

# Evaluation de l'efficacité pédagogique immédiate de deux méthodes classiques en formation médicale continue

Guy LLORCA\*

**Résumé** L'efficacité pédagogique des différentes méthodes de formation a rarement été étudiée de manière systématique en formation continue. **But** : comparer deux méthodes pédagogiques classiquement utilisées en formation continue au moyen d'un outil étalonné évaluant l'efficacité pédagogique, en termes de progression des connaissances mesurée par questions à choix multiples (QCM), et en termes de satisfaction mesurée par une échelle numérique analogique. **Matériel et Méthodes** : le sujet de la formation évaluée a porté sur les lombalgies communes. Le plan expérimental a comporté trois étapes : 1) Validation et analyse pédagogique du pouvoir discriminatif et de la stabilité des réponses avec cet outil chez 66 personnes de niveaux de connaissances différents, 2) Comparaison, à l'aide de cet outil validé, d'une méthode classique en exposés brefs (A) à une méthode d'entraînement par cas cliniques (B) chez 46 médecins généralistes et spécialistes, 3) Comparaison des progressions des généralistes et des spécialistes. **Résultats** : l'outil utilisé retrouve bien les niveaux de connaissance présumés et la stabilité des réponses, confirmant sa validité et sa fiabilité. Les deux méthodes testées sont globalement comparables en termes de progression des connaissances et de satisfaction. Les connaissances testées sont différentes chez les généralistes et les spécialistes. **Conclusions** : l'utilisation de QCM peut permettre d'étudier l'évolution immédiate des connaissances bien définies en formation médicale continue. Les méthodes classiquement utilisées semblent comparables en termes d'efficacité pédagogique. Les comportements différents des médecins généralistes et spécialistes plaident pour des formations continues spécifiquement adaptées.

**Mots clés** évaluation ; formation médicale continue ; questions à choix multiples ; QCM ; méthodes pédagogiques, lombalgies.

**Summary** The educational efficiency of training method formats has rarely been systematically studied in continuing medical education. **Aim** : To compare two methods of continuing graduate medical training in words of educational efficacy, ie in words of cognitive progression (evaluated with multiple choice question test) and of satisfaction (evaluated with numeric analogical scale). **Material and methods** : Low back pain was the subject of training. The prospective compared plan was divided in three steps : 1) Knowledge level analysis in 66 participants with presupposed different initial levels, and analysis of stability of scores, 2) Comparison, with this validated tool, between a classical method of training in brief lectures (A method) and a case-based training (B method) in 46 general practitioners and specialists, and 3) Comparison of performances of generalists and specialists. **Results** : presupposed levels and stability of scores are well documented by this approach. A and B methods are not significantly different in words of cognitive progression and satisfaction. Knowledge tested are different between physicians and specialists. **Conclusion** : MQC evaluation seems a valuable tool to study cognitive progress in continuing graduate education. Classical methods of training seem to be equivalent in words of educational efficacy. The difference observed between physicians and specialists commends specific educational program trainings.

**Keywords** evaluation; continuing graduate medical training; multiple choice questions; MQC; educational methods; low back pain.

*Pédagogie Médicale 2003 ; 4 : 26-32*

\* Laboratoire de Thérapeutique, 8, av. Rockefeller – 69373 Lyon cedex 08 et Service de Rhumatologie, Centre Hospitalier Lyon-Sud – 69495 Pierre-Bénite cedex – Tél. : +33 478 86 12 81 – Télécopie : +33 478 86 33 42 – <mailto:guy.llorca@chu-lyon.fr>

## **Introduction**

Les questions à choix multiples (QCM) sont largement utilisées en formation médicale du fait de leur relative simplicité d'élaboration et de leur facilité de correction. Ce mode d'évaluation, bien que largement critiqué<sup>1,2,3,4</sup>, sert volontiers de support à de nombreuses évaluations tant théoriques que techniques<sup>5,6</sup>. Très utilisée en formation médicale initiale, cette modalité d'évaluation l'est aussi en formation médicale continue<sup>7</sup> et dans le domaine de l'éducation des patients<sup>8</sup>. Elle a dû, enfin, s'adapter aux normes des nouvelles techniques d'information, des nouvelles méthodes pédagogiques et des exigences de la médecine fondée sur des preuves<sup>9,10,11</sup>. Cependant, sa valeur en formation médicale continue n'a pas été, jusque là, l'objet de nombreuses études systématiques<sup>12</sup>. Le but de cette étude a donc été de savoir si ce type d'évaluation était valide en formation médicale continue et, si c'était le cas, de l'utiliser pour comparer l'efficacité de deux méthodes pédagogiques différentes, en termes d'efficacité pédagogique immédiate, et pour des publics de participants différents.

