

Profils d'auto-estimation et états de connaissance partielle chez les étudiants de première candidature (1995-1996) de la Faculté de Psychologie et des Sciences de l'Éducation de l'Université de Liège

Jean-Luc GILLES

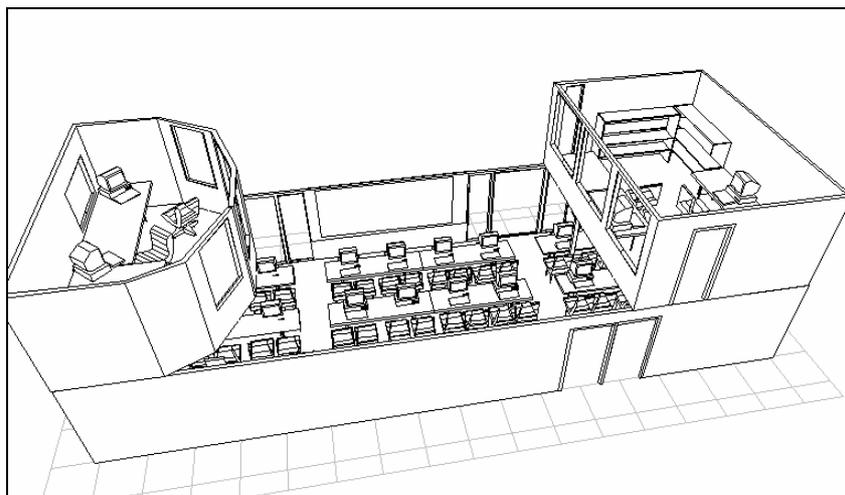
Centre d'AutoFormation et d'Evaluation Interactives Multimédias de la Faculté de Psychologie et des
Sciences de l'Éducation (C.A.F.E.I.M. - FA.P.S.E.) - Université de Liège - Belgique
Tél. : + 32-4 -366.20.78 - Fax : + 32-4 -366.29.53 - Internet : jl.gilles@ulg.ac.be
WWW : <http://www.ulg.ac.be/cafeim/>

1. Introduction

a) Le contexte de la Faculté de Psychologie et des Sciences de l'Éducation

En 1990, le Professeur LECLERCQ (Service de Technologie de l'Éducation) avec le soutien de la Faculté de Psychologie et des Sciences de l'Éducation (FA.P.S.E.) et en collaboration avec les professeurs CRAHAY (Service de Pédagogie Expérimentale), HENRY (Service de Développement et Evaluation de Programmes Pédagogiques) et REGINSTER-HANEUSE (Ecole de Santé Publique) créait le Centre d'Auto-Formation et d'Evaluation Interactives Multimédias (C.A.F.E.I.M.).

Depuis l'année académique 1994-1995, le C.A.F.E.I.M.¹, sous l'impulsion du Doyen de la FA.P.S.E., le Professeur DE KEYSER, aide les enseignants à réaliser leurs examens en mettant à leur disposition un dispositif ayant recours aux techniques d'évaluation informatisées, notamment par Questions à Choix Multiples avec Solutions Générales Implicites² (QCM-SGI) et degrés de certitude (D. LECLERCQ, 1975, 1983, 1986, 1993; E. BOXUS et al, 1991). Ces interventions ont lieu en priorité au premier cycle, là où les étudiants sont les plus nombreux : plus de 400 en première année et plus de 200 en 2ème année. Depuis deux ans c'est plus de septante évaluations qui ont ainsi été organisées par le C.A.F.E.I.M. à l'Université de Liège, dont une quarantaine dans le cadre des cours de candidature de la FA.P.S.E.



En octobre 1996, grâce au soutien des autorités académiques de notre université, le C.A.F.E.I.M. a été rééquipé et déménagé dans de nouveaux locaux. Le centre dispose désormais 36 PC-Pentium multimédia intégrés dans un réseau Novell et connectés sur l'Internet. En outre, de nouveaux équipements sont venus compléter le dispositif : lecteur optique de marques, vidéoconférence, vidéo/data-projection.

¹ Le C.A.F.E.I.M. (Président : Prof. D. LECLERCQ, directeur : J.-L. GILLES) s'est fixé pour objectifs :

- de promouvoir l'apprentissage autonome et coopératif recourant aux multimédias;
- de faciliter les évaluations de grands groupes d'étudiants en recourant aux technologies appropriées (QCM, lecture optique, etc.);
- de permettre les animations et/ou les évaluations interactives individuelles ou collectives (théâtre électronique);
- de développer une activité de recherche scientifique en matière d'application des technologies de l'information à l'apprentissage, l'enseignement et l'évaluation;
- de favoriser des activités de conception et de réalisation de produits de formation.

² Les Solutions Générales Implicites (LECLERCQ, 1986) autorisent, en plus des solutions habituellement proposées, les quatre possibilités suivantes : Rejet (aucune solution proposée n'est correcte), Toutes (toutes sont correctes), Manque (il manque des données dans l'énoncé pour que l'on puisse choisir UNE solution comme correcte), Absurdité (il y a une contrevérité dans l'énoncé à dénoncer en priorité !).

Jean-Luc GILLES, *Profils d'autoévaluation et d'états de connaissance partielle chez les étudiants de première candidature (1995-1996) de la Faculté de Psychologie et des Sciences de l'Éducation de l'Université de Liège, Colloque de l'Association Internationale de Pédagogie Universitaire (AIPU), Tunis, 4, 5 et 6 novembre 1996.*

b) Le C.A.F.E.I.M. propose deux dispositifs d'évaluation automatisée

1. La lecture optique de marques

Une façon d'évaluer les étudiants consiste à les rassembler tous dans un auditoire et à leur demander de répondre à des QCM-SGI. Les questionnaires sont alors accompagnés de feuilles de réponses spéciales où ils cochent les cases correspondant à leurs solutions choisies et à leurs degrés de certitude.

Une fois lues les feuilles destinées à la lecture optique, les réponses sont traitées par le logiciel CERT, réalisé en 1991 avec l'aide, entre autres, de la Commission des Communautés européennes dans le cadre du programme EUROTECNET (BOXUS et al., 1991).

Le logiciel CERT fournit l'analyse de chaque question de l'épreuve en trois lignes par question. Pour chaque solution proposée (de 1 à 9 pour les QCM-SGI) sont en effet fournis : (1) le pourcentage d'étudiants qui ont choisi la proposition, (2) le coefficient $r.bis^3$ et (3) la certitude moyenne. Voici comment ce présente l'écran ou le listing (les statistiques concernant la réponse correcte sont encadrées de deux barres verticales).

		Certm									
		Processus : Pas de sélection.				Matière : Pas de sélection.					
		SOL: 0	1	2	3	4	5	6	7	8	9
(1)	Q 26	0.00	26.40	3.37	7.30	33.71	5.62	11.24	1.69	1.12	9.55
(2)	Rbis	0.00	0.01	0.11	-0.27	0.25	-0.03	-0.03	-0.04	-0.12	-0.14
(3)	Cmoy	0.00	41.44	32.50	42.31	49.21	33.50	37.63	29.17	45.00	49.71
	Q 27	0.56	17.42	1.69	24.16	8.99	10.67	18.54	14.04	2.81	1.12
	Rbis	-0.18	0.07	0.00	0.09	-0.05	-0.13	0.04	0.00	-0.09	0.05
	Cmoy	12.50	37.26	37.50	38.95	36.41	31.84	44.09	44.30	36.50	12.50
	Q 28	0.00	1.12	2.81	44.94	26.97	10.67	8.99	0.00	1.69	2.81
	Rbis	0.00	-0.08	-0.10	0.25	-0.14	-0.14	0.06	0.00	-0.03	-0.04
	Cmoy	0.00	25.00	12.50	59.03	53.13	40.00	46.41	0.00	46.67	32.00
	Q 29	0.00	3.93	2.25	3.93	8.99	68.54	3.93	0.00	0.56	7.87
	Rbis	0.00	-0.24	-0.09	-0.15	-0.11	0.39	-0.13	0.00	-0.09	-0.12
	Cmoy	0.00	42.50	56.25	40.36	47.97	61.37	51.79	0.00	60.00	56.96

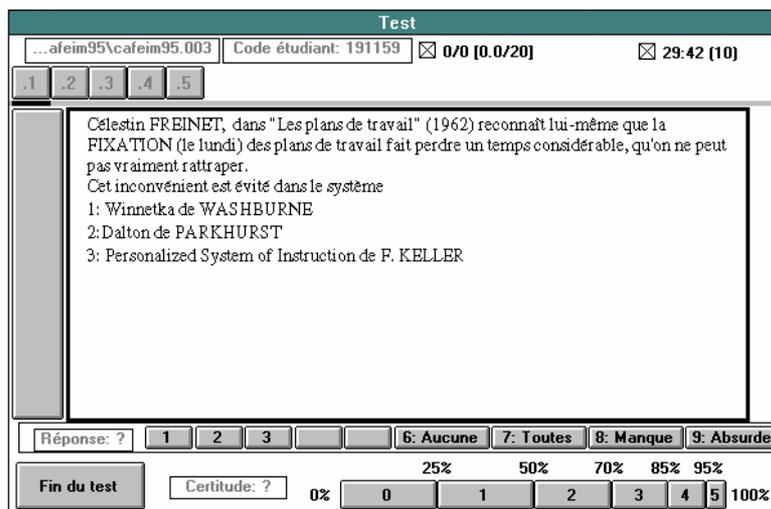
Il est également possible d'exporter les résultats des évaluations dans les logiciels EXCEL et STATISTICA en vue de traitements plus poussés.

2. Le testing interactif

Grâce à WINCHECK, un logiciel d'évaluation interactive développé par le Service de Technologie de l'Education et fonctionnant dans l'environnement Windows, nos étudiants peuvent être évalués en prenant rendez-vous pour l'examen (dans des plages horaires déterminées) en réservant un ordinateur au C.A.F.E.I.M. Ce programme permet l'utilisation du « Double Check », procédure d'évaluation interactive qui consiste à poser une question en deux volets *prim* et *bis*. L'étudiant reçoit une première (*prim*) question (QCM-SGI) où la réponse correcte attendue peut, par exemple, être « 8. Manque de données dans l'énoncé ». Après avoir répondu, l'étudiant reçoit cette réponse puis la deuxième partie de la question (*bis*), par exemple : « Quelle donnée manque ? ». Bien sûr, la question *prim* peut avoir aussi comme réponse correcte une Absurdité (comme dans l'exemple ci-après) :

³ Le coefficient $r.bis$ est une corrélation entre le choix des étudiants à chaque QCM et le score total à l'épreuve obtenue.

