

# Impact de deux entraînements à l'utilisation des degrés de certitude chez les étudiants de 1<sup>ère</sup> candidature à la Faculté de Psychologie et des Sciences de l'Education de l'ULG

Jean-Luc GILLES<sup>1</sup>

Centre d'AutoFormation et d'Evaluation Interactive Multimédias de la Faculté de Psychologie et des Sciences de l'Education (CAFEIM-FAPSE) - Université de Liège (ULG) – Belgique<sup>2</sup>

Tél. : +32-(0)4-366.20.78 Fax : +32-(0)4-366.29.53 Email : jl.gilles@ulg.ac.be

## 1. Introduction

Depuis l'année académique 1994-1995, le Centre d'AutoFormation et d'Evaluation Interactives Multimédias de la Faculté de Psychologie et des Sciences de l'Education (CAFEIM-FAPSE) de l'Université de Liège propose aux enseignants de la FAPSE des Services Méthodologiques d'Aide à la Réalisation de Tests (SMART). Les procédures d'évaluation proposées permettent soit un questionnement informatisé où les étudiants encodent leurs réponses à l'ordinateur à l'aide d'un logiciel de testing interactif intitulé WINCHECK<sup>3</sup>, soit un questionnement ayant recours à la Lecture Optique de Marques (LOM) où les étudiants cochent leurs réponses en auditoire sur des formuLOMs<sup>4</sup>.

Dès le 1er semestre, dans le cadre du *partim I* « *Approche technologique de l'éducation* » du cours *Introduction aux sciences de l'éducation*, les étudiants de 1<sup>ère</sup> candidature sont systématiquement entraînés à l'emploi du logiciel WINCHECK, au maniement des formuLOMs et à l'utilisation des degrés de certitude.

## 2. Conditions d'utilisation de la procédure d'auto-évaluation de ses compétences à l'aide des degrés de certitude

### A. Une procédure intégrée dans un dispositif plus large de Services Méthodologiques d'Aide à la Réalisation de Tests (SM@RT)

Le modèle de gestion des examens dans l'enseignement supérieur de GILLES et LECLERCQ (1995) présenté ci-après décrit le processus cyclique de réalisation des évaluations telles qu'elles se pratiquent à la FAPSE-ULG lorsque les enseignants font appel aux Services Méthodologiques d'Aide à la Réalisation de Tests (SM@RT) proposés par le CAFEIM.

Une série de recommandations en vue d'augmenter la validité et la fidélité des examens sont présentées dans les cadres rectangulaires ombrés qui entourent le schéma. Les recommandations soulignées en continu sont mises en place de façon systématique dans nos évaluations. Celles qui sont soulignées en pointillés sont, à notre avis, encore trop peu suivies par les enseignants, certaines ne le sont pas du tout et ne sont pas soulignées dans les cadres.

Le schéma montre aussi la place qu'occupe l'entraînement aux procédures d'évaluation dans le cycle de réalisation des examens.

---

<sup>1</sup> L'auteur tient à remercier :

- le Professeur Dieudonné LECLERCQ pour sa précieuse relecture, ses encouragements, ses judicieux conseils et ses fécondes précisions terminologiques;
- Myriam DPONT, Assistante au Service de Mathématiques Appliquées de la FAPSE, pour les éclaircissements apportés au niveau du traitement des données initiales.

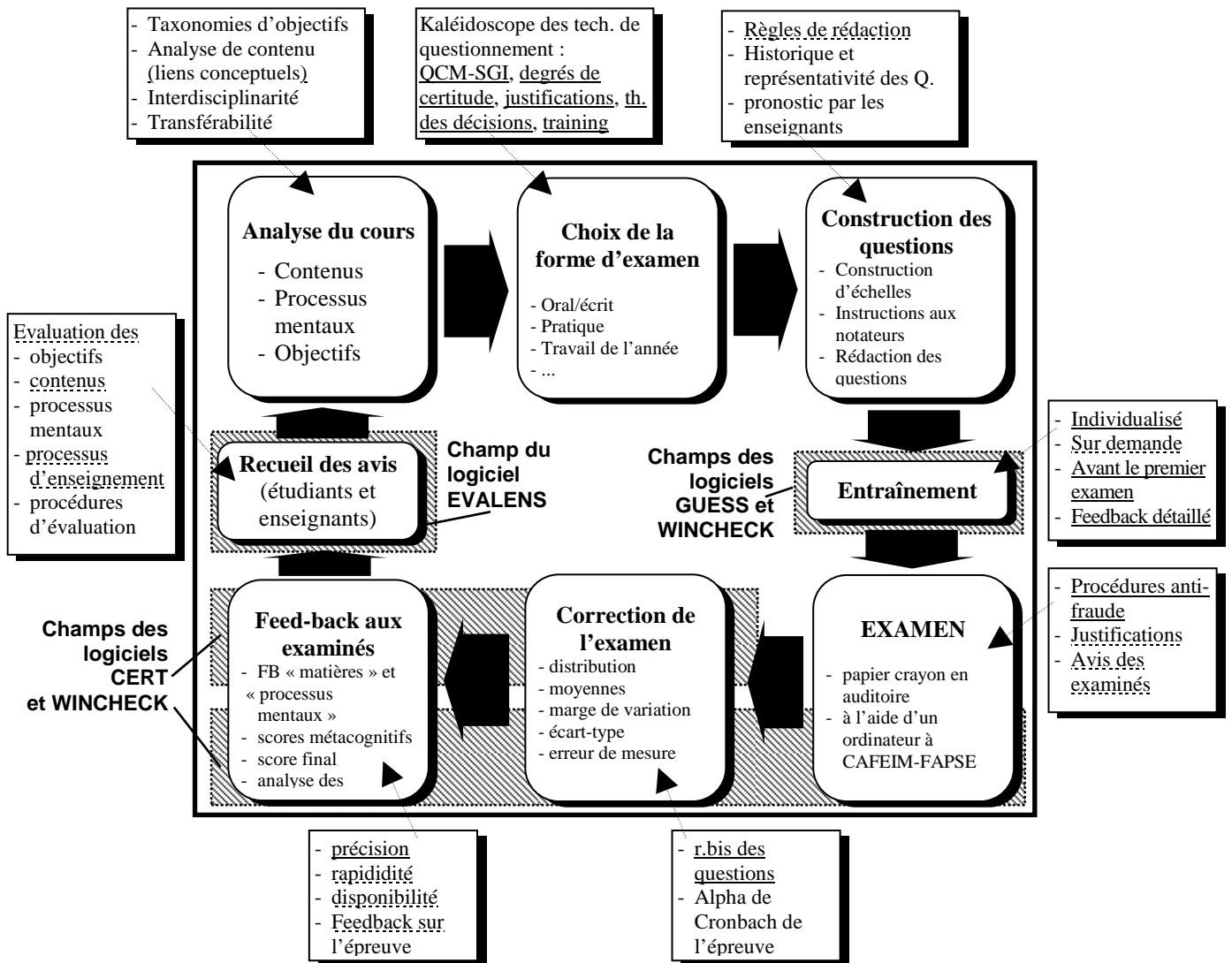
<sup>2</sup> Le C.A.F.E.I.M. - FA.P.S.E.- ULG (Président : Prof. D. LECLERCQ, directeur : J.-L. GILLES) s'est fixé pour objectifs :