## **Matériel et méthodes**

### **Les sujets de l'étude**

66 personnes ont participé à l'étude de validité et de fiabilité de l'outil. Il s'agissait de 40 médecins généralistes, 15 spécialistes rhumatologues (dont 6 rhumatologues exerçant en ville, 4 rhumatologues hospitaliers seniors et 5 internes en rhumatologie en fin de cursus), 6 étudiants en médecine en fin de stage de rhumatologie et 5 secrétaires médicales d'un service de rhumatologie. 46 d'entre elles (les médecins généralistes et les rhumatologues de ville) ont participé à la recherche proprement dite (comparaison des méthodes).

### **Le thème de l'étude**

Il a porté sur les lombalgies communes par conflit mécanique, sujet sélectionné pour sa fréquence et ses incidences en matière de santé publique et de coût<sup>13</sup>.

### **Les critères de jugement**

L'efficacité pédagogique immédiate a été analysée à partir de deux critères principaux : le gain cognitif et le degré de satisfaction. Les connaissances des données médicales sur les lombalgies ont été recueillies à l'aide d'un questionnaire de connaissance comportant 75 éléments de réponses fermées (vrai/faux<sup>14</sup>). Ce questionnaire a été élaboré par les experts rhumatologues respon-

sables de la formation, à partir des données scientifiques probantes du moment de l'étude<sup>13</sup>. Les questions ont porté sur les objectifs de base en matière d'épidémiologie, de physiopathologie, de coût de santé publique, de sémiologie clinique et radiologique, ainsi que d'attitudes thérapeutiques pratiques devant des cas habituels. La réponse à ces questions a été l'objet d'une notation (1 point par réponse exacte), exprimée en pourcentage de réponses exactes. L'évolution du pourcentage de réponses exactes a représenté le gain cognitif immédiat. Les données de satisfaction ont été recueillies à l'aide d'un questionnaire d'opinion portant sur la satisfaction globale des participants, leur perçu de l'intérêt, de la qualité et de l'utilité de la formation, les conditions pratiques de la formation (adéquation aux objectifs, densité de l'information, conditions matérielles, propension à poursuivre la formation). Chaque critère a été l'objet d'une réponse exprimée en échelle analogique numérique allant de 0 (inacceptable) à 10 (exceptionnel). La moyenne des réponses a représenté le degré de satisfaction immédiat.

### **Les méthodes pédagogiques testées**

Deux méthodes ont été comparées. La première méthode consistait en une séance d'une journée de formation effectuée à partir d'exposés brefs présentant les données correspondant aux objectifs de la formation, suivis de discussions avec les participants (méthode A). La seconde méthode a comporté 6 séances de 4 heures d'entraînement à la résolution de cas cliniques en petits groupes, cas élaborés à partir des objectifs de la formation. Cette seconde méthode (méthode B) était animée par le même formateur que la première.

### **Le plan expérimental de l'étude**

Le plan de l'étude a comporté trois étapes :

1- Étude de validité et de fidélité par « test-retest » : cette première étape a consisté à administrer le questionnaire de connaissances à un groupe de personnes susceptibles de correspondre à des niveaux différents de connaissances. Le questionnaire a été administré une première fois sans information (Test 1) aux 66 personnes participant à l'étude, puis, une seconde fois, à 24 heures d'intervalle (Test 2), toujours sans information ni commentaires sur les résultats de la veille, à 26 personnes (secrétaires, étudiants, internes et rhumatologues hospitaliers).

2- Comparaison des méthodes pédagogiques : la seconde étape a consisté à administrer le questionnaire de connaissances aux 46 médecins en formation (23 personnes par groupe dont 20 médecins généralistes

et 3 rhumatologues spécialistes), avant la formation (Prétest), puis à la fin de la formation (Post-test). La satisfaction a été analysée au moyen du questionnaire d'opinion administré à chaque groupe en fin de formation. Les critères de jugement, à cette phase, ont été représentés par les taux de progression des connaissances (Post-test - Prétest) exprimés en pourcentage, et par le degré de satisfaction des participants.

