



Évaluation clinique objective en stations multiples en anatomie pathologique: analyse critique et étude de perception

Objective multi-station clinical evaluation in pathological anatomy : critical analysis and perception

Faten Limaïem¹, Faten Farah², Oussama Belkacem¹, Saadia Bouraoui¹, Wafa Achour³

1- Hôpital Mongi Slim La Marsa / Université de Tunis El Manar, Faculté de Médecine de Tunis,

2- Institut Kassab D'orthopédie, Ksar Said. / Université de Tunis El Manar, Faculté de Médecine de Tunis,

3- Centre National de Greffe de Moelle osseuse / Université de Tunis El Manar, Faculté de Médecine de Tunis,

RÉSUMÉ

Introduction : l'évaluation clinique objective en stations multiples (ECOSM) est un outil équitable et authentique d'évaluation formative et normative réputé pertinent pour son efficacité, sa fiabilité et sa validité lors de l'évaluation clinique. Cependant, aucune expérience concernant son application en anatomie pathologique n'a été publiée.

But : Notre objectif était de réaliser une analyse critique et une étude de perception de l'ECOSM en anatomie pathologique des étudiants en 1^{ère} année du deuxième cycle des études médicales (DCM1).

Méthodes : Il s'agissait d'une étude prospective transversale et multicentrique dans laquelle nous avons développé et testé une ECOSM en anatomie pathologique sur 32 étudiants de DCM1. L'analyse docimologique des 12 stations et des 28 questions des différentes stations a été faite par le fichier « AnItem.xls ».

Résultats : Le score total moyen obtenu à l'ECOSM était de 36,2/50 (extrêmes de 29 à 41). La majorité des stations était de difficulté acceptable (92%) et de mauvaise discrimination (92%). La majorité des questions était facile (57%) et de mauvaise discrimination (75% du total des questions). Une question sur 28 était « idéale », soit 4%. L'ECOSM avait une grande hétérogénéité (alpha de Cronbach moyen de 0,29), et discriminait bien entre les étudiants moyens (courbe de distribution des scores platykurtique, étalée à gauche, à asymétrie négative). L'analyse de la perception de l'épreuve a révélé que 27 étudiants (84%) étaient globalement satisfaits de l'ECOSM et pensaient qu'elle était proche de la réalité des stages effectués.

Conclusion : L'ECOSM testée était facile et discriminait surtout entre les étudiants moyens. Elle était bien perçue par les étudiants et proche de la réalité du stage. Une analyse continue permettra d'améliorer encore plus la qualité des ECOSM.

Mots-clés : évaluation, pédagogie, compétence clinique, études médicales.

SUMMARY

Introduction: Learning clinical teaching is an alternative to traditional forms of teaching. Its objective is to lead the student to acquire a relevant **Background :** multi-station objective clinical evaluation (MSOCE) is an equitable and genuine formative and normative assessment tool that is deemed relevant to its effectiveness, reliability and validity in clinical evaluation. However, no experience with its application in pathological anatomy has been published.

Aim : The aim of the present study was to carry out a critical analysis and a perceptual study of the MSOCE in pathological anatomy in medical students.

Methods : This was a cross-sectional, multicenter prospective study in which we developed and tested a MSOCE on 32 medical students of the 1st year of the second cycle of medical studies. Docimological analysis of the 12 stations and the 28 questions of the different stations was done by the file « AnItem.xls ».

Results : The mean total score obtained at the MSOCE was 36.2/50 (range : 29 to 41). The majority of stations were of acceptable difficulty (92%) and of poor discrimination (92%). The majority of questions were easy (57%) with poor discrimination (75% of the total questions). One question on 28 was « ideal », 4%. The MSOCE had a large heterogeneity (average Cronbach alpha of 0.29), and discriminated well between the average students (platykurtic, left-spread, negatively asymmetric distribution curve). The perception analysis showed that 27 students (84%) were satisfied with the OMSCE and thought that it was close to the reality of the clinical training.

Conclusion : The OMSCE tested was easy and discriminated mostly between average students. It was well perceived by the students and close to the reality of the clinical training. Ongoing analysis will further improve the quality of MSOCE.

Key words: evaluation, pedagogy, clinical competence, medical studies.

