation et agir en fonction de buts diagnostiques, d'investigation ou de traitement. Les scripts de diagnostic contiennent les associations que le clinicien a établies entre l'entité pathologique et ses différents attributs cliniques (les signes et symptômes notamment). » (Charlin et al., 2002, p.136).

L'ACF part des mêmes postulats, mais évalue la mobilisation (ou disponibilité mentale) des associations, ce que ne fait pas le TCS. Sur ce point, l'ACF est au TCS ce qu'une question implicite est à une question explicite.

4.3. Validité Informative

« Le TCS donne la possibilité d'identifier des stratégies telles que la tendance à éviter les réponses extrêmes ou à éviter les réponses centrales, ou à donner la même réponse à chaque item, ou encore à répondre au hasard » (Vanbelle et al., 2007, p.73). En termes d'Edwards (1967), on pourrait parler de la détection de la capacité à traiter l'information (à modifier son degré d'incertitude), sachant que les humains sont sous-bayesiens (ont besoin de plus d'information ou d'évidences pour modifier leurs

⁵ Selon la formule de Hedges (1981, cité par Kirk, 1996, p 750) :
$$d = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}}$$

hypothèses préalables que l'information contenue dans les données) et en faisant l'hypothèse que les novices le sont plus que les experts. (A). Le TCS permet de mettre en évidence des progressions dans l'expertise clinique après formation, par stage par exemple (Gibot et Bollaert, 2008, p.13). (B)

L'ACF, dans sa forme actuelle, ne permet pas de mettre en évidence des données du type A, mais seulement de mettre en évidence la diversité (opposée à l'uniformité ou l'étroussée de la gamme) des ressources mobilisées par un étudiant, que ce soit en diagnostics, en projets d'action et en justifications des deux.

4.4. Validité Conséquentielle

Nous n'avons pas de données sur ce point, pas plus en ce qui concerne le TCS que l'ACF.

4.5 Validité Prédicative ou concourante

Plusieurs études (ex : Gibot et Bollaert, 2008, p.7) montrent que Le TCS est mieux réussi par les experts que par les internes et que ces derniers le réussissent à leur tour mieux que les étudiants. Les mêmes auteurs observent une corrélation entre les scores au TCS et aux ECN (Epreuves Classantes Nationales) de 0,55.

Actuellement, pour expliquer la faiblesse des corrélations entre les résultats aux ACF et aux SST et CJI, nous ne pouvons trancher entre deux hypothèses : (1) Les résultats des trois épreuves se distribuent selon des courbes en « j » (2) Ces 3 tests pourraient mesurer des compétences différentes.

4.6. Validité de Reproductibilité (Reliability – Fidélité).

Les TCS varient très fort en matière de fidélité ou consistance interne. Vanbelle et al. (2007, p.74) rapportent un alpha de Cronbach de 0,14 pour un TCS en médecine générale de 34 items jugés par 11 experts et répondus par 39 étudiants. Gibot et Bollaert (2008, p.10) rapportent une fidélité de 0,83 pour un TCS en réanimation de 26 questions réparties en 7 cas jugées par 9 experts et répondues par 31 étudiants.

Nous n'avons pas encore calculé la fidélité interjuges ni la consistance interne de nos ACF.

4.7. Validité d'Acceptabilité

Le TCS semble bien « reçu » tant par les constructeurs que par les étudiants. Le coût-efficacité semble acceptable.

Il n'en va pas de même pour l'ACF en version papier-crayon et à visée sanctionnante. La conception des différents cas contextualisés (pour ACF) aussi bien que leur correction sont assez lourdes.

Il ne nous est pas encore possible de dire si :

- (1) ces lourdeurs sont dues à la complexité incompressible des situations psychologiques, plus grande encore que celle des situations médicales,
- (2) les situations psychologiques constituent des problèmes encore plus mal définis que les problèmes médicaux,
- (3) c'est par manque de modèles théoriques intégrateurs qu'il suffirait de décliner selon des paramètres connus pour créer des cas,
- (4) c'est dû au caractère réellement séquentiel du principe de testing.

Il serait peut-être fécond de rapprocher l'ACF de la « méthode des cas programmés » (De Waele, 1975 ; VandenBrande, 1994 ; Leclercq et Van den Brande, 1997) où l'étudiant doit deviner, dans chaque épisode d'une biographie, lequel des actes a été posé par la personne biographiée. Nous avons assorti cette méthode de degrés de certitude et nous rejoignons ainsi certaines possibilités informatives (diagnostiques) du TCS. L'avantage de la méthode des cas programmés est que, puisqu'il s'agit d'un cas authentique, il y a UNE réponse correcte à chaque question (épisode).

4.8. La validité Déontologique

On a vu que le TCS découle de la difficulté qu'ont les experts (en médecine) à s'accorder sur un diagnostic, un traitement. Ce principe met les praticiens à l'abri de bien de recours juridiques, puisque les enseignants eux-mêmes reconnaissent que « les experts divergent ».

En acceptant de nombreuses réponses (ouvertes), la méthode ACF respecte les interprétations individuelles et les projets d'action... mais le prix à payer (en temps de correction) est lourd.

