

L'examen clinique objectif structuré (ECOS) comme examen validant des compétences cliniques des étudiants en médecine français : 13 ans d'expérience rouennaise

Objective Structured Clinical Examination as a validating examination of French medical students' clinical skills: a 13-year experience at Rouen University's School of Medicine

Louis SIBERT¹, Pascale SCHNEIDER², Agnès LIARD^{3,*}, Antoine LEFEVRE-SCHELLES⁴, Jean-François MENARD⁵, Isabelle AUQUIT-AUCKBUR⁶, et Benoît VEBER⁷

¹ Medical Training Center Normandie Rouen, 76000 Rouen, France

² Responsable de l'Unité d'Enseignement Compétences Cliniques, UFR Santé Rouen, Université de Rouen, 76000 Rouen, France

³ UFR Santé Rouen, Université de Rouen, 76000 Rouen, France

⁴ Responsable du CESU 76 (Centre d'Enseignement des Soins d'Urgence), 76000 Rouen, France

⁵ Département de Biostatistique, UFR Santé Rouen, Université de Rouen, 76000 Rouen, France

⁶ CHU de Rouen, Service de chirurgie plastique et reconstructrice, 76000 Rouen, France

⁷ Doyen, UFR Santé Rouen, Université de Rouen, 76000 Rouen, France

Manuscrit reçu le 19 novembre 2020 ; commentaires éditoriaux formulés aux auteurs le 13 octobre et le 24 novembre 2021 ; accepté pour publication le 25 novembre 2021

Résumé - Contexte : La réforme du second cycle des études médicales en France va introduire l'examen clinique objectif structuré (ECOS) au sein des épreuves nationales ouvrant l'accès au 3^e cycle. **But :** Rapporter treize ans d'expérience d'ECOS d'une faculté française et en décrire les points clés de son développement et implantation. **Méthodes :** Les dispositifs de planification et d'administration des ECOS sont décrits. Les scores moyens obtenus à chaque session, aux habiletés cliniques, leur fidélité (coefficient alpha), la distribution des scores au sein de chaque cohorte (coefficients de Kurtosis, Skewness) sont rapportés. Les performances aux ECOS et aux épreuves classantes nationales (ECN) ont été comparées pour les cohortes 2018 et 2019. **Résultats :** Un ECOS (7,4 stations en moyenne) a été administré consécutivement de 2008 à 2020 à des promotions de 200 étudiants en moyenne (extrêmes : 145–236). La durée moyenne des circuits était de 68 minutes (extrêmes 48–97). Les indices de fidélité variaient de 0,52 [IC5% : 0,41–0,58] à 0,73 [IC5% : 0,67–0,77] pour les scores aux stations, et de 0,65 [IC5% : 0,57–0,70] à 0,82 [IC5% : 0,78–0,85] par habiletés cliniques. Les coefficients de Kurtosis et de Skewness variaient respectivement de $2,36 \pm 0,5$ à $5,56 \pm 1,69$ et de $-0,10 \pm 0,11$ à $-0,96 \pm 0,22$. Les performances aux ECOS n'étaient pas corrélées aux ECN (coefficients de Spearman et de Pearson). **Conclusion :** L'implantation de l'ECOS au sein de notre faculté comme examen validant de fin de second cycle s'est avérée pérenne. Malgré des limites docimologiques, il permet un classement et l'identification des étudiants en difficulté lors de la pratique clinique. Les points clés sont un soutien institutionnel, la rédaction des stations, la formation des observateurs. L'entraînement aux habiletés cliniques doit s'inscrire dans la réflexion plus globale sur la place de la simulation en formation initiale en France.

Mots clés : ECOS, compétence clinique, évaluation, étudiant en médecine, niveau pré-gradué

Abstract. Context: The Objective Structured Clinical Examination (OSCE) will be used in the final National Ranking Examination (NRE) at the end of year 6 of medical studies in France. **Aim:** To report the 13-year experience of OSCEs in a French faculty of medicine and to assess the key features of its development and establishment. **Methods:** The organization and administration of the OSCEs are described. The mean scores and the mean clinical skills scores obtained, their fidelity (alpha coefficient) and the distribution of scores within

*Correspondance et offprints : Louis SIBERT, Medical Training Center Rouen Normandie, 20, rue Marie-Curie, 76000 Rouen, France.
Mailto : Louis.Sibert@chu-rouen.fr.

each cohort (Kurtosis and Skewness coefficients) were analyzed. OSCE and NRE scores were compared for the 2018 and 2019 cohorts. **Results:** An OSCE (mean 7.4 stations) was administered consecutively between 2008 and 2020 to year groups of approximately 200 students (range: 145–236). The mean duration of the circuit was 68 minutes (range: 48–97). Mean fidelity coefficients ranged from 0.52 [IC5%: 0.41–0.58] to 0.73 [IC5%: 0.67–0.77] for station scores and from 0.65 [IC5%: 0.57–0.70] to 0.82 [IC5%: 0.78–0.85] for clinical skills scores. Mean Kurtosis and Skewness coefficients respectively ranged from 2.36 ± 0.5 to 5.56 ± 1.69 and to -0.10 ± 0.11 to -0.96 ± 0.22 . Performance on OSCE was not correlated with NRE (Person and Spearman Coefficients). **Conclusion:** Implementation of the OSCE within our faculty as a validating second cycle end-of-cycle exam has proven to be sustainable. Despite statistical limits, it allows the classification of students and identification of students with difficulties in clinical practice. The key points are institutional support, designing of stations, relevant training for observers. Clinical skills training must be part of a more global reflection on the role of simulation in undergraduate medical studies in France.

Keywords: OSCE, clinical competence, assessment, medical student, undergraduate level

Introduction

La réforme du second cycle des études médicales en France

La réforme du second cycle des études médicales qui va être mise en place en France à partir de 2023 va entraîner de profondes modifications des activités d'enseignement et d'évaluation [1]. D'une approche traditionnelle « par objectifs », le second cycle va basculer vers une « approche par compétences ». Le concept d'approche par compétences n'est pas nouveau [2]. Il s'agit d'un courant pédagogique développé à partir des années 1980 pour lequel la compétence est une mobilisation et combinaison de ressources de divers domaines (cognitif, affectif, sensorimoteur), de différents types (connaissances déclaratives et connaissances d'action) et différentes origines (savoirs codifiés et savoirs d'expérience), dont la mobilisation aboutit à un « savoir-agir » adapté à l'accomplissement d'une tâche spécifique. Cette approche par compétences implique le développement d'apprentissages contextualisés et des processus d'évaluation authentique.

Ainsi, les épreuves classantes nationales informatisées (ECNi) vont être supprimées et remplacées par un système algorithmique d'affectation à l'entrée du troisième cycle, qui permettra d'attribuer à chaque étudiant un poste en se basant sur un trépied de critères: les connaissances théoriques, les compétences cliniques et le parcours de formation et d'expériences extra-universitaires. Les connaissances théoriques seront évaluées par le biais d'épreuves nationales dont le contenu fera l'objet d'une diversification docimologique majeure, avec l'introduction, entre autres, de questions à choix multiples (QCM) à contexte enrichi, de questions par point clés (*Key-Feature Problems*) et de tests de concordance de script. L'évaluation des compétences cliniques sera réalisée par le biais d'une épreuve cadrée nationalement se basant sur l'examen clinique objectif structuré (ECOS) [3].

