

GUESS
un logiciel pour entraîner à l'auto-estimation
de sa compétence cognitive¹

D. LECLERCQ² & J.L. GILLES³

Service de Technologie de l'Education (S.T.E.)
Faculté de Psychologie et des Sciences de l'Education (F.A.P.S.E.)
de l'Université de Liège

Communication présentée au Colloque
ESIEE

¹ Les auteurs tiennent à remercier Michel HURARD, licencié en informatique qui a programmé le logiciel dans le langage TOOLBOOK.

² Service de Technologie de l'Education, Université de Liège.

³ Assistant à la Faculté de Psychologie et des Sciences de l'Education, Université de Liège.

A. LA METACOGNITION : UN DESERT PEDAGOGIQUE ?

Habituellement, les enseignants ne s'occupent pas de la métacognition des étudiants
C'est selon nous parce qu'ils manquent de théories et de techniques *ad hoc*.

Dans son article de 1993, Emir SHUFORD met en exergue trois maximes :

<i>The most useful piece of learning for the uses of life is to unlearn what is untrue</i>	L'apprentissage le plus utile pour la vie est de désapprendre ce qui est faux
ANTHISTENES (445-365 B.C.) ⁴	
<i>It's not what you don't know that hurts you. It's what you thing is so that isn't.</i>	Ce qui nous pose des problèmes, ce n'est pas ce que nous ignorons. C'est ce que nous savons, mais qui est faux .
Mark TWAIN	
<i>Beware of false knowledge : It is more dangerous than ignorance.</i>	Méfiez vous de la connaissance erronée : Elle est plus dangereuse que l'ignorance.
Georges Bernard SHAW	

Nombreux sont maintenant les auteurs tels que BRUNO (1993), HUNT (1993), VAN LENTHE (1993), FABRE (1993), LECLERCQ (1993), DIRKZWAGER (1993) qui, à la suite de pionniers tels que DE FINETTI (1965), VAN NAERSSSEN (1965), SHUFORD et al. (1966) pensent qu'il est inconcevable de traiter de la même façon

- l'ignorance (avouée)
- la connaissance insuffisante parce que trop partielle (réponse correcte mais peu sûre)
- la connaissance satisfaisante (réponse correcte et assez sûre)
- la connaissance totale (réponse correcte et très sûre)
- la méconnaissance sur base d'idées fausses ou *misconceptions* (réponse incorrecte et sûre).

Nous avons démontré (LECLERCQ, 1983 et 1993) que c'est inadmissible du point de vue de l'attribution des points (car on favorise des pratiques telles que le *guessing*⁵, même quand on utilise la fameuse "*correction for guessing*"⁶). HUNT(1993) et BRUNO (1993) ont démontré que l'ignorance du formateur de l'état de doute de l'étudiant était très préjudiciable quant à la possibilité de diagnostic et donc de remédiation.

C'est pourquoi, nous demandons systématiquement aux étudiants d'accompagner chacune de leur réponse d'un degré de certitude indiquant la confiance (probabilité subjective) qu'il lui accorde. La procédure doit en plus être conforme à la théorie des décisions.

Il existe de nombreuses façons de recueillir cette certitude ou ce doute.. Nous les avons décrites ailleurs (LECLERCQ, 1993). Quelques unes d'entre elles constituent ce que SHUFORD et al (1966) appellent des "Admissible Probability Measurement Procedures" où la consigne est probabiliste (Non pas "peu sûr", mais "certitude comprise entre 25 et 50%") et où le barème des tarifs avantage le fait d'exprimer son doute (ou sa certitude) sans biais.

⁴En anglais : "Before Christ"

⁵*To guess* signifie "deviner" et non "répondre au hasard" !

⁶Cette correction consiste, quand on donne un point par réponse correcte, à retirer $1/(k-1)$ par erreur, k étant le nombre de solutions proposées. Ainsi avec une question VRAI-FAUX(2 solutions), on retire 1 point; avec une QCM à 3 solutions, on retire 1/2 point, avec 4 solutions, on retire 1/3 point, etc.

B. CONSIGNE CONCERNANT LES DEGRES DE CERTITUDE

Vous devez accompagner chaque réponse d'un des degrés de certitude suivants :



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	Incorrecte

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. LA JUSTIFICATION DU BAREME DE TARIFS

Le **barème des tarifs** peut vous paraître "bizarre". Il a été calculé de manière à ce que

- DIRE LA VERITE soit la stratégie qui rapporte le plus de points;
- ceux qui s'auto-évaluent bien, c'est-à-dire qui sont REALISTES (ni surestimation flagrante, ni sous-estimation énorme) gagnent le plus de points.

