

Évaluation d'une grille de correction révisée de l'exercice de rédaction de résumé pour l'examen de lecture critique d'article scientifique dans le cadre des épreuves classantes nationales en France

Evaluation of a modified assessment grid of a written summary aptitude for the scientific article critical reading test in the context of the national ranking exam in France

Fabrice MUSCARI¹, Jean-Louis MONTASTRUC² et Gilles FOURTANIER¹

1 Service de Chirurgie Digestive, CHU Rangueil, Toulouse, France

2 Service de Pharmacologie Médicale et Clinique, Faculté de Médecine, Toulouse, France

Mots clés :

Lecture critique d'article ; grille de correction ; résumé ; mots clés

Résumé – But : Tester une nouvelle grille de correction des résumés afin d'obtenir une réduction de la variation inter-correcteurs des scores. **Matériel et Méthodes :** La grille de correction permettait d'élaborer un score sur la base d'une évaluation « globale » et d'une évaluation de « mots clés », ce qui en faisait l'originalité. Cinquante copies ont été recueillies pour ce travail. Chaque copie a fait l'objet d'une correction par cinq correcteurs différents. Les critères de jugement étaient : la dispersion des cinq scores obtenus par chaque copie en calculant la moyenne ; le pourcentage de variation des notes en fonction des correcteurs ; la dispersion des notes de chacun des scores pour chaque correcteur par rapport à la moyenne calculée sur l'ensemble des copies. **Résultats :** Cette grille de correction modifiée avec l'adjonction de mots clés n'entraîne pas de diminution des variations des notes en fonction du correcteur. La notation de la partie mots clés entraîne même une grande variation des notes d'un correcteur à l'autre : 50 % des copies avaient une variation de note inter-correcteur de plus de 40 %. De même, cette nouvelle grille de correction entraîne une grande dispersion des notes en fonction de l'expérience du correcteur. **Conclusion :** Cette nouvelle grille de correction du résumé diminue peu les variations de scores inter-correcteurs pour une même copie. Il faut donc tester d'autres grilles de correction du résumé de la lecture critique pour assurer une équité de correction pour les étudiants.

Keywords:

Scientific article critical reading;

Abstract – Aim of study: To evaluate a new assessment grid of written summary in order to obtain a reduction of the variation of scores between different correctors. **Material and Method:** The assessment grid allowed to elaborate a score based on a “global”

assessment grid ;
summary ;
key word

evaluation and a score obtained from “key words”, which made it original. Fifty copies were collected for this work. Five different correctors collected each copy. the criteria of judgement were the following: 1) dispersion of the five scores obtained for each copy by calculating the average; 2) percentage of variation of the scores according to correctors; and 3) dispersion of the value of each score for each corrector compared to the average calculated from all copies. **Résultats :** This modified assessment grid with the addition of key the words does not lead to a reduction of scores variation according to the corrector. The key word scoring leads even to a greater variation of scores between corrector: 50% of copies had a variation of scores inter-corrector of more than 40%. Also, this new assessment grid conducts to a great dispersion of scores according to the corrector experience. **Conclusion:** This new assessment grid of written scientific article critical reading summaries weakly diminishes the variation of scores between correctors for the same copy. It is thus necessary to test of other assessment grids to ensure equity of correction for students.

Introduction

En France, depuis 2004, tous les étudiants en médecine sont soumis à l’obligation de passer des épreuves classantes nationales (ECN) anonymes à l’issue de leur deuxième cycle d’études, c’est-à-dire à la fin de leur cursus « pré-gradué »^[1,2]. Cet examen détermine à la fois la filière de formation spécialisée à laquelle pourra accéder chaque étudiant ainsi que l’université et le centre hospitalier universitaire de rattachement, au sein desquels il recevra sa formation spécialisée ; il s’agit donc d’un examen dont le résultat est associé à de forts enjeux personnels mais aussi sociaux^[3]. Ces épreuves reposent à l’heure actuelle sur l’analyse de neuf dossiers cliniques, à propos desquels les capacités des étudiants dans les champs du diagnostic et de la thérapeutique sont explorées à l’aide de quatre à 10 questions à réponses ouvertes.