Correspondance

Limaïem Faten

Hôpital Mongi Slim La Marsa / Université de Tunis El Manar, Faculté de Médecine de Tunis

e-mail : fatenlimaïem@gmail.com

INTRODUCTION

Le système d'évaluation des compétences cliniques a été profondément modifié depuis une trentaine d'années au profit de techniques plus proches de la pratique médicale [1]. Ces procédures visent à augmenter la validité et la fiabilité de l'évaluation clinique, comparées aux examens conventionnels écrits et aux examens oraux. Parmi ces techniques, l'évaluation clinique objective en stations multiples (ECOSM), équivalent de l'«examen clinique par objectifs structurés» (ECOS), a été introduite à la Faculté de Médecine de Tunis (FMT) en 1998. L'ECOSM est une épreuve basée sur la succession de stations avec pour chacune d'elles, des problèmes à résoudre en temps limité. Chaque compétence est ainsi découpée en tâches facilement évaluable. L'ECOSM permet d'évaluer la réalisation d'objectifs relevant des domaines du savoir, du savoir faire et même certains aspects du savoir être. C'est un outil équitable et authentique d'évaluation formative et normative réputé pertinent pour son efficacité, sa fiabilité et sa validité lors de l'évaluation clinique. Cependant, aucune expérience concernant son application en anatomie pathologique n'a été publiée. L'objectif du présent travail était de réaliser une analyse critique et une étude de perception de l'ECOSM en anatomie pathologique des externes en 1^{ère} année du deuxième cycle des études médicales (DCEM1).

MÉTHODES

Présentation de l'étude

Nous avons mené une étude prospective transversale multicentrique dans trois centres hospitalo-universitaires: Hôpital Mongi Slim de la Marsa, Hôpital Charles Nicolle de Tunis et Hôpital Militaire Principal d'Instruction de Tunis. Notre étude a porté sur l'ECOSM en anatomie pathologique lors du deuxième semestre de l'année universitaire 2016 - 2017.

Population étudiée :

Public cible : Notre étude a concerné trois groupes d'étudiants de DCEM1 parmi six ayant passé l'ECOSM sanctionnelle en anatomie pathologique programmée par la FMT. A la fin de cette épreuve, nous leur avons proposé de passer une 2^{ème} ECOSM pour participer à une étude de recherche en pédagogie médicale. Le nombre des candidats ayant participé à l'ECOSM était de 32.

Critères d'inclusion : Nous avons inclus tous les étudiants ayant accepté de participer à l'étude.

Préparation de l'épreuve :

La planification de l'épreuve a été faite deux semaines à l'avance et comprenait le choix des objectifs à évaluer, la rédaction des stations et leur pondération, des consignes pour les étudiants et les examinateurs, des feuillets de réponse, des grilles de correction et des grilles de satisfaction (évaluation de l'ECOSM par les étudiants). Par ailleurs, nous avons préparé le matériel pédagogique nécessaire aux stations et le schéma du circuit. Nous avons également informé les personnes impliquées.

Choix des objectifs à évaluer à partir du carnet de stage :

Le choix des stations était en rapport avec les objectifs de stage, en pertinence avec le profil du médecin généraliste ou médecin de famille tunisien et le contenu des activités d'enseignement. Neuf objectifs parmi 17 ont été évalués relevant du domaine du savoir (n=3) et du savoir faire (n=6). Ils se rapportaient aux habiletés techniques et à la solution de problèmes de santé.

Rédaction des documents : Les stations étaient construites sur la base des compétences cliniques retenues pour la formation du futur médecin généraliste ou médecin de famille. De chaque compétence découle une série d'objectifs que l'étudiant doit atteindre pour maîtriser la compétence visée. Les cas étaient proches de la pratique courante tout en étant conformes aux objectifs de l'apprentissage figurant dans le carnet de stage.

Préparation des salles et du schéma du circuit : Nous avons pris soin que toutes les salles d'examen dans lesquelles les stations d'ECOSM étaient programmées, soient opérationnelles et adéquates pour le jour de l'examen. Les schémas de circuits préparés dans les trois services sont illustrés par la figure 1.

Recueil des données : Les données issues de l'ECOSM ont été recueillies à partir des différents fiches de correction et de satisfaction. La saisie des données a été réalisée à l'aide de Microsoft Office Excel 2007.

Analyse docimologique :

Chiffrier « Anltem.xls » : Nous avons utilisé pour l'analyse docimologique des questions le fichier « Anltem.xls ». Il s'agit d'un fichier Excel, proposé et utilisé par le docimologue « Serge Normand », au bureau d'évaluation de la Faculté de Médecine de l'Université de Montréal

à partir de l'an 2000. Il calcule rapidement les indices statistiques permettant l'analyse des questions. Ce fichier est téléchargeable gratuitement sur internet [2].

