

Suivi d'une réforme des contrôles de connaissances à la faculté de médecine de Rouen, France

Francis ROUSSEL¹, Jacques WEBER²

Résumé *Contexte* : A Rouen, dans le cadre de notre réforme pédagogique, les contrôles des connaissances ont été modifiés cinq années après la réforme des enseignements. Ces nouveaux examens ont maintenant cinq ans de recul. **But** : Décrire l'impact d'une nouvelle procédure impliquant la préparation collective des épreuves. **Matériel et méthodes** : L'ensemble des résultats d'examen de la seconde année du premier cycle des études médicales a été colligé sur neuf cohortes consécutives totalisant 926 étudiants, quatre années avant la modification et cinq années après. Depuis celle-ci, il était recommandé que les épreuves fussent préparées collectivement, que les questions et leur difficulté prévisible fissent l'objet d'une discussion préalable et qu'un rapport écrit fût produit. Une commission distincte du jury évaluait le processus. **Résultats** : Sur dix modules, deux ont parfaitement respecté les consignes et quatre partiellement. Le respect de la nouvelle procédure apportait une homogénéisation des pratiques et une stabilité des résultats. Dans les deux modules respectant les consignes, les questions rédactionnelles difficiles disparaissaient et le coefficient de fidélité dans le temps augmentait de 14 %. **Conclusion** : Malgré son intérêt scientifique et pédagogique, la nouvelle procédure, suivie dans 20 % des modules seulement, a rencontré des difficultés pour se généraliser.

Mot clés Enseignement médical ; premier cycle ; étudiant en médecine ; examens ; évaluation

Abstract *Context*: In Rouen, as part of our pedagogic reform, the assessment were re-evaluated five years after the teaching methods reform. These revised examinations were initially introduced into the program in 1998. **Aim**: To describe the impact of new proceedings including a collective preparation of the tests. **Material and methods**: All the examination results of the second year of the Medical Studies First Cycle were collected from nine successive student cohorts totalizing 926 students, four years before and five after the assessment reform. This reform required a collective meeting to prepare the tests. Any question, including its presumed difficulty, should have been discussed, and a written report produced. A specific committee, distinct from the examining board had evaluated the process. **Results**: Out of ten teaching units, two units completely fulfilled the requirements, and four only partially. The complete adherence to the procedure resulted in an homogenization of the practices, and a greater stability of the results. In the two fully compliant teaching units, the difficult questions disappeared, and the fidelity coefficient was increased by 14%. **Conclusion**: Despite their scientific and pedagogic interest, the revised proceedings experienced difficulties in generalizing, as currently in use in only 20% of teaching units.

Key words Medical education; pre-clinical curriculum; medical students; assessments ; evaluation.

Pédagogie Médicale 2005 ; 6 : 25-31

1- Laboratoire d' Histologie - Ancien responsable pédagogique du PCEM2 - Faculté de médecine de Rouen

2- Laboratoire de Neurophysiologie - Ancien vice-doyen à la pédagogie - Faculté de médecine de Rouen

Correspondance : Francis Roussel - Laboratoire d' Histologie - CHU de Rouen -1 rue de Germont - F76031 Rouen cedex -

Tél : 02 32 88 80 67 - mailto:francis.rousseau@chu-rouen.fr

Introduction

Les contrôles de connaissances exercent une influence évidente sur l'activité des étudiants^{1,2}. Ils exercent également une influence moins visible mais tout aussi prégnante sur l'ensemble du système pédagogique et, en particulier, sur l'activité des enseignants.

Dans le cadre de la réforme pédagogique progressive mise en œuvre à la faculté de médecine de Rouen³, les contrôles de connaissances ont fait l'objet de modifications durant l'année 1998, entre les cohortes 1997-1998 et 1998-1999, soit cinq ans après le début de la réforme pédagogique. Les séances d'apprentissages par problèmes⁴ (APP) ont été mises en place en 1993. Les examens n'avaient pas été modifiés d'emblée pour montrer que, dans un système de référence bien connu, les performances des étudiants n'étaient en rien diminuées par l'introduction d'une nouvelle pédagogie.