L'ECOS

L'ECOS constitue une approche de l'évaluation de la compétence clinique dans laquelle toutes les composantes de la compétence sont évaluées de façon planifiée et structurée avec une attention particulière portée à l'objectivité et à la standardisation de l'examen. Un

ECOS comprend une série de situations cliniques décrites dans des stations de quelques minutes chacune, qui forment un circuit à travers lequel les étudiants font la rotation. Dans chaque station, les étudiants accomplissent une tâche spécifique parmi les éléments importants de la compétence clinique en interagissant avec un patient simulé et standardisé. Leurs performances sont évaluées par un observateur qui dispose d'une grille préétablie. Depuis son avènement il y a plus de 40 ans [4], l'ECOS est devenu le format d'évaluation de la compétence clinique le plus populaire au niveau international. Il est administré de façon courante depuis plusieurs décades au sein des cursus de santé, médicaux et non médicaux, à tous les niveaux de formation [5] dans les pays anglo-saxons, sud-américains, asiatiques et en Europe, aussi bien en tant qu'examen formatif que validant [5–8] et sous de très nombreux formats, en présence ou à distance [9]. Il a fait l'objet d'une très abondante littérature qui a fourni un grand nombre de données probantes sur son impact pédagogique positif, son acceptabilité [5,7], sa fidélité, sa validité [10–16] (et sur les règles de construction pour y parvenir) [17–20].

En France, par contre, le développement de l'ECOS est resté quasi-confidentiel jusqu'à l'avènement de la réforme actuelle. La validation des compétences cliniques était effectuée jusqu'en 2012 sous forme d'un examen oral dans le cadre du Certificat de synthèse clinique et thérapeutique. Actuellement l'examen validant est le Certificat de compétences cliniques dont l'obtention est indispensable pour valider le deuxième cycle [21]. Si, officiellement, ce certificat a pour but de vérifier les compétences cliniques des étudiants et leurs capacités à synthétiser les connaissances acquises, en pratique, il consiste le plus souvent en un examen clinique au lit d'un patient suivi d'une synthèse devant un jury. L'hétérogénéité des situations cliniques proposées et l'absence de standardisation des jurys rendent compte des biais de cette approche évaluative.

Le contexte académique et institutionnel de l'évaluation des compétences cliniques au sein de l'Unité de formation et de recherche (UFR) Santé de l'Université de Rouen

Contrairement à la majorité des UFR Santé françaises, la première expérience d'ECOS au sein de notre UFR date de plus de 20 ans et a été menée avec succès dans le

domaine de la médecine générale (18 stations dont 15 avec patients simulés administrés à 50 internes). L'ECOS a été ensuite intégré au processus officiel de validation du diplôme d'études spécialisées (DES) de médecine générale [22]. Parallèlement, nous avons développé avec succès plusieurs expérimentations d'ECOS au sein de DES de spécialités chirurgicales [23,24]. Compte tenu de la nécessité d'optimiser la reproductibilité et la standardisation de l'examen du Certificat de compétences cliniques, examen à fort enjeux car dernier examen facultaire avant les ECNi, nos expériences probantes d'ECOS nous ont fait adopter dès 2008 cette modalité comme examen de validation des compétences cliniques. Cette modification a été validée en conseil de gestion de l'UFR et figure depuis au contrat pédagogique de la troisième année du Diplôme de formation approfondie en sciences médicales (DFASM3).

L'organisation de l'examen a été confiée à une équipe spécifique comprenant le responsable du Certificat de compétences cliniques, le Vice-Doyen à la pédagogie et le responsable du centre de simulation du Campus Santé. Cette équipe est responsable du développement de l'ECOS et coordonne en collaboration avec la scolarité de l'UFR l'organisation logistique de l'examen et l'exploitation des résultats. Depuis 2008, l'ECOS est réalisé chaque année universitaire dans les locaux de l'UFR. Les étudiants ajournés passent un examen de rattrapage sous forme d'un examen clinique au lit du patient, suivi d'une synthèse orale dans les services ayant participé à la réalisation des stations.

Objectifs

Pour préparer au mieux ce changement de paradigme au niveau national, le ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation a listé les objectifs à mettre en place au sein de chaque faculté dès septembre 2021 afin de former les enseignants et les étudiants aux ECOS, à savoir : organiser des ECOS facultaires sur le plan logistique, mettre en place des ECOS formatifs, mettre en place des ECOS sommatifs à l'échelle d'une promotion [3,25].

Dans ce contexte de la réforme du second cycle des études médicales, il nous a semblé intéressant de rapporter notre expérience de 13 ans d'ECOS en tant qu'examen facultaire officiel au sein de notre UFR, d'en faire un bilan en termes de faisabilité et d'apport pédagogique.

Les objectifs spécifiques concernent la description des processus de développement, d'implantation et d'organisation au fil de nos années d'expérience d'ECOS validant, les difficultés organisationnelles et pédagogiques rencontrées et les analyses pour y remédier. Dans ce cadre, l'énoncé de données chiffrées concernant nos ECOS et des résultats obtenus par les étudiants n'ont qu'une valeur purement illustrative et descriptive, dans l'unique but d'être le plus informatif possible sur le plan pratique vis-à-vis de la collectivité éducative, à l'heure où ce dispositif est en train d'être adopté par l'ensemble des facultés de médecine françaises.

Méthodes

Type de l'étude

Il s'agit d'une étude de cohorte, rétrospective et mono centrique, menée sur treize ans.

Population concernée

Tous les étudiants en médecine inscrits en DFASM3 au sein de notre UFR entre 2008 et 2020 ont été concernés par l'étude.

Développement de l'examen

Le déploiement de notre processus de planification a été progressif sur plusieurs années mais avec d'emblée le souci de respecter au maximum les principales étapes de la planification d'un ECOS, telles qu'elles étaient préconisées dans la littérature disponible au début de notre expérience [12–16] et largement établies par la suite [17,18,20], à savoir :

- l'établissement d'un rétro planning au moins cinq à six mois à l'avance, en début d'année universitaire, qui est un point clé du processus d'implantation. Il faut identifier très tôt les dates souhaitées de livraison de la rédaction du synopsis des stations, des séances de relecture de ces synopsis, des sessions d'entraînement des patients simulés, des sollicitations des observateurs, de préparation du matériel pédagogique et des salles, en lien avec le service de scolarité de l'UFR.
- la détermination du contenu de l'examen, à savoir le choix des disciplines impliquées, les thèmes à traiter par ces disciplines, les habiletés cliniques à évaluer à travers les stations. Cette étape d'établissement d'une table de spécification est primordiale car elle donne une vision globale de l'examen, elle participe à la pertinence du contenu de l'examen et permet de décider de façon documentée du nombre de stations qui composeront l'ECOS.
- la rédaction du synopsis des stations, qui est assurée par des enseignants sollicités spécifiquement et suit un canevas pré établi [14,16,20]. Pour chaque cas, les rédacteurs doivent identifier et pondérer les habiletés cliniques évaluées, identifier et rédiger le problème ou le symptôme clinique de présentation, le contexte de présentation, les caractéristiques des patients, les instructions aux étudiants en spécifiant les tâches qu'ils doivent accomplir, le scénario et les instructions pour les patients simulés, les instructions pour les observateurs et la grille d'évaluation, le matériel pédagogique nécessaire pour simuler le contexte clinique décrit dans la station. Cet aspect rédactionnel est un point majeur pour tendre vers une standardisation acceptable de ce type d'évaluation et constitue une source potentielle de biais de mesure [15,17,18,20]. Pour tenter d'y pallier, à la suite des premières sessions, nous diffusons à tous les responsables de stations un exemple de synopsis de station ECOS avec les règles de rédaction à suivre.