Jean-Luc GILLES, *Profils d'autoévaluation et d'états de connaissance partielle chez les étudiants de première candidature (1995-1996) de la Faculté de Psychologie et des Sciences de l'Education de l'Université de Liège, Colloque de l'Association Internationale de Pédagogie Universitaire (AIPU), Tunis, 4, 5 et 6 novembre 1996.*



L'évaluateur peut se constituer des banques de questions (QCM-SGI) et créer des tests en sélectionnant les questions en fonction de critères liés aux contenus, aux processus mentaux et aux objectifs. Avant le premier examen ayant recours à ce logiciel, tous les étudiants du premier cycle sont invités à s'entraîner (exercice sans sanction) à l'utilisation du programme WINCHECK

2. Les enjeux liés à l'auto-estimation de ses compétences

Les avantages qu'il y a à récolter les degrés de certitude ont été développés dans une précédente étude (LECLERCQ et GILLES, 1996). Rappelons ici succinctement une série de raisons sociales, morales, cognitives, épistémologiques et éducatives qui nous poussent à recueillir les degrés de certitude.

L'incompétence est une situation normale de la vie. Les domaines dans lesquels chacun de nous est compétent sont bien moins nombreux que ceux où il est ignorant et ... *il faut vivre avec...* Nous pensons qu'amener les étudiants à prendre conscience de leur degré d'incompétence peut les aider à augmenter leur niveau de compétence.

L'ignorance reconnue n'est pas dangereuse. L'ignorance avouée n'a pas de conséquence sociale négative, par contre, l'ignorance ignorée, elle, est dangereuse ! Mieux vaut ne pas s'improviser médecin, pharmacien, secouriste, pilote d'avion, ... et reconnaître les limites de ses compétences.

L'ignorance dissimulée est dangereuse. Habituellement, on considère qu'il est honteux de ne pas savoir. Cependant, dans maintes situations, c'est le fait que des personnes aient tenté de dissimuler leur ignorance qui amena des catastrophes, et non le fait d'avoir avouer son incompétence.

Le doute est le moteur même de la connaissance. La prise de conscience de son incompétence, de son incertitude favorise chez l'apprenant une rupture d'équilibre qui peut l'amener à rechercher l'information, à interroger son environnement afin de réduire cette incertitude.

La production de jugements est un des niveaux d'objectif les plus élevés... et, paradoxalement, des moins évalués ! La taxonomie d'objectifs pédagogiques de BLOOM (1956) propose au sommet de la hiérarchie le niveau « évaluation » qui comprend la production de jugements qualitatifs ou quantitatifs (DE LANDSHEERE, 1979). Force est de constater que ce niveau taxonomique n'est guère entraîné et évalué dans notre système éducatif.

La connaissance n'est pas affaire de tout ou rien. On considère habituellement une réponse fournie à une question de façon binaire : soit elle est correcte, soit elle est incorrecte, sans se préoccuper de nuances liées à la conviction avec laquelle l'étudiant a répondu. Les différents états de connaissance partielle qui découlent de l'association d'une réponse et d'un degré de certitude autorisent un diagnostic plus subtil et par là différents niveaux de remédiation. Cette amélioration de la sensibilité de l'outil d'évaluation contribue également à une mesure plus subtile des modifications intra-individuelles.

L'auto-évaluation s'apprend par l'expérience personnelle. Il n'y a pas, à notre connaissance, de règles et principes d'auto-estimation de ses compétences qu'on puisse enseigner. Par contre, l'apprentissage de cette habileté métacognitive se fait par l'ajustement de nos comportements d'auto-

estimation après avoir été confronté aux conséquences de nos jugements (d'où l'importance d'associer aux degrés de certitude un barème de tarifs conforme à la théorie des décisions).

Pour en finir avec la correction for guessing... Citons sur ce point DE LANDSHEERE (1979, p. 76) : "La correction pour divination fait l'objet de nombreuses critiques. Elle repose notamment sur l'hypothèse gratuite que tous les sujets ont également deviné. De plus, on n'établit pas la distinction entre l'élimination de certains choix sur base de connaissances réelles et la divination au pur hasard. Une correction beaucoup plus adéquate est assurée quand le sujet indique dans quelle mesure il est certain de sa réponse". LECLERCQ (1988, p. 306) cite cinq raisons d'abandonner la *correction for guessing* classique et de la remplacer par les degrés de certitude : (1) elle basée sur un modèle théorique faux, (2) elle est injuste, (3) elle n'est pas formative, (4) elle n'est pas informative, et enfin, (5) elle est restreinte aux QCM.

3. Les conditions de récolte des données liées à l'auto-estimation de ses compétences

a) La consigne et le barème de tarifs utilisés

Voici la consigne mise au point par LECLERCQ (1983, 1993, 1995) et utilisée à la FA.P.S.E.

Si vous considérez que votre réponse a une probabilité d'être correcte comprise entre	Ecrivez	Vous obtiendrez les points suivants en cas de	
		réponse correcte (RC)	réponse incorrecte (RI)
0 % et 25 %	0	+ 13	+ 4
25 % et 50 %	1	+ 16	+ 3
50 % et 70 %	2	+ 17	+ 2
70 % et 85 %	3	+ 18	+ 0
85 % et 95 %	4	+ 19	- 6
95 % et 100 %	5	+ 20	- 20

Les coupures sur l'axe ne sont pas équidistantes ce qui permet une expression du degré de certitude plus nuancée à l'extrémité supérieure de l'échelle puisqu'il a été démontré (LECLERCQ, 1983) que les humains en sont capables.. Ainsi, l'étudiant peut faire la distinction entre 90 % (valeur centrale de la certitude 4) et 97,5 % (valeur centrale de la certitude 5) bien que la différence soit de 7,5 % seulement. Dans le premier cas (90 %) il n'a qu'1 chance sur 10 (1/10) de se tromper tandis que dans le second (97,5 %) il n'a qu'1 chance sur 40 (1/40), soit 4 fois moins. Etablir la même différence au milieu de l'échelle, par exemple entre 40 % (1/1,7) et 47,5 % (1/1,9), n'est pas pertinent car nous ne sommes pas capable de distinguer ces deux derniers « rapports »...

b) Deux indices métacognitifs particulièrement étudiés : le réalisme et la centration

1. L'indice de réalisme

Le réalisme des étudiants peut être mesuré à l'aide d'un indice global basé sur les différences entre les taux d'exactitude et les valeurs centrales des intervalles de probabilité. Il existe différentes variantes (LICHTENSTEIN & al., 1975; LECLERCQ, 1975, 1983; LECLERCQ & al., 1993). Dans le cadre de cette étude nous avons adapté la formule de LECLERCQ afin que le minimum de réalisme soit égal à 0 et le maximum égal à 1:

$$\text{REALISME} = \left(\left(1 - \sqrt{\text{MEM}} \right) - \beta \right) \alpha$$

avec :
$$\text{MEM} = \frac{\sum_i (\text{TE}_i - \text{VC}_i)^2 \text{NU}_i}{\text{NR}}$$

TE_i = Taux d'exactitude de la certitude i, VC_i = Valeur centrale, NU_i = Nombre d'utilisations, NR = Nombre total de réponses et i = indice (ici de 0 à 5) des degrés de certitude.
β = correction pour l'erreur minimale, dans notre cas β = 0,025.

α = correction d'étendue, dans notre cas, pour étendre la plage⁴ de 0,95 à 1, $\alpha = 1,0526$

2. L'indice de centration

Le calcul de l'indice de centration est quant à lui basé sur la différence entre la certitude moyenne et le taux d'exactitude moyen. Le signe du score obtenu par l'étudiant à cet indice détermine la surestimation (+) ou la sous-estimation (-).

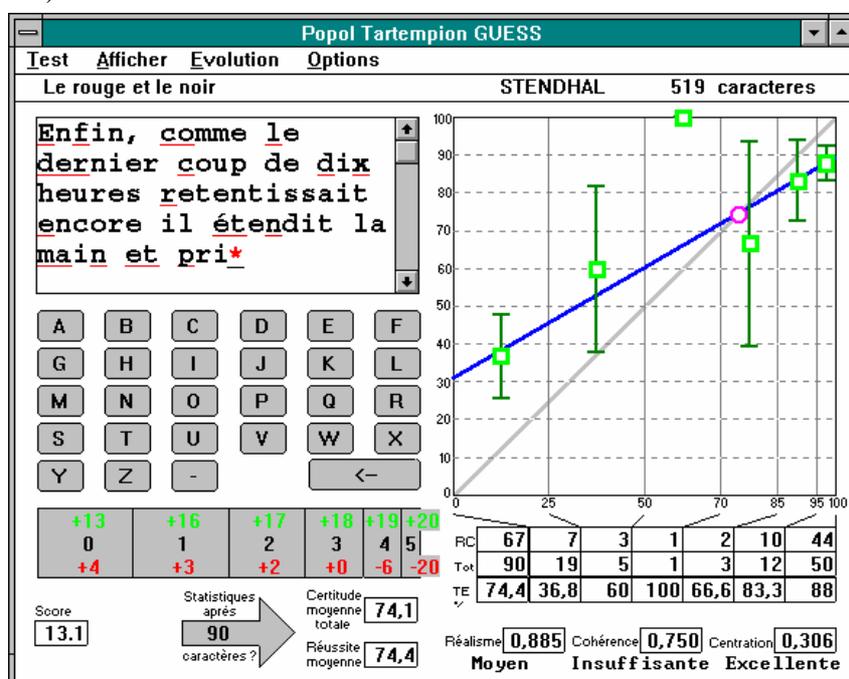
3. Le spectre des compétences

L'expression du doute à l'aide de degrés de certitude permet aussi de mettre à jour différents états de connaissance partielle. LECLERCQ & PLUNUS (1996) ont présenté dans un spectre des compétences les différents états répartis en 6 catégories : compétence parfaite (RC⁵ et DC⁶ élevé : 4 ou 5) - compétence insuffisante (RC et DC moyen : 2 ou 3) - compétence ignorée (RC et DC faible : 0 ou 1) - ignorance reconnue (RI⁷ et DC faible) - ignorance insuffisante (RI et DC moyen) - ignorance ignorée (RI et DC élevé).

d) L'entraînement des sujets à la procédure d'évaluation

1. Le logiciel GUESS

Avant d'employer les degrés de certitude lors d'examen, les étudiants se familiarisent avec leur utilisation à l'aide du logiciel GUESS. Le programme GUESS⁸ a été décrit en détail par LECLERCQ & GILLES (1994). Chaque étudiant s'entraîne individuellement à l'utilisation des degrés de certitude à l'aide d'un jeu où il doit deviner les lettres successives d'un texte (jeu inspiré des travaux de SHANNON, 1951 & d'ATTNEAVE, 1959).