En France, le premier cycle des études médicales (PCEM) dure deux ans et correspond aux apprentissages précliniques. Il fait suite à un concours de recrutement de fin de première année de PCEM, organisé suivant des directives nationales portant sur les matières, les volumes horaires, les modalités du contrôle des connaissances et incluant un *numerus clausus* par région. A Rouen, la seconde année du PCEM commence par trois mois de modules fondamentaux, globalement centrés sur une discipline ou un groupe de disciplines, enseignés sur le mode magistral. Il se poursuit par des modules d'APP centrés sur un appareil et complétés par des stages de sémiologie et quelques séminaires.

Dans l'ancien système de contrôle des connaissances, les modules faisaient l'objet d'un examen global par semestre.

La réforme des contrôles de connaissances était sous-tendue par trois principes : l'indépendance des modules ; la discussion collégiale des questions avant leur rédaction définitive ; l'évaluation du processus et de son résultat final par une commission spécifique, distincte du jury d'examen. Les procédures de fonctionnement étaient spécifiées par écrit. Les modules étaient devenus indépendants. Les étudiants devaient maintenant réussir chaque module. Certains d'entre eux avaient, par ailleurs, été scindés et d'autres modifiés de façon marginale. Chaque enseignant était, dans le respect des objectifs définis collectivement, responsable de l'enseignement magistral, de la question posée à l'examen, et de sa correction.

Il disposait⁵ de quatre formats de questions : questions à choix multiple d'une minute (QCM), questions à réponse ouverte courte de deux minutes et demie (QROC), questions rédactionnelles simples de 10 minutes (QR10) et questions rédactionnelles de synthèse de 20 minutes (QR20). La durée totale des épreuves était proportionnelle au volume horaire du module. Une réunion d'évaluation *a priori* de la validité de l'épreuve devait rassembler tous les enseignants ayant effectivement fait des cours magistraux dans les modules traditionnels, ou les experts (un anatomiste, un histologiste, un physiologiste, un sémiologiste) ayant préparé les APP. Pour les questions rédactionnelles, les enseignants devaient fournir les libellés, les corrigés type et les grilles pondérées. Ils devaient préciser la difficulté prévisible de la question en prédisant le score de passage et le pourcentage de la cohorte supposé la dépasser. Toutes ces données étaient critiquées puis faisaient l'objet d'un rapport écrit permettant d'apprécier d'éventuelles divergences avec le résultat réel des épreuves.

L'évaluation *a posteriori* de la qualité réelle de l'épreuve réunissait les membres de la commission d'évaluation qui préparait la délibération du jury. L'instance qui évaluait l'épreuve était donc différente dans son rôle et dans sa composition de celle qui évaluait les étudiants. Elle disposait de l'ensemble des résultats traités statistiquement, en s'attachant aux moyennes et écarts types des cohortes, mais aussi à l'allure de la distribution des résultats individuels (*cf. Figure 1*).

Elle disposait également de l'ensemble des prédictions de résultats établies par les enseignants et de l'ensemble des enquêtes anonymes d'opinion menées systématiquement auprès des étudiants dès la fin des enseignements, pour en évaluer la qualité ressentie. Avant la réforme des examens, elles n'étaient transmises qu'aux responsables de module. Depuis, elles étaient également utilisées pour interpréter les divergences entre prévisions de résultats et leur réalité. Au vu de l'ensemble des éléments, la commission établissait des propositions écrites transmises au jury et au doyen. La délibération avait lieu ensuite. Le jury restait libre de ne pas suivre les conclusions de la commission d'évaluation, mais devait expliciter son attitude dans son rapport.

Ce travail descriptif a pour but de décrire l'impact de la nouvelle procédure impliquant la préparation collective des épreuves ; il est limité à la seconde année du premier

cycle des études médicales (PCEM2). Nous présentons une série consécutive de résultats aux questions rédactionnelles (QR) correspondant à neuf cohortes encadrant la mise en place des nouveaux examens.