- l'élaboration des grilles d'évaluation, qui est un élément clé pour la reproductibilité des scores. Les consignes données aux rédacteurs étaient d'inclure dans la grille uniquement les items jugés essentiels pour la résolution de la tâche clinique demandée et de les pondérer selon leur pertinence et leur signification clinique, sans dépasser le nombre maximum de 20 items par grille. Pour toutes les stations, les grilles ont été rédigées sous forme de liste de vérification binaire pour les habiletés cliniques observables. Deux items mesurant les habiletés organisationnelles et relationnelles, cotés selon une échelle qualitative en cinq points, étaient systématiquement intégrés. Un exemple d'instruction aux étudiants et de grille d'évaluation est présenté sur la [figure 1](#). Nous avons établi comme consigne de ne traiter qu'un à trois objectifs d'évaluation par station et de bien vérifier que ceux-ci et les habiletés techniques testées dans les stations soient bien en adéquation avec le programme officiel du second cycle des études médicales [21].
- le recrutement des patients simulés, qui est assuré par les responsables de station. Il s'agit de personnels hospitaliers ou de l'université ou de la société civile. Les critères de recrutement suivent les recommandations publiées sur la thématique [17–19]. Le recrutement des patients devait être effectif au plus tard un mois avant l'examen, pour permettre aux responsables de stations d'assurer auprès des patients l'explication du scénario et des instructions, les séances de répétition. Cet entraînement est primordial, notamment pour assurer la standardisation des différents patients qui devront jouer le même rôle pour la même station dans des circuits différents. À titre de dédommagement, les patients simulés mineurs ou retraités se voyaient attribuer un bon d'achat, les patients simulés majeurs étaient payés en vacances grâce à des crédits pédagogiques de l'UFR.
- la formation des observateurs, qui est assurée par les responsables de stations qui vérifient la bonne compréhension de chaque item de la grille et des performances attendues de la part des étudiants. Le même processus que pour les patients est appliqué pour assurer la standardisation des différents observateurs qui interviennent dans la même station dans des circuits différents.
- le planning horaire précis, qui est établi par le service de scolarité de l'UFR. Ce planning concerne les convocations des étudiants, leur répartition dans les circuits, les horaires de début et de fin des circuits, les horaires précis d'entrée dans les circuits de chaque étudiant et la station d'entrée. Les agents du service de scolarité prennent également en charge la préparation des circuits et installent dans chaque salle le matériel pédagogique et le mobilier nécessaires à la reconstitution de la situation clinique simulée, la saisie des grilles, les instructions aux candidats.

Administration de l'ECOS

Les ECOS se déroulent sur une ou deux journées selon le nombre de stations, habituellement courant février,

dans les locaux de l'UFR en plusieurs circuits simultanés. Les étudiants, préalablement répartis en groupes correspondant à chaque rotation, sont convoqués une heure avant le début de leur circuit pour assister à un briefing concernant les consignes sur le déroulement de l'ECOS. Ils sont amenés sur le circuit et répartis devant les salles accueillant les stations par un responsable de circuit qui veille au bon déroulement des rotations, associé à un responsable du minutage qui donne le signal d'entrée et de sortie dans les stations. Pour chaque circuit, un membre de la scolarité veille au bon fonctionnement des rotations sur le plan matériel et pour la saisie des données. Un enseignant supervise chaque étage où se déroulent plusieurs circuits en parallèle. À chaque étage, deux salles de détente sont dédiées aux observateurs, aux patients et aux étudiants qui sont en station repos. Plusieurs personnes-ressources pour remplacer un observateur ou un patient simulé sont également convoquées pour le début des rotations. À l'issue de chaque rotation, un débriefing est proposé aux étudiants pour analyser leurs ressentis et remarques.

Collecte des résultats et analyses statistiques

Pour chaque station, le score maximal était de 100 points et le score minimal de 0. Les scores étaient ensuite ramenés à 20. Chaque station avait le même poids. Le score final de chaque étudiant était établi en faisant la moyenne de la somme des scores obtenus à chaque station.

Pour chaque ECOS, les résultats sont exprimés en moyenne, écart-type, médiane, maximum et minimum, nombres et pourcentage d'étudiants ajournés. À noter que le seuil de passage a été arbitrairement établi à 10 sur 20 de 2008 à 2018 puis à 12 sur 20 à partir de 2019.

Pour examiner la pertinence du contenu pédagogique de nos ECOS (c'est-à-dire pour vérifier si les objectifs d'évaluation décrits dans les stations sont bien adaptés aux objectifs du programme pédagogique officiel de fin de second cycle des études médicales [21]), nous avons identifié le poids accordé à chaque habileté clinique, exprimé en pourcentage du score maximal, au sein de chaque ECOS, et en pourcentage moyen pour l'ensemble des sessions. Puis nous avons établi les scores moyens et écarts-types pour chacune des habiletés cliniques.

L'objectif de cet article n'était pas de produire une analyse psychométrique de validité et de fidélité complète de nos ECOS mais plutôt de décrire notre expérience au sein d'un milieu naïf de ce genre d'examen, y compris pour le processus de notation et les possibilités de classement des étudiants. Comme nos ECOS sont des examens facultaires officiels, nous avons employé les analyses statistiques descriptives utilisées par la plateforme du Système inter-universitaire dématérialisé d'évaluation en santé (SIDES) pour analyser les résultats des étudiants. SIDES est la plateforme numérique déployée depuis janvier 2014 pour toutes les UFR Santé françaises, qui leur permet de réaliser et corriger tous leurs examens validant du deuxième cycle en complète autonomie [26].

INSTRUCTIONS POUR LE CANDIDAT**Contexte :** Garde aux urgences**Nom :** Frédéric D.**Âge :** 53 ans**Motif de consultation :** Douleur thoracique

Situation : Vous êtes nouvellement interne aux urgences du CHU de Rouen. L'IDE d'accueil vous sollicite à 9 h du matin alors pour interroger un patient de 55 ans qui se plaint d'une oppression thoracique. Il est particulièrement agité et angoissé. Ses constantes sont les suivantes : PA : 170/110 mmHg au bras droit, 165/105 mmHg au bras gauche, FC : 75/min, température 37,2 °C

Instructions : Vous avez 7 minutes pour :

1. Interroger le patient
2. Prescrire le bilan complémentaire immédiat
3. Interpréter l'examen complémentaire fournis

VOUS N'AVEZ AUCUN EXAMEN PHYSIQUE À PRATIQUER À CETTE STATION**ATTENTION :** Au signal sonore, passez à la station suivant**GRILLE D'OBSERVATION****Fait Non fait Points**

Sémiologie de la douleur	Durée de la douleur			5
	Intensité de la douleur avec évaluation de l'EVS et/ou EVA			5
	Recherche de la localisation (rétro sternale)			5
	Recherche des irradiations (bras et/ou mâchoire et/ou transfixiante)			5
	Facteur aggravant (effort et/ou inspiration profonde et/ou position allongée)			5
	Facteur améliorant (position penchée en avant)			5
Facteurs de risque	Cotation du tabagisme en paquet/année			5
	Hypertension artérielle			5
	Diabète			5
	ATCD familiaux coronarien et/ou vasculaire			5
	Dyslipidémie ou hypercholestérolémie			5
Bilan prescrit	ECG			10
	Créatinine ou fonction rénale			5
	Troponine			5
ECG	Fibrillation atriale			5
	Sus décalage dans le territoire inférieur			5
	Ne pas attendre les résultats biologiques pour traiter			5

Habiletés relationnelles**Habiletés organisationnelles**

Inadapté

Très adapté

0 1 2 4 5

0 1 2 4 5

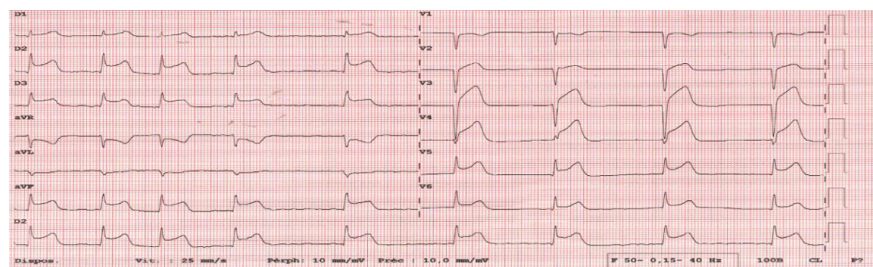
**Fig. 1.** Exemple d'instructions pour l'étudiant et de grille d'évaluation.