3- Comparaison entre médecins généralistes et spécialistes : la troisième étape a consisté à comparer, au sein de chaque groupe de participants, le comportement des médecins généralistes et celui des médecins spécialistes participants, en termes de progression des connaissances et de degré de satisfaction.

L'analyse des données a porté sur la comparaison des moyennes des taux de réponse en matière de connaissances et de satisfaction, ainsi que sur l'analyse des corrélations existant entre les différents participants et les différents objectifs pédagogiques. L'étude de la significativité a été calculée en retenant les seuils suivants : ns pour  $p > 0,05$ , \* pour  $p \leq 0,05$ , \*\* pour  $0,01 > p > 0,001$  et \*\*\* pour  $p \leq 0,001$ .

## Résultats

### Validation de l'outil

Le test 1 permet d'identifier correctement 4 niveaux (Figure 1), avec une différence statistique nettement significative entre les secrétaires médicales (niveau le plus bas), les médecins généralistes et les étudiants qui obtiennent des moyennes non significativement différentes et plus élevées que celle des secrétaires, puis les rhumatologues exerçant en ville et les internes qui obtiennent des moyennes également non significativement différentes et encore plus élevées, et, enfin, les rhumatologues hospitaliers seniors qui ont les moyennes les plus élevées et diffèrent significativement des autres niveaux.

Le test 2 de l'étude test-retest retrouve la même gradation avec des écarts pour chaque niveau (test 2-test 1) non significatifs par rapport au test 1 (Figure 2) à l'exception de celui constaté pour les secrétaires médicales ( $p < 0,001$ ). L'analyse de corrélation de ces écarts met en évidence cependant une progression linéaire de l'écart du niveau le plus élevé au plus faible ( $r = 0,79$  ;  $p < 0,001$ ) et, pour chaque niveau, avec l'ancienneté dans la fonction ( $r = 0,68$  ;  $p = 0,034$ ). L'écart moyen n'est supérieur à 5 % que pour le groupe des secrétaires médicales et peut être retenu comme la limite de validité des écarts observés à partir desquels une différence peut être valablement prise en compte comme effective.

### Comparaison des méthodes pédagogiques

Les deux groupes testés progressent très significativement (Figure 3). La comparaison du taux de progression en fonction de la méthode n'a pas de conséquence pratique significative, puisqu'elle ne dépasse pas le seuil de 5 % observé chez les secrétaires à l'occasion du retest, mais la progression globale est plus importante et plus homogène avec la méthode A. La progression est plus importante pour les participants qui obtiennent les résultats les plus bas au prétest et pour les objectifs thérapeutiques avec la méthode B ( $p = 0,01$ ).

L'analyse de satisfaction est bonne pour les deux groupes et non statistiquement significativement différente (Tableau I).

### Comparaison des médecins généralistes et spécialistes

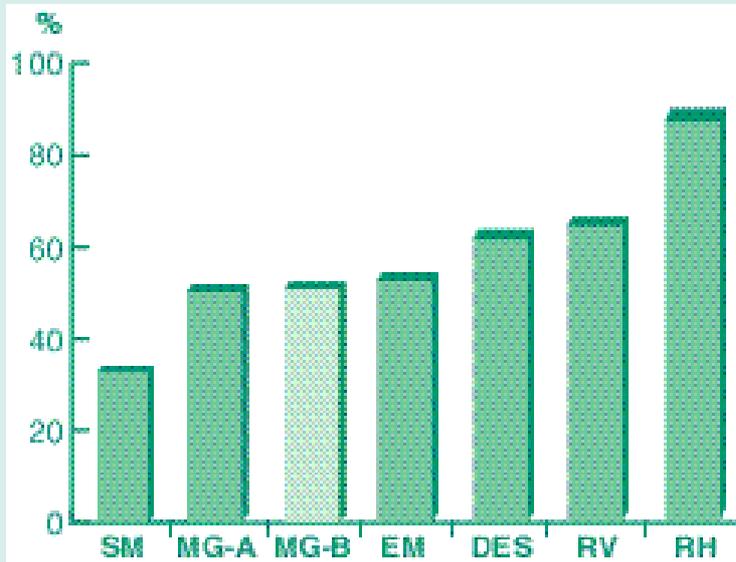
Les taux de progression sont significatifs pour les deux populations testées et supérieurs au seuil de 5 %. La progression des médecins généralistes est supérieure à celle des spécialistes (Figure 4). L'analyse des réponses par objectif met en évidence une bonne connaissance de base et des prescriptions courantes pour les médecins généralistes (avec un taux moyen de réussite toujours supérieur à 50 % pour ces objectifs). Les médecins spécialistes possèdent de meilleures connaissances des sémologies atypiques et des indications chirurgicales. Les deux groupes ont des taux de réponse insuffisants en ce qui concerne les objectifs de santé publique (particulièrement les données épidémiologiques générales et les données économiques pour lesquelles les taux de réussite sont inférieurs à 30 %) mais progressent significativement ( $p = 0,027$  \*).