Sujets et méthodes

L'ensemble des résultats chiffrés de neuf cohortes successives correspondant aux promotions 1994-1995 (99 étudiants), 1995-1996 (91 étudiants), 1996-1997 (90 étudiants), 1997-1998 (91 étudiants), 1998-1999 (90 étudiants), 1999-2000 (99 étudiants), 2000-2001 (109 étudiants), 2001-2002 (119 étudiants), 2002-2003 (138 étudiants) ont été analysés.

Les données présentées ici sont les données brutes de la première session d'examen. Elles découlent directement de la correction des copies avant la délibération de jury. La situation d'échec à laquelle on se réfère dans cet article est donc celle d'échec potentiel, à distinguer de celle d'échec réel, sanctionné par la délibération de jury. Pour l'analyser, nous avons comparé cet échec potentiel à la prévision établie par l'enseignant dans le cadre de la préparation des épreuves.

Les résultats de la session de rattrapage de septembre n'ont pas été pris en compte ici.

Dans ce travail, on a déterminé le coefficient de fidélité dans le temps en rapprochant les résultats moyens des cohortes aux questions rédactionnelles, classées en trois catégories selon le degré de respect des procédures (respect complet, respect partiel, absence de respect), des modules dont elles étaient extraites. On a appelé question difficile toute question rédactionnelle ayant entraîné le résultat moyen d'une cohorte en deçà de la moyenne de la totalité des questions rédactionnelles, sur l'ensemble de la période considérée, diminuée de 2 écarts-type ($M - 2Et$). On a étudié l'influence de ces questions difficiles sur les ajournements, soit en les supprimant, soit en les neutralisant.

Résultats

Suivi des procédures

Des réunions de préparation des épreuves se sont déroulées dans six des dix modules. La procédure a été régulièrement et intégralement respectée dans deux d'entre eux. Quatre modules étaient dans des situations intermédiaires.

Qualité des modules

Le respect de la procédure influait fortement sur la dispersion des résultats individuels des étudiants (*Figure 1-A et B*). Il apparaissait un rapprochement de ces courbes, traduisant une diminution de la variabilité intersessions pour un même module. Par exemple, dans le module B, où la procédure n'était pas respectée (*Figure 1-B*), les scores atteints par 50 % de chacune des cinq cohortes fluctuaient entre 4,90/10 et 6,40/10 (soit 1,5 point). Pendant le même temps, dans le module A, où la procédure était, au contraire, parfaitement respectée, les mesures fluctuaient entre 5,75/10 et 6,25/10 (soit 0,5 point) (*figure 1-A*). Le respect de la procédure influait plus globalement sur la fidélité dans le temps des résultats collectifs des cohortes, avec une amélioration de 10 % en cas de respect partiel de la procédure et de 14 % en cas de respect total.

Qualité des questions rédactionnelles

Les moyennes obtenues par les différentes promotions aux questions rédactionnelles des 9 premières sessions avaient une distribution normale. Dans les modules où la procédure était respectée, leur moyenne était à 5,54 / 10 ($Et : 0,83$) pour 5,48 / 10 ($Et : 1,38$) dans les modules où elle ne l'était pas. Ont donc été considérées comme difficiles, les questions entraînant une moyenne de cohorte inférieure à 3,88/10 ($M - 2Et$). Dans les modules où la procédure n'était pas respectée, leur pourcentage atteignait 14 %. Il était de 0 % dans les modules où elle l'était parfaitement.

Les éliminations et les neutralisations des questions difficiles ont donné le même résultat. Elles ont montré sur les 7 premières années, que seuls trois des 76 ajournements (4 %) auraient été liés à une ou plusieurs des questions difficiles.