Tableau I. Scores globaux obtenus aux examens cliniques objectifs structurés (ECOS) 2008–2020 par promotion.

Année	Nombre d'étudiants	Nombre de stations	Moyenne (écart-type)	Médiane	Score maximum	Score minimum	Nombre (et %) d'étudiants ajournés
2008	145	8	12,74 (1,51)	12,61	17,70	5,74	11 (7 %)
2009	160	8	12,99 (1,53)	13,23	16,17	9,23	15 (9 %)
2010	164	8	13,78 (1,43)	13,96	16,45	8,83	12 (7 %)
2011	192	8	12,95 (1,59)	13,10	16,60	8,24	13 (6 %)
2012	196	6	13,52 (2,22)	14,03	18,53	5,13	18 (9 %)
2013	219	6	13,88 (1,49)	13,92	18,08	8,67	2 (1 %)
2014	204	6	11,90 (1,64)	12,03	15,62	6,80	21 (10 %)
2015	236	5	13,81 (1,66)	13,91	18,08	9,98	16 (8 %)
2016	213	7	13,26 (1,84)	13,49	17,36	6,47	20 (9 %)
2017	216	8	12,84 (1,37)	13,00	15,90	8,90	23 (10 %)
2018	213	8	12,73 (1,43)	12,80	16,31	8,01	11 (5 %)
2019	224	8	12,95 (1,52)	12,97	16,31	8,64	23 (10 %)
2020	218	10	13,31 (1,37)	13,37	17,27	9,94	36 (16 %)

Les performances des différentes promotions aux ECOS sont exprimées en scores globaux calculés sur l'ensemble des stations et sont exprimés en moyenne, écart-type, médiane, score maximum et minimum pour chaque cohorte.

Chaque station a le même poids et est notée sur 100 points. Les scores obtenus sont calculés ensuite sur 20. La quantité d'étudiants ajournés est exprimée en nombre et pourcentage de chaque cohorte.

Ainsi, la fidélité des scores a été appréciée par la mesure du coefficient Alpha de Cronbach. Celui-ci a été calculé pour le score total de chaque session en prenant comme items, respectivement, le nombre de stations et les habiletés cliniques évaluées dans les stations. Celles-ci comprenaient : l'interrogatoire, l'examen clinique, les investigations complémentaires, le diagnostic, la prise en charge thérapeutique, les habiletés relationnelles et relationnelles. Les habiletés techniques n'ont pas été prises en compte.

Pour évaluer les capacités de l'ECOS à classer les étudiants, nous avons étudié la distribution des scores obtenus au sein de chaque cohorte à l'aide des coefficients de Kurtosis et de Skewness, fournis par SIDES. Le coefficient de Kurtosis mesure le degré d'aplatissement de la distribution des scores. Pour une distribution des scores suivant la loi gaussienne normale, la valeur du coefficient doit être voisine de 3. Le coefficient de Skewness mesure le degré d'asymétrie de la distribution des scores. Pour une distribution des scores normale, sa valeur doit être voisine de 0. Lorsque le coefficient de Skewness est négatif, la courbe de distribution des scores permet d'identifier les étudiants les moins performants de la cohorte [27].

De plus, pour vérifier l'hypothèse que les ECOS et les ECNi mesurent deux éléments différents de la compétence clinique, nous avons comparé les performances des étudiants des cohortes 2018 et 2019 aux ECOS et aux ECNi à l'aide des coefficients de Pearson et de Spearman.

Résultats

Nombre d'étudiants

Treize ECOS ont été administrés consécutivement de 2008 à 2020 à des promotions de 200 étudiants en moyenne avec un minimum de 145 en 2008 et un maximum de 236 en

2015 (Tab. I). Au total, 2600 étudiants de DFASM3 ont passé un ECOS au sein de notre UFR. Pour pouvoir administrer l'ECOS à une promotion entière de façon standardisée, les rotations étaient effectuées sur quatre circuits administrés simultanément, répartis sur deux étages du même bâtiment, matin et après-midi sur la même journée. L'ECOS 2020 a été administré sur sept circuits simultanés répartis sur plusieurs bâtiments et sur deux demi-journées.

Nombre de stations par ECOS et durée des circuits

La durée des stations était de sept minutes chacune avec une minute entre chaque station pour effectuer la rotation. Le nombre moyen de stations par session était de 7,4, avec un minimum de cinq stations pour la session 2015 et un maximum de 10 lors de la session 2020 (Tab. I). À chaque circuit était incorporée une station repos. La session 2020 a été divisée en deux parties réalisées en deux jours différents pour permettre d'administrer le même circuit à tous les étudiants de façon simultanée. La première journée, le circuit comprenait les stations 1 à 5, la seconde journée, les stations 6 à 10. Le schéma des circuits et des stations composant l'ECOS 2020 est présenté sur la figure 2. Au total, 96 situations cliniques ont été rédigées pour l'ensemble des sessions, représentant 35 disciplines (Tab. II). La durée moyenne des ECOS était de 68 minutes, avec un minimum de 48 minutes pour le circuit 2015 et un maximum de 97 minutes pour la session 2020.

Moyennes des scores aux ECOS

Les scores obtenus par chaque promotion figurent sur le tableau I. Le score moyen de l'ensemble des promotions était de 13,12 sur 20. La moins bonne moyenne était obtenue en 2014 [11,90] et la meilleure en 2015 [13,81]. Le