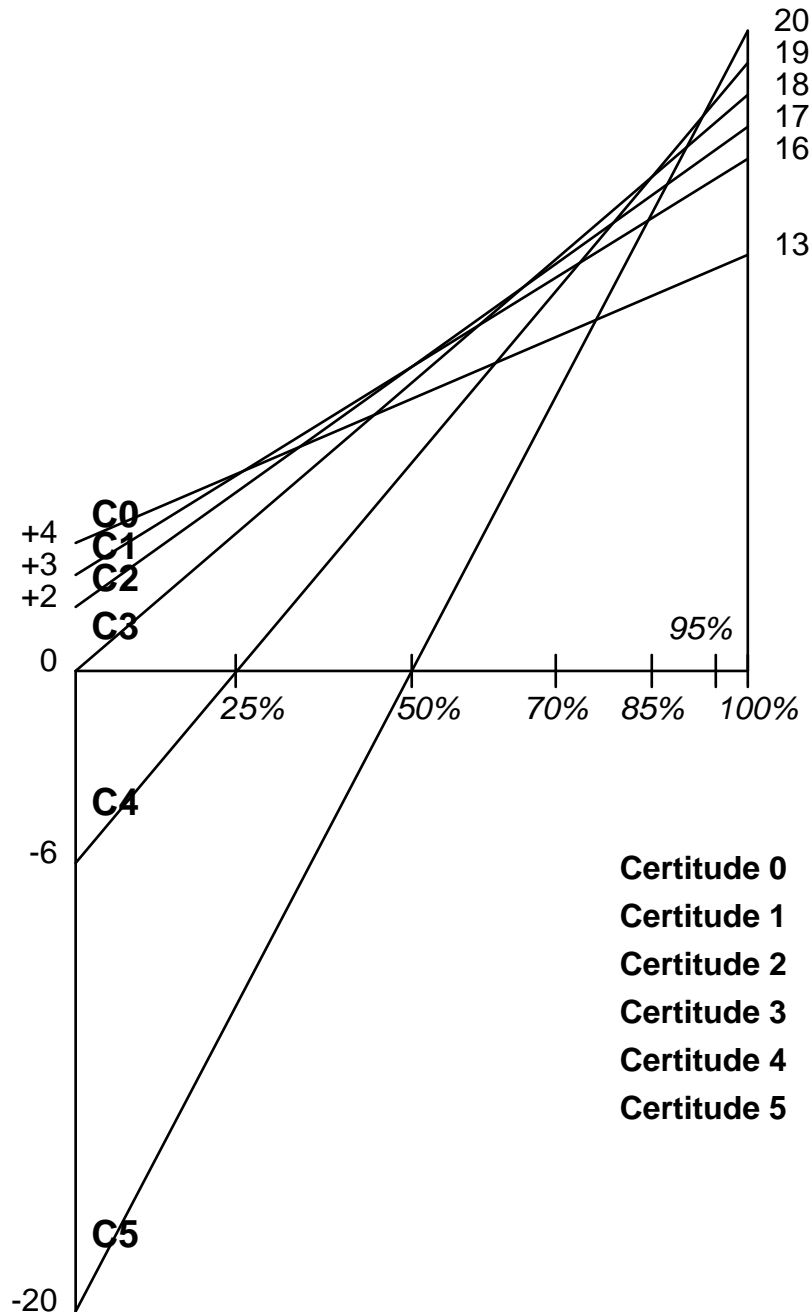
Attention ! 4/20 est un score positif, mais il est cependant l'indication d'une note insuffisante (puisque inférieur à 10/20).

La majorité des étudiants, s'auto-estimant bien (avec réalisme), sont généralement avantagés par les degrés de certitude c'est-à-dire que leur score ainsi calculé est meilleur que s'il avait été calculé sur la seule base de la correction for guessing, c'est-à-dire le nombre de réponses correctes diminué de la pénalité $1/k-1$ (où k = le nombre de solutions proposées) pour chaque erreur, c'est-à-dire

- retrait de
- 1 point en cas de VRAI/FAUX
 - 1/2 point pour 3 solutions proposées
 - 1/3 point pour 4 solutions proposées
 - 1/4 point pour 5 solutions proposées
 - 1/5 pour 6 solutions proposées
 - etc.

Les six obliques ci-après constituent les espérances mathématiques du score à une question pour chacun des six degrés de certitude choisis étant donné le continuum (de 0 à 100 %) des certitudes ressenties. Chacune des six obliques est obtenue en joignant le tarif en cas de réponse correcte au tarif en cas de réponse incorrecte.

Remarquez pour quelle zone (de 0 à 100) chaque oblique surplombe toutes les autres !



D. LE NON HASARD ET LA NECESSITE (de s'entraîner à l'estimer)

Les étudiants n'ont pas l'habitude de répondre à des QCM, encore moins en les accompagnant de degrés de certitude.

Parce que, hélas, on ne les y a pas habitués comme l'auraient dit les Soldats de La Palice ⁷ !

Les **météorologistes** eux, sont quotidiennement obligés de faire des prédictions dont ils ont la confirmation (ou l'infirmité) le lendemain, et ce 365 fois par an.

Une des premières études systématiques sur le **réalisme des prédictions** a d'ailleurs été effectuée par COOKE (1906) sur 1951 prédictions du temps du lendemain par des météorologistes, prédictions accompagnées d'un degré de certitude sous la forme d'expressions verbales telles que "A peu près certain" (code 5), "Probabilité normale" (code 4), "Il existe un doute" (code 3), etc..

Les prédictions accompagnées de la certitude 5 se confirmaient à 98%, la certitude 4 à 94%, la certitude 3 à 77%. (d'après LICHTENSTEIN et al., 1977). On voit que ces météorologistes se sont montrés **COHERENTS** : plus ils sont sûrs, plus leur taux d'exactitude est élevé (la cohérence se mesure par la corrélation entre ces deux variables : certitude et réussite).

Par contre, il est impossible de savoir s'ils étaient **REALISTES**, et ce par faute de la consigne, qui n'était, hélas, qu'**ordinaire**. Il aurait fallu que la consigne demande de préciser non pas si la prédiction était "à peu près certaine", mais **par exemple** selon la consigne décrite en section B, qui, elle, est une consigne **métrique** !

Mais aucun d'entre nous n'a été amené à se prêter systématiquement à ce genre d'exercice.

Quel serait notre réalisme dans un tel cas ? Nous sous-estimerions-nous ? Nous surestimerions-nous ? Pour le savoir, chacun doit avoir essayé.

C'est pourquoi nous avons développé le jeu GUESS qui se joue à l'aide d'un ordinateur. Ce jeu est inspiré du SHANNON Guessing Game (1951, décrit dans LECLERCQ, 1983). Il a été programmé par Michel HURARD, licencié en informatique de l'Université de Liège.

E. LES REGLES DU JEU GUESS

Le joueur doit deviner les lettres successives d'un texte (d'au moins 100 lettres, souvent plus) extrait d'un écrit de la littérature française. L'étudiant doit taper à chaque fois une des 26 lettres de l'alphabet au clavier. Il doit accompagner cette réponse (cette "prédiction") d'un degré de certitude (avec la consigne ci-dessus).

Il est ensuite informé de la réponse correcte, c-à-d LA lettre à deviner, qui vient s'afficher au bon endroit du texte sur l'écran.

Il doit ainsi deviner toutes les lettres une à une en connaissant les lettres précédentes. Dans ce jeu, certaines lettres sont, évidemment faciles et d'autres difficiles à deviner.

Dès la deuxième lettre à deviner, il faut tenir compte d'une 27e possibilité : l'espace entre deux mots, ou un signe de ponctuation (virgule, point, etc..). Dans tous ces cas, la réponse à fournir est de presser sur la barre d'espacement.

⁷Maréchal de France tué à Pavie (1525). La chanson que ses soldats composèrent pour célébrer sa vaillance et qui se terminait par ces vers "Un quart d'heure avant sa mort, il vivait encore" est restée célèbre par sa naïveté, qu'on a injustement attribuée à La Palice lui-même. (d'après le Dictionnaire Robert 2).