Ces dernières années, dans le sillage du courant de la médecine fondée sur des données probantes (*evidence-based medicine* – EBM)^[4], il est devenu clair que la capacité à analyser de façon critique les nouvelles informations scientifiques était une dimension de la compétence médicale dont il convenait de se préoccuper explicitement, tant en formation médicale initiale qu’en formation

médicale continue^[5]. Ceci a conduit à envisager dès 2001 de compléter les épreuves de l’internat puis les ECN par la mise en place d’une épreuve de lecture critique d’article (LCA), avec un coefficient qui devait représenter 20 % de la note totale de l’examen^[2,6]. Cette nouvelle épreuve de trois heures concerne un article original au sujet duquel deux types d’exercices sont demandés aux étudiants. D’une part, six à dix questions rédactionnelles explorent les aspects méthodologiques du travail de recherche rapporté ; d’autre part, les étudiants doivent élaborer un résumé structuré de l’article, dans la limite maximale de 250 mots. La perspective de l’introduction de cette épreuve a suscité dès le début des réserves concernant les risques d’inégalité pour une épreuve à fort coefficient dont les modalités de correction n’avaient pas été vérifiées, notamment en ce qui concerne la fidélité inter-correcteurs^[3,6]. Un rapport de mission commandé par le ministre de l’enseignement supérieur et de la recherche^[7] a néanmoins confirmé récemment (avril 2007) la légitimité de cette épreuve dans le cadre des ECN, son introduction effective étant simplement différée à juin 2009, pour permettre aux différentes facultés d’y préparer de manière satisfaisante leurs étudiants. Peu de travaux^[6,8,9] se sont intéressés à la reproductibilité de la correction de l’épreuve de LCA, mais il semble que les plus grandes variations d’un

correcteur à l'autre^[10,11] sont surtout observées lors de la correction de la rédaction du résumé, des variations du score pouvant aller jusqu'à 60 % pour la même copie ayant été constatées^[8].

En 2005, Louis-Sylvestre *et al.*, à l'issue d'un travail étudiant l'aide qu'apportait l'utilisation d'une grille de correction, suggéraient que les variations inter-correcteurs lors de la correction du résumé pourraient être diminuées en optimisant la grille de correction grâce à l'ajout de mots clés. C'est cette hypothèse que nous avons voulu tester dans le cadre du présent travail.

Matériel et méthodes

Matériel

Ce travail a porté sur la première épreuve de lecture critique d'article organisée au sein des deux facultés de médecine de Toulouse, en février 2005. Cet examen a concerné 180 étudiants. Nous leur avons proposé un article d'ophtalmologie^[12] rapportant une étude de validation d'un test diagnostique de la sécheresse oculaire. Le résumé avait été effacé. Conformément aux modalités prévues dans le cadre des ECN, les étudiants devaient répondre à 10 questions et faire le résumé de l'article dans un temps limité. Sur la base du programme des cours de pathologie dispensés au cours de l'année, nous avons postulé que l'article était analysable et compréhensible par tous. La technique du résumé avait été enseignée de façon formelle durant l'année. Nous avons limité l'analyse des copies à celles des étudiants d'une seule des deux facultés, soit 80 copies. Nous avons exclu les copies dont la longueur des réponses dépassait le nombre de mots imposés (250), afin d'éviter les biais supplémentaires liées à d'éventuelles différences de pénalisation des copies pour ce motif d'un correcteur à l'autre, comme cela a été évoqué dans le travail de François *et al.*^[10], dans la mesure où aucun barème de pénalité pour le dépassement du nombre de mots n'est clairement défini par le centre national des concours d'interne (CNCI) et que ceci est laissé au choix du jury

qui corrigera l'épreuve. Au total, nous avons exclu 30 copies (37 %) et avons travaillé à partir de 50 copies.

Méthodes

Les 50 copies ont fait l'objet d'une correction par cinq correcteurs différents. Les correcteurs n'étaient pas des ophtalmologistes et avaient une expérience en terme de correction d'examen très variable. Il s'agissait d'un correcteur expérimenté (un professeur des universités ayant déjà participé à la correction des ECN), de trois correcteurs occasionnels (un praticien hospitalier et deux assistants-chefs de clinique des universités) et d'un correcteur novice n'ayant jamais corrigé de copies (interne de chirurgie). Ce choix, délibéré, cherchait à constituer une représentation des correcteurs « réaliste ».