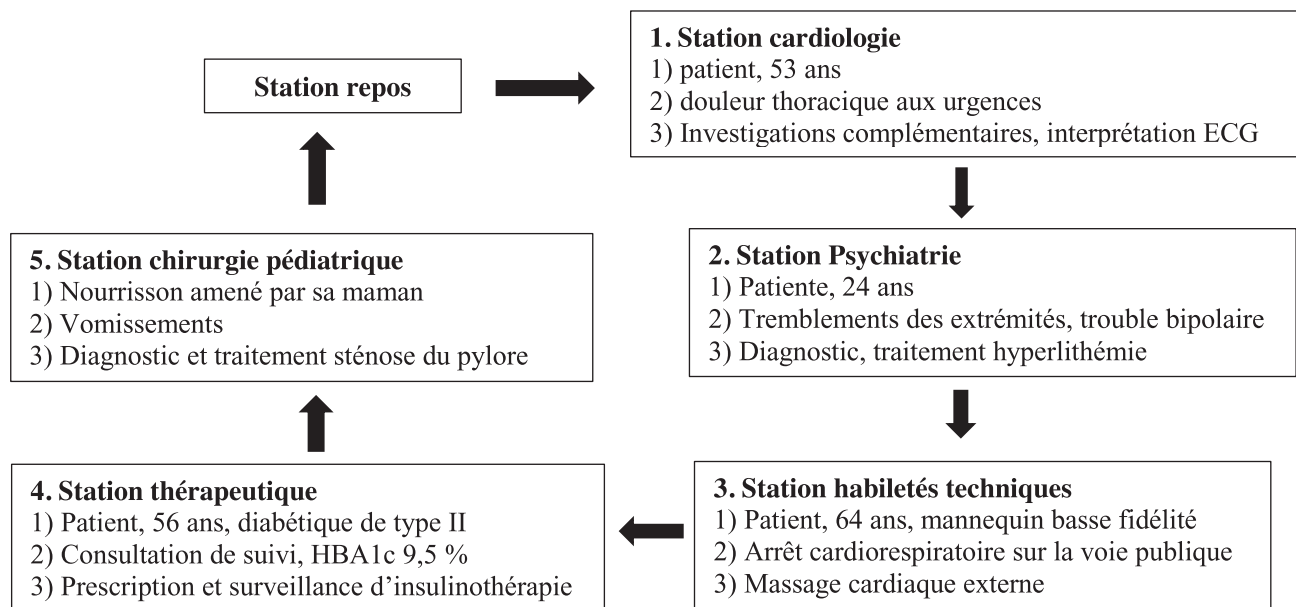
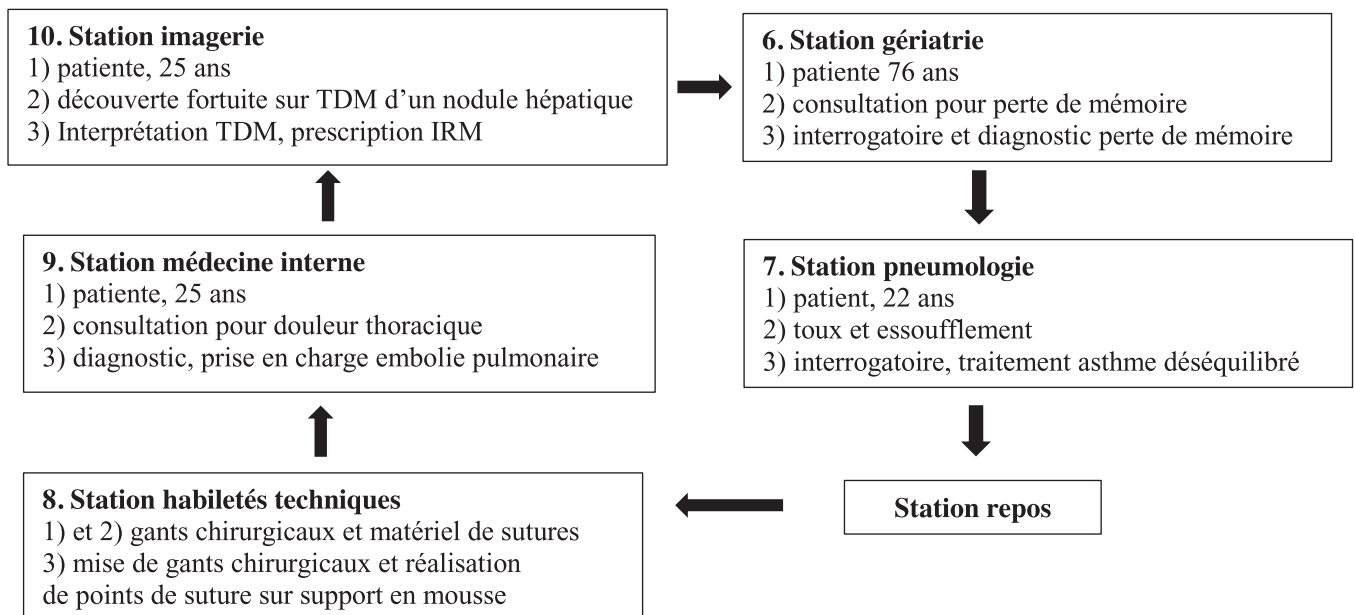
Circuit 1 : stations de 1 à 5Circuit 2 : stations de 6 à 10

Fig. 2. Schéma du circuit des examens cliniques objectifs structurés (ECOS) 2020.

ECOS de 10 stations administré sur 2 jours : stations 1 à 5 le premier, stations 6 à 10 le second.

Sept circuits en parallèle. Durée de chaque station : 7 min, durée entre chaque station : 1 min.

Pour chaque station sont indiqués : en gras la discipline ; 1) l'âge et le sexe du patient simulé ; 2) le contexte clinique de présentation ; 3) la ou les tâches à accomplir.

Tableau II. Répartition des disciplines et des stations au sein des sessions d'examens cliniques objectifs structurés (ECOS) 2008–2020.

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	TOTAL
Anesthésie – réanimation	x			x										2
Cardiologie	x				x								x	3
Chirurgie – sutures											x	x	x	3
Chirurgie digestive			x			x				x				3
Chirurgie maxillo-faciale							x							1
Chirurgie pédiatrique	x					x							x	3
Chirurgie plastique – main		x		x	x									3
Chirurgie vasculaire			x				x							2
Communication										x	x			2
Dermatologie		x										x		2
Diabétologie		x						x						2
Endocrinologie											x			1
Gériatrie													x	1
Gynéco – obstétrique		x		x					x			x		4
Hématologie		x				x			x					3
Hépatogastro		x						x						2
Imagerie				x						x			x	3
Injection IM											x			1
Maladies infectieuses		x			x				x			x		4
Médecine du travail		x			x				x					3
Médecine interne	x			x			x						x	4
Médecine légale			x					x			x			3
Néonatalogie				x										1
Néphrologie			x				x			x				3
Neurochirurgie			x			x				x				3
Neurologie					x						x			2
Nutrition				x								x		2
ORL			x				x			x				3
Orthopédie				x			x							2
Pédiatrie			x					x			x			3
Pneumologie	x					x			x				x	4
Psychiatrie			x				x			x			x	4
Rhumatologie	x					x			x			x		4
Thérapeutique	x									x	x	x	x	5
Urgences					x				x			x	x	4
Urologie	x													1
Nombre de stations	8	8	8	8	6	6	6	5	7	8	8	8	10	96

pourcentage d'étudiants ajournés était compris entre cinq et 10 % durant la période où le score de passage était fixé à 10 sur 20 pour l'ensemble des stations. Pour un score de passage fixé à 12 sur 20, le pourcentage d'étudiants ajournés était de 10 % en 2019 puis de 16 % en 2020, l'année où nous avons augmenté le nombre de stations à 10.

Pondération des habiletés cliniques

Les pondérations accordées à chaque habileté clinique évaluée au travers des stations au sein de chaque ECOS sont représentées sur le [tableau III](#). S'agissant d'évaluer les compétences cliniques d'étudiants de fin de second cycle,

la consigne donnée était d'attribuer le poids le plus important aux habiletés de prise en charge thérapeutique. Les habiletés relationnelles et les habiletés organisationnelles étaient dotées d'un poids de 10 % et 6 %, respectivement. À noter que pour les trois dernières sessions, nous avons développé des stations mesurant la réalisation de gestes techniques simples (sutures simples, pose de gants chirurgicaux, massage cardiaque externe), avec une pondération moyenne de 15,8 %.

À titre informatif, les scores moyens par habiletés cliniques obtenus pour l'ensemble des sessions figurent sur le [tableau IV](#).

Tableau III. Pondération des habiletés composantes de la compétence clinique au sein des examens cliniques objectifs structurés (ECOS).

Année	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	% moyen
Nombre d'étudiants	145	160	164	192	196	219	204	236	213	216	213	224	218	
Nombre de stations	8	8	8	8	6	6	6	5	7	8	8	8	10	
I	23,1 %	29,6 %	23,9 %	21 %	19 %	23,8 %	22,7 %	35 %	16,9 %	15,5 %	27,5 %	15,8 %	18,9 %	22,5 %
EC	12,3 %	5,6 %	19 %	16,9 %	14,5 %	25,3 %	15,8 %	16,6 %	5 %	6,4 %	2,5 %	6,8 %	3,5 %	11,5 %
IC	5 %	9,4 %	1,9 %	8,1 %	7,8 %	14,3 %	6,5 %	5 %	9,6 %	20 %	5,4 %	3,6 %	17,5 %	8,8 %
D	8,8 %	13,4 %	16,6 %	10,6 %	10,5 %	10,8 %	17,8 %	6 %	10,6 %	7,9 %	4,1 %	7,8 %	2,5 %	9,8 %
TT	29 %	18,9 %	20 %	27,5 %	31 %	6,7 %	23,3 %	20,4 %	40 %	27,5 %	28,3 %	38,8 %	31,6 %	26,4 %
HR	12,1 %	17,5 %	15,5 %	12,6 %	9 %	10,8 %	10,3 %	16 %	8,6 %	17,6 %	4,9 %	9,4 %	6,5 %	11,6 %
HO	9,8 %	5,6 %	3,1 %	3,3 %	8,2 %	8,2 %	3,5 %	1 %	9,4 %	5,1 %	4,9 %	5,5 %	7 %	5,7 %
HT	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	22,5 %	12,5 %	12,5 %	15,8 %

Éléments de la compétence clinique évalués dans les stations ECOS : I : interrogatoire ; EC : examen clinique ; IC : investigations complémentaires ; D : diagnostic ; TT : prise en charge thérapeutique ; HR : habiletés relationnelles ; HO : habiletés organisationnelles ; HT : habiletés techniques ; NE : non évaluables.