- 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.

<sup>3</sup> WINCHECK a été développé par le Service de Technologie de l'Education de l'Université de Liège dans le cadre du projet EUROFORM CERT.

<sup>4</sup> Formulaire destinés à la lecture optique de marques

Soulignons également que l'entraînement dont il est question dans cette étude concerne l'utilisation des degrés de certitude et qu'il existe d'autres modalités pratiques relatives aux examens auxquelles nous entraînonons aussi les étudiants, par exemple l'utilisation des feuilles de lecture optique de marques.



## B. Garanties pour un recueil sans biais de l'auto-estimation

Une série de conditions méthodologiques doivent être rencontrées pour garantir un recueil sans biais des données liées à l'auto-estimation de ses compétences. Le non respect des règles ci-après explique l'échec de nombreuses études menées aux Etats-Unis entre 1960 et 1975. Le lecteur trouvera des informations plus complètes à ce propos dans LECLERCQ (1993, pp. 141-143) et BOXUS & al. (1991, pp. 56-70).

### a) Le barème de tarifs doit être conforme à la théorie des décisions

Il s'agit de gratifier une réponse correcte accompagnée d'un degré de certitude élevé d'un meilleur score que si elle était accompagnée d'une certitude faible et inversement pour les réponses incorrectes. Les tarifs du barème des points doivent être calculés de manière à favoriser une seule stratégie : celle qui consiste à dire la vérité. Le barème des points ci-après garantit que l'expression de son intime conviction rapporte plus de points que tout autre stratégie.

### b) La consigne doit être "probabiliste".

Demander à l'étudiant d'indiquer sa certitude par des termes vagues du type "peu sûr", "moyennement sûr", "très sûr", etc. est à proscrire car ces expressions recouvrent des réalités différentes en fonction des sujets. De plus,

Jean-Luc Gilles (1997), Impact de deux entraînements à l'utilisation des degrés de certitude chez les étudiants de 1<sup>ère</sup> candidature à la faculté de psychologie et des sciences de l'éducation de l'ULg, in Boxus, Jans, Gilles et Leclercq (Eds), *Stratégies et médias pédagogiques pour l'apprentissage et l'évaluation dans l'enseignement supérieur*, Actes du XV<sup>ème</sup> Colloque de l'AIPU, Liège, 1997, pp. 311-326.

avec des termes aussi flous la variabilité est telle chez un même étudiant qu'on ne peut même pas recourir à des traitements ordinaux intra-sujets. Voici la consigne mise au point par LECLERCQ (1983, 1993) et qui est utilisée à la FA.P.S.E - ULG.

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

### **c) Des zones correspondant à la précision humaine**

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. 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 »...

### **d) Le calcul d'indices métacognitifs doit être possible**

La consigne utilisée autorise le calcul d'un indice de réalisme basé sur les différences entre les taux d'exactitude et les valeurs centrales des intervalles de probabilité ainsi que le calcul d'un indice de centration basé sur la différence entre la certitude moyenne et le taux d'exactitude moyen et dont le signe détermine la surestimation (+) ou la sous-estimation (-). Les procédures de calcul associées à ces indices métacognitifs sont reprises au point 7 ci-après.

## **3. Le dispositif d'entraînement**

Dans leur grande majorité, les étudiants du premier cycle de la FAPSE-ULG n'ont jamais rencontré au cours de leur cursus scolaire de procédure d'évaluation ayant recours aux degrés de certitude. Dès lors, il est indispensable de les y entraîner avant le premier examen.

Lors des mois d'octobre et de novembre 1996, ces étudiants de 1<sup>ère</sup> candidature ont été répartis en 3 groupes soumis chacun à deux modalités d'entraînement :

- 1) GUESS : axé sur l'emploi des degrés de certitude et effectué à l'aide d'un logiciel<sup>5</sup> qui permet de fournir en peu de temps de nombreuses réponses accompagnées de degrés de certitude;
- 2) WINCHECK : axé sur l'entraînement à l'utilisation du logiciel de testing interactif et la pratique des questions à choix multiple avec degrés de certitude.

Ces entraînements ont eu lieu au Centre d'AutoFormation et d'Evaluation Interactives Multimédias de la Faculté de Psychologie et des Sciences de l'Education (CAFEIM-FAPSE).

### **A. L'entraînement GUESS**

L'entraînement à l'auto-estimation de sa compétence cognitive à l'aide du logiciel GUESS 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 d'au moins cent lettres (inspiré de SHANNON, 1951 & 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 s'affiche ainsi à l'écran ainsi qu'un graphique de réalisme.

<sup>5</sup> GUESS a été conçu par le Prof. D. LECLERCQ et programmé dans le langage ToolBook par M. HURARD, licencié en informatique. Il est possible d'obtenir des licences d'utilisation du logiciel GUESS à des conditions « éducation » en s'adressant au Service de Technologie de l'Education, tél. +32-41-66.20.72, fax +32-41-66.29.53.