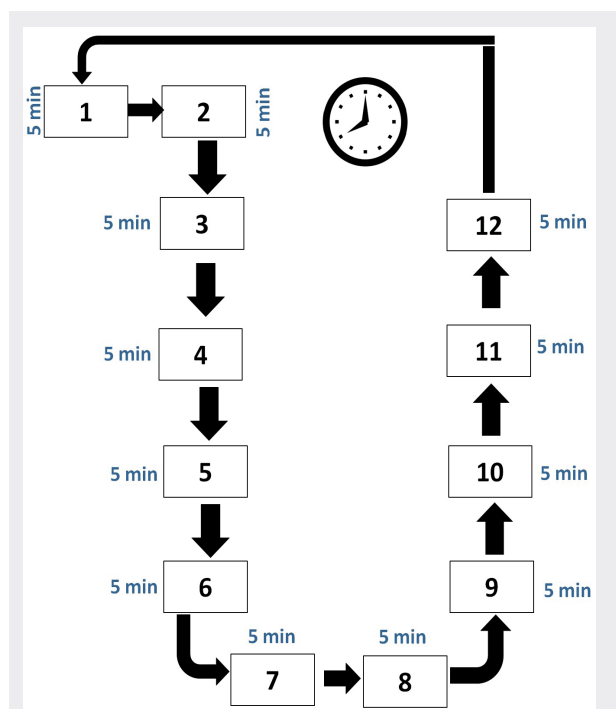


Figure 1. Schéma du circuit de l'ECOSM.

Analyse des questions et des stations d'ECOSM :

Nous avons analysé toutes les questions des stations d'ECOSM. Ainsi, deux fichiers « AnItem.xls » ont été créés pour analyser les stations d'ECOSM d'une part et toutes les questions de chaque station d'autre part. Dans notre étude, le chiffrier « AnItem.xls » utilisé nous a permis le traçage de la courbe de distribution des scores.

Paramètres analysés : Les principaux paramètres analysés étaient : l'écart-type, le coefficient alpha de Cronbach, l'indice de difficulté, l'indice de discrimination, la corrélation difficulté discrimination.

Coefficient alpha de Cronbach : Cet indice statistique varie entre 0 et 1. Il traduit un degré d'homogénéité (ou consistance) interne d'autant plus élevé que sa valeur est proche de 1. Préférentiellement, les valeurs de cet indice vont de 0,7 à 0,9

Difficulté : C'est le pourcentage de réussite à la question. L'indice de difficulté peut prendre des valeurs allant de 0 à 1. Au niveau du fichier « AnItem.xls », les questions dont l'indice de difficulté est inférieur à 0,50 sont marquées automatiquement dans un encadré pour attirer l'attention du lecteur.

Discrimination : La valeur de la discrimination prend des valeurs variables de (-1) à 1. Lorsque l'indice de discrimination est nul, la question ne produit aucune discrimination. Lorsqu'il est négatif, il révélerait l'existence d'une sorte d'anomalie, car la question serait globalement mieux réussie par les étudiants ayant les résultats les plus faibles sur l'ensemble de l'épreuve que par les étudiants qui présentent les résultats les plus élevés.

Corrélation difficulté discrimination : Elle permet d'étudier les caractéristiques d'une éventuelle relation linéaire entre la difficulté et la discrimination de l'ensemble des questions d'une épreuve donnée. Elle est obtenue par le coefficient « r » de Pearson. La valeur de cet indicateur varie de (-1) à (+1). Une valeur proche de 0 montre l'absence de relation linéaire entre les deux variables (mais n'exclut pas l'existence d'une relation autre que linéaire).

Question idéale : la question idéale est définie comme étant une question de difficulté acceptable et de bonne ou excellente discrimination.

Déroulement de l'épreuve :

Durée : l'examen s'est déroulé sur une matinée. La durée de l'épreuve était de 60 minutes et chaque station durait 5 minutes. Le temps de déplacement d'une station à une autre était de quelques secondes.

Plan détaillé chronologique : dans chaque service d'anatomie pathologique, le sens du parcours de l'ECOSM était affiché. Il était indiqué par les numéros de stations et organisé dans un seul sens. Une numérotation et un parcours étaient définis avec l'indication pour chaque étudiant de sa « station de départ » et du parcours à suivre (temps par station ; unité de lieu). Un signal sonore à la fin de chaque station indiquait à l'étudiant de passer à la station suivante. Chaque station comportait une fiche d'instruction pour l'étudiant, du matériel pédagogique si nécessaire et un feuillet de réponse. Les étudiants étaient amenés à répondre à une série de questions relatives au thème de la station.