Voici l'écran tel que le voit le joueur :

(1) (2) (3)

(4) Les Lions du Panshir Ken FOLLETT 478 caracteres

Jane, jeune étudiante anglaise qui vit à Paris, découvre que*

(5) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z - <-

(6) +13 0 +4 +16 1 +3 +17 2 +2 +18 3 +0 +19 4 -6 +20 5 -20

(7) Score 11.6

Statistiques après 60 caractères ?

Certitude moyenne totale 85,5 Réussite moyenne 71,6

(8) (9) (10) (11) (12) (13)

(14) RC 43 0 0 4 11 0 28 Tot 60 0 1 10 14 0 35 TE% 71,6 0 40 78,5 80

Réalisme 0,865 Cohérence 0,951 Centration 13,91

Moyen Bonne Faible

(15)

Commentaires de cet écran :

- (1) Titre du texte original que le joueur essaye de deviner.
- (2) Nom de l'auteur.
- (3) Nombre total de caractères.
- (4) Fenêtre d'affichage du texte (les réponses correctes attendues). Les lettres soulignées rappellent que la réponse du joueur avait été INCORRECTE.
- (5) Clavier (où la touche "-" vaut pour tous les espaces et signes de ponctuation; la flèche "arrière" sert à corriger la lettre choisie (avant d'avoir introduit le degré de certitude).
- (6) Affichage des codes de certitude et du barème des tarifs en cas de réponse correcte (au-dessus) et incorrecte (en-dessous). Le joueur fournit son degré de certitude en pressant une de ces 6 zones.
- (7) Score moyen ramené sur 20 (quel que soit le nombre de questions déjà posées).
- (8) Nombre de réponses (caractères) déjà fournies.
- (9) Certitude moyenne (en %), ici 46 %.
- (10) Réussite moyenne (en %), ici 50 % (le joueur s'est donc TRES LEGEREMENT sous-estimé).

- (11) Indice de **Réalisme** (et commentaire sur sa qualité : maximum = 1).
 (12) Indice de **Cohérence** (et commentaire sur sa qualité : maximum = 1).
 (13) Indice de **Centration** (négative si le joueur se SURESTIME, positive s'il se SOUS-ESTIME, nulle dans l'idéal). Dans cet exemple :

$$\text{Centration (4)} = \text{Réussite moyenne (50)} - \text{Certitude moyenne (46)}$$

- (14) Tableau de répartition où

RC = réponses correctes avec un degré de certitude donné.

RA = nombre de réponses fournies avec un degré de certitude donné.

TE % = taux d'exactitude, en % (TE % = (RC/TOT) x 100) par degré de certitude.

- (15) Graphique de

- **Cohérence** : La droite (des moindres carrés) du joueur montre sa "tendance". Plus ses taux d'exactitude (les carrés) sont proches de cette droite (ou "s'alignent sur cette droite"), plus le joueur est COHERENT avec lui-même. L'indice (mathématique) de cohérence est calculé par la corrélation des taux d'exactitude.
- **Centration** : Le petit cercle, défini par la certitude moyenne (en abscisse) et l'exactitude moyenne (en ordonnée). L'indice de centration est calculé par la différence entre ces deux valeurs.
- **Réalisme** : Les carrés sont-ils proches de la diagonale (lieu où la réalité et la prédiction coïncident) ? Si la zone du carré (taux d'exactitude) d'un degré de certitude, entouré de ses limites de confiance (c'est-à-dire une Erreur de Mesure ajoutée vers le haut ET une erreur de mesure ajoutée vers le bas) ne recouvre pas la diagonale, il y a sous estimation ou surestimation. Des taux d'exactitude de 0 et 100 % ne permettent pas de calculer d'erreur de mesure (et donc ne permettent pas d'afficher des limites de confiance).

La formule de l'indice de réalisme est $1 - (5ERR)$, avec

$$ERR = \sqrt{\frac{\sum_i (TE_i - CV_i)^2 \cdot NU_i}{NA}}$$

où

TE_i = Taux d'Exactitude de la certitude i .

VC_i = Valeur centrale.

NU_i = Nombre d'utilisations.

NR = Nombre d'erreurs.

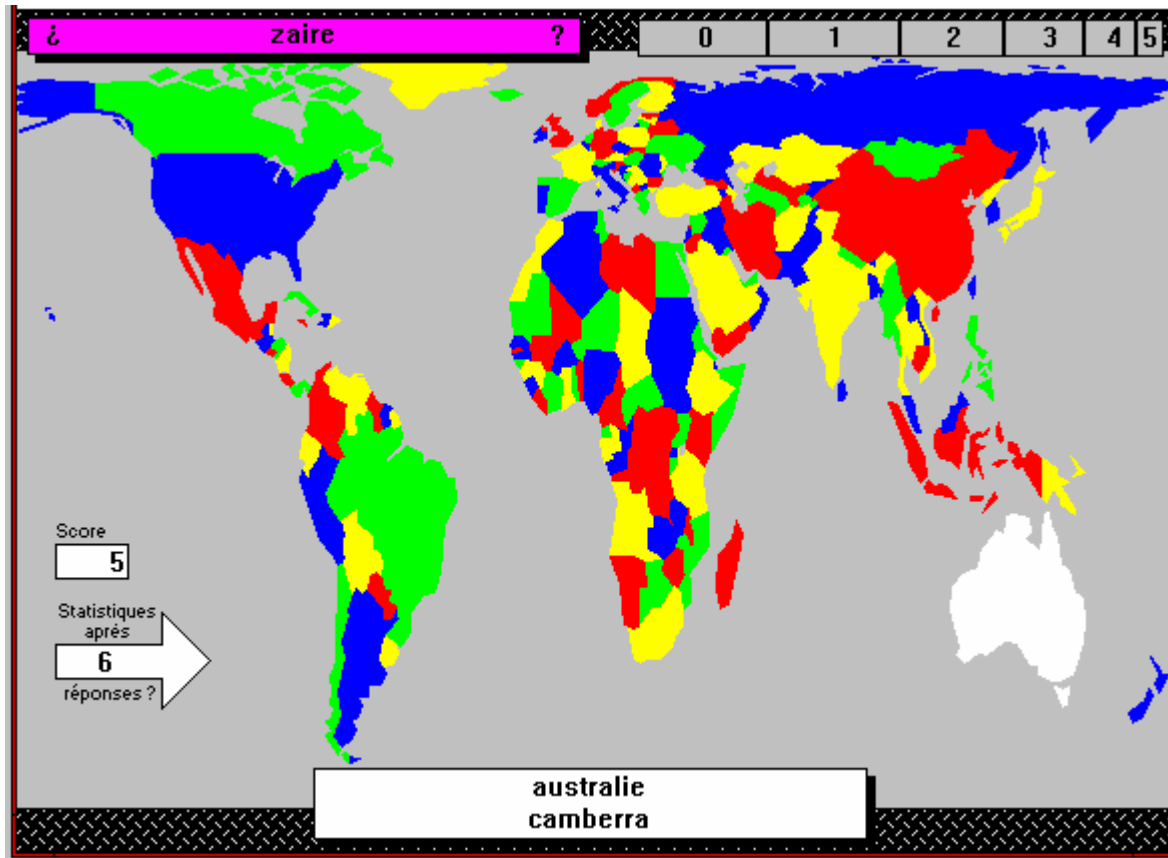
ERR = Somme des carrés des erreurs.