L'analyse de satisfaction est bonne pour les deux groupes et non statistiquement significativement différente (Tableau I).

**Tableau I : Analyse du degré de satisfaction des participants selon la méthode pédagogique utilisée.**

Méthode	Moyenne/10 ± DS	Max. Min.	P
A	8,2 ± 0,6	10 – 6,5	ns
B	8,5 ± 0,5	10 – 7,5	

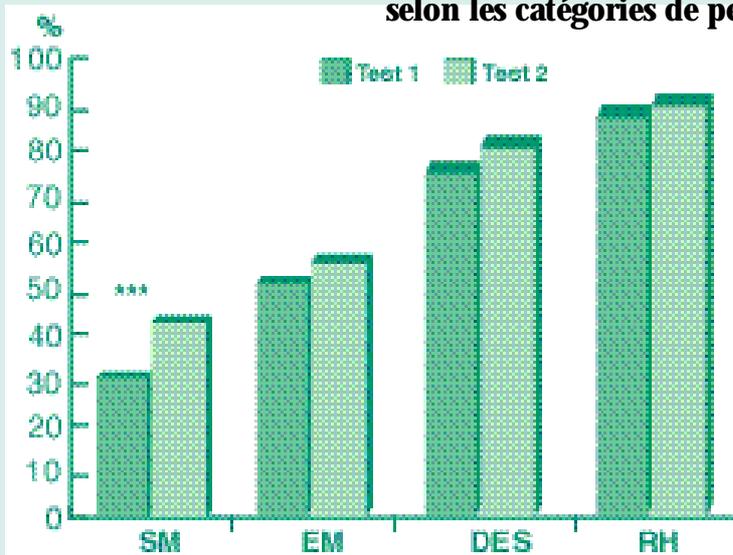
**Figure 1 :**  
**Comparaison des taux de réponses exactes des différentes populations testées**



*SM : secrétaires médicales*  
*MG-A et MG-B : médecins généralistes participants aux méthodes A et B*  
*EM : étudiants en médecine*  
*DES : internes de spécialité*  
*RV : rhumatologues exerçant en ville*  
*RH : rhumatologues hospitaliers*

\*\*\* :  $p < 0,001$

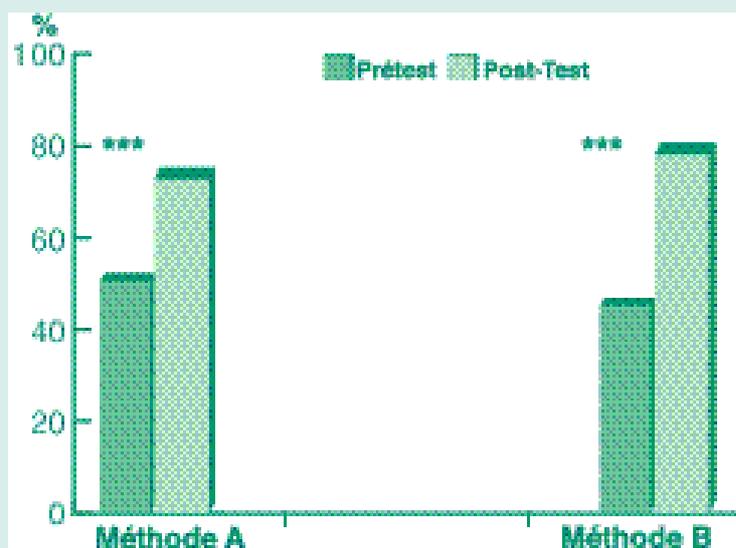
**Figure 2 :**  
**Comparaison des taux de réponses exactes à 24 heures d'intervalle selon les catégories de personnes**



*SM : secrétaires médicales*  
*EM : étudiants en médecine*  
*DES : internes de spécialité*  
*RH : rhumatologues hospitaliers*