Le joueur effectue une prédiction en tapant une lettre qu'il accompagne de la probabilité subjective de réussite exprimée à l'aide d'un degré de certitude. Il est ensuite informé de la réponse correcte qui s'affiche dans la zone réservée au texte. Lettre par lettre, le texte d'au moins cent lettres s'affiche ainsi à l'écran. Evidemment, le début des mots est plus difficile à deviner que leur fin (c'est ce que SHANNON voulait démontrer). Après un nombre donné de réponses un graphique de réalisme se construit dans le coin supérieur droit de l'écran, l'étudiant peut y observer la justesse des probabilités subjectives de réussites qu'il a attribuées à ses prédictions (en abscisse) en les confrontant aux taux de réussite (en ordonnée) pour chaque degré de certitude. Les étudiants du premier cycle sont invités dès le 1er semestre à s'entraîner avec GUESS

⁴ Dans la procédure d'utilisation des degrés de certitude définie par LECLERCQ, ERR MAX = 0,975 pour un sujet qui se trompe systématiquement avec le degré de certitude 5, et, ERR MIN = 0,025 pour un sujet qui répond correctement à toutes les questions avec un degré de certitude 5, ceci détermine une plage de 0,95 (0,975 - 0,025).

⁵ RC = réponse correcte

⁶ DC = degré de certitude

⁷ RI = réponse incorrecte

⁸ Le logiciel GUESS a été programmé dans le langage OpenScript de TOOLBOOK par M. HURARD, ingénieur informaticien.

Jean-Luc GILLES, *Profils d'autoévaluation et d'états de connaissance partielle chez les étudiants de première candidature (1995-1996) de la Faculté de Psychologie et des Sciences de l'Éducation de l'Université de Liège, Colloque de l'Association Internationale de Pédagogie Universitaire (AIPU), Tunis, 4, 5 et 6 novembre 1996.*

(c'est une activité obligatoire dans le cadre du partim I *Approche technologique de l'éducation et de la formation* du cours *Introduction aux sciences de l'éducation*).

4. Une recherche empirique exploratoire

Dans le cadre de cette recherche nous ne tenterons pas d'interpréter nos observations en vue d'y rechercher des causes à effets de type « A influence B ». L'objectif de cette étude n'est pas de manipuler une série de variables dans le contexte d'un plan expérimental⁹. Par contre, nous chercherons à mettre en évidence des variables qui pourraient bien être en relation avec les phénomènes d'auto-estimation chez les étudiants (par exemple la performance académique, la difficulté de l'épreuve, ...). Nous essayerons aussi de décrire des tendances générales en espérant contribuer à une compréhension plus fine de la problématique de l'auto-estimation de ses compétences. Ce type d'étude exploratoire devrait contribuer à lancer de nouvelles pistes de recherches probablement plus expérimentales. Dans ce domaine comme dans d'autres, recherches empiriques et expérimentales se nourrissent mutuellement.

a) Variables de population prises en compte

Nous avons observé les performances aux indices de réalisme et de centration dans la population des étudiants de 1ère candidature qui ont participé aux examens de huit cours figurant au programme de la première année en psychologie et sciences de l'éducation. Cette population a été catégorisée en plusieurs sous-groupes définis en fonction de la sanction académique : échec ou réussite, avec, dans ce dernier cas, le grade obtenu en première ou en seconde session. Ce qui donne :

- Le groupe « R1SESS » (N=78), étudiants qui ont réussi en première session. Ce groupe a été subdivisé en 3 sous-groupes : grande distinction « R1SESS-GD », distinction « R1SESS-D » et satisfaction « R1SESS-S »;
- Le groupe « R2SESS » (N=65), étudiants qui ont réussi en seconde session. Sous-groupes en fonction du grade obtenu : « R2SESS-GD », « R2SESS-D », « R2SESS-S »;
- Le groupe « ECHABAN » (N=210), étudiants qui ont été ajourné en 2ème session et qui ont donc échoué ou abandonné.

Nous avons aussi différencié les étudiants inscrits pour la première fois en 1995 de ceux qui répétaient leur 1ère candidature en 1995-1996. Ces groupes ont été codés respectivement « I95 » (N=286) et « REP » (N=67).

Enfin, nous avons également considéré les performances des étudiants qui ont effectué en début d'année académique (octobre et novembre 1995) l'entraînement à l'utilisation des degrés de certitude à l'aide du logiciel GUESS. Bien que nous ayons recommandé cet entraînement à tous les étudiants, seule une partie des étudiants de 1ère candidature ont saisi l'opportunité de l'exercice (55,1%). Les étudiants entraînés ont été codés « GUESS » (N=195) et les autres « NONGUESS » (N=158).

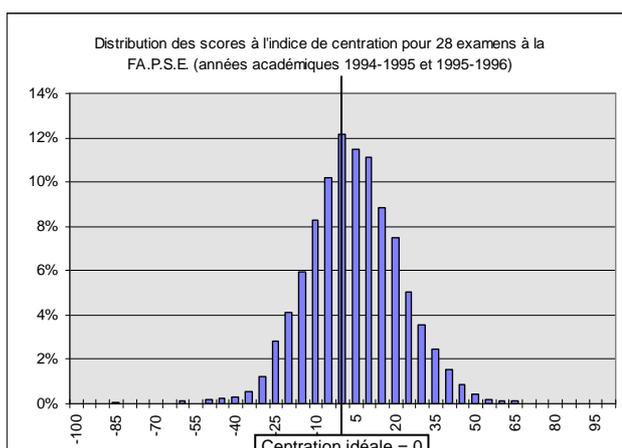
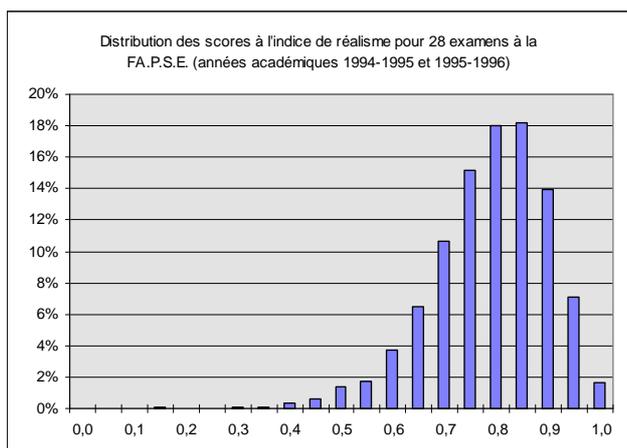
b) Repères pour l'interprétation des indices de réalisme et de centration

Dans une précédente étude (GILLES, 1996), nous avons établi pour les indices de réalisme et de centration des normes locales à la FA.P.S.E. qui nous permettent de comparer les performances d'un individu ou d'un groupe avec les performances moyennes des étudiants de la faculté. Ces normes sont basées sur les performances relevées lors de 24 examens qui ont eu lieu en 1994-1995 et 1995-1996. Ci-dessous figurent les distributions des scores des étudiants de la faculté aux indices de réalisme et de centration.

⁹ Nous menons actuellement une recherche où nous essayons d'évaluer l'effet de l'entraînement à l'utilisation des degrés de certitude sur les performances à l'indice de réalisme lors d'exercices en fin de cours (quizz). Dans le cadre de cette recherche nous avons mis en place un plan expérimental en carré latin avec un groupe témoin.

Jean-Luc GILLES, *Profils d'autoévaluation et d'états de connaissance partielle chez les étudiants de première candidature (1995-1996) de la Faculté de Psychologie et des Sciences de l'Éducation de l'Université de Liège, Colloque de l'Association Internationale de Pédagogie Universitaire (AIPU), Tunis, 4, 5 et 6 novembre 1996.*

Notons la distribution très décentrée (par son mode) vers la droite, donc plutôt en J, du graphique



de gauche relatif à l'indice de réalisme.

En ce qui concerne la distribution des scores de centration, on observe une minorité d'étudiants dont l'auto-estimation est excellente c'est à dire entre -4 et 4 (12,1%) ainsi qu'un pourcentage plus élevé d'étudiants qui se sur-estiment (52,2%) par rapport aux étudiants qui se sous-estiment (34,1%). Suite à cette étude, nous recommandons les interprétations suivantes pour les valeurs obtenues aux indices de réalisme et de centration :

L'INDICE DE REALISME EST		
IDEAL si il est compris entre	0,96 et 1	Performance accomplie par 1,6 % des étudiants
EXCELLENT	0,91 et 0,95	7,1 %
TRES BON	0,86 et 0,90	13,9 %
BON	0,81 et 0,85	18,2 %
MOYEN	0,76 et 0,80	18 %
FAIBLE	0,71 et 0,75	15,2 %
INSUFFISANT	0,66 et 0,70	10,7 %
TRES INSUFFISANT	0 et 0,65	14,7 %

L'INDICE DE CENTRATION REFLETE :

UNE AUTO-ESTIMATION EXCELLENTE (SANS TROP DE SUR- ou SOUS-ESTIMATION)

Si la valeur obtenue est comprise entre **-4 et +4** Performance accomplie par **12,1 %** des étud^{ts}
(avec surestimation très modérée si signe positif, ou sous-estimation très modérée si signe négatif)

UNE SOUS-ESTIMATION

TRES LEGERE	si elle est comprise entre -5 et -9	Performance accomplie par	10,2 % des étud ^{ts}
LEGERE	-10 et -14		8,3 %
ASSEZ FORTE	-15 et -19		6 %
FORTE	-20 et -24		4,1 %
TRES FORTE	-25 et -29		2,8 %
EXTREME	-30 et -100		2,7 %

UNE SURESTIMATION

TRES LEGERE	si elle est comprise entre +5 et +9	Performance accomplie par	11,5 % des étud ^{ts}
LEGERE	+10 et +14		11,1 %
ASSEZ FORTE	+15 et +19		8,9 %
FORTE	+20 et +24		7,5 %
TRES FORTE	+25 et +29		5 %
EXTREME	+30 et +100		9,2 %

5. Questions relatives aux performances métacognitives des étudiants des sous-groupes décrits ci-avant

a) Décèle-t-on des différences significatives aux indices de réalisme et de centration chez les étudiants qui réussissent ou qui échouent ?