## B. L'entraînement WINCHECK

WINCHECK est un logiciel d'évaluation interactive fonctionnant dans l'environnement Windows. Les questions proposées aux étudiants sont à choix multiple avec Solutions Générales Implicites<sup>6</sup> (QCM-SGI). Le système permet également l'utilisation du « double check<sup>7</sup> ». Voici l'écran d'une question posée à l'aide de WINCHECK :

Ces deux types d'entraînements, ont chacun leurs avantages et leurs inconvénients. Grâce au logiciel GUESS, beaucoup de réponses accompagnées de certitudes sont fournies en peu de temps et la visualisation d'un feed-back immédiat est possible. Cependant, l'exercice proposé est assez éloigné de la situation d'examen qui, elle, comporte des QCM à 9 solutions maximum. L'entraînement WINCHECK, lui, permet de poser des questions qui du point de vue formel ressemblent fort à celles qui sont posées à l'examen. Par contre, cet entraînement ne permet pas en une séance de fournir beaucoup de réponses accompagnées de degrés de certitudes (une vingtaine).

## 4. Le dispositif d'observation des performances d'auto-estimation des étudiants

### A. Les Quizz

Les étudiants de la faculté de Psychologie et des Sciences de l'Education ont aussi la possibilité de s'entraîner à l'utilisation des degrés de certitude lors de quizz organisés en fin de cours où ils sont invités à répondre à quelques questions (entre 5 et 10). Ils répondent sur des feuilles spéciales fournies par le C.A.F.E.I.M.-FA.P.S.E. sur lesquelles ils cochent les réponses qu'ils accompagnent d'un degré de certitude (comme à l'examen). Après le ramassage des feuilles de réponses, les réponses correctes sont communiquées. Les feuilles sont traitées dans la semaine à l'aide du dispositif de lecture optique de marques. Au cours suivant, le professeur commente les statistiques de résultats de l'ensemble du groupe, commente les erreurs les plus fréquentes et réexplique éventuellement la matière mal comprise. Enfin, chaque étudiant reçoit une feuille de feed-back personnalisée avec entre autres informations : les réponses et les certitudes qu'il a fournies, les réponses correctes attendues, son graphique de réalisme, les indices de son réalisme et de sa centration. Cette procédure et les résultats ont été décrits par GILLES & LECLERCQ (1996) et GILLES (1995).

Dans le cadre de cette étude, ce sont les résultats des 3 premiers quizz du *partim I* « Approche technologique de l'éducation » du cours *Introduction aux sciences de l'éducation*, qui ont été traités. Les entraînements GUESS et WINCHECK ont eu lieu entre le 2<sup>ème</sup> et le troisième quizz.

QUIZZ N°1 (01/10/96) N = 293			
QUIZZ N°2 (08/10/96) N = 275			
ENTRAINEMENTS PHASE I (du 15/10 au 21/10/96)	GUESS	WINCHECK	TEMOIN
		Etudiants « A-E » N=39	Etudiants « F-M » N=56
QUIZZ N°3 (22/10/96) N = 289			

Afin que tous les étudiants puissent bénéficier des deux entraînements proposés en phase I, deux autres phases d'entraînements ont été proposées : du 22/10 au 28/10/96 et du 29/10 au 04/11/96.

Les 3 quizz comportaient chacun 10 questions à choix multiple en rapport avec le contenu du cours qui venait d'être donné par le professeur D. LECLERCQ.

### B. La population étudiée

Sur les 250 étudiants qui ont effectué les 3 quizz, 95 ont participé à la 1<sup>ère</sup> phase d'entraînement : 39 pour l'entraînement GUESS et 56 pour l'entraînement WINCHECK.

<sup>6</sup> 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 !).

<sup>7</sup> Il s'agit d'une procédure d'évaluation interactive mise au point par D. LECLERCQ (1993) et 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 la réponse puis la deuxième partie de la question (*bis*), par exemple : « quelle donnée manque ? ». Suivent à nouveau une série de propositions. Les performances des étudiants se présentent alors selon différents cas de figure qui peuvent ensuite donner lieu à des procédures de remédiation adaptées selon le diagnostic.

<sup>8</sup> Il s'agit des étudiants qui lors de la seconde phase d'entraînement, à partir du 29/10/96, ont effectué l'entraînement GUESS.

- Les étudiants ont été répartis en 3 groupes en fonction de la 1<sup>ère</sup> lettre du nom de famille :
- de A à E entraînement GUESS ;
  - de F à M entraînement WINCHECK ;
  - de N à Z, pas d'entraînement (groupe témoin).

En ce qui concerne le groupe témoin, nous avons pris en compte les 36 étudiants qui à la phase suivante allaient participer à l'entraînement qui leur serait proposé (GUESS).

Sur les 135 étudiants<sup>9</sup> des 3 groupes considérés, 9 étudiants s'étaient inscrits avant 1996. Nous avons préféré extraire ces sujets de notre échantillon car ils étaient probablement déjà sensibilisés à la procédure d'évaluation.

Enfin, nous avons extrait au hasard 1 étudiant du groupe GUESS et 17 étudiants du groupe WINCHECK en vue d'effectuer nos comparaisons sur 3 groupes de 36 sujets.

## 5. Mesure des performances d'auto-estimation

### A. Le réalisme (R)

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 utilisé la formule des écarts des taux d'exactitude aux valeurs centrales des probabilités subjectives de réussite, adaptée par GILLES (1996) afin que le minimum de réalisme soit égal à 0 et le maximum égal à 1:

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

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

$\text{TE}_i$  = Taux d'exactitude de la certitude  $i$ ,  $\text{VC}_i$  = Valeur centrale,  $\text{NU}_i$  = Nombre d'utilisations,  $\text{NR}$  = Nombre total de réponses et  $i$  = indice (ici de 0 à 5) des degrés de certitude.

$\beta$  = correction pour l'erreur minimale, dans notre cas  $\beta = 0,025$ .

$\alpha$  = correction d'étendue, dans notre cas, pour étendre la plage<sup>10</sup> de 0,95 à 1,  $\alpha = 1,0526$

#### 1) Mesure des Gains Bruts de Réalisme (GBR)

Il s'agit d'évaluer l'effet des entraînements par le calcul de la différence entre les scores de réalisme obtenus au 2<sup>ème</sup> quizz et au 3<sup>ème</sup> quizz.

$$\text{GBR} = \text{Post-test} - \text{Prétest}$$

#### 2) Mesure des Gains Relatifs (GRR) et Pertes Relatives de Réalisme (PRR)

On établit le rapport entre la progression en réalisme d'un étudiant et la progression maximale qui aurait été possible pour cet étudiant. C'est la formule proposée par McGUIGAN (1967) qui sera utilisée :

$$\text{GRR} = \frac{\text{Post - test} - \text{Prétest}}{1 - \text{Prétest}} \times 100$$

D'HAINAUT (1973) conseille en cas de perte, de prendre en compte le rapport de ce qui a été perdu avec ce qu'il était possible de perdre. La formule devient dès lors :

$$\text{PRR} = \frac{\text{Post - test} - \text{Prétest}}{\text{Prétest}} \times 100$$

<sup>9</sup> Le grand nombre d'étudiants absents peut s'expliquer en partie par le fait que les redoublants sont déjà sensibilisés à la procédure d'évaluation.