Évaluation de l'ECOSM par les étudiants : À la fin de l'épreuve, les étudiants ont complété une grille de satisfaction pour exprimer leurs opinions sur le déroulement de l'ECOSM et son contenu.

Correction de l'ECOSM par l'enseignant : La correction de l'épreuve a été effectuée par un seul enseignant. La notation finale de l'ECOSM se faisait par addition de toutes les évaluations des stations réalisées par l'étudiant.

RÉSULTATS

Description de la population incluse : Trente-deux étudiants de troisième année de deuxième cycle des études médicales DCEM1 ont été inclus dans notre étude dont 28 de sexe féminin et 4 de sexe masculin soit un sex-ratio H/F de 0,14. Les étudiants étaient répartis en 3 groupes d'effectif respectif de 11, 12 et 9 étudiants en fonction du lieu de l'épreuve.

Analyse docimologique :

Scores des étudiants : le score total moyen obtenu à l'issue de l'évaluation des 12 stations d'ECOSM était de 36,2 sur 50 (extrêmes : 29 - 41). Le tableau I résume les indices docimologiques des différentes stations et questions de l'ECOSM.

Dispersion : l'analyse a révélé une homogénéité des résultats des étudiants pour toutes les questions (puisque l'écart-type des notes brutes de toutes les questions égal à 3,4 était <1/2 moyenne).

Difficulté : la majorité des questions, a été à facile (57%). Le taux des questions à difficulté acceptable était de l'ordre de 39% (Figure 2). Aucune question à indice égal à 0 n'a été retrouvée. Par contre, un indice de difficulté égal à 1 a été retrouvé dans trois questions. La majorité des stations d'ECOSM était de difficulté acceptable (59%). Toutes stations confondues, l'indice de difficulté moyen a été de 0,72.

Discrimination : les questions à mauvaise discrimination étaient majoritaires (75% du total des questions analysées) (Figure 3). Des questions à indice de discrimination négatif ont été retrouvées dans 11 cas soit 39%. Les stations à mauvaise discrimination étaient majoritaires (92% du total des stations).

Tableau 1. Indices docimologiques des différentes stations et questions de l'ECOSM.

			Indice de difficulté	Indice de discrimination	Alpha de Cronbach
Station 1			0,88	0,08	0,00
Q2	Q1		1,00	0,00	0,29
		0,80	0,08	0,23	
Station 2			0,64	0,12	0,05
	Q1		1,00	0,00	0,29
		0,84	-0,06	0,30	
Q2					
Q3		0,50	0,35	0,25	
Q4		0,33	-0,04	0,30	
Station 3			0,55	0,10	0,05
	Q1		0,55	0,10	0,26
Station 4			0,62	0,14	-0,04
Q2	Q1		0,70	0,07	0,24
		0,31	0,06	0,27	
Station 5			0,60	0,05	-0,11
Q2	Q1		0,70	0,26	0,20
Q3		0,69	-0,06	0,29	
		0,22	0,22	0,25	
Station 6			0,65	0,19	-0,51
Q2	Q1		0,71	0,33	0,11
		0,55	0,43	0,04	
Station 7			0,92	0,37	-0,06
Q2	Q1		0,85	0,27	0,24
Q3		1,00	0,00	0,29	
		0,89	0,46	0,23	
Station 8			0,89	-0,23	0,17
Q2	Q1		0,94	-0,08	0,30
Q3		0,83	-0,03	0,29	
		0,91	-0,09	0,30	
Station 9			0,64	-0,24	0,17
Q2	Q1		0,77	0,06	0,26
		0,50	-0,37	0,36	
Station 10			0,67	-0,11	0,11
	Q1		0,67	-0,11	0,31
Station 11			0,84	-0,09	0,11
	Q1		0,76	-0,13	0,31
		0,98	-0,10	0,30	
Q2					
Q3		0,72	0,14	0,28	
Station 12			0,67	0,15	-0,03
Q2	Q1		0,49	0,19	0,25
		0,79	-0,03	0,28	

* Q : question.

Corrélation difficulté discrimination : Le coefficient de corrélation entre les indices moyens de difficulté et de discrimination de toutes les questions a été de -0,17. Le coefficient de corrélation entre les indices moyens de difficulté et de discrimination de toutes les stations d'ECOSM a été de -0,023.