En d'autres termes, les performances métacognitives de nos étudiants sont-elles corrélées avec leur réussite académique ? La réponse à cette question est importante, car dans l'affirmative elle impliquerait qu'on établisse par des études expérimentales ultérieures si cette corrélation cache une relation de cause à effet, et, dans le cas où cette relation serait établie, qu'on en mesure le poids dans l'analyse des causes de l'échec en 1ère candidature à l'université.

b) Observe-t-on des performances métacognitives constantes lors des différents examens ?

Les scores obtenus à l'indice de réalisme et de centration sont-ils relativement indépendants des épreuves où les degrés de certitude sont récoltés ? Autrement dit, les étudiants font-ils preuve d'un niveau de performance métacognitive relativement constant à ces indices, ou, au contraire, leurs performances fluctuent-elles d'une évaluation à l'autre ? Les coefficients de corrélation entre les scores de réalisme et de centration à une épreuve confrontée à toutes les autres peuvent nous éclairer sur ce point (la difficulté réside ici dans le fait que tous les étudiants ne sont pas forcément présents à tous les examens : des répétants peuvent être dispensés, d'autres abandonnent en cours d'année, etc.).

c) Le réalisme s'améliore-t-il au fil des examens réalisés au cours de l'année académique ?

Nous avons recueilli des données métacognitives dans le cadre de 8 examens qui se sont succédés au cours de l'année 1995-1996. On peut se demander si lors des derniers examens les étudiants obtenaient de meilleurs résultats à l'indice de réalisme que lors des premières épreuves. Avec l'habitude, les étudiants s'auto-estiment-ils de façon de plus en plus réaliste ?

d) Observe-t-on une corrélation entre les scores de réalisme et de centration lors de l'entraînement GUESS et lors des 8 examens envisagés ?

L'entraînement à l'utilisation des degrés de certitude est indépendant du contenu des cours en 1ère candidature (texte de la littérature française dont il faut deviner les lettres des mots), dès lors, les scores métacognitifs « neutres » de cet entraînement se corrélaient-ils avec les scores obtenus lors des examens ? Nous ne pourrions effectuer cette mesure que pour une partie des étudiants : ceux qui ont effectué l'entraînement.

e) Observe-t-on des différences entre les scores métacognitifs des filles et des garçons ?

Jean-Luc GILLES, Profils d'autoévaluation et d'états de connaissance partielle chez les étudiants de première candidature (1995-1996) de la Faculté de Psychologie et des Sciences de l'Éducation de l'Université de Liège, Colloque de l'Association Internationale de Pédagogie Universitaire (AIPU), Tunis, 4, 5 et 6 novembre 1996.

Dans une étude précédente (GILLES, 1995) nous avons relevé une tendance à la sous-estimation lors d'un entraînement à l'utilisation des degrés de certitude plus forte chez les filles que chez les garçons : 13% des filles se sous-estimaient (indice de centration inférieur à -1) alors qu'on observait à peine 6 % des garçons dans la même situation. Qu'en est-il dans le cadre des données de cette étude ?

f) Dans quelle mesure le spectre des compétences fluctue-t-il d'une session à l'autre ?

La ventilation des réponses dans les différentes gammes du spectre est intéressante à observer chez les étudiants qui échouent en 1ère session et qui se représentent en seconde session. Quelles différences observe-t-on entre les spectres de première session et de seconde session chez ceux qui réussiront après un examen de repêchage et ceux qui échoueront ?

6. Analyse des données de 8 examens qui ont eu lieu à la FA.P.S.E. en 1ère candidature lors de l'année académique 1995-1996

a) Différences entre les scores métacognitifs des étudiants qui réussissent ou qui échouent

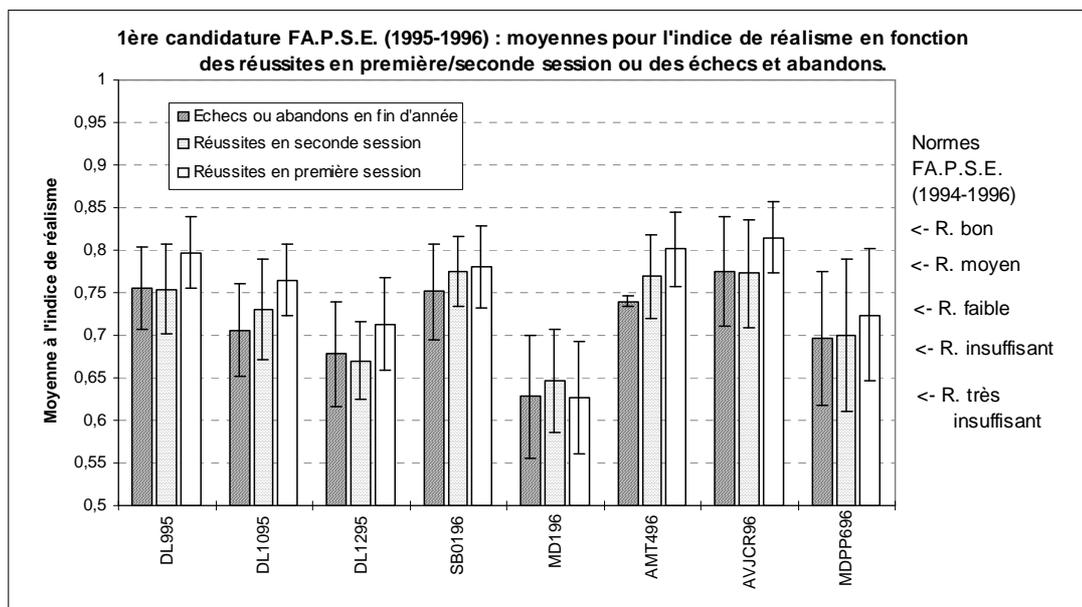
Les étudiants de 1ère candidature de l'année académique 1995-1996 ont été répartis en fonction de leurs résultats en fin d'année académique. Rappelons les trois profils d'étudiants : ceux qui réussissent en première session (R1SESS), en seconde session (R2SESS), ou qui échouent ou abandonnent (ECHABAN).

1. Différences liées à l'indice de réalisme

Les scores moyens de réalisme aux 8 examens sont repris en gras pour les 3 groupes dans le tableau ci-dessous. La dernière colonne (TEexam) reprend les taux d'exactitude (% de réponses correctes) pour les 8 épreuves considérées (nous verrons plus loin que les taux d'exactitude qui traduisent la difficulté d'une épreuve, semblent être en relation avec les performances à l'indice de réalisme).

Code examen :	Date :	1ère C. PSE	ECHABAN	R2SESS	R1SESS	TEexam
DL995	03/10/95	M = 0,766 ET = 0,097 (N = 247)	M = 0,756 ET = 0,096 (N=139)	M = 0,754 ET = 0,106 (N=48)	M = 0,797 ET = 0,084 (N=60)	M = 63%
DL1095	24/10/95	M = 0,724 ET = 0,097 (N = 290)	M = 0,706 ET = 0,110 (N=173)	M = 0,730 ET = 0,118 (N=52)	M = 0,765 ET = 0,085 (N=65)	M = 51%
DL1295	19/12/95	M = 0,683 ET = 0,116 (N = 237)	M = 0,678 ET = 0,123 (N=148)	M = 0,670 ET = 0,091 (N=43)	M = 0,713 ET = 0,108 (N=46)	M = 44%
SB0196	09/01/96	M = 0,762 ET = 0,105 (N = 281)	M = 0,751 ET = 0,113 (N=169)	M = 0,775 ET = 0,082 (N=48)	M = 0,781 ET = 0,096 (N=64)	M = 56%
MD196	11/01/96	M = 0,630 ET = 0,138 (N = 295)	M = 0,628 ET = 0,145 (N=175)	M = 0,646 ET = 0,122 (N=52)	M = 0,627 ET = 0,131 (N=65)	M = 39%
AMT496	23/04/96	M = 0,763 ET = 0,109 (N = 265)	M = 0,740 ET = 0,119 (N=135)	M = 0,769 ET = 0,099 (N=58)	M = 0,801 ET = 0,087 (N=72)	M = 50%
AVJCR96	26/04/96	M = 0,784 ET = 0,119 (N = 301)	M = 0,775 ET = 0,128 (N=161)	M = 0,773 ET = 0,127 (N=64)	M = 0,815 ET = 0,083 (N=76)	M = 54%
MDPP696	04/06/96	M = 0,702 ET = 0,160 (N = 196)	M = 0,697 ET = 0,157 (N=131)	M = 0,700 ET = 0,178 (N=41)	M = 0,724 ET = 0,155 (N=33)	M = 57%

Mises en graphique, les données du tableau ci-avant montrent bien les meilleurs scores de réalisme pour les étudiants qui réussissent en 1ère session (à une exception près : l'évaluation codée «MD196»).



Les scores des étudiants qui réussissent en seconde session se situent dans une zone intermédiaire. Ce sont en général les étudiants qui échouent ou abandonnent qui effectuent les moins bonnes performances à l'indice de réalisme. Remarquons également que lorsqu'on compare les scores de réalisme du meilleur groupe (les étudiants qui réussissent en 1ère session) aux normes établies pour la FA.P.S.E. (4.b) ci-avant), ce groupe n'atteint pas la zone « Idéale » (comprise entre 0,96 et 1) ni « Excellente » (comprise entre 0,91 et 0,95).

Les différences observées entre les 3 groupes sont-elles significatives ou dues au fluctuations du hasard ? Le tableau ci-dessous reprend en synthèse les résultats des analyses de variance des scores à l'indice de réalisme pour la variable « sanction académique » (R1SESS, R2SESS, ECHABAN) confrontée à chacun des 8 examens envisagés.

Analyse de la Variance (Effets significatifs marqués en gras à $p < ,05$)

	Effet SC	Effet dl	Effet Vb	Erreur SC	Erreur dl	Erreur Vw	F	p
DL95	,076593	2	,038297	2,244321	244	,009198	4,163554	,016671
DL1095	,161716	2	,080858	3,278308	287	,011423	7,078711	,000998
DL1295	,052541	2	,026270	3,099569	234	,013246	1,983262	,139926
SB0196A	,051942	2	,025971	3,063102	278	,011018	2,357048	,096589
MD196B	,014464	2	,007232	5,582788	292	,019119	,378264	,685385
AMT496A	,175577	2	,087788	2,983278	262	,011387	7,709832	,000558
AVJCR496	,094539	2	,047270	4,156806	298	,013949	3,388752	,035057
MDPP696	,019172	2	,009586	4,972586	193	,025765	,372059	,689807

Bien que la variabilité due aux différences entre les groupes (V_b) est supérieure à la variance d'erreur interne aux groupes (V_w) dans 6 cas sur 8, on observe des différences significatives que pour 4 examens sur 8.

