

Facteurs liés à la concordance des corrections d'une épreuve de lecture critique d'article

Factors Linked to the Reliability of the Marking Process Intended for the Scientific Paper Critical Analysis Assessment.

Pierre CZERNICHOW¹, Jean-Paul FILLASTRE², Alain LAVOINNE³,
Jean-François LEMELAND⁴, Francis ROUSSEL⁵

Résumé **Contexte** : Une épreuve de lecture critique d'article (LCA) sera introduite en 2008 dans l'examen classant national (ECN). Les rares travaux consacrés à cette épreuve suggèrent une concordance médiocre des corrections. **But** : Identifier certains facteurs liés à la concordance des corrections d'une épreuve de LCA. **Sujets et méthodes** : 59 étudiants volontaires de DCEM 2 et DCEM 3 ont été soumis à une épreuve de LCA dans les conditions prévues pour l'ECN. Trois correcteurs ayant préparé l'épreuve (choix d'article, questions et réponses attendues selon 16 questions, grilles de correction (s), pondérations), ont corrigé chaque copie de façon indépendante, sans grille puis avec une grille. L'accord entre les 3 couples de correcteurs, avec ou sans grille, pour les 16 questions, soit 96 couples, a été mesuré par les coefficients de corrélation intra-classe (CCI). La valeur médiane du CCI pour chaque facteur (questions, grille ou non, questions ou résumé, couples de correcteurs) a été comparée aux autres CCI par le test de Mann-Whitney ; les facteurs indépendants de variations du CCI ont été déterminés par une analyse de régression multiple. **Résultats** : 6 questions sur 16, l'usage de la grille, et l'un des couples de correcteurs étaient liés de façon indépendante aux CCI. **Conclusion** : La concordance des corrections d'une épreuve de LCA semble dépendre fortement de la qualité de l'épreuve et des correcteurs ; le recours à une grille de correction pourrait représenter un facteur d'amélioration d'impact modeste de cette concordance.

Mots clés Lecture critique d'article ; évaluation ; concordance inter-juges ; examen national classant.

Abstract **Context**: Early from 2008, a test of scientific paper critical analysis (SPCA) ability will be implemented in the French National Medical Sorting Exam (NSE). Few papers have been already published on this topic, suggesting a poor reliability of the test marking process. **Aim**: To identify factors associated with the reliability of SPCA test marking. **Subjects and methods**: On a voluntary basis, 59 medical students from the 4th or the 5th year at the Rouen University Medical School, completed a SPCA examination under the same conditions as for the NSE. Three teachers who prepared the assessment (choice of the article, questions and expected answer according to 16 items, correction grid and weighting) blindly marked twice questions and the abstract, without and then with a correction grid. The agreement among the three pairs of correctors, with and without a correction grid, for the 16 items (or 96 pairs) was assessed according to the intra class-correlation coefficients (ICC). Median CCI for pairs defined by each factor (items, using a grid or not, questions or abstract, pairs of teachers) was compared with the remaining pairs using a Mann-Whitney test. Independent factors were identified using multiple regression analysis. **Results**: Six items out of 16 using a correction grid and one pair of teachers out of 3 were independently associated with ICC. **Conclusion**: The reliability of this SPCA test marking method highly depends on the quality of the test and the correctors. Using a correction grid could only marginally improve the reliability.

Key words Paper critical appraisal ; assessment; reliability; national ranking exam.

Pédagogie Médicale 2006;7:82-90

Introduction

La lecture critique d'article (LCA) vise à préparer les étudiants en médecine à l'exercice de leur profession en leur apprenant à critiquer¹ de façon autonome leur documentation professionnelle. L'acquisition de cette habilité doit être évaluée ; cette évaluation en condition de concours est cependant mal connue. Or, une épreuve de LCA doit figurer à partir de 2008 dans l'examen classant national (ECN), qui détermine en France l'orientation des futurs médecins vers les spécialités possibles.

Le modèle d'épreuve choisi par le Centre national de concours de l'internat (CNCI) associe six à dix questions à réponse ouverte et la rédaction d'un résumé de moins de 250 mots portant sur un article original rédigé en français².

But

Dans des travaux précédents, la concordance des corrections sans grille d'une épreuve de LCA avait semblé médiocre^{3,4} ; cependant, certains facteurs qui pourraient influencer cette concordance sont susceptibles d'être maîtrisés par les organisateurs d'une telle épreuve : choix des questions, utilisation d'une grille de correction, choix des correcteurs. Le but de ce travail est d'identifier parmi ces facteurs ceux liés de façon indépendante à la concordance des corrections.