## B. La centration

### 1) Mesure de l'Erreur de Centration (EC)

Le calcul de l'indice de centration est basé sur la différence entre la certitude moyenne (CM) et le taux d'exactitude (TE). Les scores des Erreurs de Centration (EC) montrent en fonction du signe la tendance à se surestimer (+) ou à se sous-estimer (-). L'étendue des valeurs possibles va de -100 à +100. Une performance idéale vaut 0.

$$EC = CM - TE$$

### 2) Mesure de l'Erreur Absolue de Centration (EAC)

Le signe des scores des Erreurs Absolues de Centration (EAC) n'étant pas pris en compte, c'est la « distance absolue » qui sépare les performances mesurées de la performance idéale (EAC = 0) qui est ici envisagée.

$$EAC = | EC |$$

### 3) Mesure des Gains Absolus (GA) : diminutions de l'Erreur Absolue de Centration

La formule utilisée ici correspond à la valeur absolue de l'erreur de centration mesurée au prétest à laquelle on soustrait l'erreur absolue de centration obtenue au post-test. Cette façon de procéder fonctionne à l'inverse de la mesure du gain brut habituelle et s'explique par le fait que la valeur maximum dans le cas du calcul de l'erreur de centration vaut « 0 » (donc il y a un gain lorsqu'il y a diminution).

$$GA = EAC \text{ Prétest} - EAC \text{ Post - test}$$

### 4) Mesure des Gains et Pertes Relatifs/Relatives (GR ou PR) : diminution/augmentation de l'Erreur Absolue de Centration par rapport au gain possible

Lorsqu'on envisage la mesure des gains ou des pertes<sup>11</sup> par rapport au gain ou à la perte possible (donc de façon relative), le traitement des gains est à différencier de celui des pertes.

En cas de gain le formule est :

$$GR = \downarrow EAC = \frac{EAC \text{ Prétest} - EAC \text{ Post - test}}{EAC \text{ Prétest}} \times 100$$

En cas de perte elle devient :

$$PR = \uparrow EAC = \frac{EAC \text{ Prétest} - EAC \text{ Post - test}}{100 - EAC \text{ Prétest}} \times 100$$

Remarquons que PR est négatif.

---

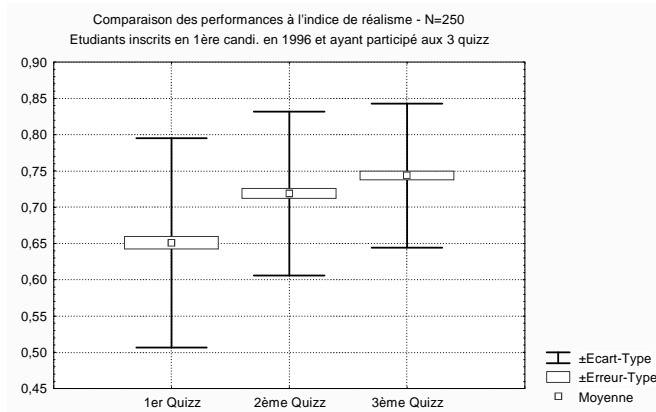
<sup>10</sup> Dans la procédure d'utilisation des degrés de certitude définie par LECLERCQ, ERR MAX = 0,975 pour un sujet qui se trompe systématiquement avec le degré de certitude 5, et, ERR MIN = 0,025 pour un sujet qui répond correctement à toutes les questions avec un degré de certitude 5, ceci détermine une plage de 0,95 (0,975 - 0,025).

<sup>11</sup> Quand la valeur du Gain Absolu (GA = EAC Prétest - EAC Post-test) est négative.

## 6. Progression des performances d'auto-estimation de l'ensemble des étudiants ayant participé aux 3 quizz (N=250)

A titre de comparaison avec les performances des 3 groupes de 36 étudiants chacun présentés plus loin voici les scores obtenus par l'ensemble des étudiants qui ont participé aux 3 quizz (N=250).

### A. Indice de réalisme (R) :



t de Student pour échantillons appariés

dl : 249	2 <sup>ème</sup> quizz	3 <sup>ème</sup> quizz
1 <sup>er</sup> quizz	<b>5,9 : TS</b> <b>P : .00000</b>	<b>8,8 : TS</b> <b>P : .00000</b>
2 <sup>ème</sup> quizz		<b>2,6 : TS</b> <b>P : .009</b>

N=250	1 <sup>er</sup> Quizz		2 <sup>ème</sup> Quizz		3 <sup>ème</sup> Quizz	
	Moyennes	Ec.-Type	Moyennes	Ec.-Type	Moyennes	Ec.-Type
Réalisme	0,650	0,144	0,719	0,113	0,744	0,099
Taux d'exactitude	61,8	16,2	61,9	15,8	69,9	9,5

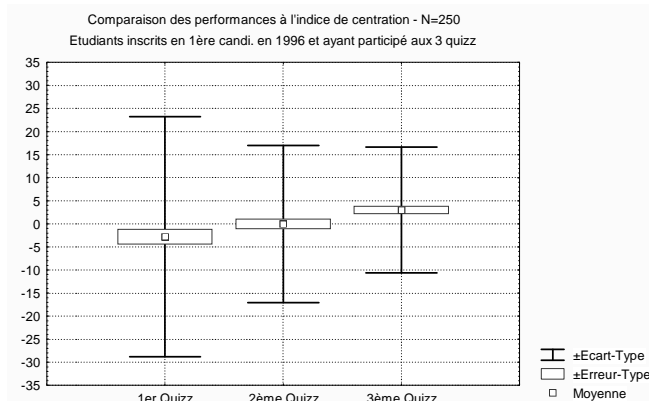
Le réalisme s'améliore au fil des épreuves. La moyenne des étudiants passe de 0,65 au 1<sup>er</sup> quizz à environ 0,75 au 3<sup>ème</sup> quizz. La moyenne relativement basse récoltée lors du 1<sup>er</sup> quizz peut s'expliquer par un relatif « effet de surprise », les étudiants n'ayant été exposé qu'à une seule utilisation des degrés de certitude lors d'une séance d'accueil (il s'agissait d'un quizz avec degrés de certitude organisé dans le but de rendre plus attrayante les explications aux nouveaux étudiants de la faculté). Remarquons que l'écart-type se resserre autour de la moyenne au fil des quizz. Les différences observées sont statistiquement très significatives. Ces données confirment les résultats obtenus par LECLERCQ (1993) et GILLES (1996).