Cependant, comme le montre le tableau ci-dessous, la moyenne des scores à l'indice de réalisme pour l'ensemble des examens révèle de nettes différences entre les groupes ECHABAN, R2SESS et R1SESS. Les étudiants qui réussissent en 1ère session obtiennent de meilleurs scores à l'indice de réalisme que les étudiants qui réussissent en seconde session, qui eux-mêmes obtiennent de meilleurs scores que les étudiants qui échouent.

	ECHABAN (N=210)	R2SESS (N=65)	R1SESS (N=78)
Moyenne à l'ensemble des examens (nombre moyen d'examen = 6)	0,712	0,734	0,763
Ecart type	0,053	0,064	0,071

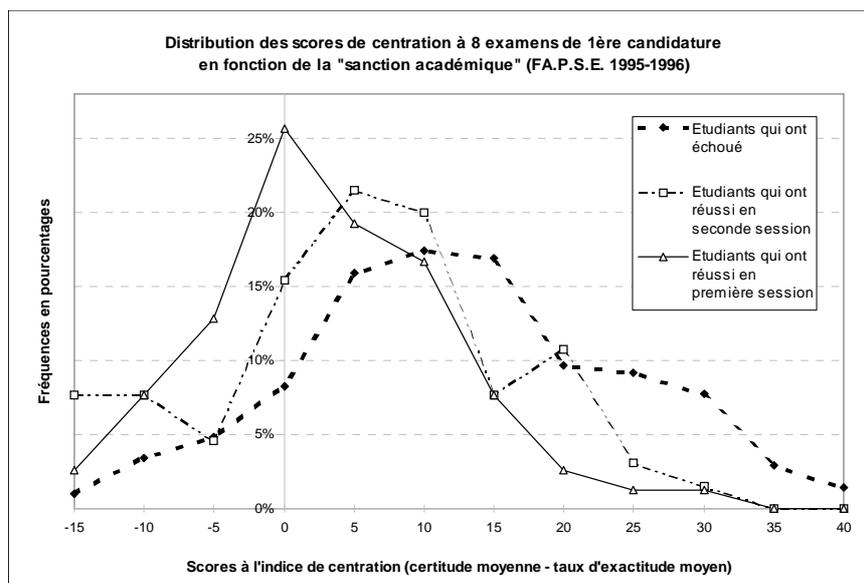
Lorsqu'on envisage l'ensemble des 8 examens, et non plus chacun séparément, l'analyse de la variance pour les trois groupes de la « sanction académique » indique une variabilité entre les groupes (V_b) beaucoup plus élevée que la variance d'erreur interne (V_w). Le résultat du rapport de la variance inter-groupe sur la variance d'erreur est très significatif ($F = 17,07$). Le risque que les différences entre les groupes puissent être attribuées au hasard est très faible.

Analyse de la Variance (Effet significatif marqué en gras à $p < ,01$)

	Effet SC	Effet dl	Effet V_b	Erreur SC	Erreur dl	Erreur V_w	F	p
Moy. Réalisme	,156066	2	,078033	1,599681	350	,004571	17,07317	,000000

2. Différences liées à l'indice de centration

La moyenne des indices de centration des 8 examens a été calculée pour chaque étudiant. Le graphique ci-dessous a été établi en distribuant ces scores en fonction des trois groupes d'étudiants définis par la « sanction académique ». Rappelons que la centration idéale vaut 0 (écart minimal entre la certitude moyenne et le taux d'exactitude moyen) et que l'indice révèle la sous-estimation lorsqu'il est négatif et la sur-estimation lorsqu'il est positif (voir normes pour la F.A.P.S.E. au point 4. b) ci-avant).



Il est frappant de constater que le « pic » de centration des étudiants qui ont réussi en 1ère session est aligné sur la centration idéale (0). On observe également une tendance à se sur-estimer qui croît en fonction du niveau de réussite : les étudiants qui réussissent en seconde session, dans l'ensemble, se sur-estiment plus que les étudiants qui réussissent en première session mais moins que ceux qui échouent. Enfin, on constate que la tendance générale à la sur-estimation se vérifie une fois de plus.

b) Degré de régularité des performances à l'indice de réalisme lors des différents examens

Des matrices de corrélation ont été calculées pour les résultats de l'ensemble des étudiants de 1ère candidature ainsi que pour chacun des trois groupes d'étudiants constitués en fonction de la performance. On observe des corrélations significatives mais peu élevées dans seulement 6 cas sur 28.

Matrice des corrélations (N = 109 - suppr. des observ. à VM) Corrélations significatives marquées à $p < ,05000$

	DL995	DL1095	DL1295	SB0196A	MD196B	AMT496A	AVJCR496	MDPP696
DL995	1,00							
DL1095	0,11	1,00						
DL1295	0,22	0,19	1,00					
SB0196	-0,10	-0,22	-0,06	1,00				
MD196	-0,01	-0,01	0,10	0,08	1,00			
AMT496	0,21	-0,03	0,08	0,22	0,08	1,00		
AVJCR496	0,06	0,12	0,03	0,06	-0,03	0,17	1,00	
MDPP696	0,18	-0,08	0,21	-0,06	0,13	-0,16	0,02	1,00

En ce qui concerne l'examen codé « MD196 », les corrélations sont à prendre avec prudence étant donné les différences significatives établies ci-après (test t) entre le groupe des étudiants de l'examen

(N = 295) et l'échantillonnage utilisé pour réaliser la table des corrélations (N = 109). Dans les autres cas les différences ne sont pas significatives et l'échantillonnage peut être considéré comme représentatif.

Test t pour des Ech. Indépendants (Variables traitées comme des éch. indépend.)

	Moyenne Groupe1	Moyenne Groupe2	Valeur t	dl	p	N Actifs Groupe1	N Actifs Groupe2	Ec.-type Groupe1	Ec.-type Groupe2
DL995 vs. 109	,765714	,755167	,908008	354	,364492	247	109	,097132	,109352
DL1095 vs. 109	,723864	,709733	1,192403	397	,233815	290	109	,109102	,095107
DL1295 vs. 109	,683125	,687188	-,310866	344	,756091	237	109	,115570	,106951
SB0196A vs.109	,762299	,762778	-,039735	388	,968325	281	109	,105476	,110260
MD196B vs. 109	,630847	,596607	2,268795	402	,023810	295	109	,137979	,125088
AMT496A vs109	,762967	,744170	1,468627	372	,142779	265	109	,109386	,119717
AVJCR496 vs109	,784420	,778010	,573455	494	,566598	301	195	,119043	,125437
MDPP696 vs109	,701843	,690652	,693694	389	,488288	196	195	,159996	,158990

Nous avons également envisagé la matrice de corrélation pour chaque groupe d'étudiant. Matrice des corrélations pour le groupe des étudiants qui réussissent en 1ère session :

Corrélations R1SESS (N=19 - Suppression des Observ. à VM) Corrélations significatives marquées à $p < ,05000$

	DL995	DL1095	DL1295	SB0196A	MD196B	AMT496A	AVJCR496	MDPP696
DL995	1,00							
DL1095	,19	1,00						
DL1295	,06	,32	1,00					
SB0196A	-,14	-,08	-,34	1,00				
MD196B	,10	,22	,17	,02	1,00			
AMT496A	,05	-,08	-,10	,36	,15	1,00		
AVJCR496	-,47	,15	-,21	,39	-,16	,12	1,00	
MDPP696	,24	-,04	,35	-,16	-,30	-,51	-,31	1,00

Seulement 2 corrélations significatives sur 28.

Matrice des corrélations pour le groupe des étudiants qui réussissent en 2ème session :

Corrélations R2SESS (N=23 - Suppression des Observ. à VM) Corrélations significatives marquées à $p < ,05000$

	DL995	DL1095	DL1295	SB0196A	MD196B	AMT496A	AVJCR496	MDPP696
DL995	1,00							
DL1095	,05	1,00						
DL1295	-,07	,07	1,00					
SB0196A	,10	-,21	-,22	1,00				
MD196B	-,19	,05	,45	-,21	1,00			
AMT496A	,44	-,37	-,01	,02	,07	1,00		
AVJCR496	,16	-,11	-,26	,29	-,27	-,01	1,00	
MDPP696	-,23	-,36	,02	,25	,19	-,09	,21	1,00

Seulement 2 corrélations significatives sur 28.

Matrice des corrélations pour le groupe des étudiants qui échouent :

Corrélations ECHABAN (N=67 - Suppression des Observ. à VM) Corrélations significatives marquées à $p < ,05000$

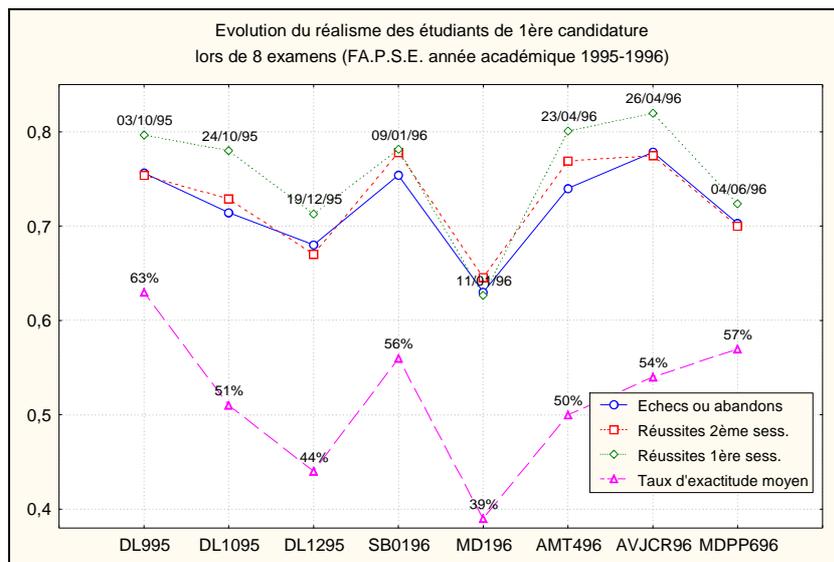
	DL995	DL1095	DL1295	SB0196A	MD196B	AMT496A	AVJCR496	MDPP696
DL995	1,00							
DL1095	,10	1,00						
DL1295	,31	,16	1,00					
SB0196A	-,21	-,28	-,03	1,00				
MD196B	,04	-,03	,05	,18	1,00			
AMT496A	,14	,08	,12	,17	,07	1,00		
AVJCR496	,04	,16	,07	-,06	,03	,15	1,00	
MDPP696	,30	-,01	,22	-,14	,27	-,12	,02	1,00

Seulement 2 corrélations significatives sur 28.