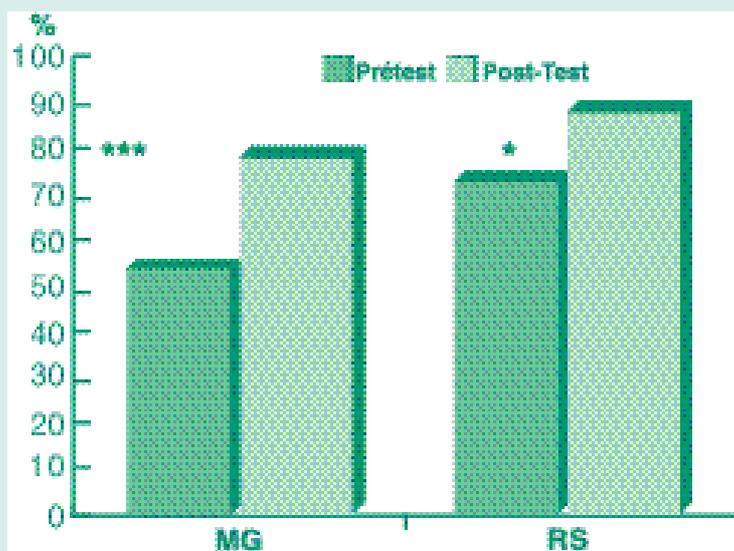
\*\*\* :  $p < 0,001$

**Figure 3 :**  
**Comparaison de la progression des connaissances selon la méthode exprimée en pourcentage de réponses exactes aux tests**



\*\*\* :  $p < 0,001$

**Figure 4 :**  
**Comparaison de la progression des médecins généralistes (MG) et des rhumatologues spécialistes (RS), exprimée en pourcentage de réponses exactes aux tests**



\* :  $p < 0,05$

\*\*\* :  $p < 0,001$

## Discussion

L'utilisation de tests de connaissances construits en QCM est valide, en ce qui concerne l'analyse d'efficacité pédagogique immédiate de formations médicales, si le taux de progression enregistré est supérieur à la variabilité spontanée, évaluée pour cette étude à 5 %, taux représentant la variation spontanée la plus forte enregistrée en dehors de toute intervention pédagogique<sup>15</sup>. Ce type de questionnement ne peut toutefois valablement explorer que des connaissances factuelles, conceptuelles ou procédurales à haut niveau de preuve<sup>16</sup>.

L'utilisation de ces questionnaires en mode Pré-Post immédiat ne renseigne valablement que sur l'efficacité immédiate. L'efficacité à moyen et long terme doit être explorée par des post-test tardifs<sup>17</sup>. L'efficacité pédagogique évaluée ne représente pas le degré d'appropriation des connaissances qui, lui, ne peut valablement être exploré qu'à l'aide d'autres méthodes, tels les audits de pratique<sup>12</sup>.

Si le style d'apprentissage semble influencer les performances des étudiants<sup>18</sup>, les deux méthodes testées semblent comparables en matière d'efficacité pédagogique moyenne immédiate. La méthode A, plus traditionnelle, semble cependant privilégier les éléments de connaissance factuels et conceptuels. La méthode B, plus interactive, semble favoriser la transmission des éléments de connaissance procéduraux et semble permettre d'obtenir des taux de progression supérieurs pour les participants dont le niveau de départ est le plus bas. La satisfaction des participants est comparable mais, dans notre étude, la répartition des participants, effectuée sur la base d'une demande des deux groupes, a été réalisée sur un mode non aléatoire pour des raisons pratiques, ce qui peut représenter un éventuel biais de sélection.

En ce qui concerne le thème de l'étude, les médecins généralistes et les médecins spécialistes ne répondent pas de la même façon aux différents objectifs. Les données de santé publique représentent les objectifs les moins bien réussis par tous les participants testés. Ces données apparaissent certainement comme moins importantes aux yeux des praticiens par rapport à celles qui concernent directement le soin quotidien. Les connaissances des deux populations étudiées, sous réserve du faible effectif des spécialistes testés, semblent bien différentes. Ces données plaident pour une spécificité des programmes de formation continue qui méritent d'être adaptés au public concerné.