Les pondérations données aux différents éléments de la compétence clinique sont exprimées en pourcentage des points attribués pour chaque circuit. Chaque station a le même poids pondéral. Le score maximum par station est de 100, recalculé sur 20.

Tableau IV. Résultats globaux obtenus aux examens cliniques objectifs structurés (ECOS) par habiletés cliniques.

Année	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Nombre d'étudiants	145	160	164	192	196	219	204	236	213	216	213	224	218
Nombre de stations	8	8	8	8	6	6	6	5	7	8	8	8	10
I	12,4 (1,6)	11,7 (1,5)	13,8 (1,5)	12,7 (1,8)	13,3 (2,2)	12,8 (1,7)	13 (1,8)	13,5 (2)	12,2 (2,1)	12,9 (2,4)	12,3 (1,9)	12,3 (2,0)	13,4 (2,0)
EC	12,3 (2,4)	11,3 (4,8)	13,7 (2,0)	14,2 (2,3)	12,7 (3,2)	14,3 (1,9)	13,3 (2,2)	11,4 (3)	11,4 (3)	13,8 (2,6)	12,2 (3,8)	11,4 (2,6)	12,7 (4,1)
IC	11,3 (2,4)	14,3 (2,9)	15 (5,8)	11,9 (3,3)	14,2 (3,3)	15,4 (2,1)	9,0 (4,4)	10,7 (6)	12,4 (3,8)	13,2 (2,2)	12 (4,1)	12 (4,9)	14 (2,1)
D	14,9 (3,6)	15,7 (2,0)	14,3 (2,3)	16,5 (2,3)	12,8 (3,3)	13,5 (1,2)	9,7 (3,2)	17,5 (1,9)	15,9 (2,7)	14,2 (3,4)	17,7 (3,7)	14,8 (2,8)	6,1 (4,2)
TT	12,4 (2,1)	12,6 (2,5)	12,6 (3,1)	11,2 (2,5)	13,9 (3,1)	12,9 (4,2)	11 (2,5)	14,4 (2,8)	13,4 (2,1)	10,2 (2,2)	11,8 (2,3)	12,1 (2,0)	11,5 (1,8)
HR	14,5 (2,5)	13,8 (2,5)	15,6 (2,5)	13,8 (2,3)	15,8 (3,7)	15 (2,8)	14,7 (2,9)	15,9 (2,7)	14,5 (2,7)	15,2 (3,0)	14 (1,9)	13,5 (3,3)	15,2 (2,5)
HO	12,6 (2,6)	11,6 (3)	10,3 (6,4)	12,8 (2,7)	12,7 (4,6)	13,4 (3,0)	13,2 (4,5)	12,6 (6,5)	13,7 (2,7)	14,1 (1,8)	14,2 (2,0)	12,7 (2,3)	14 (2,4)
HT	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	13,2 (2,1)	16,6 (3,1)	17,1 (1,9)

Éléments de la compétence clinique évalués dans les stations ECOS : I : interrogatoire ; EC : examen clinique ; IC : investigations complémentaires ; D : diagnostic ; TT : prise en charge thérapeutique ; HR : habiletés relationnelles ; HO : habiletés organisationnelles ; HT : habiletés techniques ; NE : non évaluables.

Les scores obtenus aux différents éléments de la compétence clinique sont exprimés en score moyen sur 20 et écart-type (entre parenthèse).

Tableau V. Évolution du coefficient de fidélité Alpha de Cronbach (par stations et par habiletés cliniques).

Années	Nombre d'étudiants	Nombre de stations	Coefficient Alpha par stations	Coefficient Alpha par habiletés cliniques
2008	145	8	0,61 [IC5% : 0,50–0,67]	0,78 [IC5% : 0,71–0,80]
2009	160	8	0,66 [IC5% : 0,57–0,71]	0,76 [IC5% : 0,69–0,80]
2010	164	8	0,61 [IC5% : 0,51–0,67]	0,68 [IC5% : 0,60–0,74]
2011	192	8	0,61 [IC5% : 0,53–0,67]	0,78 [IC5% : 0,73–0,81]
2012	196	6	0,69 [IC5% : 0,62–0,74]	0,81 [IC5% : 0,76–0,84]
2013	219	6	0,57 [IC5% : 0,47–0,63]	0,74 [IC5% : 0,69–0,78]
2014	204	6	0,56 [IC5% : 0,45–0,62]	0,70 [IC5% : 0,63–0,71]
2015	236	5	0,52 [IC5% : 0,41–0,58]	0,68 [IC5% : 0,61–0,72]
2016	213	7	0,73 [IC5% : 0,67–0,77]	0,82 [IC5% : 0,78–0,85]
2017	216	8	0,54 [IC5% : 0,44–0,61]	0,65 [IC5% : 0,57–0,70]
2018	213	8	0,51 [IC5% : 0,67–0,77]	0,72 [IC5% : 0,66–0,75]
2019	224	8	0,61 [IC5% : 0,41–0,58]	0,72 [IC5% : 0,66–0,76]
2020	218	10	0,61 [IC5% : 0,52–0,66]	0,75 [IC5% : 0,73–0,81]

Le coefficient Alpha de Cronbach mesure l'homogénéité d'un test (fidélité) en établissant la corrélation entre les scores de chaque question avec le score global du test. Pour des examens à forts enjeux, une valeur de 0,7 ou plus est considérée comme acceptable. Il est exprimé avec écart-type et intervalle de confiance.

Coefficient de fidélité

Les coefficients de fidélité, présentés dans le [tableau IV](#), variaient de 0,52 [IC5% : 0,41–0,58] à 0,73 [IC5% : 0,67–0,77] en prenant comme unité de mesure les scores obtenus aux stations. Ils variaient de 0,65 [IC5% : 0,57–0,70] à 0,82 [IC5% : 0,78–0,85] en prenant comme unité de mesure les scores obtenus aux habiletés cliniques. Il est à noter que les coefficients les plus bas ont été obtenus durant les années 2014 et 2015, années où le nombre d'étudiants était très élevé, avec un circuit ECOS ne comportant respectivement que six et cinq stations

Distributions des scores

Le coefficient de Kurtosis variait de $2,36 \pm 0,5$ à $5,56 \pm 1,69$ avec une valeur moyenne pour toutes les sessions de 3,27 ([Tab. VI](#)), suggérant que les ECOS permettaient de classer les étudiants en fonction de leurs performances suivant des courbes gaussiennes normales. Le coefficient de Skewness variait de $-0,10 \pm 0,11$ à $-0,96 \pm 0,22$ avec une valeur moyenne de $-0,45$. La valeur constamment négative à toutes les sessions (extrêmes : $-0,10$ – $-0,96$), confirme que les scores obtenus aux ECOS permettaient d'identifier les étudiants les moins performants. À titre d'exemple, la courbe de distribution des scores de l'ECOS 2020 est représentée sur la [figure 3](#). L'évolution des coefficients de distribution des scores entre 2008 et 2020 est illustrée sur la [figure 4](#). Les distributions des scores des trois dernières sessions semblent se rapprocher des valeurs normales.