Ces matrices ne montrent guère de corrélations positives et significatives pour les scores à l'indice de réalisme. Ceci semble indiquer qu'en ce qui concerne le réalisme les performances sont peu constantes d'un examen à l'autre.

c) Evolution du réalisme au fil des examens et parallélisme avec la difficulté des épreuves

Les 8 examens considérés ont eu lieu à des moments différents au cours de l'année académique 1995-1996. Le graphique ci-après montre l'évolution du réalisme en fonction des examens présentés dans leur ordre chronologique.



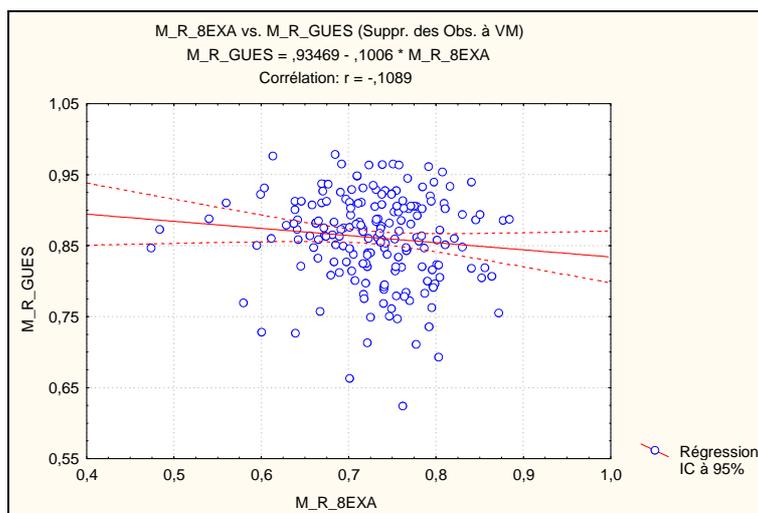
L'hypothèse d'une amélioration des scores de réalisme au fil des examens présentés par les étudiants semble devoir être rejetée à la lecture du graphique ci-avant. Lors d'une étude précédente (GILLES, 1996) nous avons observé une relation entre la facilité d'un examen (exprimé en pourcentage de réponses correctes) et les scores moyens à l'indice de réalisme. Cette relation entre la facilité d'un examen et le réalisme des étudiants se vérifie pour les 8 examens envisagés. La corrélation entre le taux d'exactitude moyen et le réalisme moyen pour les 8 examens vaut : $r = 0,45$

d) Relations entre les performances métacognitives à l'entraînement GUESS et aux examens

1. Corrélations non significatives pour l'indice de réalisme

La corrélation pour les scores de réalisme des étudiants inscrits en 1ère candidature et qui ont participé à l'entraînement GUESS (N = 195), et, la moyenne des scores de réalisme pour les 8 évaluations vaut :

$$r = -0,11$$



Notons que tous les étudiants n'effectuent pas la totalité des 8 examens (certains ont abandonné en cours d'année, d'autres ont été absents,...), le sous-ensemble des étudiants qui a participé à l'entraînement a effectué 6,4 examens en moyenne.

Jean-Luc GILLES, *Profils d'autoévaluation et d'états de connaissance partielle chez les étudiants de première candidature (1995-1996) de la Faculté de Psychologie et des Sciences de l'Éducation de l'Université de Liège, Colloque de l'Association Internationale de Pédagogie Universitaire (AIPU), Tunis, 4, 5 et 6 novembre 1996.*

Les corrélations obtenues pour chacune des épreuves, montrent également peu de relation entre les performances à l'indice de réalisme lors de l'entraînement et lors des examens.

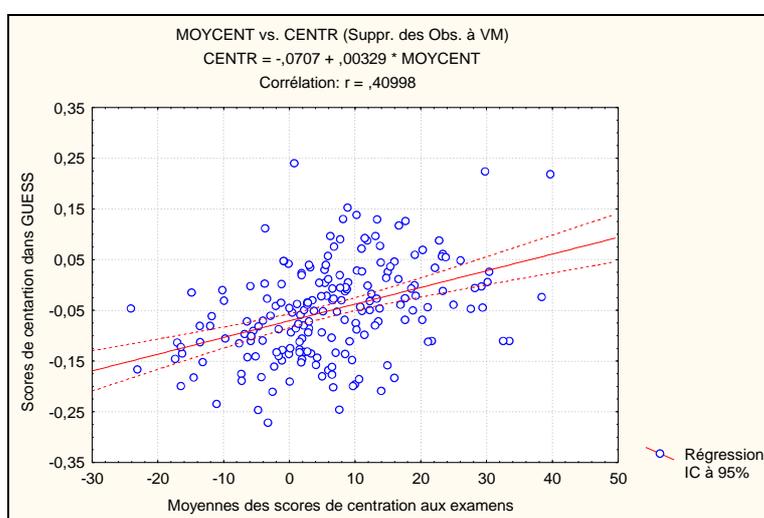
Corrélations significatives marquées à $p < ,05000$
(Suppression des observations à valeurs manquantes)

	REAL_GUES	P=	N
DL995	0,03	0,690	149
DL1095	0,02	0,777	172
DL1295	-0,19	0,027	141
SB0196A	0,08	0,325	161
MD196B	-0,13	-0,133	166
AMT496A	-0,12	-0,112	166
AVJCR496	0,03	0,724	170
MDPP696	-0,21	0,025	113

2. Corrélations significatives pour l'indice de centration

La corrélation entre les scores de centration des étudiants (1ère candidature) lors de l'entraînement GUESS (N = 195) et la moyenne des scores de centration pour les 8 examens vaut :

$$r = 0,41$$



Corrélations significatives marquées à $p < ,05000$
(Suppression des observations à valeurs manquantes)

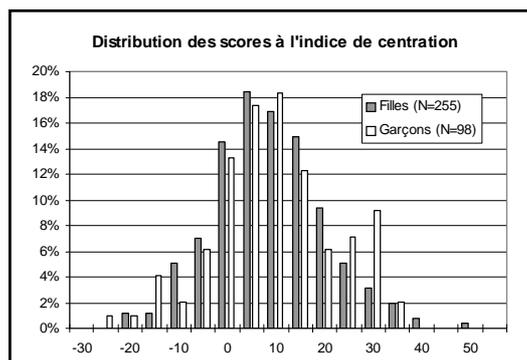
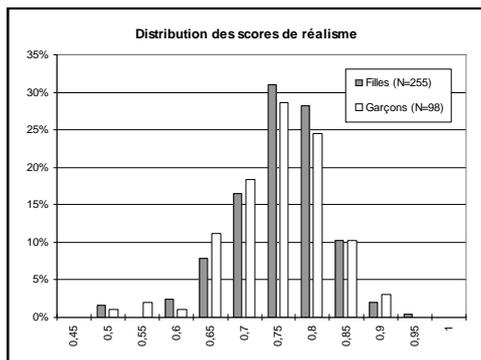
	CENTRA_GUES	P=	N
DL995	0,31	0,000	146
DL1095	0,23	0,001	194
DL1295	0,12	0,094	194
SB0196A	0,26	0,000	194
MD196B	0,29	-0,000	194
AMT496A	0,13	-0,072	194
AVJCR496	0,28	0,000	194
MDPP696	-0,03	0,754	194

Nous constatons qu'à l'inverse de l'indice de réalisme, la corrélation entre l'entraînement et l'ensemble des examens est élevée ($r = 0,41$). De plus, les corrélations obtenues pour chacune des épreuves et l'entraînement sont significatives dans 6 cas sur 8.

Notons également que c'est là où les corrélations étaient significative pour le réalisme, qu'elles ne le sont plus pour la centration et inversement. Est-ce une coïncidence, ou cela cache-t-il une relation entre les deux indices ?

e) Différences métacognitives entre filles et garçons

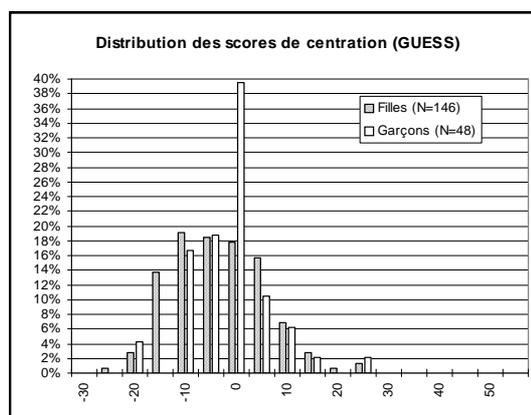
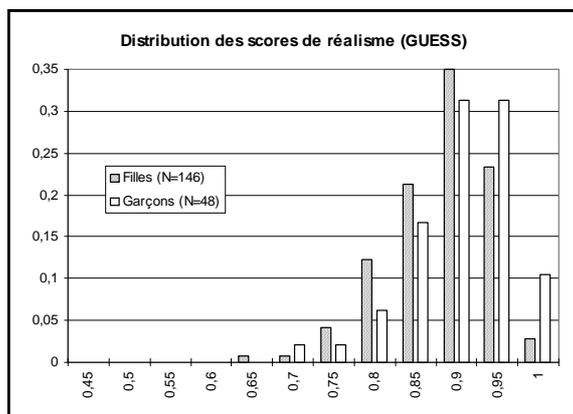
1. Lors des examens



Les scores de réalisme sont légèrement meilleurs pour les filles ($M = 7,29$) que pour les garçons ($M = 7,22$) lors des examens. En ce qui concerne la centration, les filles sont plus proches de la valeur idéale ($M = 6,3$) que les garçons ($M = 7,1$). La moyenne des scores de centration est positive pour les filles ainsi que pour les garçons ce qui est indicatif d'une tendance générale à la sur-estimation. Remarquons également le pic de sur-estimation des garçons : environ 11% d'entre-eux obtiennent un score de sur-estimation situé au-delà de 30 qui correspond à la catégorie « Sur-estimation extrême » (voir normes au point 4. b) ci-avant), tandis qu'environ 6% des filles entrent dans cette catégorie.