### B. Indices de centration :

#### 1) Erreur de Centration (EC)

Basé sur la différence entre la certitude moyenne (CM) et le taux d'exactitude (TE), l'Erreur de Centration (EC) vaut idéalement 0. Une valeur positive indique une tendance à la surestimation et une valeur négative une sous-estimation.

EC	1 <sup>er</sup> Quizz		2 <sup>ème</sup> Quizz		3 <sup>ème</sup> Quizz	
	Moyennes	Ec.-Type	Moyennes	Ec.-Type	Moyennes	Ec.-Type
	-2,8	26	-0,03	17,0	3,0	13,6



t de Student pour échantillons appariés

dl : 249	2 <sup>ème</sup> quizz	3 <sup>ème</sup> quizz
1 <sup>er</sup> quizz	1,44 : NS P : .15	<b>3,31 : TS</b> <b>P : .001</b>
2 <sup>ème</sup> quizz		<b>2,31 : S</b> <b>P : .021</b>

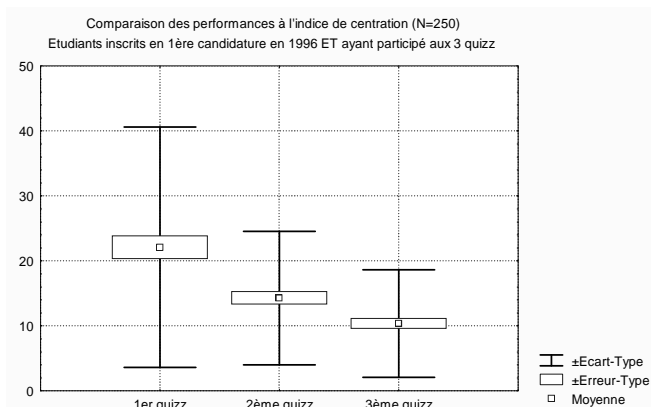
Jean-Luc Gilles (1997), Impact de deux entraînements à l'utilisation des degrés de certitude chez les étudiants de 1<sup>ère</sup> candidature à la faculté de psychologie et des sciences de l'éducation de l'ULg, in Boxus, Jans, Gilles et Leclercq (Eds), *Stratégies et médias pédagogiques pour l'apprentissage et l'évaluation dans l'enseignement supérieur*, Actes du XV<sup>ème</sup> Colloque de l'AIPU, Liège, 1997, pp. 311-326.

En ce qui concerne la tendance à la sur- ou à la sous-estimation, le graphique ci-dessus permet d'observer d'abord une sous-estimation, suivie ensuite d'une calibration quasi parfaite au 2<sup>ème</sup> quizz (-0,03), et finalement, une sur-estimation<sup>12</sup> au 3<sup>ème</sup> quizz. A noter, l'écart-type qui se resserre au fil des quizz. Les différences sont statistiquement significatives, sauf pour les 1<sup>er</sup> et 2<sup>ème</sup> quizz.

## 2) Erreur Absolue de Centration (EAC)

Rappelons que la performance idéale à l'indice de centration vaut « 0 ».

N=250 EAC	1 <sup>er</sup> Quizz		2 <sup>ème</sup> Quizz		3 <sup>ème</sup> Quizz	
	Moyennes	Ec.-Type	Moyennes	Ec.-Type	Moyennes	Ec.-Type
	20,1	16,7	13,5	10,3	11,4	7,9



t de Student pour échantillons appariés		
dl : 249	2 <sup>ème</sup> quizz	3 <sup>ème</sup> quizz
1 <sup>er</sup> quizz	<b>5,46 : TS</b> <b>P : .00000</b>	<b>7,23 : TS</b> <b>P : .00000</b>
2 <sup>ème</sup> quizz		<b>2,46 : S</b> <b>P : .014</b>

Les moyennes des erreurs de centration calculées en valeurs absolues montrent une amélioration des performances au fil des quizz; les valeurs moyennes se rapprochent du 0 idéal. Toutes les différences sont statistiquement significatives.

<sup>12</sup> La tendance à la surestimation a souvent été soulignée dans la littérature (PITZ, 1974 ; BRUNO, 1993 ; FABRE, 1993 ; GATHY et DENEFF, 1993, GILLES, 1996).

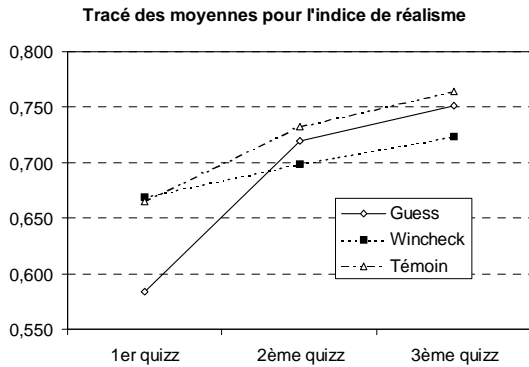
Jean-Luc Gilles (1997), Impact de deux entraînements à l'utilisation des degrés de certitude chez les étudiants de 1<sup>ère</sup> candidature à la faculté de psychologie et des sciences de l'éducation de l'ULg, in Boxus, Jans, Gilles et Leclercq (Eds), *Stratégies et médias pédagogiques pour l'apprentissage et l'évaluation dans l'enseignement supérieur*, Actes du XV<sup>ème</sup> Colloque de l'AIPU, Liège, 1997, pp. 311-326.