L'étude a porté uniquement sur le résumé rédigé par les étudiants. Chacun était jugé de façon indépendante par les cinq correcteurs à partir d'une grille de correction type (annexe 1). Cette grille de correction a été élaborée en tenant compte des suggestions formulées dans le travail de Louis-Sylvestre *et al.*^[8] ; elle comportait donc une partie dite « globale » permettant au correcteur de calculer un score pour chaque paragraphe du résumé (section « introduction » cotée sur 8 points, section « matériel et méthodes » cotée sur 8 points, section « résultats » cotée sur 8 points, section « conclusion » cotée sur 6 points), soit un score total sur 30 points, et une partie dite « mots clés » où à chaque paragraphe étaient assignées une liste de 20 mots clés rapportant un point chacun. Cette dernière partie (mots clés) fait l'originalité de cette grille. Nous avons coté la totalité du résumé sur 50 points avec 30 points pour l'évaluation « globale » et 20 points pour l'évaluation à l'aide des mots clés. La répartition des points a été déterminée de façon arbitraire puisque aucune grille semblable n'a été publiée à ce jour. Le répertoire des mots clés avait été validé consensuellement par les personnes ayant choisi l'article ; les mots clés sont mentionnés en gras sur le résumé princeps présenté en annexe 2.

Tableau I. Scores moyens (partiels et totaux) attribués par les différents correcteurs à l'égard de la rédaction du résumé.

	Score Total	<i>p</i>	Score d'évaluation	<i>p</i>	Score d'évaluation à	<i>p</i>
	Moyenne +/- écart type		« Globale »		l'aide des	
			Moyenne +/- écart type		« Mots Clés »	
					Moyenne +/- écart type	
Correcteur 1 (PU)	27,6 +/- 2,8		19,4 +/- 1,9		8,0 +/- 1,4	
Correcteur 2 (CCA*)	24,5 +/- 3,4		18,5 +/- 2,2		6,0 +/- 2,0	
Correcteur 3 (PH)	25,0 +/- 4,5		16,4 +/- 2,8		8,6 +/- 2,2	
Correcteur 4 (CCA)	26,6 +/- 3,9	<0,0001	18,9 +/- 2,4	<0,0001	7,8 +/- 2,0	<0,0001
Correcteur 5 (IdC)	28,5 +/- 4,9		20,0 +/- 3,0		8,3 +/- 2,2	
Ensemble des correcteurs	26,5 +/- 3,9		18,6 +/- 2,5		7,8 +/- 1,9	

PU : professeur des universités

CCA : chef de clinique-assistant

PH : praticien hospitalier

IdC : interne de chirurgie

* Correcteur à l'origine de la grille.

Critères de jugement

Les différences des moyennes observées pour chaque correcteur, et pour chaque catégorie, ont fait l'objet d'une analyse de covariance. Après qu'un rang ait été affecté à chaque copie selon son score, un test de concordance de Kendall a été effectué pour vérifier que tous les correcteurs classaient les copies dans le même ordre sans forcément donner le même score.

Nous avons étudié dans un premier temps la dispersion des cinq scores attribués à chaque copie respectivement par les cinq correcteurs en calculant la moyenne (+/- écart type) et un pourcentage de variation (différence entre les deux scores extrêmes divisée par la moyenne). Ces pourcentages de variation ont été classés en trois classes (<20 %, 20 à 40 %, >40 %), pour le score total ainsi que pour les scores spécifiques de chacune des composantes

(évaluation « globale » et évaluation à l'aide des « mots clés »).

Nous avons ensuite étudié la dispersion des trois types de scores (score total, score de l'évaluation « globale » et score de l'évaluation à l'aide des « mots clés ») selon les correcteurs en dénombrant, pour chacun d'entre eux, la proportion de copies ayant reçu un score en deçà (sous-cotation) ou au-delà (sur-cotation) de la fourchette de +/- un écart-type par rapport à la moyenne calculée sur l'ensemble des copies.