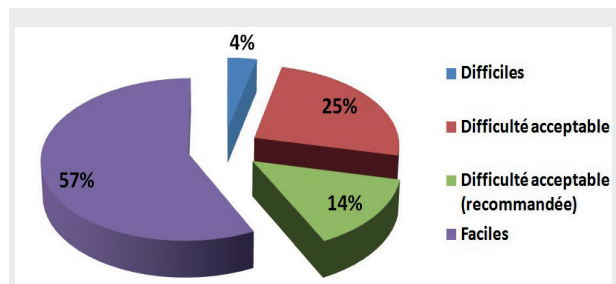


Figure 2. Répartition des questions en fonction de l'indice de difficulté.

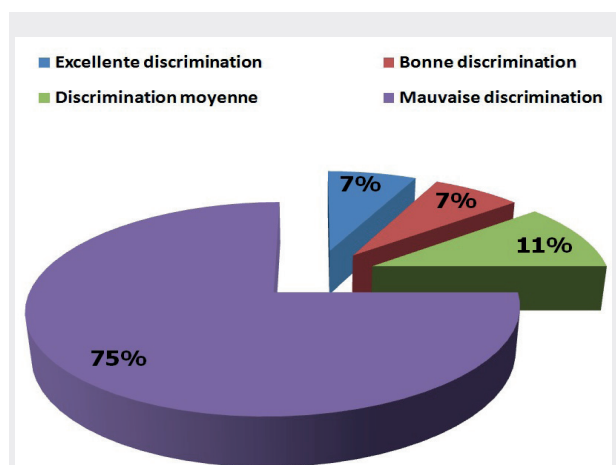


Figure 3. Répartition des questions en fonction de l'indice de discrimination

Questions idéales : La question 2 de la station 6 était une question idéale, de difficulté acceptable (indice de difficulté = 0,55) et d'excellente discrimination (indice de discrimination = 0,43).

Homogénéité interne : Toutes questions confondues, l'alpha de Cronbach moyen a été de 0,29 (homogénéité interne non acceptable) (grande hétérogénéité). Pour les stations d'ECOSM prises dans leur globalité, l'alpha de Cronbach moyen a été de 0,10 (homogénéité interne non acceptable) (grande hétérogénéité).

Courbes de distribution des scores : le coefficient d'aplatissement est négatif égal à -0,94. La distribution

était platykurtique (figure 4). Elle était basse et relativement aplatie. Son pic était plus arrondi autour de la moyenne avec des extrémités plus courtes et resserrées où il est difficile de différencier les individus. Le coefficient d'asymétrie était égal à -0,31. L'épreuve avait une courbe de distribution à asymétrie négative (moyenne < médiane < mode). Elle était étalée à gauche. C'est la représentation graphique d'un effet plafond : l'épreuve est facile.

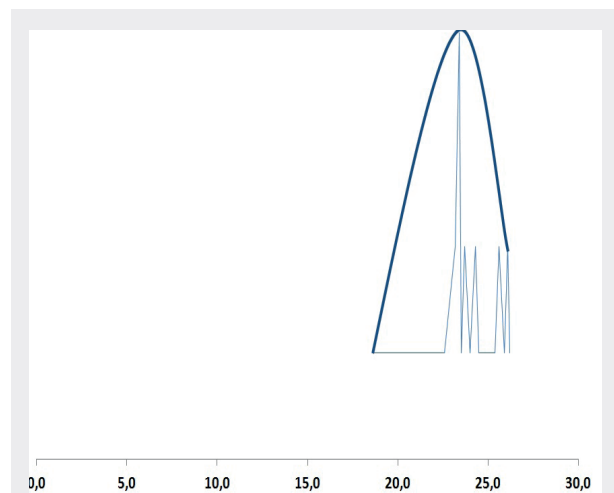


Figure 4. Courbe de distribution des scores dans les stations d'ECOSM

Évaluation de l'ECOSM par les étudiants : l'analyse des grilles de satisfaction a révélé que la majorité des étudiants étaient satisfaits des conditions de déroulement de l'épreuve soit 94%. Vingt-sept étudiants (84%) pensaient que l'épreuve était proche de la réalité des stages effectués.

DISCUSSION

Dans notre étude, nous avons développé et testé une ECOSM en anatomie pathologique sur 32 étudiants de DCEM1 ayant effectué un stage d'anatomie pathologique d'une durée de trois semaines durant le deuxième semestre de l'année universitaire 2016 - 2017. Le score total moyen obtenu à l'issue de l'ECOSM était de 36,2 sur 50 avec des extrêmes allant de 29 à 41. La majorité des stations était de difficulté acceptable (92%). Aussi, les questions à difficulté acceptable et faciles ont représenté respectivement 39% et 57%, soit 96% du total des questions. Les questions à mauvaise discrimination étaient majoritaires, représentant

75% du total des questions analysées. Une question sur 28 était « idéale », soit 4%. L'ECOSM avait une courbe de distribution platykurtique à asymétrie négative.