Corrélation des scores et des classements des étudiants aux ECOS et aux ECNi

La comparaison des scores et des classements aux ECOS et aux ECNi de la promotion 2018 n'a pas mis en évidence de corrélation significative, avec un coefficient de

Tableau VI. Évolution des valeurs des coefficients de distribution des scores aux examens cliniques objectifs structurés (ECOS).

Années	Coefficient de Kurtosis (écart-type)	Coefficient de Skewness (écart-type)
2008	5,56 (1,69)	−0,40 (0,57)
2009	2,66 (0,23)	−0,39 (0,13)
2010	3,70 (0,49)	−0,74 (0,18)
2011	2,66 (0,26)	−0,37 (0,14)
2012	4,61 (0,66)	−0,96 (0,22)
2013	3,02 (0,41)	−0,32 (0,18)
2014	3,06 (0,29)	−0,31 (0,15)
2015	2,36 (0,15)	−0,10 (0,11)
2016	3,36 (0,51)	−0,59 (0,18)
2017	2,63 (0,22)	−0,36 (0,12)
2018	3,14 (0,33)	−0,27 (0,18)
2019	2,93 (0,23)	−0,21 (0,13)
2020	2,86 (0,23)	−0,13 (0,14)
Moyenne	3,27	−0,45

Le coefficient de Kurtosis mesure le degré d'aplatissement de la distribution des scores. Pour une distribution des scores suivant la loi gaussienne normale, la valeur du coefficient doit être voisine de 3.

Le coefficient de Skewness mesure le degré d'asymétrie de la distribution des scores. Pour une distribution normale, le coefficient d'aplatissement a pour valeur 0. Lorsque ce coefficient est négatif, la courbe de distribution des scores permet d'identifier les étudiants les moins performants de la cohorte.

Spearman entre les scores ou entre les classements, identique, à 0,28, IC5% = [0,16–0,40] et un coefficient de Pearson à 0,29, IC5% = [0,16–0,40], $p < 0,0004$ pour les scores et à 0,27 IC5% = [0,15–0,39], $p < 0,0004$ pour les classements.

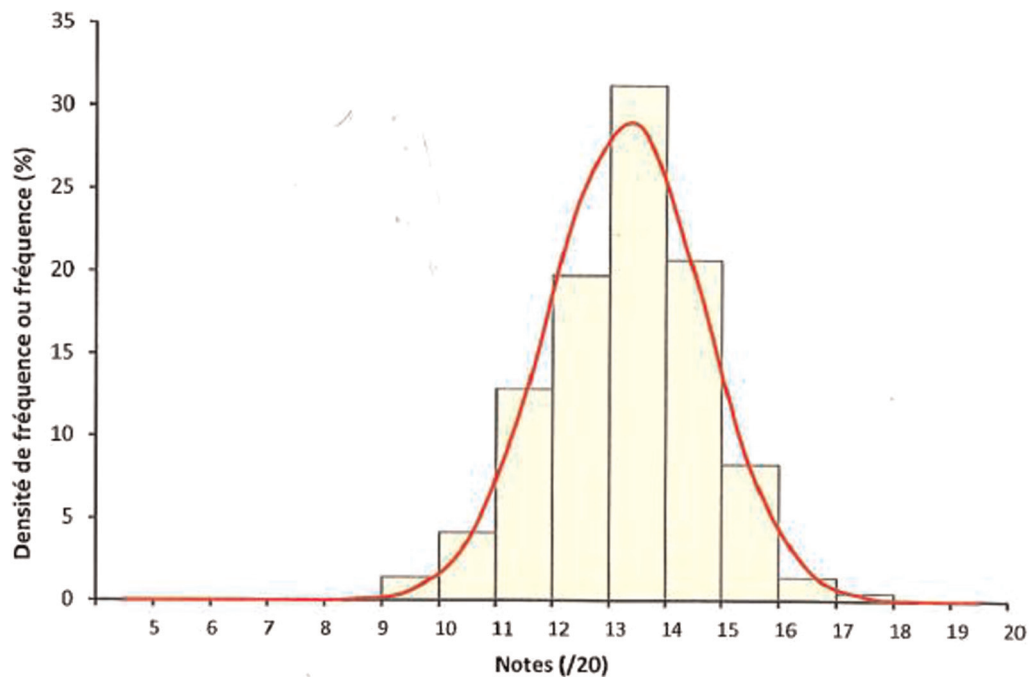


Fig. 3. Courbe de distribution des scores du circuit des examens cliniques objectifs structurés (ECOS) 2020.

Le circuit ECOS 2020 comporte 10 stations, administrées à 218 étudiants.

Abscisse : score global obtenu à l'ECOS sur 20.

Ordonnées : % d'étudiants de la cohorte.

Coefficient de Kurtosis 2020 (écart type) : $2,86 \pm 0,23$.

Coefficient de Skewness 2020 (écart-type) : $-0,13 \pm 0,14$.

Le coefficient de Kurtosis mesure le degré d'aplatissement de la distribution des scores. Pour une distribution des scores suivant la loi gaussienne normale, la valeur du coefficient doit être voisine de 3.

Le coefficient de Skewness mesure le degré d'asymétrie de la distribution des scores. Pour une distribution normale, le coefficient d'aplatissement a pour valeur 0. Lorsque ce coefficient est négatif, la courbe de distribution des scores permet d'identifier les étudiants les moins performants de la cohorte.

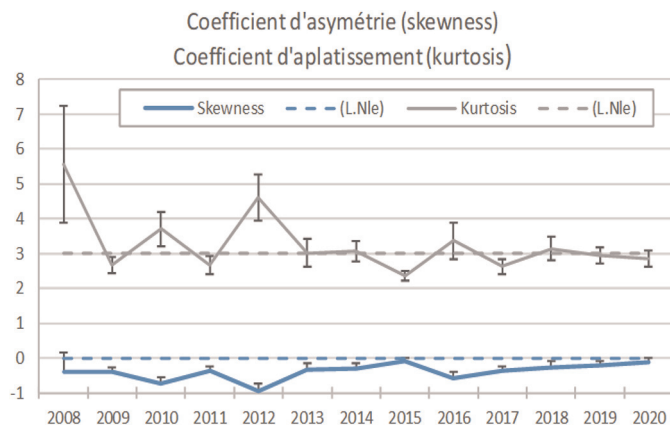


Fig. 4. Évolution des coefficients de distribution des scores.

Le coefficient de Kurtosis mesure le degré d'aplatissement de la distribution des scores. L.Nle : Pour une distribution des scores suivant la loi gaussienne normale, la valeur du coefficient doit être voisine de 3.

Le coefficient de Skewness mesure le degré d'asymétrie de la distribution des scores. L.Nle : Pour une distribution normale, le coefficient d'aplatissement a pour valeur 0. Lorsque ce coefficient est négatif, la courbe de distribution des scores permet d'identifier les étudiants les moins performants de la cohorte.

La comparaison des scores et des classements aux ECOS et aux ECNi de la promotion 2019 n'a pas mis non plus en évidence de corrélation avec un coefficient de Spearman entre les scores ou entre les classements, identique, à 0,24, $IC5\% = [0,11-0,35]$ et un coefficient de Pearson à 0,25 $IC5\% = [0,12-0,36]$ $p = 0,0001$ pour les scores et à 0,22, $IC5\% = [0,10-0,34]$, $p = 0,0007$ pour les classements (Fig. 5).