5. Conclusions

L'ACF, nouvelle méthode d'évaluation en pédagogie, comporte à la fois des qualités (triple concordance, cas contextualisé, lien théorie-pratique, ouverture aux interprétations divergentes, aux actions originales...), et des faiblesses (lourdeur de la conception et de la correction entre autres).

Or nous n'en sommes qu'à sa première mise en œuvre et nous allons tenter de palier ces faiblesses afin de permettre aux futurs enseignants de conserver leur vision de la théorie comme un savoir vivant qui peut les aider dans les différents aspects de leur métier.

Références bibliographiques

- Charlin, B., Gagnon, R., Sibert, L., & Van der Vleuten, C. (2002). Le test de concordance de script, un instrument d'évaluation du raisonnement clinique. *Pédagogie Médicale*, 3, 135-144.
- Charlin, B., Tardif, J., & Boshuizen, H.P.A. (2000). Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Acad. Med.*, 75, 182-190.
- De Waele, J.P. (1971). *La méthode des cas programmés en criminologie*. Bruxelles : Dessart.
- Delcomminette, S. (2008). *Donner plus de sens à une formation de formateurs et en mesurer l'impact : Portfolio professionnel du Master complémentaire en Pédagogie de l'Enseignement Supérieur (Form@sup)*. Université de Liège.
- Edwards, W. (1967). *Probabilistic information processing by men and man-machine systems : La simulation du comportement humain*. Paris : Dunod.
- Feltovich, P.J., & Barrows, H.S. (1984). Issues of generality in medical problem solving. In: H.G. Schmidt & M.L. De Volder (Eds). *Tutorials in problem-based learning: A new direction in teaching the health professions*. Assen (Holland): Van Gorcum, 128-142.
- Gagnon, R., Charlin, B., Lambert, C., Carrière, B., & Van der Vleuten, C. (2008). Script Concordance Testing: More Cases or More Questions? *Adv in Health Sci Educ* 2008.
- Gibot, S., & Bollaert, P.E. (2008). Le test de concordance de script comme outil d'évaluation formative en réanimation médicale. *Pédagogie Médicale*, 9, 7-18.
- Giordan, A., & De Vecchi, G. (1996). *L'enseignement scientifique : comment faire pour que « ça marche » ?* Nice, France : Z'éditions.
- Hedges, L.V. (1981), Distributional theory for Glass's estimator of effect size and related estimators, *Journal of Educational Statistics*, 6, 107-128.
- Kirk, R.E. (1996). Practical significance: a concept whose time has come. *Educational and Psychological Measurement*, vol 56, 5, 746-759.
- Leclercq, D. (1995). *Conception d'Intervention et Construction de Produits de formation*. Editions de l'Université de Liège
- Leclercq, D. (2006). L'évolution des QCM. In G. Figari & L. Mottier-Lopez. *Recherches sur l'évaluation en Education*. Paris : L'Harmattan, 139-146.
- Leclercq, D. (2008). *A la recherche de la Triple Concordance*. IFRES - ULG
- Leclercq, D., & Delcomminette, S. (2008, mai). Conception et expérimentation d'une Séquence d'Analyse de situations-problèmes et de Conception d'Actions basée sur la psychologie éducationnelle en formation d'enseignants à l'enseignement secondaire supérieur. *Actes du colloque ABC Education 2008*.
- Leclercq, D., & Poumay, M. (2005). *The 8 Learning Events Model and its principles. Release 2005-1*. Retrieved November 14, 2007, from <http://www.labset.net/media/prod/8LEM.pdf>
- Leclercq, D., & Vanden Brande, L. (1997). Une méthode pour la formation universitaire clinique en criminologie : les cas programmés, in E. Boxus, V. Jans, J.L. Gilles & D. Leclercq, Stratégies et médias pédagogiques pour l'apprentissage et l'évaluation dans l'enseignement supérieur. *Actes du 15^e colloque de l'Association Internationale de Pédagogie Universitaire (AIPU)*, Liège : STE-Affaires Académiques, 635-644.
- Meterissian, S., Zabolotny, B., Gagnon, R., & Charlin, B. (2007). Is the script concordance test a valid instrument for assessment of intraoperative decision-making skills? *The American Journal of Surgery*, 193, 248-251.
- Poumay, M. (2000). Case Based learning. Retrieved August 1st, 2008, from <http://www.labset.ulg.ac.be/projets/hp/pgTHEORIESSchemata.htm>
- Schmidt, H.G., Norman, G.R., & Boshuizen H.P.A. (1990). A cognitive perspective on medical expertise: Theory and implications. *Acad. Med.*, 65, 611-621.

- Tardif, J. (2006). *L'évaluation des compétences : documenter le parcours de développement*. Montréal : Chenelière éducation.
- Tooth, A. (2008). *Mise en ligne de l'évaluation par l'Analyse de Cas Fractionnée formative et recueil des avis des étudiants : Rapport de stage de licence en Sciences de l'Education*, Université de Liège, Service de Technologie de l'Education.
- Van Den Brande, L. (1994) *Training in getting to know a person. The method of programmed biographies*, Thèse de doctorat en psychologie, Université de Liège
- Vanbelle, S., Massart, V., Giet, D., & Albert, A. (2007). Test de concordance de script : un nouveau mode d'établissement des scores limitant l'effet du hasard. *Pédagogie médicale*, 8(2), 71-81.