Les ECOSM ont fait l'objet de nombreuses publications et travaux dans diverses spécialités. A notre connaissance, une évaluation par ECOSM n'a à ce jour jamais été publiée dans le domaine de l'anatomie pathologique. Cependant, notre travail avait plusieurs limites notamment :

- Le faible effectif des participants et des stations ne permettant pas d'obtenir une bonne validité du contenu.
- Le système de notation adopté était arbitraire et ne pouvait donc pas juger de la compétence réelle de chaque candidat.
- Les terrains de stage des étudiants ayant passé l'ECOSM étaient différents avec des degrés variables d'encadrement et d'atteinte des objectifs d'apprentissage.
- L'absence de stations testant le savoir être des étudiants. Le patient standardisé offre une qualité d'interaction qui permet de développer des scénarios intégrant des valeurs telles qu'humanisme, éthique et déontologie.
- La fatigue ou le manque de motivation de certains étudiants vu que cette épreuve a succédé l'ECOSM sanctionnelle de 15 minutes.

L'ECOSM bénéficie d'une grande reconnaissance dans les milieux de formation médicale et est considérée comme étant un examen adéquat pour évaluer les compétences cliniques des étudiants [2]. Cet examen a la particularité d'être objectif puisque tous les candidats sont évalués à l'aide d'exactly les mêmes stations avec la même grille et le même schéma de notation. L'ECOSM est considérée comme étant une méthode d'évaluation équitable, car les situations cliniques sont les mêmes pour tous les étudiants soumis à l'ECOSM [3]. De plus, grâce à l'ECOSM, les étudiants peuvent s'exercer dans un environnement sécuritaire où il n'y a pas de risque encouru pour les patients simulés [4]. S'ajoute aussi l'avantage de pouvoir évaluer simultanément plusieurs étudiants dans un circuit ECOSM [5]. Dans notre étude, la durée de l'épreuve était de 60 minutes et toutes les stations duraient 5 minutes. Une ECOSM est composée d'une séquence de cas de patients, de situations cliniques, présentées dans différentes stations, de même durée, constituant un circuit à travers lequel les étudiants font la rotation. L'ECOSM classique comporte une vingtaine de stations d'une même durée (4 à 5 minutes) [6]. Toutes les cinq minutes, une

sonnerie retentit et les étudiants ont 30 secondes pour changer. Il y a deux types de postes, selon qu'il s'agit d'exécuter des tâches ou de répondre à des questions [6]. La note globale est substituée par un score, construit à partir d'une check-list d'actions séparées ou d'échelles globales [6]. Les examinateurs se sont mis d'accord au préalable sur ce qu'ils attendent de l'étudiant. Si le candidat a documenté le maximum d'items, il a un bon score. Le total des notes peut être assorti d'une pondération [6]. La validation des stations est la dernière phase de construction et s'effectue avec l'aide de pairs professionnels de terrain. Après l'examen, un feedback est donné aux enseignants et aux étudiants [7]. Le débriefing est un processus d'apprentissage réflexif intentionnel où enseignants et étudiants réexaminent ensemble la situation clinique en favorisant le développement du raisonnement clinique et les capacités de jugement de l'apprenant [8]. Il doit être conçu pour mettre en synergie, renforcer et transférer des connaissances à partir d'une expérience d'apprentissage. Dans notre étude, l'évaluation de l'ECOSM a été faite par l'analyse docimologique. Cette analyse comporte trois volets : la vérification de la congruence entre les questions posées et les objectifs, le calcul de différents indices statistiques et la décision de réviser, remplacer ou conserver les questions. Dans notre étude, la majorité des stations d'ECOSM était de difficulté acceptable (92%) et la majorité des questions était facile (57% du total des questions). Une épreuve de qualité en termes de difficulté, est une épreuve à difficulté acceptable. C'est pour cela qu'on opte plutôt pour les questions à difficulté acceptable. Ceci dit, l'épreuve doit comporter également des questions faciles permettant de distinguer les étudiants de faible niveau, ainsi que des questions difficiles permettant de différencier les étudiants de niveau de compétence plus élevé [9]. Une fois l'indice de difficulté de la question a été déterminé, c'est la bonne interprétation de ce paramètre qui semble être la tâche la plus délicate vu qu'il est influencé par plusieurs facteurs dont il faut tenir compte [10, 11]. Les caractéristiques du groupe d'étudiants ayant passé l'examen, par exemple, semblent influencer cet indice. Leur niveau de compétence réelle, leur degré de préparation à l'examen conditionnent en partie leurs réponses aux questions. D'ailleurs, certains auteurs ont stipulé même que l'indice de difficulté pour un groupe bien préparé peut aller de 0,7 à 1 [12]. Nos étudiants étaient très bien préparés à cette ECOSM qui a directement succédé l'examen. Aussi, les résultats retrouvés pour