Résultats

Variations inter-correcteurs

Les scores moyens attribués respectivement par chaque correcteur (tableau 1) pour chaque

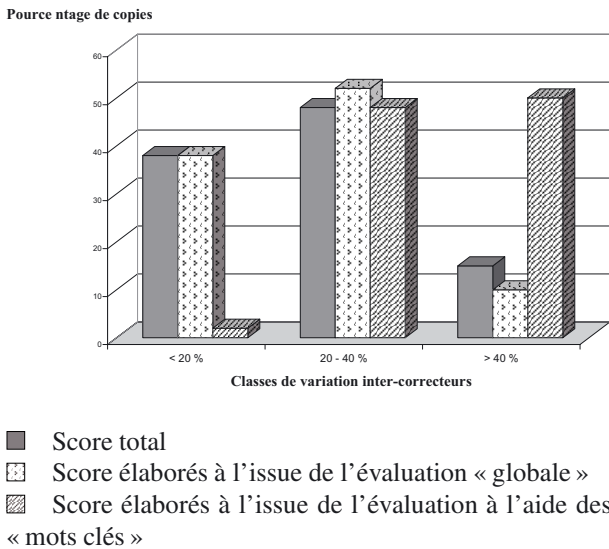


Fig. 1. Répartition du nombre de copies dans chacune des trois classes de variation inter-correcteurs pour le score total.

composante de la correction (score total, score pour l'évaluation « globale », score pour l'évaluation à l'aide des « mots clés ») sont statistiquement différents à l'analyse de covariance. Les scores attribués par chaque correcteur pour l'évaluation « globale » sont plus élevés que les scores élaborés à l'aide des « mots clés ». La moyenne des scores totaux attribués par le correcteur à l'origine de la grille de correction était la plus basse et celle des scores totaux attribuée par le correcteur novice la plus élevée.

Les répartitions des copies selon les trois classes de pourcentage de variation inter-correcteurs pour le score total, d'une part, et pour les sous-scores issus de l'évaluation « globale » et de l'évaluation avec les « mots clés », d'autre part, sont respectivement illustrées sur la figure 1. Les variations inter-correcteurs pour l'évaluation à l'aide des « mots clés » s'avèrent plus importantes que les variations pour l'évaluation « globale » (figure 1) : les scores obtenus à l'aide des « mots clés » de près d'une copie sur deux étaient caractérisés par une fourchette de notes extrêmes supérieure à 40 % de la moyenne des scores.

Profils de cotation selon les correcteurs

Le score moyen attribué par les différents correcteurs était de 26,5 \pm 3,9 (sur 50) pour le score total ; de 18,6 \pm 2,5 pour le score calculé avec l'évaluation « globale » (sur 30) et de 7,8 \pm 1,9 (sur 20) pour le score calculé avec les « mots clés ». Nous avons étudié pour chaque correcteur la dispersion de ses scores par rapport à ces moyennes, pour chercher et mesurer une sous- ou une sur-notation.

Concernant le score total, la proportion de copies sous-cotées variait, en fonction des correcteurs, de 0 à 15 % (médiane de 8 %) ; la proportion de copies sur-cotées variait pour sa part de 15 à 52 % (médiane de 24 %). Le correcteur ayant le plus sur-coté, était le correcteur novice n'ayant jamais corrigé. *A contrario*, le correcteur le plus expérimenté n'a fait aucune sous-cotation par rapport à la moyenne. Les proportions de copies respectivement sous- ou sur-cotées concernant le score issu de l'évaluation « globale » sont similaires à celles observées pour le score total.

Concernant les scores élaborés à l'aide des « mots clés », nous avons retrouvé en fonction des correcteurs, une proportion de 6 (pour le correcteur 1) à 36 % (pour le correcteur 5) des scores hors de la fourchette de \pm un écart type par rapport à la moyenne. On constate que l'expérience du correcteur entraîne moins de dispersion des scores par rapport à la moyenne. Seul le correcteur le plus expérimenté n'avait sous-coté aucune copie. Pour les autres correcteurs, une sous-cotation concernait de 4 à 26 % des copies, avec une sous-cotation plus fréquente pour le correcteur concepteur de la grille. La sur-cotation des copies concernait 8 à 32 % de celles-ci. Par contre, le correcteur inexpérimenté n'avait pas la plus grande dispersion des notes pour cette partie (à la différence des notes observées pour la partie « aspect global » et le score total).