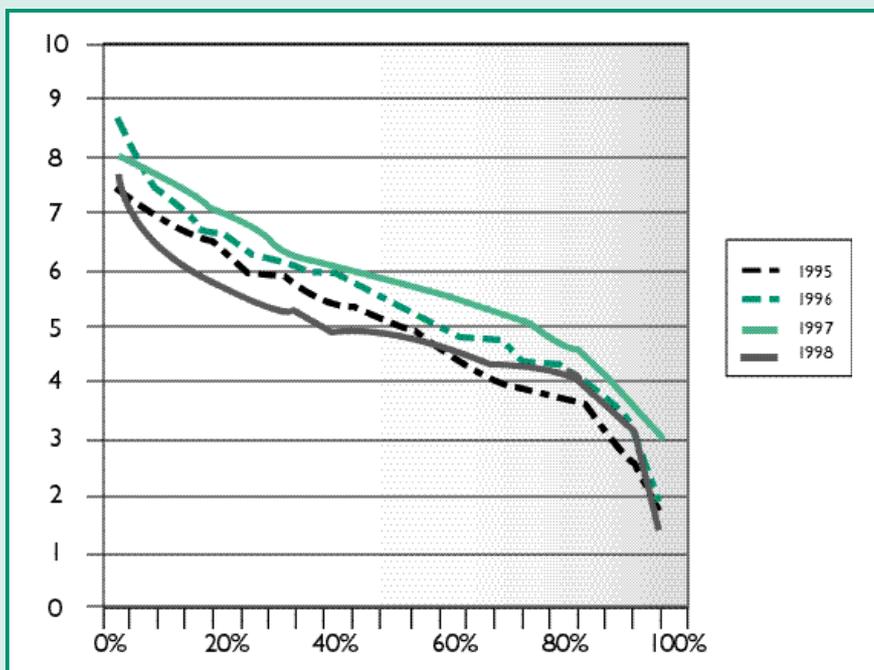
Performances des étudiants

Les scores moyens ont augmenté de façon variable suivant le degré de respect de la procédure. Dans les modules la respectant parfaitement, la moyenne des scores passait de 5,25/10 à 6,04/10, soit une augmentation de 15 %, alors que dans les modules ne la respectant pas, elle passait de 5,59/10 à 5,94/10, soit une augmentation de 6 %. Le pourcentage d'ajournés potentiels passait de 20 % avant réforme à 14 % après réforme dans les modules respectant parfaitement la procédure.

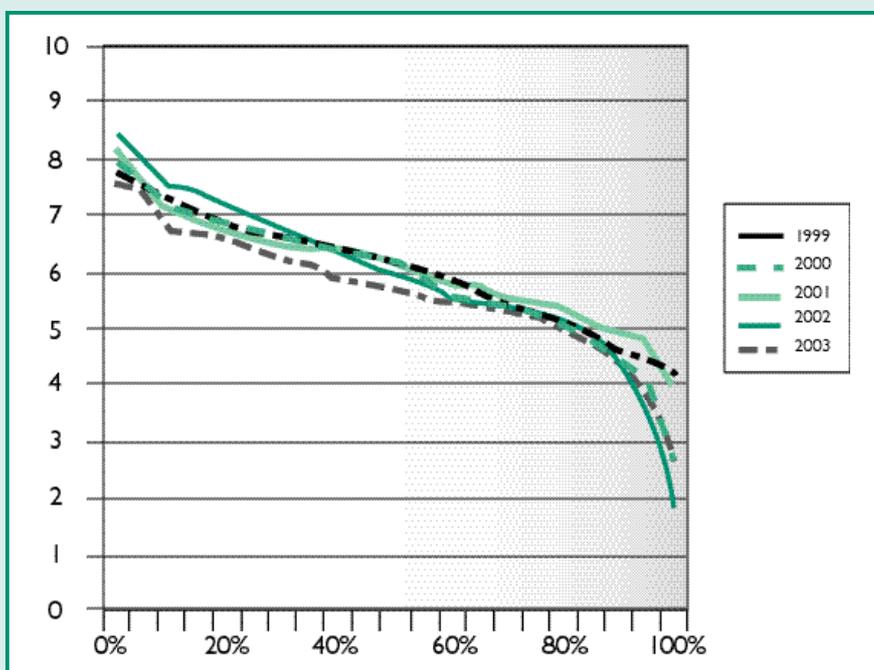
Recherche et Perspectives

Figure 1-A :
Influence du réglage de l'appareil de mesure sur la distribution des résultats individuels