Cette étude met en évidence la validité des questionnaires utilisant des QCM pour tester l'efficacité à court terme de formations continues. Ces questionnaires ne sont cependant probablement pas adaptés à l'analyse de

l'efficacité en matière de problèmes complexes faisant intervenir des connaissances essentiellement conditionnelles, connaissances certainement mieux explorées par des outils d'évaluation plus élaborés. Leur validité dans l'analyse de l'efficacité pédagogique à moyen ou long terme reste à préciser. Enfin, leur intérêt dans l'analyse de l'appropriation des savoirs est certainement limité, puisqu'ils ne sont le reflet que de connaissances théoriques ou d'intentions, et non le reflet de comportements réels en condition de décision.

## Conclusion

L'évaluation de l'efficacité pédagogique de formations continues en utilisant des questionnaires à réponses en choix multiples est possible, si l'on prend la précaution de s'assurer, au préalable, de leur pouvoir discriminatif et des limites de reproductibilité de cet outil. Cette étude met en évidence les qualités métrologiques de tels questionnaires dont la mise en œuvre est simple, mais dont les limites sont liées à la rigidité du mode de correction qui exclut de l'analyse les éléments complexes de la connaissance B

## Références

1. Veloski JJ, Rabinowitz HK, Robeson MR, Young PR. Patients don't present with five choices: an alternative to multiple-choice tests in assessing physicians' competence. *Acad Med* 1999 ; 74: 539-46.
2. Williams C, Trigwell P, Yeomans D. Pass the Royal College examinations. MCQ technique. *British J Hosp Med* 1996 ; 55: 479-81.
3. Vahalia KV, Subramaniam K, Marks SC, DeSouza EJ. The use of multiple-choice tests in anatomy: common pitfalls and how to avoid them. *Clin Anat* 1995 ; 8: 61-5.
4. McGaghie WC. Evaluating Competence for Professional Practice. In : Curry L, Wergin JF (Eds). *Educating Professionals. Responding to new expectation for competence and accountability*. San Francisco (Ca) : Jossey-Bass, 1993.

5. Bloch R, Burgi H. *The Swiss catalogue of learning objectives. Med Teach* 2002 ; 24: 144-50.
6. Ali J, Campbell JP, Gana T, Burns PN, Ochsner MG. *Swine and dynamic ultrasound models for trauma ultrasound testing of surgical residents. J Surg Res* 1998 ; 76: 17-21.
7. Chan DH, Leclair K, Kaczorowski J. *Problem-based small-group learning via the Internet among community family physicians a randomized controlled trial. MD Comput* 1999 ; 16: 54-8.
8. Mulloy E, Donaghy D, Quigley C, McNicholas WT. *Prospective audit of an asthma education programme in an out-patient setting. Ir Med J* 1996 ; 89: 226-8.
9. Elves AW, Ahmed M, Abrams P. *Computer-assisted learning: experience at the Bristol Urological Institute in the teaching of urology. Br J Urol* 1997 ; 80, Suppl 3: 59-62.
10. Ali J, Cohen R, Adam ., Gana T, Pierre I, Bedaysie H, Ali E, West U, Winn J. *Teaching effectiveness of the advanced trauma life support program as demonstrated by an objective structured clinical examination for practicing physicians. World J Surg* 1996 ; 20: 1121-5.
11. Hammond EJ, McIndoe AK, Sansome AJ, Spargo PM. *Multiple-choice examinations: adopting an evidence-based approach to exam technique. Anaesthesia* 1998 ; 53: 1105-8.
12. Llorca G. *La Formation Médicale (Aspects conceptuels). Méditations, Lyon, 1999.*
13. Bourgeois P, Charlot J, Derriennic F, Lebrun T, Leclerc A, Meyer J.P, D'Houtaud A. et Phelip X. *Rachialgies en milieu professionnel. Quelles voies de prévention? Editions INSERM, expertise collective, Paris, 1995.*
14. Bandaranayake R, Payne J, White S. *Using multiple response true-false multiple choice questions. Aust NZ J Surg* 1999 ; 69: 311-5.
15. De Landsheere G. *Evaluation continue et examens. Précis de docimologie. Editions Labor éducation 2000, Bruxelles, 1992 (6° édition).*
16. Tardif J. *Pour un enseignement stratégique. L'apport de la psychologie cognitive. Les éditions Logiques, Montréal, 1997.*
17. Ali J, Adam R, Pierre I, Bedaysie H, Josa D, Winn J. *Comparison of performance 2 years after the old and new (interactive) ATLS courses. J Surg Res* 2001 ; 97: 71-5.
18. Lynch TG, Woelfl NN, Steele DJ, Hansen CS. *Learning style influences student examination performance. Am J Surg* 1998 ; 176: 62-6.