Le test de Kendall montre que les correcteurs sont concordants dans leur manière de classer les copies mais avec un faible coefficient de concordance, de 0,489 pour le score total, de 0,510 pour le score

issue de l'évaluation « globale » et 0,236 pour le score issu de l'évaluation à l'aide des « mots clés ».

Discussion

L'addition d'un répertoire de mots clés à la grille d'évaluation de la rédaction du résumé lors de l'épreuve de LCA dans le cadre des ECN ne semble pas de nature à optimiser la fidélité inter-correcteurs lors de la correction de l'épreuve. En effet, 10 % des copies sont affectées de scores variant de plus de 40 % d'un correcteur à l'autre lorsque les performances de cinq correcteurs sont comparées. On observe aussi, contrairement à ce qui était attendu^[4,8] que l'adjonction pour le calcul du score total d'un score partiel élaboré à l'aide de l'évaluation de l'occurrence de « mots clés » ne diminue pas les variations des scores selon les correcteurs. L'élaboration de ce score partiel à l'aide de « mots clés » est même grevée d'une plus grande variation inter-correcteurs : près de 50 % des copies se répartissent respectivement dans la classe de variation relative de 20–40 % et autant dans celle supérieure à 40 %. Lorsqu'on étudie la dispersion par rapport à la moyenne des scores partiels issus de l'évaluation à l'aide de « mots clés » en fonction des correcteurs, on constate que le correcteur le plus expérimenté a le moins de dispersion avec une proportion de copies sur cotées de 6 %. Pour les autres correcteurs, la proportion de copie sous- ou sur-cotées varie de 20 à 50 % avec une sous-cotation pour le correcteur à l'origine de la grille et une sur-cotation pour les autres. Cette dispersion suggère que, en l'absence de consignes explicites, la prise en compte des « mots clés » dans les copies n'est pas effectuée de façon univoque. En réalité, les mots clés retenus étaient parfois des mots isolés mais plus souvent des groupes de mots, des locutions, voire des phrases quasi-complètes, de sorte que les difficultés de leur exploitation étaient en partie prévisibles. Il conviendrait peut-être de codifier de façon plus formelle la prise en compte de l'occurrence formelle de chaque « mot clé » ou celle de locutions voisines

dont sens implicite pourrait être tenu pour équivalent mais il n'est pas *a priori* certain qu'une telle mesure n'augmente pas encore la variabilité des scores en raison d'une extension des différences d'interprétation possibles. Une double correction avec calcul de la moyenne des deux scores et discussion entre les deux correcteurs lorsque la différence des deux scores dépasse 15 % pourrait être suggérée, comme cela a été fait dans le travail de Maruani *et al.*^[11]. Une telle disposition ne prémunirait cependant pas contre le risque qu'une copie victime de deux erreurs de mesure se produisant dans le même sens et avec le même ordre de grandeur ne soit pas révisée^[3]. Enfin, lorsqu'on examine le classement des copies dans le bon ordre, indépendamment des scores absolus attribués, on constate que, bien que la concordance de classement entre les correcteurs, telle qu'appréciée par le test de Kendall, soit significative, cette concordance est moins bonne pour le classement selon les scores issue de l'évaluation à l'aide des « mots clés » que pour celui effectué selon les scores issus de l'évaluation « globale ». Dans la perspective d'épreuves à finalité classante, cette « moins-value » n'est évidemment pas acceptable.

Dans la littérature, peu de travaux ont évalué la correction de l'épreuve de LCA telle qu'elle est prévue par le CNCI^[4,8-11]. Tous concluent à une variation inter-correcteurs qui est maximale pour la correction de la section dédiée à la rédaction des résumés. D'autres travaux^[13-16] ont évalué l'aptitude à la lecture critique de l'information médicale selon des formats très différents, les méthodes utilisées varient du simple questionnaire^[13] à une situation simulée avec un acteur dans le rôle d'un patient^[15].