Module A respectant de façon optimale la procédure de préparation des épreuves



**Module A
avant réforme**

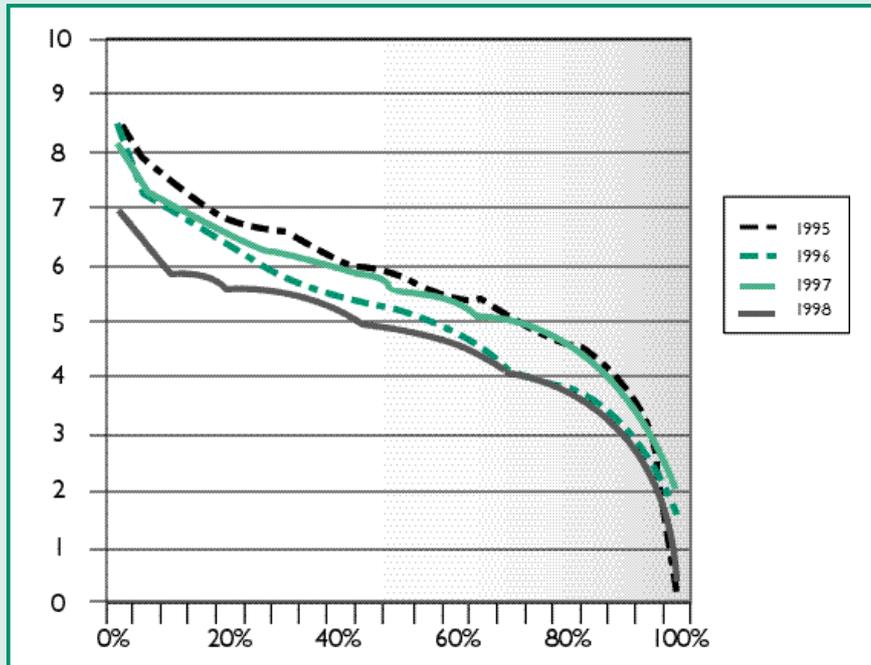


**Module A
après réforme**

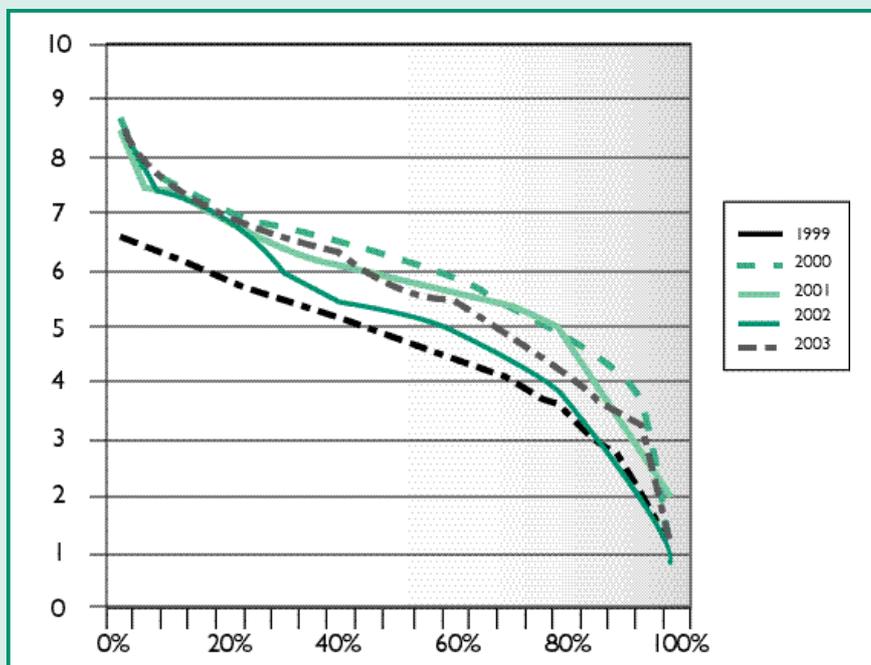
Chaque diagramme présente comparativement plusieurs cohortes successives avec en ordonnées, le score (sur 10) atteint par un pourcentage donné d'étudiants ; le module A était parfaitement réglé ; ce respect de la procédure stabilise les résultats individuels des étudiants.