Matériel et méthodes

La méthodologie détaillée de cette étude a déjà été présentée³. Brièvement, 59 étudiants volontaires inscrits dans le 2^e cycle des études médicales à l'unité de formation et de recherche de médecine et de pharmacie de Rouen [25 en deuxième (DCEM 2) et 34 en troisième année du 2^e cycle des études médicales (DCEM 3)] ont été soumis à une épreuve test de LCA dans les conditions de l'ECN. Cette épreuve était consacrée à un article⁵ accompagné de 12 questions ou sous-questions portant sur le texte de l'article et à la rédaction du résumé en quatre parties ; elle a été préparée par six enseignants, puis validée par un groupe national d'experts auprès du CNCI³. Les étudiants n'avaient pas fait l'objet d'une formation spécifique à ce

type d'épreuve⁶, en particulier pour l'élaboration d'un résumé, et n'avaient pas préparé ce test.

Trois correcteurs ayant préparé l'épreuve (choix d'article, questions et réponses attendues) ont corrigé de façon indépendante les copies rendues anonymes : une première fois sans grille de correction, en confrontant les réponses avec la correction-type élaborée préalablement. Après plusieurs semaines, du fait du recueil des scores et des échanges des copies entre correcteurs, chaque correcteur a effectué une seconde correction, en utilisant cette fois une grille de correction détaillée comportant les réponses précises attendues aux 12 questions et aux quatre parties du résumé. Aucune concertation entre correcteurs n'est intervenue en cas de divergence de notation, même importante.

Les coefficients de corrélation intra-classe (CCI)⁷ ont été calculés pour chacun des trois couples de correcteurs, chacun avec ou sans grille, avec les scores totaux obtenus pour l'ensemble des questions (hors résumé), d'une part, et l'ensemble du résumé, d'autre part ; dans un deuxième temps, ce même calcul a été effectué avec les couples de scores portés à chacune des 16 questions ou parties de résumé, pour chaque couple de correcteur, avec ou sans grille, soit en tout 96 CCI.

La valeur médiane des CCI obtenus avec chaque variable étudiée (chacune des 16 questions ou partie de résumé, le fait d'utiliser une grille ou pas, le fait qu'il s'agisse d'une question ou d'une partie du résumé, chacun des trois couples possibles de correcteurs) a été comparée à celle des autres CCI à l'aide d'un test de Mann-Whitney ; enfin pour déterminer parmi les variables précédentes, celles liées de façon indépendante aux CCI, une analyse de régression multiple a été réalisée.

Résultats

Les scores établis à l'issue des corrections des questions et du résumé de l'épreuve de LCA chez les 59 étudiants, respectivement par les trois correcteurs, avec et sans grille de correction sont fournis en *annexe*. Le *tableau 1* montre les CCI calculés pour les trois couples de correcteurs, sans grille et avec grille, en distinguant le score total affecté à l'ensemble des questions hors résumé et celui affecté au résumé ; l'utilisation de la grille améliorerait la concordance

1- Département d'Epidémiologie et de Santé Publique - Faculté de Médecine - 76000 Rouen - France

2- Service de Néphrologie; même adresse

3- Laboratoire de Biochimie; même adresse

4- Département de Microbiologie; même adresse

5- Département d'Anatomie et de Cytologie Pathologiques; même adresse.

Correspondance : Pierre Czernichow - Département d'Epidémiologie et de Santé Publique - CHU de Rouen - F76031 Rouen Cedex France - Tel : 02 32 88 80 50 - Mailto: Pierre.Czernichow@chu-rouen.fr

Tableau 1 :
Coefficients de corrélation intra-classe (CCI) et intervalle de confiance à 95 % (IC 95 %)
des corrections de 59 copies d'épreuve de lecture critique d'article par 3 couples de correcteurs