## 7. Comparaison des progressions des 3 groupes : « GUESS », « WINCHECK » et « TEMOIN »

### A. Indice de réalisme (R)

	N	1 <sup>er</sup> QUIZZ		GB	2 <sup>ème</sup> QUIZZ		GB	GR	3 <sup>ème</sup> QUIZZ	
		Moyennes	Ec.-Type		Moyennes	Ec.-Type			Moyennes	Ec.-Type
GUESS	36	0,584	0,142	0,135	0,719	0,111	0,032	16,8	0,751	0,107
WINCH	36	0,669	0,124	0,030	0,699	0,111	0,024	14,3	0,723	0,108
TEMOIN	36	0,665	0,178	0,067	0,732	0,089	0,032	16,1	0,764	0,090
TsGrpes	108	0,640	0,153	0,077	0,717	0,104	0,029	15,8	0,746	0,102



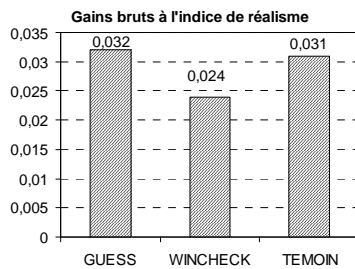
Comparaison des différences entre groupes au 3<sup>ème</sup> quizz (après entraînement) :

t de Student pour échantillons indépendants

dl 70	WINCHECK	TEMOIN
GUESS	1,07 : NS P : .28	0,55 : NS P : .58
WINCHECK		1,71 : NS P : .09

Les trois groupes améliorent leur réalisme, mais on ne peut affirmer qu'un groupe progresse plus qu'un autre au 3<sup>ème</sup> quizz, les différences étant statistiquement non significatives.

Les différences observées entre les gains bruts de réalisme ne sont pas significatives.

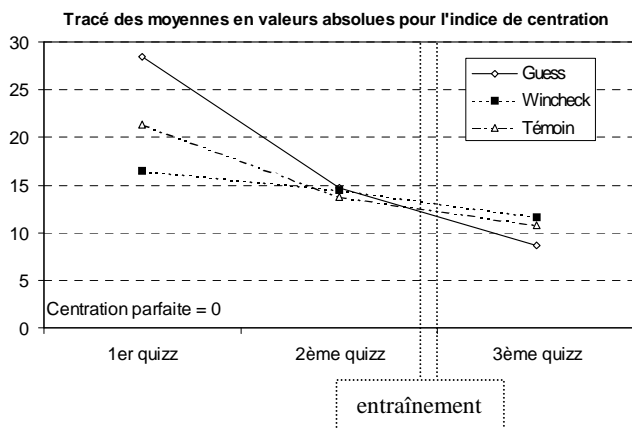


t de Student pour échantillons indépendants

Gains Bruts	WINCHECK	TEMOIN
GUESS	0,207 : NS P : .75	0,022 : NS P : .98
WINCHECK		0,213 : NS P : .83

## B. Erreurs Absolues de Centration (EAC)

	N	1 <sup>er</sup> QUIZZ				2 <sup>ème</sup> QUIZZ				3 <sup>ème</sup> QUIZZ	
		Moyennes	Ec.-Type	GB	GR	Moyennes	Ec.-Type	GB	GR	Moyennes	Ec.-Type
GUESS	36	28,5	19,6	13,8	51%	14,7	10,5	6	41%	8,7	8,4
WINCH	36	16,5	14	2	12%	14,5	10,4	2,9	20%	11,6	8,6
TEMOIN	36	21,3	19,8	7,6	35%	13,7	10,1	2,9	21%	10,8	7,1
TsGrpes	108	22,1	17,8	7,8	35%	14,3	10,3	3,9	27%	10,4	8,2



t de Student pour échantillons indépendants (dl : 70)

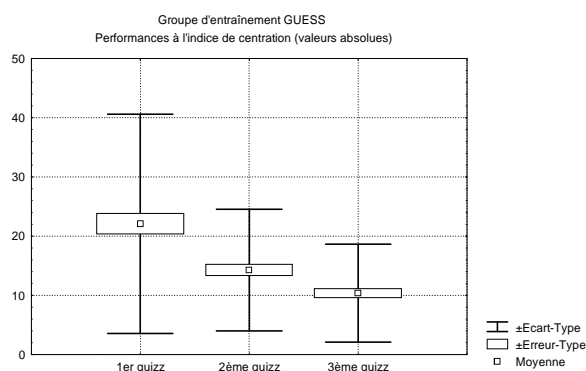
	2 <sup>ème</sup> quizz		3 <sup>ème</sup> quizz	
	WINCH.	TEMOIN	WINCH.	TEMOIN
GUESS	NS : 0,08 P : .93	NS : 0,39 P : .69	NS : 1,49 P : .14	NS : 1,15 P : .25
WINCH.		NS : 0,31 P : .76		NS : 0,41 P : .68

Les différences observées entre les moyennes des centrations des étudiants entraînés et le groupe témoin n'apparaissent pas statistiquement significatives pour le troisième quizz (après les deux entraînements). Cependant, on observe des gains relatifs deux fois plus élevés pour le groupe GUESS (41%) par rapport aux groupes WINCHECK (20%) et TEMOIN (21%).

## 8. Comparaison de l'évolution des Erreurs Absolues de Centration (EAC) au sein de chacun des 3 groupes

Au sein d'un même groupe d'étudiants, les moyennes des Erreurs Absolues de Centration (EAC) récoltées au 2<sup>ème</sup> et au 3<sup>ème</sup> quizz sont-elles significativement différentes ?

### A. Groupe d'entraînement GUESS (N = 36)



	Moyenne	Minimum	Maximum	Ec-Type
1 <sup>er</sup> Quizz	28,5	1,7	67,5	19,6
2 <sup>ème</sup> Quizz	14,6	0,25	39,0	10,5
3 <sup>ème</sup> Quizz	8,6	0,0	37,0	8,4

t de Student pour échantillons appariés

dl : 35	2 <sup>ème</sup> quizz	3 <sup>ème</sup> quizz
1 <sup>er</sup> quizz	<b>3,98 : TS</b> <b>P : .0003</b>	<b>5,53 : TS</b> <b>P : .000003</b>
2 <sup>ème</sup> quizz		<b>2,62 : S</b> <b>P : .012</b>

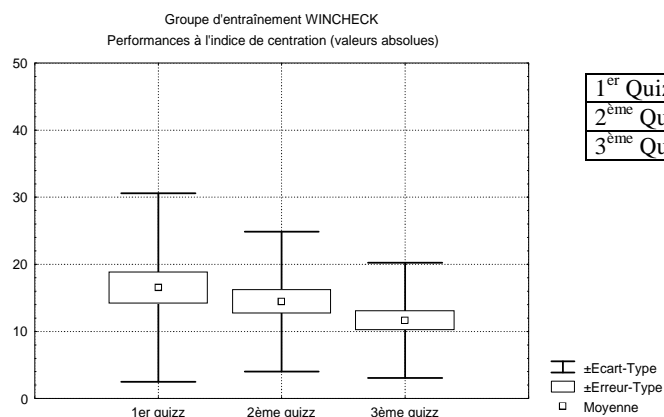
Les moyennes des Erreurs Absolues de Centration (EAC) obtenues par le groupe GUESS se rapprochent de la valeur idéale 0. D'un point de vue statistique les moyennes sont significativement différentes lorsqu'on compare le 2<sup>ème</sup> et le 3<sup>ème</sup> quizz (avant et après entraînement).