REALISME = 1 - (5ERR)

F. UNE VARIANTE

A côté du jeu GUESS-lettres existe une variante : GUESS PAYS, créée par Michel HURARD. La question apparaît en haut à gauche de l'écran, les certitudes en haut à droite.

Le joueur fournit sa réponse en désignant un pays (clic + déclic) à l'aide de la souris. N.B. : le clic maintenu (sans déclic) permet d'identifier précisément les pays (qui s'illuminent).



Contrairement au jeu précédent, le graphique de réalisme n'apparaît pas dans la même fenêtre que la fenêtre de questions. L'une fait disparaître l'autre.

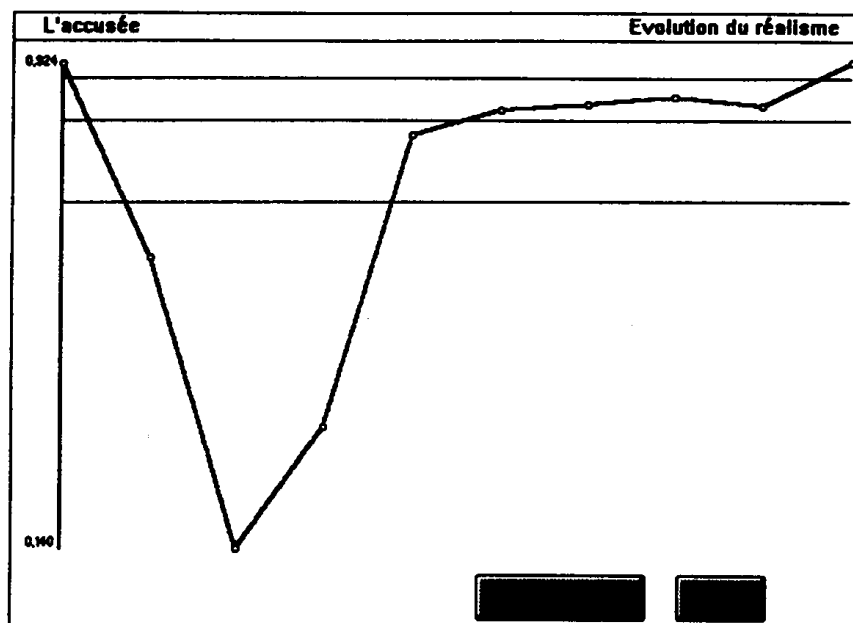
G. LE GRAPHIQUE D'EVOLUTION DE L'INDICE DE REALISME

Lorsque le joueur émet sa première réponse, son réalisme peut être très bon ou très mauvais, mais, au cours des réponses successives, on observe une tendance à se stabiliser vers des valeurs élevées.

C'est pourquoi il vaut mieux ne visualiser son réalisme qu'après 10 essais seulement.

En fait, le "maître du jeu" peut seul fixer après combien d'essais on peut voir l'évolution. L'écran ci-dessous montre l'évolution de l'indice de réalisme d'un joueur après 10 essais.

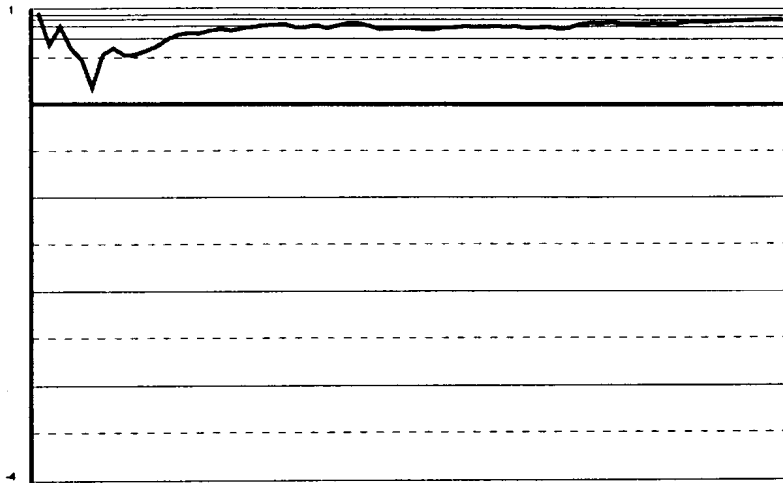
En abscisse apparaissent le nombre de réponses fournies et en ordonnée, les valeurs de l'indice de réalisme. Celui-ci peut aller à + 1 (parfait) à - 4 (catastrophique).