	Correcteurs					
	Paire XY		Paire XZ		Paire YZ	
	CCI	IC95 %	CCI	IC95 %	CCI	IC95 %
Questions						
Sans grille	0,79	0,67 - 0,87	0,53	0,32 - 0,69	0,28	0,03 - 0,50
Avec grille	0,83	0,74 - 0,90	0,84	0,74 - 0,90	0,76	0,63 - 0,85
Résumé						
Sans grille	0,49	0,27 - 0,66	0,47	0,25 - 0,65	0,72	0,57 - 0,82
Avec grille	0,75	0,61 - 0,84	0,72	0,57 - 0,82	0,60	0,42 - 0,74
Total						
Sans grille	0,69	0,53 - 0,81	0,71	0,56 - 0,82	0,48	0,26 - 0,65
Avec grille	0,80	0,68 - 0,87	0,84	0,75 - 0,90	0,72	0,57 - 0,82

des corrections pour deux des trois couples de correcteurs. La valeur médiane des CCI différait pour plusieurs des facteurs étudiés (*Tableau 2*) : pour six des 16 questions (trois questions ayant un CCI significativement supérieur aux autres et trois autres significativement inférieur aux autres), l'ensemble des questions du résumé, l'un des trois couples de correcteurs (XY) et l'utilisation d'une grille.

Enfin, l'analyse en régression multiple montrait que pour tous les facteurs identifiés comme associés aux variations de CCI à l'analyse unifactorielle, à l'exception des questions du résumé, cette liaison était indépendante (*Tableau 3*).

Discussion

Les résultats de cette étude réalisée dans le contexte d'un faculté de médecine particulière, portant sur un seul article et une seule épreuve, organisée par un petit nombre d'enseignants motivés et administrée à des étudiants volontaires, ne peuvent être examinés qu'avec précaution. Leur généralisation à une réalité bien plus vaste, comme celle des quelques 5000 étudiants en médecine qui achèveront leur 2^e cycle en France à partir de 2008, est discutable ; d'autres travaux de même type sont nécessaires avant de conclure. Toutefois, les spécificités même de cette étude pouvaient faire attendre une bonne concordance des corrections ; en outre, l'attention consacrée à la préparation de cette épreuve, assurée par les enseignants mêmes qui l'ont corrigée, étaient des

raisons supplémentaires de bonne concordance entre correcteurs.

Or, la concordance observée globalement dans ces circonstances est médiocre. On peut donc craindre que l'impact de facteurs tels que ceux que nous avons étudiés, sur une épreuve organisée à une échelle plus large, soit encore plus important lorsque la diversité des épreuves et des facultés fera intervenir une source de variabilité supplémentaire.

Le résultat essentiel de notre travail est le constat d'une influence majeure du choix des questions sur la concordance des corrections des épreuves de LCA, malgré les efforts consacrés à sa préparation ; cette influence peut être favorable, en améliorant cette concordance, comme nous l'avons constaté pour trois questions, ou bien défavorable, en la réduisant, pour trois autres questions.

On peut s'interroger sur les raisons de cette influence : caractère analytique ou synthétique de la réponse attendue, objectif poursuivi par la question (exploration de connaissances générales, compréhension des méthodes ou des résultats de l'article, discussion et interprétation de ces résultats), ces objectifs posant des difficultés croissantes. Il est également possible que le choix même de l'article ait une influence sur la concordance des corrections, ce que nous ne pouvions pas tester. Ce résultat suggère le besoin d'analyser plus avant les caractéristiques des questions d'une telle épreuve, liées à la concordance de leurs corrections.

Un autre résultat préliminaire est le constat que la concordance des corrections pourrait être également influencée par les correcteurs ; ce résultat, établi sur un échantillon

Tableau 2 :
Facteurs associés aux variations des coefficients de corrélation intra-classe (CCI)
issus des corrections de 59 copies d'épreuve de lecture critique d'article par 3 couples de correcteurs :
résultats de l'analyse unifactorielle

Facteur étudié	Nombre de CCI ^a	CCI (médiane)	Minimum	Maximum	p ^b
Questions et parties de résumé					
Question 1	6	0,67	0,19	0,84	ns
Autres questions ou parties	90	0,56	-0,07	0,98	
Question 2	6	0,70	0,30	0,89	ns
Autres questions ou parties	90	0,56	-0,07	0,98	
Question 3	6	0,28	0,00	0,65	*
Autres questions ou parties	90	0,57	-0,07	0,98	
Question 4	6	0,86	0,73	0,90	***
Autres questions ou parties	90	0,57	-0,07	0,98	
Question 5	6	0,31	0,19	0,84	ns
Autres questions ou parties	90	0,57	-0,07	0,98	
Question 6	6	0,36	0,27	0,52	*
Autres questions ou parties	90	0,58	-0,07	0,98	
Question 7	6	0,44	0,36	0,81	ns
Autres questions ou parties	90	0,57	-0,07	0,98	
Question 8	6	0,47	0,27	0,67	ns
Autres questions ou parties	90	0,57	-0,07	0,98	
Question 9	6	0,85	0,56	0,87	**
Autres questions ou parties	90	0,55	-0,07	0,98	
Question 10	6	0,69	0,56	0,77	ns
Autres questions ou parties	90	0,55	-0,07	0,98	
Question 11	6	0,58	0,10	0,76	ns
Autres questions ou parties	90	0,56	-0,07	0,98	
Question 12	6	0,73	0,45	0,98	ns
Autres questions ou parties	90	0,56	-0,07	0,98	