toutes les questions, n'ont pas été trop dispersés indiquant l'homogénéité du niveau des étudiants. Le niveau cognitif mesuré par la question peut également influencer sa difficulté. En effet, une question faisant appel au savoir diffère forcément de celle nécessitant un esprit de synthèse ou d'analyse [13]. Dans notre étude, trois questions faisaient appel au savoir et neuf questions faisaient appel au savoir faire. Des problèmes liés à la formulation de la question ainsi qu'à son contenu, peuvent influencer l'indice de difficulté. Une question classée statistiquement difficile peut être de formulation ambiguë ou prêtant à confusion, admettre plus qu'une réponse, ou avoir une clé de réponse carrément incorrecte [14]. En contrepartie, une question classée statistiquement facile peut concerner un objectif important sur lequel les encadreurs référents ont trop insisté et les étudiants s'y sont intéressés à la révision. Dans notre étude, les questions à mauvaise discrimination étaient majoritaires (75% du total des questions analysées). L'indice de discrimination est la différence entre l'indice de difficulté de la question calculé pour le groupe dit « fort ou supérieur » et l'indice de difficulté pour le groupe dit « faible ou inférieur ». Plus l'écart est grand, plus la question discrimine entre les étudiants **ayant** eu un résultat total élevé à l'épreuve et les élèves ayant eu un résultat total faible [10]. Dans notre étude, la majorité des questions était facile, le nombre de questions était réduit et le nombre d'étudiants n'était pas important. En effet, un nombre d'examinés réduit (< 30 étudiants), peut faire accroître les variations aléatoires de l'indice de discrimination. Un nombre de questions également réduit peut nuire à la discrimination. Une question très facile (ou très difficile) altère le pouvoir discriminant. Aussi, l'homogénéité du groupe d'étudiants, diminue le pouvoir discriminant des questions de l'examen. Dans notre étude, un indice de difficulté égal à 1 a été retrouvé dans trois questions. Il n'est pas intéressant d'avoir une question réussie par tout le monde (indice de difficulté égal à 1 donc question à facilité absolue) ou par personne (indice de difficulté égal à 0 donc question à difficulté absolue) car de telles questions ne permettent pas de discriminer les étudiants entre eux. Dans notre étude, l'homogénéité interne a été estimée par le coefficient alpha de Cronbach. C'est l'outil statistique le plus puissant en matière d'homogénéité interne. Le chiffrier « AnItem.xls » permet en plus de l'estimation de la consistance interne d'une épreuve donnée, de calculer la consistance de l'épreuve si une

question désignée est omise. Ceci permet une analyse plus spécifique des questions et pointe de façon précise les questions sources d'hétérogénéité au sein de l'examen. Dans notre étude, l'homogénéité interne était jugée non acceptable avec un alpha de Cronbach moyen de 0,29 reflétant une grande hétérogénéité. Cet indice statistique varie entre 0 et 1. Il traduit un degré d'homogénéité (ou consistance) interne d'autant plus élevé que sa valeur est proche de 1. Préférentiellement, les valeurs de cet indice vont de 0,7 à 0,9. Cependant, dans notre étude le nombre réduit des stations altère cet indice. Dans notre étude, l'épreuve avait une courbe de distribution à asymétrie négative (moyenne < médiane < mode) et donc était facile. Elle était platykurtique étalée à gauche et donc elle discriminait bien entre les étudiants moyens mais peu entre les faibles et les forts. Dans notre étude, nous avons testé une ECOSM sur un groupe d'étudiants avant de l'introduire pour l'évaluation sanctionnelle. Le but de l'évaluation de l'ECOSM était d'identifier les problèmes observés lors du déroulement de l'épreuve, de mesurer la réussite de chaque station et de l'ensemble de l'épreuve, de proposer des améliorations pour les prochaines évaluations et de proposer des rectifications dans la formation des étudiants en fonction de la réussite de chaque station. L'étude de perception a montré que la majorité des étudiants (85%) étaient satisfaits de l'ECOSM. Cependant, cinq étudiants (16%) n'étaient pas satisfaits du temps alloué aux stations d'ECOSM. Ainsi, deux étudiants ont signalé que l'énoncé de certaines stations (sans préciser lesquelles) devrait être allégé. Il faut donc raccourcir les énoncés de ces stations ou en prolonger la durée. Vingt-sept étudiants (84%) pensaient que l'épreuve était proche de la réalité des stages effectués.