Voici l'évolution des trois indices au cours de 100 réponses successives d'une même personne à un texte donné (ici "Le beau rivage") :

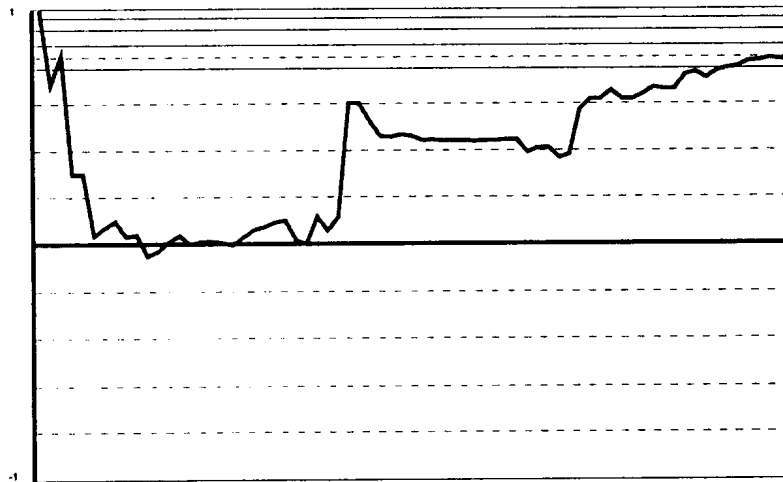
Le beau rivage

Evolution du realisme



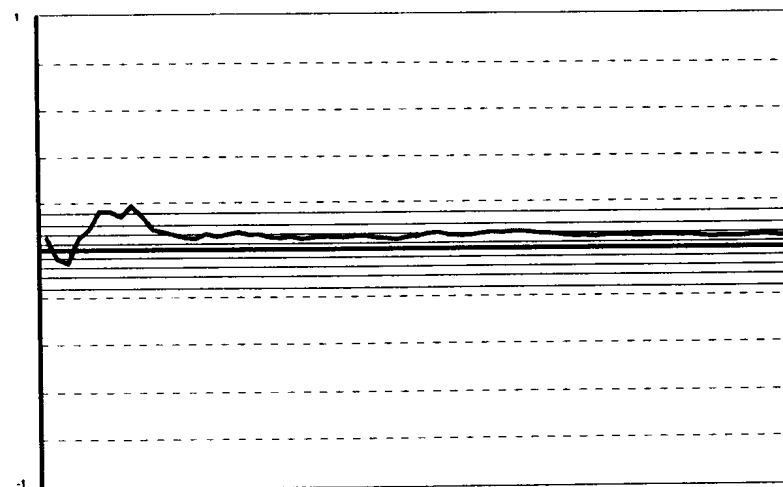
Le beau rivage

Evolution de la coherence



Le beau rivage

Evolution de la centration



H. LA SIGNIFICATION DES INDICES DE LA QUALITE DE L'AUTO-ESTIMATION

A titre de repères, voici les valeurs des indices calculés, à partir des réponses de 311 étudiants universitaires à une épreuve de Psychologie éducationnelle comportant 80 questions (D. LECLERCQ, 1992) :

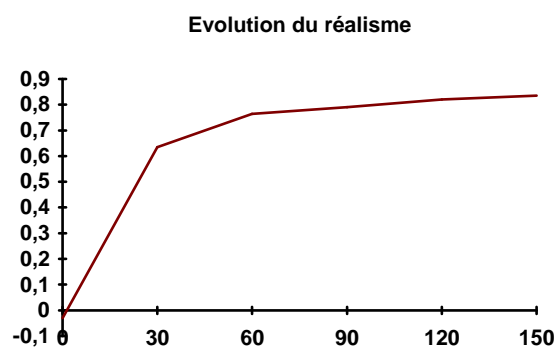
LA CENTRATION est ...	si la différence est	% sur 311 étudiants
IDEALE	= 0	1 %
EXCELLENTE	moins que 3	18 %
BONNE	de 3 à 6,99	19 %
SATISFAISANTE	de 7 à 10,99	19 %
FAIBLE	de 11 à 15,99	17 %
INSUFFISANTE	16 et plus	26 %
MINIMALE	97,5	0 %

LA COHERENCE est ...	si la corrélation est	% sur 311 étudiants
IDEALE	= 1	0 %
EXCELLENTE	supérieure à 0.97	13 %
BONNE	entre 0.93 et 0.96	18 %
SATISFAISANTE	entre 0.85 et 0.92	28 %
FAIBLE	entre 0.75 et 0.84	18 %
INSUFFISANTE	moins que 0.75	23 %
MINIMALE	- 1	0 %

LE REALISME est	si l'index est	% sur 311 étudiants
IDEAL	= 1	1 %
EXCELLENT	supérieure à 0.95	15 %
BON	entre 0.91 et 0.94	18 %
SATISFAISANT	entre 0.84 et 0.90	22 %
FAIBLE	entre 0.70 et 0.83	22 %
INSUFFISANT	moins que 0.70	22 %
MINIMAL	- 3.75	

I. L'EVOLUTION DU REALISME MOYEN (sur 10 étudiants) AU COURS DU TEST

num. étu.	nbre rép.	réalisme début	réalisme à 1/5 du nbre de réponses	réalisme à 2/5 du nbre de réponses	réalisme à 3/5 du nbre de réponses	réalisme à 4/5 du nbre de réponses	réalisme fin
257	178	-0,9	0,9	0,9	0,9	0,9	0,9
256	230	0,3	0,8	0,75	0,8	0,8	0,8
255	90	-2	0,1	0,4	0,5	0,7	0,75
254	200	0,6	0,4	0,7	0,6	0,6	0,7
253	172	0,5	0,75	0,8	0,9	0,95	0,95
252	119	0	0,75	0,8	0,75	0,8	0,8
251	114	-0,5	0,4	0,8	0,85	0,8	0,8
250	92	0,9	0,75	0,85	0,9	0,9	0,9
249	143	0,6	0,6	0,7	0,75	0,8	0,85
248	241	0,2	0,9	0,95	0,95	0,95	0,9
	158	-0,03	0,635	0,765	0,79	0,82	0,835

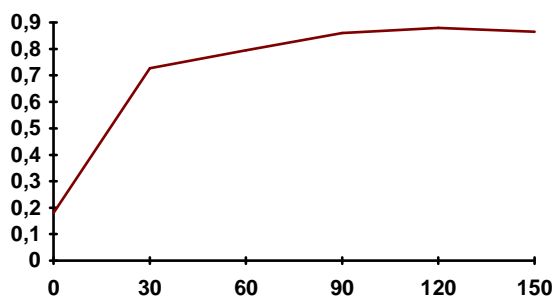


Après 3/5 du nombre de réponses, soit environ 90 questions, le réalisme atteint à peu près 0,80. On note aussi que certains étudiants ont atteint très tôt un indice de réalisme élevé. C'est le cas des étudiants 257, 256 et 248.