Recherche et Perspectives

Tableau 2 (suite)

Facteur étudié	Nombre de CCI ^a	CCI (médiane)	Minimum	Maximum	p ^b
Questions et parties de résumé					
Résumé – partie 1	6	0,58	0,42	0,72	ns
Autres questions ou parties	90	0,56	-0,07	0,98	
Résumé – partie 2	6	0,55	0,37	0,60	ns
Autres questions ou parties	90	0,57	-0,07	0,98	
Résumé – partie 3	6	0,49	0,45	0,63	ns
Autres questions ou parties	90	0,57	-0,07	0,98	
Résumé – partie 4	6	0,22	0,07	0,45	**
Autres questions ou parties	90	0,58	-0,07	0,98	
Section de l'article					
Toutes questions	72	0,60	0,00	0,98	
Résumé (Toutes parties)	24	0,48	-0,07	0,72	ns
Correcteurs					
Paire XY	32	0,68	-0,07	0,98	**
Paire XZ	32	0,46	0,11	0,89	ns
Paire YZ	32	0,52	0,00	0,857	ns
Utilisation d'une grille					
Oui	48	0,61	0,13	0,98	*
Non	48	0,48	-0,07	0,90	
Total	96	0,56	-0,07	0,98	

Les effectifs correspondent à toutes les combinaisons des 96 CCI associées aux facteurs étudiés : les 3 couples de correcteurs (n = 32 CCI chacun), chacun avec ou sans grille (n = 48 CCI à chaque fois) ont corrigé chaque question élémentaire (n = 6 CCI), les 12 questions de l'article (n = 72 CCI) et les 4 parties du résumé (n=24) ;

^b Test de Mann-Whitney :

- ns : non significatif ;
- * p < 0,05 ;
- ** p < 0,01 ;
- *** p < 0,001.

Tableau 3 :
Facteurs indépendants associés aux variations du coefficient de corrélation intra-classe (CCI) issus des corrections de 59 copies d'épreuve de lecture critique d'article par 3 couples de correcteurs : résultats de l'analyse en régression multiple

Facteur étudié	Coefficient ^a	Intervalle de confiance à 95%		Test F partiel
Résumé – partie 4	-0,32	-0,47	-0,16	16,99
Question 3	-0,24	-0,38	-0,10	11,58
Question 6	-0,17	-0,31	-0,03	5,82
Usage d'une grille	+0,11	0,05	0,18	11,56
Paire de correcteurs XY	+0,13	0,06	0,20	13,20
Question 12	+0,18	0,04	0,32	6,54
Question 9	+0,24	0,10	0,38	11,60
Question 4	+0,30	0,16	0,44	17,50

^a Ce coefficient mesure l'influence propre de chaque facteur étudié dans les variations des CCI : un signe + indique que ce facteur concerné est associé à une concordance supérieure, un signe - indique que ce facteur est associé à une concordance moindre

particulier et limité, et seulement trois correcteurs, reste à confirmer. Sous cette réserve, il serait utile d'identifier les caractéristiques des enseignants qui sont associés à la concordance de leurs corrections : expérience pédagogique, ancienneté, discipline d'enseignement, par exemple ; la connaissance de ces caractéristiques pourrait permettre, le cas échéant, de solliciter plus particulièrement certains enseignants pour corriger les épreuves de LCA. Une autre piste de recherche à explorer serait une correction multiple, impliquant pour chaque copie un nombre d'enseignants supérieur à celui utilisé dans ce travail.