2. Lors de l'entraînement

A l'entraînement, sur un contenu « neutre », les garçons obtiennent de meilleurs scores à l'indice de réalisme ($M = 0,877$) que les filles ($M = 0,856$). Leur moyenne pour l'indice de centration est également meilleure ($M = -3,6$) que celle des filles ($M = -5,6$). En moyenne, les scores à l'indice de réalisme sont meilleurs lors de l'entraînement environ 10% de garçons et 3% de filles obtiennent un score « idéal » entre 0,96 et 1.



39% des garçons ont un score idéal à l'indice de centration lors de l'entraînement contre 18% chez les filles. A l'inverse de la distribution des scores de centration à l'examen, on observe pour l'entraînement qui eu lieu en octobre-novembre 1995, plus de sous-estimations que de sur-estimations.

f) Flux dans le spectre des compétences

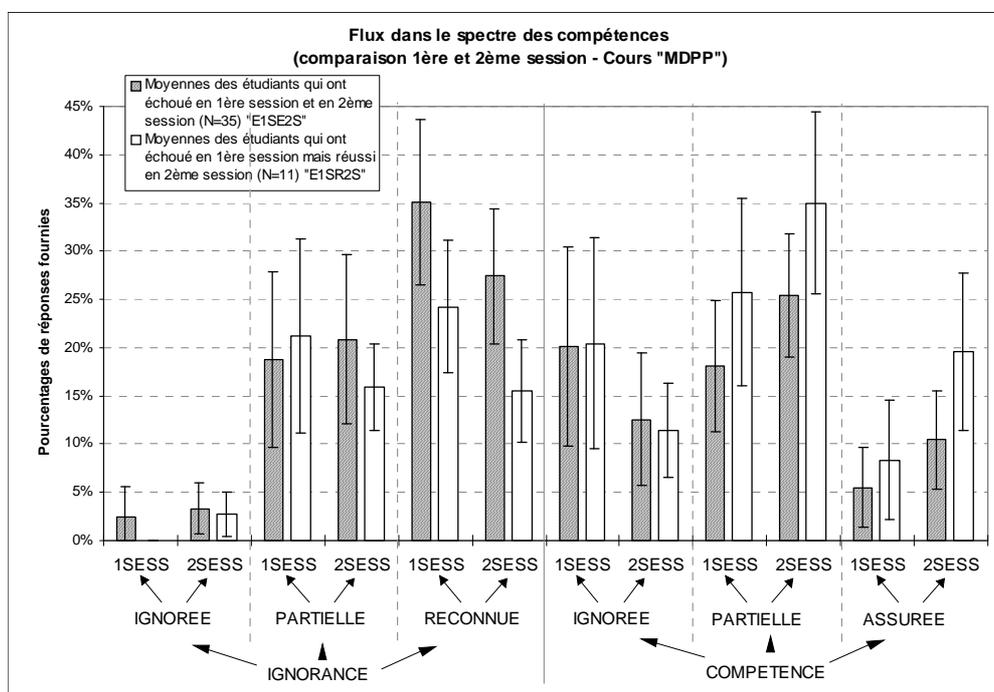
Le tableau ci-après illustre les phénomènes de flux dans les états de connaissances partielles chez les étudiants, au départ du seul examen parmi les 8 envisagés pour lequel il y eu en septembre 1996 un examen de repêchage avec QCM et degrés de certitude. Ceux qui échouent en 1ère session et qui échouent également en seconde session (1) sont environ trois fois plus nombreux ($N=35$) que les autres (2) qui réussiront en seconde session ($N=11$).

Lorsqu'on compare les performances en juin (1SESS) et en septembre (2SESS) de ceux qui échoueront en seconde session, on constate une augmentation de l'ignorance ignorée et de l'ignorance partielle avec une diminution de l'ignorance reconnue. La compétence ignorée diminue aussi (peut-être répondent-ils moins au hasard), par contre, la compétence partielle et assurée augmentent mais pas assez.

Remarquons que ceux qui réussissent en seconde session (2) avaient obtenu en juin (1SESS) plus de compétences partielles et assurées que ceux qui vont échouer en seconde session. Ils étaient également victime de moins d'ignorances ignorées, partielles et reconnue. En septembre, ils (2) creuseront l'écart, avec, globalement, encore moins d'ignorances et plus de compétences.

		IGNORANCE						COMPETENCE					
		Ignorée		Partielle		Reconnue		Ignorée		Partielle		Assurée	
		1SESS	2SESS	1SESS	2SESS	1SESS	2SESS	1SESS	2SESS	1SESS	2SESS	1SESS	2SESS
(1) ECHEC en 2ème session (N = 35)	MOY	2,4%	3,3%	18,8%	20,9%	35,1%	27,4%	20,2%	12,6%	18,1%	25,4%	5,5%	10,4%
	ET	0,06	0,05	0,18	0,18	0,17	0,14	0,21	0,14	0,14	0,13	0,08	0,10
(2) REUSSITE en 2ème session (N = 11)	MOY	0,0%	2,7%	21,2%	15,9%	24,2%	15,5%	20,5%	11,4%	25,8%	35,0%	8,3%	19,5%
	ET	0,00	0,05	0,20	0,09	0,14	0,11	0,22	0,10	0,20	0,19	0,12	0,16

Le graphique ci-dessous montre bien à quel point ceux qui réussissent parviennent à augmenter leurs compétences partielles et assurées tout en diminuant les autres états de compétence/ignorance du reste du spectre (à l'exception d'environ 2,5% d'ignorance ignorée qui apparaissent en 2ème session). L'autre groupe diminue le nombre d'ignorances reconnues et de compétences ignorées, mais sans atteindre les niveaux de ceux qui réussissent tout en augmentant, mais insuffisamment, les compétences partielles et assurées.



7. Conclusions

a) Résultats et tendances

1. Sanction académique et performances aux indices de réalisme et de centration

On observe pour l'indice de réalisme des différences pour les trois groupes « Réussite en 1ère session », réussite en seconde session » et « échec/abandon ». Cependant, l'analyse de variance ne montre des différences significatives que pour 4 examens sur 8. Néanmoins, lorsqu'on considère la moyenne des performances à l'indice de réalisme pour l'ensemble des 8 évaluations, l'analyse de variance révèle un effet très significatif : la variabilité entre les groupes est alors supérieure à la variabilité d'erreur interne.

En ce qui concerne l'indice de centration, les distributions des scores moyens à l'ensemble des examens ont été comparées. On observe trois distributions « décalées » en fonction de la performance académique. Le mode de la distribution des étudiants qui réussissent en 1ère session est situé sur la performance « idéale » de centration (différence entre la certitude moyenne et le taux d'exactitude moyen

égale à 0). Les performances des deux autres groupes sont en moyenne moins bonnes avec à chaque fois un déplacement de la distribution vers la droite. Les étudiants qui échouent en seconde session ont dans l'ensemble de moins bonnes performances à l'indice de centration que les étudiants qui échouent.

On observe également un phénomène général de sur-estimation dans les trois groupes. Ce phénomène a déjà été relevé et décrit précédemment (GILLES, 1995a et 1995b). Ce type de problématique de recherche est également étudié par DIRKZWAGER & al. (1993, 1995).

Une tendance semble donc se dégager : les étudiants les plus performants du point de vue académique sont également les plus réalistes et les plus « calibrés » du point de vue de la sur- et sous-estimation. Les performances de réalisme et de centration décroissent avec les performances académiques.

Dès lors se pose la question de la relation entre la performance académique et la performance d'auto-estimation. La qualité de l'auto-estimation influence-t-elle le degré de performance académique ? faut-il en tenir compte dans l'analyse des causes du grand nombre d'échecs en 1ère candidature de l'enseignement supérieur universitaire ?

2. Fluctuation des performances individuelles à l'indice de réalisme lors des examens

Des matrices de corrélation ont été calculées pour les résultats de l'ensemble des étudiants de 1ère candidature ainsi que pour chacun des trois groupes d'étudiants constitués en fonction de la performance académique (« réussite en 1ère session », « réussite en seconde session » et « échec/abandon »). Le nombre de corrélations significatives est à chaque fois peu élevé et semble indiquer une fluctuation des performances aussi bien pour les étudiants pris tous ensemble que pour chacun des groupes isolément. Une tendance à la fluctuation des performances de réalisme semble se dégager de l'étude des données liées aux 8 examens envisagés (1ère candidature FA.P.S.E., 1995-1996).

3. Instabilité « horizontale » des performances à l'indice de réalisme

Au début de ce travail nous pensions pouvoir assister à une amélioration des performances à l'indice de réalisme au fil des évaluations que les étudiants effectueraient. Cette hypothèse ne se vérifie pas à la lecture des données des 8 examens envisagés. Les performances fluctuent d'un examen à l'autre sans progresser véritablement (on aurait pu imaginer une amélioration des performances « en dents de scie »). Cette instabilité horizontale s'observe pour les trois groupes d'étudiants. Le graphique de l'évolution du réalisme des étudiants de 1ère candidature montre de meilleures performances pour ceux qui réussissent en 1ère session en comparaison avec les deux autres groupes très proches.

Nous constatons également une corrélation assez marquée ($r = 0,45$) entre le taux d'exactitude moyen et le réalisme moyen pour les 8 examens.

4. Une relation entre les scores obtenus à l'indice de centration lors de l'entraînement et lors des examens

La corrélation entre les scores à l'indice de réalisme lors de l'entraînement et les scores moyens calculés pour l'ensemble des examens est très basse et peut être considérée comme nulle ($r = -0,11$). Par contre la corrélation obtenue pour l'indice de centration est assez marquée ($r = 0,41$), de plus les corrélations calculées pour chacune des épreuves et l'entraînement sont significatives dans 6 cas sur 8.

L'indice de réalisme constitue une mesure « intégrale » qui reprend les performances d'auto-estimation pour chaque degré de certitude et les incorpore en une seule valeur. Une première explication pourrait être liée à ce nombre de réponses accompagnées d'un degré de certitude et qui a probablement un impact sur les mesures effectuées (comme le montre les graphiques de réalisme, plus le nombre de réponses fournies pour chaque degré de certitude est élevé, plus l'erreur de mesure diminue, voir illustration au point d)1. ci-avant). Or l'entraînement permet de donner beaucoup de réponses en peu de temps (en moyenne 200 en 40 minutes) tandis qu'aux examens ce sont en général entre 20 à 50 questions qui sont posées (en moyenne 6 à 7 fois moins qu'à l'entraînement). Une seconde explication pourrait être liée au contenu même de l'entraînement qui est « neutre », sans rapport avec les cours suivis par les étudiants, et qui pourrait expliquer les différences de scores qui engendrent la corrélation basse.