J. EVOLUTION DE LA COHERENCE (MOYENNE) AU COURS DU TEST

num. étu.	nbre rép.	cohérence début	cohérence. à 1/5 du nbre de réponses	cohérence 2/5 du nbre de réponses	cohérence3/5 du nbre de réponses	cohérence 4/5 du nbre de réponses	cohérence fin
257	178	-1	0,95	0,9	0,9	0,95	0,95
256	230	1	0,95	0,6	0,9	0,9	0,95
255	90	-0,2	0,62	0,75	0,8	0,85	0,86
254	200	0	0,85	0,9	0,95	0,9	0,85
253	172	1	0,75	0,8	0,9	0,92	0,92
252	119	1	0,6	0,65	0,55	0,58	0,6
251	114	0	0,2	0,7	0,8	0,9	0,85
250	92	0	1	1	1	0,97	0,97
249	143	0	0,55	0,75	0,85	0,87	0,85
248	241	0	0,8	0,9	0,95	0,95	0,85
	158	0,18	0,727	0,795	0,86	0,879	0,865

Evolution de la cohérence



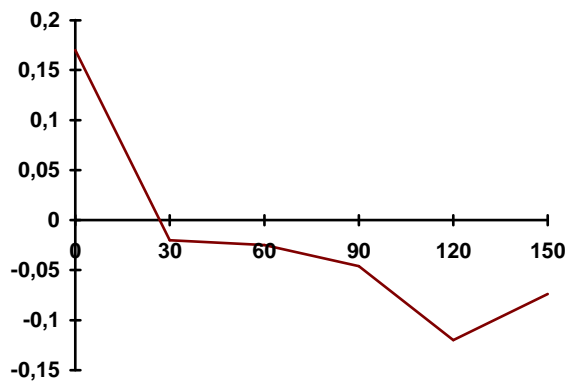
Une cohérence proche de 1 ou égale à 1 est souvent obtenue par des étudiants (l'étudiant 250 par exemple) qui n'utilisent que deux degrés de certitude et dont, par conséquent, l'acuité (non étudiée ici) est faible. Cette stratégie "gonfle" la moyenne des scores de cohérence. Cet indice ne devrait être calculé qu'à partir des résultats d'étudiants ayant une bonne "acuité".

On constate que les étudiants se répartissent en deux groupes : ceux qui dès la 30e réponse ont déjà un cohérence élevée (les étudiants 257,256, 254,248) et les autres. L'exercice bénéficie donc plus à certains étudiants qu'à d'autres.

K. EVOLUTION DE LA CENTRATION AU COURS DU TEST

num. étu.	nbre rép.	centration début	centration à 1/5 du nbre de réponses	centration à 2/5 du nbre de réponses	centration à 3/5 du nbre de réponses	centration à 4/5 du nbre de réponses	centration fin
257	178	0	0	0,05	0,05	0,08	0,08
256	230	-0,12	-0,16	-0,16	-0,16	-0,16	-0,16
255	90	0,8	0,3	0,2	0,2	0,16	0,14
254	200	0,25	-0,3	-0,2	-0,2	-0,2	-0,2
253	172	-0,16	0,12	0,08	0,07	0,04	0,04
252	119	-0,25	0	-0,2	-0,4	-0,4	0
251	114	0,4	0,12	0,06	0,12	0,13	0,14
250	92	0,13	-0,2	-0,12	-0,1	-0,8	-0,8
249	143	0,25	-0,08	-0,08	-0,12	-0,12	-0,08
248	241	0,4	0,04	0,12	0,08	0,08	0,1
	158	0,17	-0,02	-0,025	-0,046	-0,12	-0,074