Plusieurs limites de notre travail doivent être discutées. L'une d'entre elles tient au fait qu'il concerne un faible nombre de copies, qu'il a été fait sur une seule épreuve et sur une seule promotion d'étudiants. Pour rendre des résultats plus extrapolables, il faudrait réaliser cette étude sur plusieurs épreuves au cours de la même année et, dans l'idéal, sur plusieurs promotions d'étudiants afin d'avoir une

plus large population. Une autre limite est liée à son caractère « naturaliste », les épreuves ayant été corrigées par des correcteurs d'expérience variable, non spécifiquement entraînés à cette tâche ; ce parti pris était cependant délibéré et rend compte selon nous avec un certain réalisme des conditions réelles des corrections telles qu'elles seront mises en œuvre. Une réserve importante est enfin liée à l'exclusion d'un nombre important de copies (30, soit 37 % du nombre initial de copies) au motif que la longueur des réponses dépassait la limite impartie de 250 mots. Nous avons argumenté dans la section « méthodes » les raisons d'un tel choix mais il est difficile d'en inférer les conséquences de façon univoque. Dans la mesure où ont été écartées les copies les moins concises, on ne peut pas exclure que la dispersion des scores ait été encore plus grande si les 80 copies avaient été corrigées, ce qui serait de nature à accréditer davantage le défaut de fidélité inter-correcteurs lié à cette épreuve.

En dépit des réserves concernant la fidélité de sa correction, l'épreuve de LCA sera organisée dans le cadre des ECN organisées au titre de l'année universitaire 2008-09^[7], avec pour seul aménagement une diminution de son coefficient à 10 %, ce qui lui confère un poids relatif similaire à celui de chacun des neuf dossiers cliniques.

En réalité, les problèmes docimologiques constatés pour la correction de l'épreuve de LCA ne font probablement que traduire la complexité intrinsèque de l'activité de jugement critique à l'égard de l'information scientifique. D'une certaine manière, la recherche d'un exercice d'évaluation « idéal », dont la solution serait systématiquement univoque de manière à être garante d'une fidélité inter-correcteur, est relativement antinomique avec l'idée que les problèmes complexes admettent précisément, par nature, plusieurs solutions possibles. Ceci pourrait plaider pour l'utilisation de tests permettant de prendre en compte la dimension d'incertitude inhérente à la plupart des activités cognitives conduites dans le cadre de la pratique médicale. C'est le cas notamment du test de concordance de script (TCS)^[17], initialement développé pour

évaluer le raisonnement clinique et dont la validité a été documentée dans plusieurs disciplines^[18-20]. On pourrait très bien concevoir qu'une transposition adaptée à l'exploration de la capacité d'analyse critique de l'information scientifique soit développée. Ainsi, plutôt que de chercher à éliminer la propension des correcteurs à porter des jugements variables, on exploiterait au contraire cette variabilité « naturelle » pour établir un mode d'élaboration des scores. Cet examen (TCS) est facile à corriger (la correction pouvant même être informatisée) et ses qualités docimologiques, sous certaines conditions, peuvent être raisonnablement assurées^[17-20].

En conclusion, il n'est pas certain que l'adjonction d'un répertoire de mots clés aux grilles de correction usuellement utilisées pour corriger la rédaction du résumé lors de l'épreuve de LCA dans le cadre des ECN soit de nature à améliorer la concordance inter-correcteurs. Compte tenu des enjeux importants de cette épreuve, peut-être serait-il opportun d'explorer d'autres formats d'évaluation de la capacité de lecture critique de l'information scientifique, de manière à en optimiser plus significativement les qualités docimologiques, ce qui serait de nature à dissiper définitivement les réserves que cette épreuve a suscitées.

Remerciements

Nous remercions l'équipe de chirurgie digestive du CHU de Rangueil à Toulouse, qui a participé à la correction des copies : Pr Bertrand Suc, Dr Jean-Pierre Duffas (praticien hospitalier), Dr Anne Mathieu (chef de clinique – assistant des hôpitaux), et Bertrand Foppa (interne des hôpitaux).

Contributions

Fabrice Muscari a assuré la conception du travail, le recueil et l'analyse des données, l'écriture du manuscrit. Jean-Louis Montastruc et Gilles Fourtanier ont contribué à la révision du manuscrit.