La moindre concordance des corrections de quatre items de résumé, observée à l'analyse unifactorielle, a disparu à l'analyse en régression multiple ; il semble que ce soit la rédaction de la conclusion de ce résumé qui ait entraîné le plus de divergences entre correcteurs. Ce point pourrait être amélioré par une formation à la rédaction de cette partie de l'épreuve, qui n'avait pas été réalisée dans les conditions de notre travail.

Enfin, un dernier résultat important concerne l'usage d'une grille de correction : comme attendu, un effet favorable sur la concordance des corrections a été constaté et notre travail montre que cette influence est indépendante des autres facteurs de variabilité. Conformément à des travaux publiés⁴, la grille utilisée dans ce travail comportait des éléments de contenus détaillés, propres à l'épreuve ; une telle grille nous semble donc indispensable pour la

correction de cette épreuve ; toutefois nos résultats suggèrent que l'amélioration de la concordance qui y est associée pourrait être modeste : son impact semble bien plus faible que celui du choix des questions dans notre travail ; sous réserve de confirmation, la grille ne garantit pas à elle seule la concordance des corrections d'un examen classant fondé sur une telle épreuve.

Conclusion

Ce travail suggère que la concordance des corrections d'une épreuve de LCA est très dépendante des questions de l'épreuve ainsi que des correcteurs ; l'usage d'une grille spécifique et détaillée est un moyen marginal de l'améliorer. Une attention particulière devrait être consacrée à l'avenir au sein du CNCI à l'élaboration de ces épreuves et à leurs modalités de correction.

Contributions

Pierre Czernichow et Francis Roussel ont participé à l'élaboration du dossier utilisé pour l'étude, corrigé les épreuves, analysé les données et rédigé le manuscrit. Jean-Paul Fillastre a participé à l'élaboration du dossier utilisé pour l'étude, et corrigé les épreuves. Alain Lavoinne et Jean-François Lemeland ont participé à l'élaboration du dossier utilisé pour l'étude.

Annexe :
Résultats des corrections des questions (Q) et du résumé (R)
de l'épreuve de lecture critique d'article (LCA) chez 59 étudiants de l'UFR de Rouen
par 3 correcteurs (X, Y et Z), avec et sans grille de correction

Etudiant	Avec grille de correction						Sans grille de correction					
	Correcteur X		Correcteur Y		Correcteur Z		Correcteur X		Correcteur Y		Correcteur Z	
	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R
1	28	13	34	13	26	12	40	21	35	16,5	41,5	17
2	16	0	13	0	14	0	27	0	16,5	0	26	0
3	16	0	19	0	15	0	26	0	16,5	0	25,5	0
4	12	8	13	8	13	3	22	21	19	14	26,5	10
5	21	10	28	16	23	11	26	15	26,5	13,5	42,5	19
6	19	7	28	11	18	4	35,5	11	30	8,5	38	8
7	27	9	30	12	24	8	37	20,5	24,5	16,5	44,5	17
8	15	13	11	13	15	12	24	19	20,5	15	29	22
9	20	13	22	15	20	11	30,5	19	26	17	41,5	17
10	27	13	25	12	28	10	31	14	29	13,5	41,5	17
11	38	15	34	17	30	13	43,5	19	39	19	42	21
12	14	7	14	12	21	9	21	14	12	18,5	29	12
13	12	12	15	19	22	11	19	14	12,5	12,5	29	12
14	18	14	17	12	23	6	20	14	13,5	8,5	33	10
15	20	11	24	14	23	9	26	10,5	21	12	22	12
16	20	16	19	21	26	16	29	25	22,5	14	33	17
17	18	15	17	14	19	7	25	15	24	18,5	21	15
18	15	7	16	13	18	6	31	12	24	14,5	30	15
19	30	11	32	11	27	9	32	13	33	13,5	27	16
20	17	15	21	17	17	11	24	14	27,5	19	24	19
21	20	15	22	15	22	12	24,5	13	28,5	20	23	17
22	17	9	18	14	14	10	26	12	20	19,5	21	19
23	29	11	32	13	29	8	34	9	34,5	12	38	17
24	14	18	20	18	24	12	26,5	20	23	20,5	41	20
25	21	11	22	13	21	8	30	15	24	19,5	51	21
26	21	10	25	13	22	9	28	18	28	14	43	18
27	26	14	19	15	28	10	29	16,5	23,5	18	40	17
28	16	8	16	9	23	5	25	19,5	22	16,5	27	14
29	6	8	12	10	5	7	15	15,5	9,5	16	14	16