Evolution de la centration

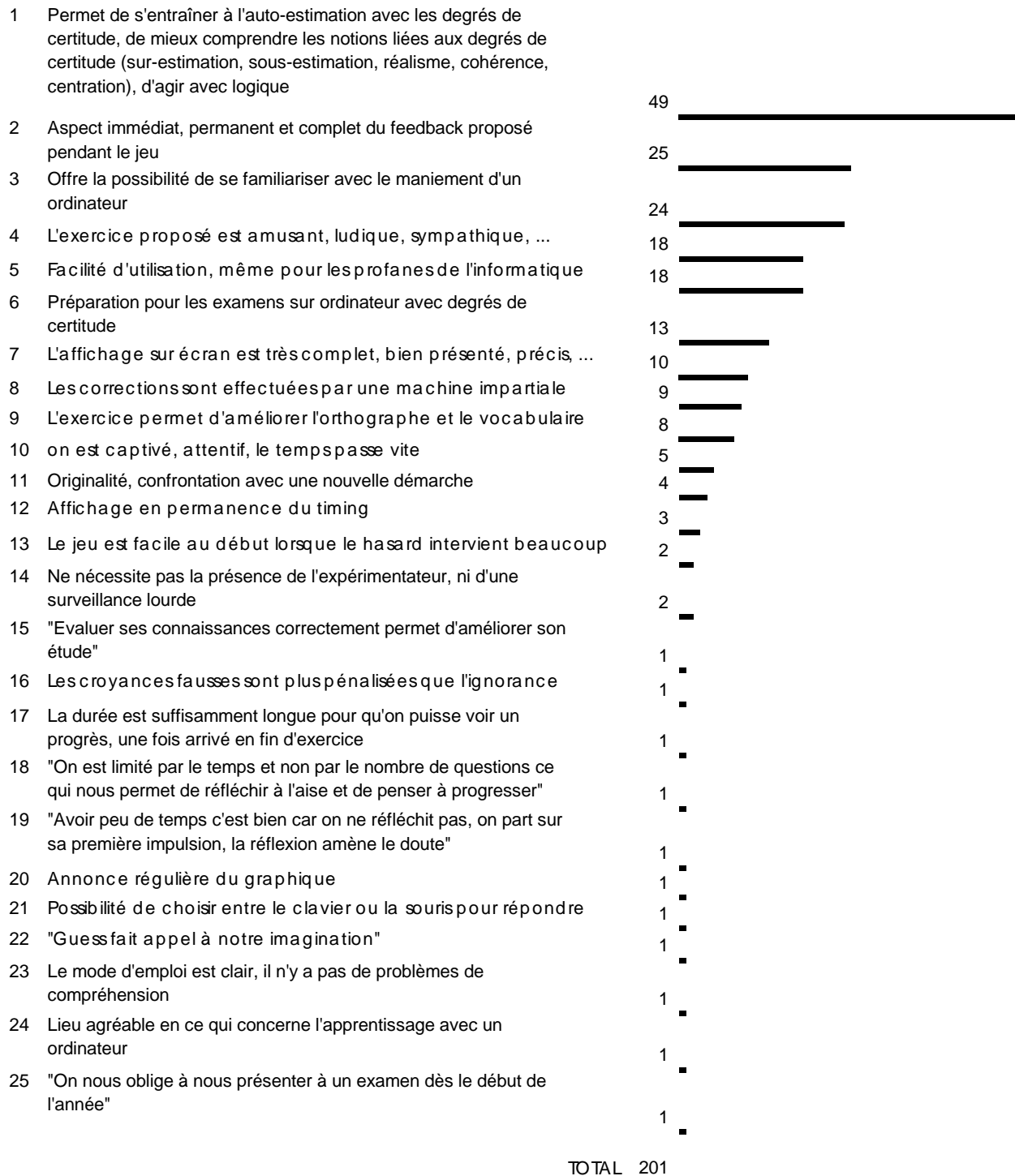


Il est intéressant de constater que la surestimation rapportée par de très nombreux auteurs dans la littérature (BRUNO, DENEFF, FABRE, GATHY, LECLERCQ, in *LECLERCQ & BRUNO, Item Banking : Self Assessment & Interactive Testing*, NATO ARW, Berlin : Springer Verlag, 1993) fait place, après 30 questions seulement à une stratégie de prudence (légère sous-estimation). Soulignons qu'il s'agit d'un testing interactif et que l'étudiant est informé de la qualité de sa performance (donc de sa certitude) après CHAQUE réponse, ce qui n'est pas le cas dans les tests papier-crayon.






















L. LES AVIS DES ETUDIANTS

Chaque étudiant a reçu un questionnaire lui demandant "les aspects positifs" et "les aspects négatifs" de cette expérience du logiciel. Nous supportons ici même des avis peu fréquents car ils pourraient servir dans la fabrication ultérieure d'un questionnaire fermé (à choix).

1. Les appréciations positives



2. Les appréciations négatives

1	Le graphique n'est pas suffisamment clair et compréhensible	18	
2	Stress dû au temps très court, à l'impression de devoir rivaliser contre une machine, à la présence d'un timing à l'écran, au feedback permanent, au nombre minimum de réponses à introduire	15	
3	Une partie des résultats obtenus au test est due au hasard (au début d'un mot on devine les lettres et la chance intervient)	14	
4	L'exercice est trop souvent interrompu par un message proposant de visualiser les résultats, cela déconcentre l'utilisateur, le coupe dans son élan	12	
5	Le tableau des résultats est difficile à comprendre	11	
6	Tout le monde n'est pas familiarisé avec le maniement de l'ordinateur, ce qui peut influencer les résultats	10	
7	Comment interpréter les résultats ? Quelle est l'utilité des différents indices ? Il n'y a pas de conclusion suffisamment claire pour donner du sens à l'exercice	10	
8	Il n'y a pas d'explication quant au but du jeu, sans la personne sur place on ne peut pas commencer l'exercice	9	
9	L'impression d'être catalogué alors qu'il n'est pas certain que l'on puisse établir le même "jugement" dans d'autres tâches et dans notre façon d'agir dans la vie de tous les jours	7	
10	L'exercice ne tient pas compte du bagage culturel de l'utilisateur, de l'étendue de son vocabulaire; c'est plus difficile pour un étranger	7	
11	La durée est trop longue ce qui provoque des problèmes de concentration chez certains : "à la fin on écrit n'importe quoi"	7	
12	Le barème : "4/20 c'est trop pour une ignorance totale", "la perte de points avec la certitude 5 est trop grande", "difficile de remonter quand on a perdu beaucoup", "une moyenne de notes supérieures à 20 peut cacher une lacune importante"	6	
13	Les textes proposés n'ont pas tous le même niveau de difficulté (certains comportent des dates, des noms propres, ...) et les résultats sont influencés par la difficulté du texte	5	
14	Il est impossible de passer une question et d'y revenir par la suite	5	
15	Après un signe de ponctuation on est certain qu'il y a un espace, ce qui entraîne toujours une réponse correcte avec un degré de certitude maximum et cela fausse les résultats	3	
16	L'exercice n'est pas pris au sérieux car c'est un jeu	3	
17	Textes peu intéressants, bizarres, simplistes	3	
18	Aspect nouveau et étrange de l'exercice	2	
19	On ne visualise pas les pourcentages liés aux 6 degrés de	2	
20	On ne peut pas justifier sa réponse	2	
21	Amène à jouer la carte de la prudence lors d'un examen	2	

22	La fenêtre d'affichage du texte est trop petite	2	■
23	Pas de définition des termes "cohérence", "réalisme", "centration"	2	■
24	Lorsqu'on utilise la souris pour encoder des lettres les temps de réaction de l'ordinateur sont plus lents	2	■
25	L'exercice n'est pas basé sur des matières de cours	2	■
26	L'écran est trop chargé	2	■
27	Le score devrait être établi sur 20	2	■
28	Les fautes d'orthographe qui subsistent dans le texte influencent le score	1	■
29	Le graphique ne permet pas de visualiser la progression pendant le jeu	1	■
30	Ne reproduit pas le contexte de l'examen où on est limité aussi par le nombre de questions	1	■
31	L'affichage du barème des points déconcentre l'utilisateur	1	■
32	Le test ne propose pas grand chose de concret	1	■
33	Quelle est la validité du résultat final étant donné la variabilité de la cohérence, du réalisme, de la centration pendant le test	1	■
34	Le jeu devrait consister à trouver un mot plutôt qu'une lettre	1	■
35	Le point de départ du jeu est trop vague et laissé au hasard, n'importe quelle lettre peut convenir	1	■
36	Jeu trop monotone	1	■
37	On devrait pouvoir effectuer une 2ème tentative avec un autre texte	1	■
38	Exercice effectué en dehors des heures de cours	1	■
39	"On joue la prudence, puis on prend trop confiance"	1	■
40	"Mes résultats n'ont pas été imprimés"	1	■
41	Les traitements informatiques ne peuvent pas être appliqués à des questions ouvertes	1	■
42	Des erreurs de programmation sont toujours possibles	1	■
43	"Il faudrait un test où on est moins sûr à la fin d'un mot"	1	■
44	Manque de contact humain	1	■
45	Pas intéressant pour les répétants	1	■
46	"J'ai perdu du temps car personne n'était là pour remettre mon ordinateur sur la bonne voie"	1	■
47	La façon dont les erreurs sont visualisées (trait rouge souligné) n'est pas assez voyante	1	■
		1	■
TOTAL		185	

Bien des critiques nous indiquent les améliorations à apporter à la consigne, aux explications de départ, au logiciel, aux textes, etc.