Annexe 1

Grille de correction utilisée pour la correction de l'épreuve de lecture critique d'article

	0	2	4	6	8
Introduction					
. aspect global *	absence	erreurs	incomplète	complète et confuse	complète et confuse
. note des mots clés ¹ :	-	-	-	-	-
Matériel et Méthodes					
. aspect global **	absence	absence de l'un ou l'autre	erreurs	incomplet ou confus	complet et claire
. note des mots clés ¹ :	-	-	-	-	-
Résultats					
. aspect global ***	absence	pas de chiffres ou interprétation	résultats du CJP sans p ou RR	CJP seul avec p ou RR	CJP et critère Iir complets
. note des mots clés ¹ :	-	-	-	-	-
Conclusion					
. aspect global	absence erronée	non prouvé par l'article	incomplète ou confuse	idem à l'article complète	-
. note des mots clés ¹ :	-	-	-	-	-

Total de l'« aspect global » = / 30 points et Total des mots clés = / 20 points

TOTAL = / 50 points

* comprend un bref rappel du problème ; énonce l'hypothèse ; se termine par l'objectif de l'étude

** décrit les patients : nombre , critères d'inclusions ; décrit le critère de jugement principal ; décrit la structure de l'étude (prospective, randomisée, double aveugle, placebo, multicentrique, ...)

*** énonce les résultats du critère principal en donnant les chiffres (RR, OR, p-value) et les résultats des critères secondaires
CJP = critère de jugement principal = « valeur seuil à 12 mm », p ou RR = « sensibilité à 59 % et spécificité à 69 % » ;

Iir = secondaire = « concordance avec le test de shirmer »

¹ Les mots clés sont mentionnés dans l'annexe 2.

Annexe 2

RÉSUMÉ DE RÉFÉRENCE

LE TEST AU FIL IMPRÉGNÉ DE ROUGE PHÉNOL : PREMIERS RÉSULTATS POUR L'ÉVALUATION DE SEUIL DE POSITIVITÉ DANS LE DIAGNOSTIC DE SÉCHERESSE OCULAIRE.

But de l'étude : Déterminer la **valeur seuil** du test au fil imprégné de rouge phénol (Zone-Quick810[®], MeniconTM) pour le **diagnostic de sécheresse oculaire** à l'aide de la méthode ROC (Receiver Operating Characteristic), et estimer la **corrélation avec le test de Schirmer** (sans anesthésie).

Matériel et méthodes : **Cinquante quatre patients** consécutifs, dont 50 femmes, souffrant de sécheresse oculaire d'origine probablement

immunitaire ont été inclus dans l'étude sur le **double critère** des signes fonctionnels d'oeil sec et des signes extra-oculaires faisant suspecter un syndrome de Gougerot-Sjögren. Le test au fil imprégné de rouge phénol (FRP) et le test de Schirmer I, lus respectivement à 15 secondes et 5 minutes, ont été réalisés dans un ordre aléatoire. Seule la plus faible des deux valeurs obtenues (droite ou gauche) pour chaque test était retenue. Le même protocole a été appliqué pour **29 volontaires sains** ne présentant aucun signe fonctionnel de sécheresse oculaire. La **répartition en âge et en sexe était comparable** dans les deux groupes de sujets.

Résultats : La méthode **ROC** a montré que la valeur **seuil de 12 mm** pour le test au FRP est celle donnant la meilleure combinaison entre **sensibilité et spécificité (56 % et 69 %** respectivement). Dans ces conditions, la **concordance avec le test de Schirmer I est statistiquement très significative** (test Kappa, $p < 10^{-3}$), mais il reste cependant 32 % des sujets avec des **résultats discordants** entre les deux tests.

Conclusion : La **valeur seuil de 12 mm** pour le test au FRP permet de combiner une **sensibilité à 56 %**, une **spécificité à 69 %** et une bonne **concordance avec le test de Schirmer I**. Cette dernière est cependant **imparfaite**. Ces deux méthodes d'exploration fonctionnelle de la sécrétion lacrymale sont donc **complémentaires**, et des études supplémentaires restent nécessaires pour mieux connaître la place à donner à chacun d'entre eux, notamment dans le cadre du diagnostic de syndrome de Gougerot-Sjögren.