Etudiant	Avec grille de correction						Sans grille de correction					
	Correcteur X		Correcteur Y		Correcteur Z		Correcteur X		Correcteur Y		Correcteur Z	
	Q	R	Q	R	Q	R	Q	R	Q	R	Q	R
30	33	19	39	19	1236	12	47,5	19	50,5	23,5	57	22
31	23	13	26	15	24	10	34	24,5	24,5	15,5	47	16
32	20	10	22	15	19	11	30,5	19,5	26	20,5	28	15
33	19	10	21	13	31	5	33	19	31	14	42	15
34	18	16	12	20	12	6	22	17	18,5	14	40	14
35	25	6	22	5	30	2	24,5	12	23,5	11,5	28	7
36	20	15	24	20	25	11	21,5	16	24	20,5	23	15
37	15	12	11	15	16	12	20,5	17	15	9	20	14
38	22	16	23	18	23	14	26,5	17	29	13,5	30	14
39	17	15	12	19	17	7	18,5	18	23,5	15	30	10
40	14	12	12	13	15	11	14,5	20	17,5	16,5	27	14
41	26	15	33	18	27	11	43	13	39	15	38	16
42	24	16	25	15	29	15	38	20	28,5	17	34	16
43	14	14	13	19	12	12	17,5	20,5	21	16,5	25	9
44	17	11	24	15	18	7	25	17	15	15	35	8
45	15	9	17	8	18	7	24	16	16	5	25	10
46	28	11	36	15	28	11	36	22	33,5	8,5	45	17
47	22	15	22	15	16	8	28	19	28,5	15	36	10
48	21	9	18	10	23	7	27,5	12	23,5	10	36	12
49	30	17	33	16	29	15	34,5	18	28	18	43	15
50	16	15	16	16	13	9	26,5	16,5	20	15	35	13
51	34	13	37	16	38	11	40	16	32,5	13	49	13
52	13	13	13	16	11	11	20	16	11,5	9	36	17
53	11	14	12	16	12	12	19	18,5	18,5	12	26	15
54	14	10	17	8	16	4	19	13,5	14	7,5	28	3
55	21	9	22	14	26	10	29	12,5	26	16	34	14
56	12	8	14	14	13	9	18	16	10,5	10	22	11
57	38	13	42	16	47	12	51	18,5	48,5	22	55	21
58	19	14	20	17	16	13	24	16	22,5	16	39	16
59	20	11	25	13	28	8	25,5	13	30,5	15	41	13
Médiane	20	12	21	15	22	10	26,5	16	24	15	34	15

Références

1. Westwood B, Westwood G. *The culture of criticism and argument in health education*. *Med Teach* 2002;24:156-61.
2. Anonyme: Arrêté du 29 Janvier 2004 (NOR: SANH0420335A) relatif à l'organisation des épreuves classantes nationales anonymes donnant accès au troisième cycle spécialisé des études médicales. *Journal Officiel de la République Française* du 5 Février 2004:2519-21.
3. Roussel F, Czernichow P, Lavoinne A, Lemeland JF, Fillastre JP. *Reproductibilité de la correction d'une épreuve de lecture critique d'article. Evaluation par une étude pilote chez 59 étudiants en médecine*. *Pédagogie Médicale* 2005;6:71-8.
4. Louis-Sylvestre C, Fuhman C, Housset B. *Difficultés de correction d'une épreuve d'analyse critique d'article scientifique: une étude exploratoire*. *Pédagogie Médicale* 2005;6:138-46.
5. Roudot-Thoraval F, Monnet E, Mercet P, Bastie A, Dhumeaux D, Miguet JP. *Stratégies de dépistage de l'hépatite virale C en Médecine générale. Résultats d'un essai randomisé bicentrique*. *Gastroenterol Clin Biol* 2000;24:1037-41.
6. Audet N, Gagnon R, Ladouceur R, Marcil M. *L'enseignement de l'analyse critique des publications est-il efficace ? Révision des études et de leur qualité méthodologique*. *Can Med Assoc J* 1993;148:945-52.
7. Fermanian J. *Mesure de l'accord entre deux juges : cas quantitatif*. *Rev Epidem Santé Publ* 1984;32:408-13.

Manuscrit reçu le 19 février 2005 ; commentaires éditoriaux formulés aux auteurs le 3 novembre 2005 ; accepté pour publication le 28 février 2006.