M. CONCLUSION

Bien que l'analyse des données soit largement incomplète (10 étudiants sur 200), il apparaît déjà que ce petit jeu remplit des fonctions formatives de façon assez satisfaisante. Les remarques et critiques des étudiants sont particulièrement précieuses pour l'améliorer.

Il faudra encore attendre que ces mêmes étudiants aient été confrontés à un test réel pour savoir s'ils y obtiennent de meilleurs indices de cohérence, de réalisme, de centration que leurs prédécesseurs, non entraînés avec GUESS. Si oui, ils devraient - à égalité de réponses correctes - recevoir plus de points, et donc être payés pour cette compétence métacognitive accrue. Déontologiquement, il nous paraît impératif d'offrir une telle possibilité d'exploration non pénalisante aux étudiants novices dans cette estimation de compétence⁸.

⁸ GUESS est disponible au STE (voir conditions auprès des auteurs).

Bibliographie

- ATTNEAVE, F., *Application of Information Theory to Psychology*, New York, Holt, Rinehart and Winston, 1959.
- BOXUS, E., CHECK : Une banque de question interactive, in *Actes du Colloque "Les questionnaires automatisables"*, Marne-La-Vallée, ESIEE, mars 1992, 31-49.
- BOXUS, E., LECLERCQ, D., OSTERRIETH, S., WUIDAR, H., *Principes communs pour évaluer les résultats cognitifs de la formation*, Bruxelles : Commission des Communautés Européennes, EUROTECNET, 1991.
- BRUNO, J.E., Using Testing to Provide Feedback to Support Instruction : A Reexamination of the Role of Assessment in Educational Organizations, in *LECLERCQ & BRUNO, Item Banking : Self-Assessment & Interactive Testing*, NATO ARW, Berlin : Springer Verlag, 1993.
- DE FINETTI, B., La décision et les probabilités, *Revue des Mathématiques Pures et Appliquées*, Bucarest, 1965a, 405-413.
- DE FINETTI, B., Methods for Discriminating levels of partial knowledge concerning a test item, *British Journal of Mathematical and Statistical Psychology* 18, 1965b, 87-123.
- DIRKZWAGER, A., The Dependability of Test Scores : Generalizability Theory and Hierarchical Linear Models, in *LECLERCQ & BRUNO, Item Banking : Self-Assessment & Interactive Testing*, NATO ARW, Berlin : Springer Verlag, 1993.
- FABRE, J.M., Subjective Uncertainty and the Structure of the Set of all Possible Events, in *LECLERCQ & BRUNO, Item Banking : Self-Assessment & Interactive Testing*, NATO ARW, Berlin : Springer Verlag, 1993.
- HUNT, D.P., Human Self-Assessment : Theory and Application to Learning and Testing, in *LECLERCQ & BRUNO, Item Banking : Self-Assessment & Interactive Testing*, NATO ARW, Berlin : Springer Verlag, 1993.
- LECLERCQ, D., Confidence Marking, its use in testing, in *POSTLETHWAITE, CHOPPIN (Eds), Evaluation in Education*, Oxford : Pergamon, 1982, vol. 6, 2, 161-287.
- LECLERCQ, D., Mesurer la connaissance partielle et le réalisme par les degrés de certitude, in *Actes du Colloque International "Formation, Evaluation, Sélection par Questionnaires Fermés"*, Marne-La-Vallée, mars 1988, vol. 1, 306-316.
- LECLERCQ, D. & GILLES, J.L., Hypermédias : teaching through assessment, in *D. LECLERCQ et J. BRUNO (Eds), Item Banking : Self Assessment & Interactive Testing*, NATO, ARW, Conference in Liège, October 1992, Berlin : Springer Verlag.
- LECLERCQ, D., Validity, Reliability and Acuity of Self-Assessment in Educational Testing, in *LECLERCQ & BRUNO, Item Banking : Self-Assessment & Interactive Testing*, NATO ARW, Berlin : Springer Verlag, 1993.

- LICHTENSTEIN, D., FISCHHOFF, B., PHILLIPS, L.D., Calibration of probabilities : the state of the art, decision making and change in human affairs, in *Proceedings of the Fifth Research Conference on Subjective Probability, Utility and Decision Making*, Darmstadt, 1-4 September, D. Reidel; 1977.
- MURPHY, A.H. & WINKLER, R.L., Subjective Probability forecasting experiments in meteorology : some preliminary results, *Bulletin of the American Meteorological Society* 55, 1974, 1206-1216.
- PITZ, G.F., Subjective Probability Distribution for imperfectly Known Quantities, in L.W. GREGG (Ed.), *Knowledge and Cognition*, New-York : Wiley, 1974.
- SHANNON, C.E., Prediction and entropy of printed English, *Bell. Syst. Techn. J.* 30, 1951, 50-64.
- SHUFORD, E. H., In Pursuit of the Fallacy : Resurrecting the Penalty, in LECLERCQ & BRUNO, *Item Banking : Self-Assessment & Interactive Testing*, NATO ARW, Berlin : Springer Verlag, 1993.
- SHUFORD, E., ALBERT, A. & MASSENGILL, N.E. (1966), Admissible probability measurement procedures, *Psychometrika* 31, 1966, 125-145.
- VAN LENTHE, J., The Development and Evaluation of ELI, an Interactive Elicitation Technique for Subjective Probability Distributions, in LECLERCQ & BRUNO, *Item Banking : Self-Assessment & Interactive Testing*, NATO ARW, Berlin : Springer Verlag, 1993.
- VAN NAERSSSEN, R.F. & VAN BEAUMONT, R. (1965), Ervaringen met een Zekerheidsaanduiding bij objectieve Tentamens, *Nederlands Tijdschrift Psychologie* 20, 1965, 208-315.