Nous observons également une tendance à la réduction de l'écart-type au fil des quizz. Le rapport des variances (test F) des quizz pris 2 par 2, montre des différences très significatives entre le 2<sup>ème</sup> et le 3<sup>ème</sup> quizz (avant et après entraînement).

Test F de Snédecor

F - dl 35	2 <sup>ème</sup> QUIZZ	3 <sup>ème</sup> QUIZZ
1 <sup>er</sup> QUIZZ	<b>3,81 - TS</b>	<b>10,9 - TS</b>
2 <sup>ème</sup> QUIZZ		<b>2,88 - TS</b>

## B. Groupe d'entraînement WINCHECK (N = 36)



	Moyenne	Minimum	Maximum	Ec-Type
1 <sup>er</sup> Quizz	16,5	0,0	59,0	14,0
2 <sup>ème</sup> Quizz	14,4	1,0	45,0	10,4
3 <sup>ème</sup> Quizz	11,6	0,75	33,5	8,6

t de Student pour échantillons appariés

dl 35	2 <sup>ème</sup> quizz	3 <sup>ème</sup> quizz
1 <sup>er</sup> quizz	0,87 : NS P : .39	1,69 : NS P : .098
2 <sup>ème</sup> quizz		1,37 : NS P : .177

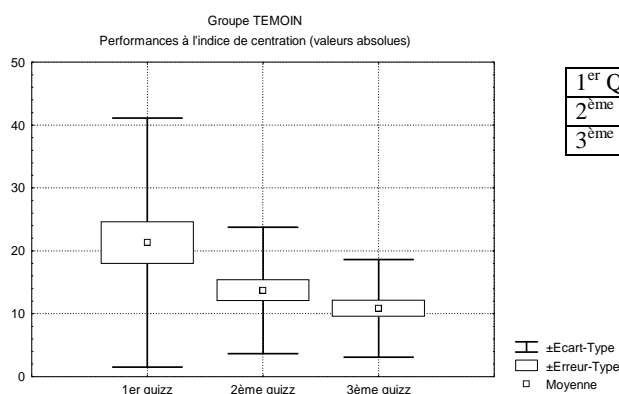
Les moyennes des Erreurs Absolues de Centration du groupe WINCHECK se rapprochent de la valeur idéale 0 au 2<sup>ème</sup> et au 3<sup>ème</sup> quizz, mais d'un point de vue statistique les moyennes ne sont pas différentes.

On observe également une réduction de l'écart-type au fil des quizz, mais le rapport des variances (test F) des quizz pris 2 par 2, montre une différence statistiquement non significative lorsqu'on compare le 2<sup>ème</sup> et le 3<sup>ème</sup> quizz.

Test F de Snédecor

F - dl 35	2 <sup>ème</sup> QUIZZ	3 <sup>ème</sup> QUIZZ
1 <sup>er</sup> QUIZZ	<b>1,81 - TS</b>	<b>2,65 - TS</b>
2 <sup>ème</sup> QUIZZ		1,46 - NS

## C. Groupe TEMOIN (N = 36)



	Moyenne	Minimum	Maximum	Ec-Type
1 <sup>er</sup> Quizz	21,3	0,0	77,5	19,8
2 <sup>ème</sup> Quizz	13,7	0,5	38,0	10,08
3 <sup>ème</sup> Quizz	10,8	0,7	37,2	7,8

t de Student pour échantillons appariés

dl 35	2 <sup>ème</sup> quizz	3 <sup>ème</sup> quizz
1 <sup>er</sup> quizz	<b>2,2 : S</b> <b>P : .034</b>	<b>3,15 : TS</b> <b>P : .003</b>
2 <sup>ème</sup> quizz		1,28 : NS P : .21

La moyenne obtenue à l'indice de centration par le groupe TEMOIN se rapproche de la valeur idéale 0 au fil des tests. La différence est significative lorsqu'on compare les performances des étudiants au 1<sup>er</sup> quizz avec celles du 2<sup>ème</sup> et du 3<sup>ème</sup> quizz. Les différences entre le 2<sup>ème</sup> et le 3<sup>ème</sup> quizz apparaissent comme moins marquées et le test statistique confirme qu'elles ne sont pas significatives.

Le rapport des variances montre également des différences très significatives, sauf lorsqu'on compare le 2<sup>ème</sup> quizz avec le 3<sup>ème</sup>.

F - dl 35	WINCHECK	TEMOIN
GUESS	<b>3,85 - TS</b>	<b>6,44 - TS</b>
WINCHECK		1,67 - TS

## 9. Conclusions

### A. Amélioration des performances moyennes au fil des quizz lorsqu'on envisage l'ensemble des étudiants (N = 250)

Au niveau de l'ensemble des 250 étudiants ayant participé aux 3 quizz, on assiste à une progression moyenne des performances aux indices métacognitifs. A chaque fois, la progression est accompagnée d'une réduction de l'écart-type.

#### 1) Amélioration des scores à l'indice de réalisme (R)

Nous observons une augmentation de la moyenne des scores obtenus à l'indice de réalisme. La comparaison des performances aux trois quizz montre des différences statistiquement très significatives. Cette progression moyenne du réalisme au fil des quizz s'accompagne d'une réduction de l'écart-type .

#### 2) Indice de centration : diminution des erreurs (EC) et erreurs absolues (EAC)

Les scores moyens aux indices de centration EC et EAC s'améliorent de quizz en quizz. A chaque fois, l'amélioration de la moyenne est accompagnée d'une réduction de l'écart-type.