→ **Les Mots clés** sont surlignés en **gras**

Références

- Décret n° 2004-67 du 16 janvier 2004 relatif à l'organisation du troisième cycle des études médicales. Journal officiel de la République française du 18 janvier 2004;1394-95
- Arrêté du 29 janvier 2004 relatif à l'organisation des épreuves classantes nationales anonymes donnant accès au 3^e cycle spécialisé des études médicales. Journal officiel de la République française du 5 février 2004;2519-21
- Jouquan J. Quelques questions posées par l'épreuve d'évaluation de la lecture critique d'articles scientifiques. *Pédagogie Médicale* 2005;6:136-7
- Evidence-based Medicine working group. Evidence-based Medicine: a new approach to teaching the practice of Medicine. *JAMA* 1992;268:2420-5
- Huguier M. Epreuve de lecture critique d'articles médicaux. *Presse Med* 2007;36:565-6
- Roussel F, Czernichow P, Lavoine A, Lemeland JF, Fillastre JP. Reproductibilité de la correction d'une épreuve de lecture critique d'article : évaluation par une étude pilote chez 59 étudiants en médecine. *Pédagogie Médicale* 2005;6:71-8
- Fagniez PL. Evaluation de la mise en place de l'enseignement de la lecture critique d'article dans les facultés de médecine et opportunité de son intégration aux épreuves classantes nationales. Rapport de mission au Ministre délégué de l'Enseignement supérieur et de la Recherche. 2007 [On-line] Disponible sur : <http://qlf-media-education-gouv-fr.aw.atosorigin.com/file/08/7/20087.pdf>
- Louis-Sylvestre C, Furchman C, Housset B. Difficultés de correction d'une épreuve d'analyse critique d'article scientifique : une étude exploratoire. *Pédagogie Médicale* 2005;6:138-46
- Czernichow P, Fillastre JP, Lavoine A, Lemeland JF, Roussel F. Facteurs liés à la concordance des corrections d'une épreuve de lecture critique d'article. *Pédagogie Médicale* 2006;7:82-90
- François P, Boyer L, Nuiroy L, Labarère J. Fiabilité et validité d'une épreuve de lecture critique d'articles médicaux scientifiques. *Presse Med* 2007;36:575-81
- Maruani A, Giraudeau B, Alison D, Bertrand P, Bourlier P, Fauchier L *et al.* Epreuve de lecture critique d'article scientifique des épreuves classantes nationales : pertinence de la correction. *Presse Med* 2007;36:571-4
- Labetoulle M, Mariette X, Joyeau L, Baudoin C, Kirsh O, Offret H *et al.* Le test au fil imprégné de rouge phénol : premiers résultats pour l'évaluation de seuil de positivité dans le diagnostic de sécheresse oculaire. *J Fr Ophtalmol* 2002;25:674-80
- Stern Dt, Linzer M, O'Sullivan PS, Weld L. Evaluating medical residents' literature-appraisal skills. *Acad Med* 1995;70:152-4

14. Green ML, Ellis PJ. Impact of an evidence based Medicine curriculum based on adult learning theory. *J Gen Intern Med* 1997;12:742-50
15. Bradley P, Humphris G. Assessing the ability of medical students to apply evidence in practice: the potential of the OSCE. *Med Educ* 1999;33:815-7
16. MacRae, H.M., Regehr, G., McKenzie, M., Brenneman, F. and McLeod, R., Assessment of Critical Appraisal Skills. *Am J Surgery* 2004;187:120-3.
17. Charlin B, Gagnon R, Sibert L, Van der Vleuten C. Le test de concordance de script, un instrument d'évaluation du raisonnement clinique. *Pédagogie Médicale* 2002;3:135-44
18. Caire F, Sol JC, Charlin B, Isidori P, Moreau J. Le test de concordance de script (TCS) comme outil d'évaluation formative des internes en neuro-chirurgie : implantation du test sur Internet à l'échelle nationale. *Pédagogie Médicale* 2004;5:87-94
19. Marie I, Sibert L, Roussel F, Hellot MF, Lechevallier J, Weber J. Le test de concordance de script : un nouvel outil d'évaluation du raisonnement et de la compétence clinique en médecine interne ? *La revue de Médecine Interne* 2005;26:501-7
20. Sibert L, Charlin B, Gagnon R, Corcos J, Khalaf A, Grisé P. Evaluation du raisonnement clinique en urologie : l'apport du test de concordance de script. *Prog Urol* 2001;11:1213-9

Correspondance et offprints : Fabrice Muscari, Service de chirurgie digestive, CHU Rangueil, 1 avenue J. Poulhès, TSA 50032, 31059 Toulouse Cedex 9, France
Mailto : muscari.f@chu-toulouse.fr