Figure 1-B :
Influence du réglage de l'appareil de mesure sur la distribution des résultats individuels

Module B ne respectant pas de façon optimale la procédure de préparation des épreuves



**Module B
avant réforme**



**Module B
après réforme**

Chaque diagramme présente comparativement plusieurs cohortes successives avec en ordonnées, le score (sur 10) atteint par un pourcentage donné d'étudiants ; le module B n'était pas parfaitement réglé ; en l'absence de respect de la procédure, les résultats individuels des étudiants ne sont pas stabilisés.

Recherche et Perspectives

Discussion

La comparaison des cohortes successives entraînait à supposer que leur qualité, représentée par la moyenne des scores, était constante. Or, la mise en place de la réforme des examens s'est accompagnée d'une augmentation générale des valeurs mesurées qui apparaît incompatible avec l'hypothèse d'invariabilité des cohortes. Cependant, l'indépendance nouvelle des modules a pu modifier la préparation des étudiants et *in fine* leur performance. On ne peut pas non plus éliminer complètement l'hypothèse d'une imperfection persistante de notre appareillage de mesure. Si ces considérations peuvent expliquer l'augmentation des scores dans les modules où la procédure n'était pas respectée, il existe un autre phénomène dans les modules où elle l'était parfaitement : non seulement les questions « difficiles » qui faisaient baisser artificiellement les moyennes étaient éliminées, mais encore, nombre de résultats moins aberrants mais tout aussi peu crédibles se trouvaient modifiés. La mesure de la performance réelle des étudiants est devenue plus représentative de celle-ci quand la procédure de réglage a été respectée. Les mesures anciennes apparaissent donc non seulement floues, mais également sous-estimées.

La procédure, avec son formalisme complet, n'a été pleinement respectée que dans 20 % des cas. Néanmoins, certains comportements ont évolué. Des modules fiers d'être les plus sélectifs (*ie* : ajournant le plus d'étudiants) se sont modifiés sensiblement, et les notations des questions rédactionnelles ont convergé. Cependant, le travail d'évaluation des procédures reste mal connu ou encore incompris des enseignants. Pour expliquer ce médiocre respect des procédures, plusieurs hypothèses peuvent être avancées : a) une compétence pédagogique insuffisante du corps enseignant ; b) un manque de disponibilité, comparativement à d'autres pays, lié aux effectifs⁶ et à la multiplicité des tâches « plein temps » (hospitalière, d'enseignement, de recherche, d'administration) ; c) un manque de motivation lié à la faible reconnaissance des activités d'enseignement.

Conclusion

Dans le domaine de la réalisation pratique des contrôles de connaissances, des progrès structurants ont été réalisés dans notre institution facultaire. Des informations auparavant inaccessibles sont maintenant disponibles, analy-

sables, ouvertes à toute discussion. C'était la condition préalable à tout progrès et, dans les modules respectant complètement la procédure, les questions « atypiques » ont disparu.

L'amélioration qui en résulte permet d'apprécier plus exactement (fidélité augmentée de 14 %) la performance collective des étudiants et de constater qu'elle était historiquement sous-évaluée. Elle permet d'améliorer sensiblement l'analyse de la variabilité entre cohortes d'étudiants. Elle permet également d'extraire du flou les résultats individuels des étudiants et d'aborder dans de bien meilleures conditions les discussions de jury sur les performances proches du score de passage.