Les différences observées lors des comparaisons des moyennes des erreurs de centration (EC) sont statistiquement significatives lorsqu'on compare le 3<sup>ème</sup> quizz avec le 1<sup>er</sup> et le 2<sup>ème</sup>. L'amélioration des performances liées à l'indice EC s'accompagne d'une légère surestimation au 3<sup>ème</sup> quizz. Les différences liées à l'erreur absolue de centration (EAC) sont statistiquement significatives pour les 3 cas de figure (voir point 6.B.2) ci-avant).

Ces conclusions liées à l'amélioration au fil des quizz des performances aux indices de réalisme et de centration confirment les résultats obtenus par LECLERCQ (1993) et GILLES (1996).

### B. Comparaison des performances des 3 groupes de 36 sujets

#### 1) Indice de réalisme (R) : pas de progression significative en faveur de l'un ou l'autre groupe

Les trois groupes améliorent leurs scores moyens à l'indice de réalisme, mais lorsqu'on compare les groupes GUESS, WINCHECK et TEMOIN au 3<sup>ème</sup> quizz, force est de constater que même si des différences apparaissent en faveur du groupe entraîné avec GUESS, celles-ci ne sont statistiquement pas significatives.

#### 2) EAC : pas d'amélioration moyenne significative en faveur de l'un ou l'autre groupe, mais gains relatifs deux fois plus élevés pour le groupe GUESS

Les gains relatifs pour le groupe GUESS (41%) sont deux fois plus élevés que pour les groupes WINCHECK (20%) et TEMOIN (21%). Les différences de progression n'apparaissent cependant pas comme étant statistiquement significatives (voir 7.B ci-avant).

### C. Diminution significative des erreurs absolues de centration (EAC) après entraînement des 36 sujets du groupe GUESS

Rappelons que les entraînements ont eu lieu entre le 2<sup>ème</sup> et le 3<sup>ème</sup> quizz.

Le tableau ci-dessous synthétise les observations liées à la comparaison de l'évolution des erreurs absolues de centration au sein des 3 groupes de 36 sujets (voir point 8 ci-avant).

		GUESS	WINCHECK	TEMOIN
Comparaison des 1 <sup>er</sup> et 2 <sup>ème</sup> QUIZZ	MOYENNE	TS	NS	TS
	VARIANCE	TS	TS	S
Comparaison des 2 <sup>ème</sup> et 3 <sup>ème</sup> QUIZZ	MOYENNE	S	NS	NS
	VARIANCE	TS	NS	NS

#### 1) L'entraînement GUESS diminue l'erreur absolue de centration (EAC)

On observe des différences statistiquement significatives d'EAC chez les sujets du groupe GUESS lorsqu'on compare les moyennes du 2<sup>ème</sup> quizz avec celles du 3<sup>ème</sup> quizz (voir 8.A). Pour les deux autres groupes, la comparaison des EAC ne révèle pas de différences significatives (voir 8.B et 8.C).

#### 2) L'entraînement GUESS diminue l'écart-type des erreurs absolues de centration (EAC)

Lorsqu'on compare les écarts-types obtenus par le groupe GUESS avant et après entraînement, une réduction statistiquement très significative est observée. Pour les deux autres groupes, WINCHECK et TEMOIN, les différences d'écarts-types « avant-après » entraînement ne sont pas significatives.

### D. Perspectives liées aux entraînements WINCHECK et GUESS

L'entraînement WINCHECK ne semble pas améliorer le réalisme<sup>13</sup> des étudiants. Cependant, d'un point de vue déontologique, il nous paraît important de continuer à familiariser les étudiants à l'utilisation de WINCHECK (qu'il n'ont jamais rencontré auparavant) avant le premier examen ayant recours à ce logiciel d'évaluation interactive.

En ce qui concerne GUESS, cette étude constitue un encouragement à la mise en place systématisée de l'entraînement. Prochainement, une version internet du programme sera disponible afin d'offrir un plus large accès à cette procédure d'entraînement à l'utilisation des degrés de certitude.

<sup>13</sup> Rappelons que dans le cadre de l'entraînement WINCHECK, les étudiants sont amenés à fournir peu de réponses (une vingtaine) accompagnées d'un degré de certitude.

Jean-Luc Gilles (1997), Impact de deux entraînements à l'utilisation des degrés de certitude chez les étudiants de 1<sup>ère</sup> candidature à la faculté de psychologie et des sciences de l'éducation de l'ULg, in Boxus, Jans, Gilles et Leclercq (Eds), *Stratégies et médias pédagogiques pour l'apprentissage et l'évaluation dans l'enseignement supérieur*, Actes du XV<sup>ème</sup> Colloque de l'AIPU, Liège, 1997, pp. 311-326.

## 9. 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 Eurotecnet.
- 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.
- D'HAINAUT, L. (1973). Etude d'une nouvelle variable pour l'analyse statistique des expériences pédagogiques, Paris, Bulletin de Psychologie, n° 305, vol. XXVI, p. 622-630.
- 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.
- GATHY, P. & DENEFF, J.-F. (1993) Self-confidence assessment during computer-assisted testing in histology. NATO ASI Series, Item Banking: Interactive Testing and Self Assessment, Berlin: Springer Verlag, 1993, Vol. 112, pp. 233-241.
- GILLES, J.-L. (1995). Entraînement à l'autoévaluation : une comparaison filles/garçons à l'université. Actes 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 FA.P.S.E. - ULG, Actes du colloque de l'ADMEE-EUROPE "Dix années de travaux de recherche en évaluation", 18, 19 et 20 septembre 1996, à paraître.
- GILLES, J.-L. et LECLERCQ, D., (1996). 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, Tunis, 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'Education, Université de Liège Institut de Psychologie et des Sciences de l'Education.
- 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.
- 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.
- McGUIGAN, F.J., (1967). The G. Statistics, An Index of Amount Learned, N.S.P.I. Journal, 1967, 69, pp. 14-16.
- PITZ, G. F. (1974). Subjective probability distributions for imperfectly known quantities. In Greg, L.W. (ed.) Knowledge and Cognition. New York: Adison Wiley, pp. 29-41.
- SHANNON, C.E. (1951). Prediction and entropy of printed english. Bell Syst. Techn. J. 30, pp. 50-64.