La corrélation obtenue pour la centration semble indiquer que les phénomènes de sur- et sous-estimation se reproduisent de façon assez proche qu'on soit en situation d'entraînement ou d'examen. La meilleure corrélation obtenue par cet indice plus « moyen » (centration = certitude moyenne - taux d'exactitude moyen) s'explique peut-être par ses caractéristiques moyennes.

5. Différences métacognitives entre filles et garçons

On observe lors des examens des scores légèrement meilleurs pour les filles à la fois à l'indice de réalisme ($M = 7,29$ contre $M = 7,22$ pour les garçons) et à l'indice de centration ($M = 6,3$ contre $M = 7,1$ pour les garçons). Cette tendance s'inverse lorsqu'on envisage les scores obtenus à l'entraînement (réalisme : M des filles = $0,856$; M des garçons = $0,877$; centration : M des filles = $-5,6$; M des garçons = $-3,6$). Ces différences s'expliquent peut-être les contextes de récolte des données : l'entraînement propose beaucoup de questions liées à un contenu « neutre » et est sans conséquences du point de vue de l'impact de la cote obtenue. L'examen, au contraire, propose beaucoup moins de questions qui, elles, sont liées à un contenu qu'il a fallu étudier, et, la cote obtenue a un impact réel sur la « sanction académique ».

6. Etats des flux dans le spectre des compétences

On considère habituellement une réponse fournie à une question de façon binaire : soit elle est correcte, soit elle est incorrecte, sans se préoccuper de nuances liées à la conviction avec laquelle l'étudiant a répondu. Des pionniers tels que DE FINETTI (1965), VAN NAERSEN (1965), SHUFORD & al. (1966), LECLERCQ, (1975) suivis par de nombreux chercheurs (BRUNO, 1993; DIRKZWAGER, 1993; FABRE, 1993; GILLES, 1995, HUNT, 1993; JANS, 1995; PLUNUS, 1996; VAN LENTHE, 1993) recommandent de traiter différemment l'ignorance reconnue (réponse incorrecte et peu sûre), la connaissance incomplète (réponse correcte mais peu sûre), la connaissance assurée (réponse correcte et très sûre), l'ignorance ignorée (réponse incorrecte et très sûre).

En ce qui concerne les étudiants qui ont été ajourné en 1ère session, nous observons des profils différents déjà en 1ère session chez ceux qui réussiront en seconde session en comparaison avec ceux qui échoueront en bout de course. Ce sont là, nous semble-t-il, des données importantes qui devront être confrontées à de nouvelles études envisageant les résultats d'autres examens au travers du spectre des compétences. Si ces profils se confirmaient, les actions préventives de lutte contre l'échec en candidature pourraient être complétées par une mise en garde systématique ainsi que des remédiations plus ciblées sur le profil des étudiants qui présentent un « spectre d'échec » en seconde session.

b) Positionner les résultats de la recherche dans un modèle pragmatique de gestion des examens

Nous espérons pouvoir positionner les apports de cette recherche, du moins une partie, dans un modèle de gestion globale des examens (GILLES et LECLERCQ, 1995). Nous pensons qu'il s'agit là d'une démarche qui permet, d'une part, d'intégrer dans une vue d'ensemble des apports ponctuels liés à des recherches avancées, et, d'autre part, de donner plus de cohérence à une série d'actions qui visent à améliorer la qualité des processus d'évaluation sur le terrain.

exemple, pour un même objectif, l'énergie et la forme de la remédiation consacrée à la correction d'une ignorance ignorée devrait être différente de celle qui permettra d'améliorer un état de connaissance incomplète. De plus, s'il se confirme que des profils à risque peuvent être dégagés, il faudra en tenir compte dans l'élaboration de procédures préventives de lutte contre l'échec en candidature à l'université.

* * *

-
- Liste 2 : les questions ayant reçu une réponse « mal informée », aucun (ou presque) étudiant n'a choisi la réponse correcte
Liste 3 : les questions ayant reçu une réponse « non informée », la majorité des étudiants avoue ne pas savoir
Liste 4 : les questions ayant reçu une réponse « incertaine », la majorité des étudiants fournit la réponse correcte mais avec un degré de certitude faible

Jean-Luc GILLES, Profils d'autoévaluation et d'états de connaissance partielle chez les étudiants de première candidature (1995-1996) de la Faculté de Psychologie et des Sciences de l'Éducation de l'Université de Liège, Colloque de l'Association Internationale de Pédagogie Universitaire (AIPU), Tunis, 4, 5 et 6 novembre 1996.

8. Bibliographie

- ATTNEAVE, F. (1959). Application of information theory to psychology. New York: Holt, Rinehart and Winston.
- BOXUS & AL. (1991). Principes communs pour évaluer les résultats cognitifs de la formation. Commissions des Communautés européennes, programme Eurotecnec.
- BRUNO, J. (1993). Using testing to provide feedback to support instruction: a reexamination of the role of assessment in educational organizations. NATO ASI Series, Item Banking: Interactive Testing and Self Assessment, Berlin: Springer Verlag, 1993, Vol. 112, pp. 190-209.
- DE LANDSHEERE, G., (1979). Dictionnaire de l'évaluation et de la recherche en éducation, Presses universitaire de France, Paris.
- DE FINETTI, B. (1965). Methods for discriminating levels of partial knowledge concerning a test item, British Journal of Mathematical and Statistical Psychology 18, pp. 87-123.
- DIRKZWAGER, A. (1993). A computer environment to develop valid and realistic predictions and self-assessment of knowledge with personal probabilities. NATO ASI Series, Item Banking: Interactive Testing and Self Assessment, Berlin: Springer Verlag, 1993, Vol. 112, pp. 146-166.
- FABRE, J.-M. (1993). Subjective uncertainty and the structure of the set of all possible events. NATO ASI Series, Item Banking: Interactive Testing and Self Assessment, Berlin: Springer Verlag, 1993, Vol. 112, pp. 99-113.
- GILLES, J.-L. (1995a). Gender comparison of metacognition : realism in self-estimation in cognitive competency with university students, à paraître.
- GILLES, J.-L. (1995b). Entraînement à l'autoévaluation : une comparaison filles/garçons à l'université. Actes du Colloque de l'AIPU « Enseignement supérieur : stratégies d'enseignement appropriées » - août 1995 - Université du Québec à Hull.
- GILLES, J.-L. (1996). Utilisation des degrés de certitude et normes de réalisme en situation d'examen et d'auto-estimation à la Faculté de Psychologie et des Sciences de l'Éducation de l'Université de Liège, Actes du Colloque de l'ADMEE-EUROPE « Dix années de travaux de en évaluation » - septembre 1996 - Université Pierre Mendès France à Grenoble, à paraître.
- GILLES, J.-L. et LECLERCQ, D., (1995). Procédures d'évaluation adaptées à des grands groupes d'étudiants universitaires - Enjeux et solutions pratiquées à la FAPSE-ULG, Actes du Symposium International sur la Rénovation Didactique en Biologie - novembre 1995 - Université de Tunis, à paraître.
- HUNT, D. (1993). Theory and application to learning and testing. NATO ASI Series, Item Banking: Interactive Testing and Self Assessment, Berlin: Springer Verlag, 1993, Vol. 112, pp. 177-189.
- JANS, V. et LECLERCQ, D. (1996). Forum : un système d'animation et d'évaluation de grands groupes universitaires, communication au colloque de l'Association Internationale de pédagogie Universitaire (AIPU), Tunis, novembre 1996, à paraître.
- LECLERCQ, D. (1975). L'évaluation subjective de la probabilité d'exactitude des réponses en situation pédagogique. Thèse de doctorat en Sciences de l'Éducation, Université de Liège Institut de Psychologie et des Sciences de l'Éducation.
- LECLERCQ, D. (1983). Confidence marking, its use in testing. Postlethwaite, Choppin (eds.) Evaluation in Education, Oxford : Pergamon, 1982, vol. 6, 2, pp. 161-287.
- LECLERCQ, D. (1986). La conception des questions à choix multiple, Bruxelles, Ed. Labor.
- LECLERCQ, D. & al (1993). The Taste approach: General implicit solutions in MCQq, open books exams and interactive testing and sel-assessment. NATO ASI Series, Item Banking: Interactive Testing and Self Assessment, Berlin: Springer Verlag, 1993, Vol. 112, pp. 210-232.
- LECLERCQ, D. & GILLES J.-L. (1994). GUESS, un logiciel pour entraîner à l'auto-estimation de sa compétence cognitive. Actes du colloque QCM et questionnaires fermés, Paris: ESIEE, 1994.
- LECLERCQ, D. & PLUNUS, G., (1996). Double Check, étude de méthodes d'innovation d'évaluations universitaires valides, formatives et efficaces, Rapport recherche-action, Université de Liège, Service de Technologie de l'Éducation, à paraître.
- LICHTENSTEIN, & al. (1975). Calibration of probabilities : the state of the art, decision making and change in human affairs Proceedings of the Fifth Research Conference on Subjective Probability, Utility and Decision Making, Darmstadt, 1-4 September, D. Reidel.
- PLUNUS, G. (1996). Pistes diagnostiques et procédures pour une évaluation universitaire valide, formative et efficace, Mémoire présenté pour l'obtention du grade de licencié en Sciences de l'Éducation, Faculté de Psychologie et des Sciences de l'Éducation, Université de Liège.
- SHANNON, C.E. (1951). Prediction and entropy of printed english. Bell Syst. Techn. J. 30, pp. 50-64.
- SHUFORD, E. & al (1966). Admissible probability measurement procedures. Psychometrika 31, pp. 125-145.
- VAN LENTHE, J. (1993). The development and evaluation of ELI, an interactive elicitation technique for subjective probability distributions. NATO ASI Series, Item Banking: Interactive Testing and Self Assessment, Berlin: Springer Verlag, 1993, Vol. 112, pp.
- VAN NAERSEN, R.F. & al (1966). Is de utiliteitscurve van examenscores een ogief ? Nederland Tijdschrift Psychologie 21(6), pp. 358-363.

Jean-Luc GILLES, Profils d'autoévaluation et d'états de connaissance partielle chez les étudiants de première candidature (1995-1996) de la Faculté de Psychologie et des Sciences de l'Éducation de l'Université de Liège, Colloque de l'Association Internationale de Pédagogie Universitaire (AIPU), Tunis, 4, 5 et 6 novembre 1996.