CONCLUSION

En conclusion, l'ECOSM testée dans notre étude était facile et discriminait surtout entre les étudiants moyens. Elle était bien perçue par les étudiants et proche de la réalité du stage. Les remarques et les besoins exprimés par les étudiants serviront pour le développement de meilleures stations ECOSM qui composeront le circuit. Malgré ses limites, notre étude a prouvé que l'ECOSM pouvait être un outil pédagogique et d'évaluation formative utilisable pour l'enseignement des étudiants en médecine. Des études complémentaires sont nécessaires pour juger de ses qualités psychométriques. Par ailleurs, une analyse

continue permettra d'améliorer encore plus leur qualité.

Pour améliorer le déroulement de l'ECOSM et optimiser sa réussite, nous proposons les solutions suivantes :

- Les stations doivent être préparées au moins un mois à l'avance de façon collégiale. Le choix des objectifs à évaluer et la rédaction des différentes stations doivent être discutés.
- Les questions produites ainsi que leurs grilles de notation respectives doivent être minutieusement élaborées afin d'éliminer au maximum tout ce qui peut nuire à leurs caractéristiques métriques avant d'être incluses dans les épreuves.
- Un "débriefing" doit succéder à l'épreuve pour donner les réponses correctes et pour avoir un feedback de tous les intervenants.

Conflits d'intérêts :

Les auteurs ne déclarent aucun conflit d'intérêts.

RÉFÉRENCES

- Bertrand C, Hodges B, Segouin C, Gagnayre R, Ammirati C., Marty J, et al. Les examens cliniques par objectifs structurés. *Le Praticien en anesthésie réanimation*. 2008;12(3):212-7.
- Mitchell ML, Henderson A, Groves M, Dalton M, Nulty D. The objective structured clinical examination (OSCE): Optimizing its value in the undergraduate nursing curriculum. *Nurse Education Today*. 2009;29(4): 398-404.
- Newble D. Assessing clinical competence at the undergraduate level. *Medical Education*. 1992;26(6):504-11.
- Barbary N. Du côté des statisticiens. Excel 2010 expert: Fonctions, simulations, bases de données. Paris: Eyrolles. 2011:401-33.
- Fox W. Les mesures de tendance centrale. *Statistiques sociales*. Québec: Les presses de l'Université Laval, 2007:69-117.
- Descargues G, Sibert J, Lechevallier, Weber JP, Marpeau L. Évaluation au cours de la formation initiale de la compétence clinique en Gynécologie Obstétrique: approche innovatrice basée sur l'observation de la performance par l'examen clinique objectif et structuré (ECOS). *Journal de Gynécologie Obstétrique et Biologie de la Reproduction*. 2001;30(3):257-64.
- Van Der Vleuten CP. Assessment of clinical skills with standardized patients : state of the art. *Teaching and Learning in Medicine*. 1990;2(2):58-76.
- Salas E, Klein C, King H, Salisbury M and Al. Debriefing medical teams : 12 evidence-based best practices and tips. *Joint Commission Journal on Quality and Patient Safety*. 2008; 34(9):518-27.
- Chartier P, Loarer E. L'analyse des items. Evaluer l'intelligence logique: Approche cognitive et dynamique. Paris: Dunod, 2008:43-65.
- Laveault D, Grégoire J. La fidélité des résultats. Introduction aux théories des tests en psychologie et en sciences de l'éducation. Bruxelles: De Boeck, 2005:110-2.
- Fouassin Dome M. Elaboration et analyse d'un test de dépistage en mathématiques. Maîtrise de l'enseignement en mathématiques, Université de Québec à Chicoutimi, Québec 1982.
- Barman A, Ja'afar R, Rahim F, Noor A. Psychometric characteristics of MCQs used in assessing phase-II undergraduate medical students of Universiti Sains Malaysia. *The Open Medical Education Journal*. 2010;3(3):1-4.
- Nguyen D, Blais J. Approche par objectifs ou approche par compétences? Repères conceptuels et implications pour les activités d'enseignement, d'apprentissage et d'évaluation au cours de la formation clinique. *Pédagogie médicale*. 2007;8(4):232-51.
- Hingorjo M, Jaleel F. Analysis of One-Best MCQs: the difficulty index, discrimination index and distractor efficiency. *J Pak Med Assoc*. 2012;62(2):142-7.