L'intérêt de la procédure de réglage des épreuves est scientifiquement et pédagogiquement évident ; néanmoins, cette nouvelle culture ne s'épanouit pleinement que dans une minorité de modules. Dans les modules où la procédure n'est pas respectée, les jurys sont encore les lieux où se déroulent et s'entremêlent toutes les étapes régulatrices⁷. Cela confirme qu'une culture d'évaluation ne se décrète pas et demande du temps pour se mettre en place, même dans un milieu *a priori* favorable.

Remerciements

Nous remercions Richard Médeiros qui a revu l'abstract.

Contributions

Francis Roussel et Jacques Weber ont participé à l'élaboration du programme de préparation des contrôles des connaissances. Francis Roussel a participé aux réunions préparatoires des épreuves et a colligé les données. Francis Roussel et Jacques Weber ont réalisé l'analyse des données ; ils ont rédigé une première version du manuscrit et en ont effectué la révision à partir des critiques des lecteurs.

Références

1. Seale JK. *The influence of assessments on students' motivation to learn in a therapy degree course.* *Med Educ* 2000; 34 : 614-621.
2. Newble DI, Jaeger K. *The effect of assessments and examinations on the learning of medical students.* *Med Educ* 1983; 17 : 165-171.
3. Weber J, Denis P, Colin R. *Rouen : un nouveau cursus des études médicales.* *Médecine et Hygiène* 1996 ; 54 : 2339-2334.
4. Colin R, Roussel F, Weber J, Denis P, Lauret P, Des Marchais JE. *Student and faculty early evaluation of the progressive adoption of a problem-based learning curriculum reform at Rouen, France.* *Education for Health* 1999 ;12: 223-232.
5. Was V, McGibbon D, Van der Vleuten C. *Composite undergraduate clinical examinations: how should the components be combined to maximalise reliability?* *Med Educ* 2001 ; 35: 326-330.
6. Barzansky B, Etzel SI. *Educational programs in US medical schools, 2003-2004.* *JAMA* 2004 ; 292 : 1025-1031.
7. Jouquan J. *L'évaluation des apprentissages des étudiants en formation médicale initiale.* *Pédagogie Médicale* 2002 ; 3 : 38-52.

Manuscrit reçu le 27 novembre 2003 ; commentaires éditoriaux formulés aux auteurs 30 décembre 2003 et le 26 mars 2004 ; accepté pour publication le 20 décembre 2004.

www.pedagogie-medicale.org un site à votre disposition

Votre revue est disponible sur Internet. Vous y retrouverez les informations générales sur la revue, les adresses de contact, les sommaires des différents numéros parus, les directives aux auteurs etc. Les éditoriaux et toutes les rubriques de la section « Vie pédagogique » y sont disponibles « *in extenso* » en format PDF (lisibles avec Acrobat Reader). Au fur et à mesure de leur publication, les fiches pratiques seront mises en ligne sur le site. Vous pourrez ainsi les télécharger, les imprimer chez vous, et vous construire progressivement votre propre manuel pratique. Divers liens vers les auteurs des articles déjà publiés et les organisations du comité de parrainage sont également disponibles.

Un moteur de recherche spécifique au site vous permet désormais de rechercher des articles ou résumés par mots-clés, auteur, etc... N'hésitez pas à transmettre vos opinions sur les articles (via le forum de discussion) ou directement à la rédaction. Les commentaires les plus intéressants seront publiés dans les pages de la revue.

Le site www.pedagogie-medicale.org se veut un lien amical et fonctionnel entre tous ceux qui sont impliqués dans la formation médicale. Le site doit leur permettre de partager leurs préoccupations et leurs solutions dans leur action pour la formation initiale et la formation continue de nos professionnels de santé.

La rédaction