Discussion

Notre travail correspond à la plus importante expérience jamais rapportée en France concernant l'ECOS en tant qu'examen facultaire validant. Le premier enseignement à tirer de notre expérience concerne sa faisabilité en tant qu'examen officiel et sa pérennisation. Nous avons pu développer un processus pérenne concernant tous les aspects organisationnels de l'épreuve. Nos expérimentations antérieures [22–24] et un fort soutien institutionnel ont été les facteurs déterminants du succès de l'implantation de l'ECOS dans notre UFR. La mise en place d'une équipe spécifique, composée d'un petit groupe d'enseignants formés

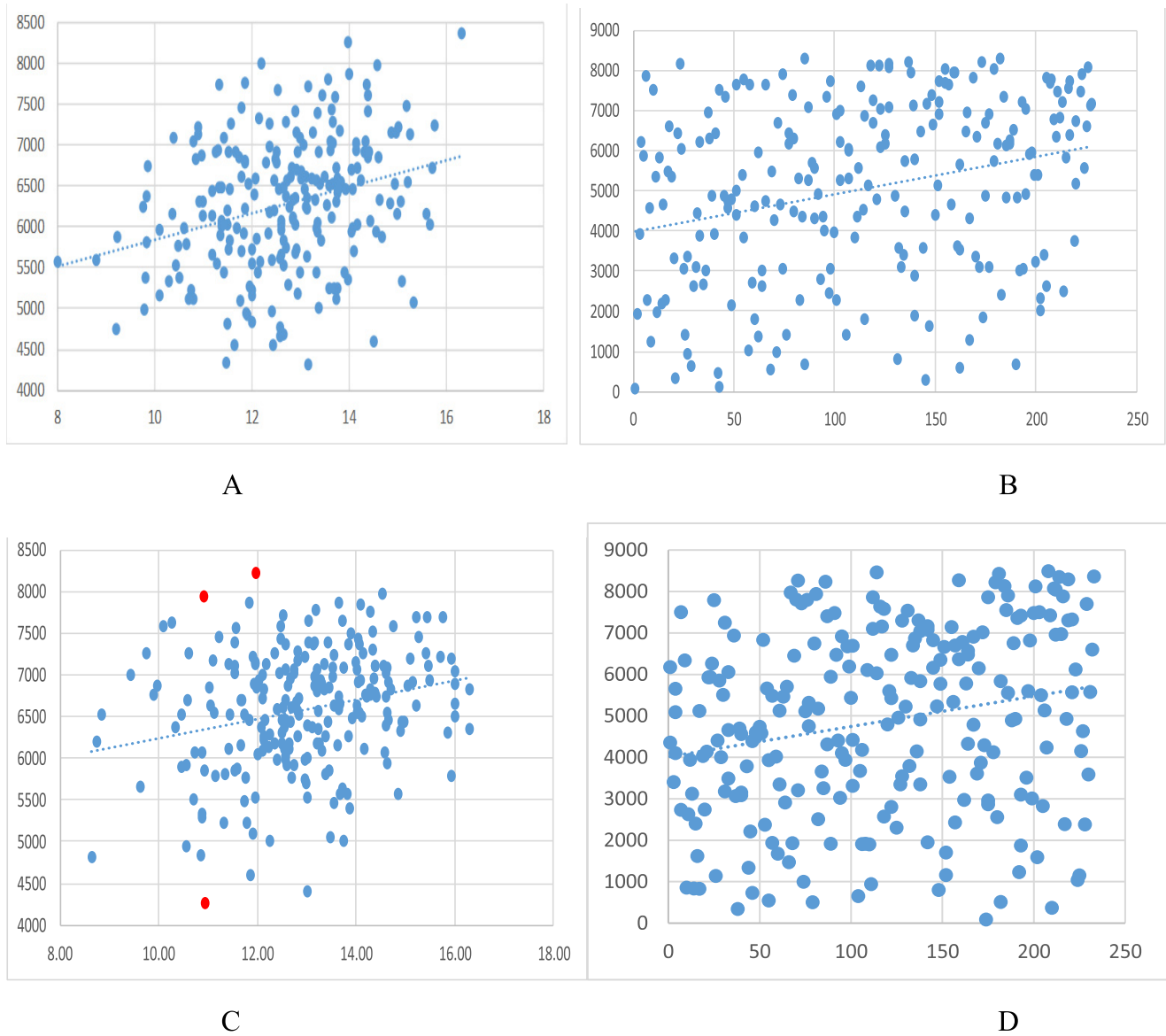


Fig. 5. Corrélations entre les performances des étudiants des promotions 2018 et 2019 aux examens cliniques objectifs structurés (ECOS) et aux épreuves classantes nationales (scores et classements). A : Corrélation scores aux ECOS / scores aux épreuves nationales classantes (ECN) de la promotion 2018. r (Pearson) = 0,29, $IC5\% = [0,16-0,40]$, $p < 10^{-4}$. B : Corrélation classement aux ECOS / classements aux ECN de la promotion 2018. r (Pearson) = 0,27 $IC5\% = [0,15-0,39]$, $p < 10^{-4}$. Les coefficients de corrélation de Spearman entre les scores ou entre les classements aux ECOS et aux ECN sont identiques : r (Spearman) = 0,28, $IC5\% = [0,16-0,40]$, $p < 10^{-4}$. C : Corrélation scores aux ECOS / scores aux ECN de la promotion 2019. r (Pearson) = 0,25, $IC5\% = [0,12-0,36]$, $p = 0,0001$. D : Corrélation classement aux ECOS / classements aux ECN de la promotion 2019. r (Pearson) = 0,22, $IC5\% = [0,10-0,34]$, $p = 0,0007$. Les coefficients de corrélation de Spearman entre les scores ou les classements aux ECOS et aux ECN sont identiques : r (Spearman) = 0,24, $IC5\% = [0,11-0,35]$, $p = 0,0003$.

à l'ECOS et mandatée pour organiser, assurer le suivi et l'administration de l'ECOS, est apparue un élément de terrain essentiel.

La principale limite de notre étude est d'ordre psychométrique. Notre but principal était de proposer un partage d'expérience le plus concret possible vis-à-vis des institutions francophones et françaises qui découvrent ce qu'est un ECOS. Nous avons ainsi volontairement traité les

données chiffrées obtenues à nos différents ECOS de la même manière que sont traitées les données obtenues pour la majorité des examens facultaires en France, en utilisant les outils statistiques purement descriptifs fournis par la plateforme SIDES, plateforme numérique utilisée par toutes les facultés de médecine françaises pour l'administration de l'ensemble de leurs examens. Ainsi les analyses produites dans cet article ne correspondent pas aux

standards internationaux en vigueur concernant les analyses psychométriques des processus d'évaluation de la compétence clinique [11,28,29]. Elles doivent être considérées comme une simple illustration de notre témoignage.

Certains aspects de nos résultats sont cependant intéressants à discuter, car ils reflètent les difficultés que nous avons rencontrées lors du processus d'implantation. Un aspect particulier concerne les modalités de développement des grilles d'évaluation. Nous avons construit des grilles à type de liste de vérification binaire plutôt que des grilles d'évaluation globale qualitative. La transformation des performances observées en scores nous a parue plus simple et plus objective avec les listes de vérification binaire. Leur appropriation par les enseignants d'une collectivité peu entraînée à développer ce type d'évaluation nous a semblée plus aisée. De plus, ces grilles ont l'avantage de permettre de recruter des observateurs non experts de la discipline concernée dans les stations, élément majeur pour la faisabilité des ECOS dans notre milieu. Le recours à des grilles d'évaluation globale pour coter les performances observées est plus difficile et si ces grilles semblent plus adaptées pour apprécier la prise en charge globale du patient, elles s'avèrent surtout utiles pour mesurer des tâches multidimensionnelles complexes. Elles nécessitent des observateurs experts des habiletés cliniques évaluées [13,30]. Elles présentent cependant une meilleure fidélité inter-juges que les listes de vérifications binaires [30]. La vérification de la fidélité inter-juges de nos grilles n'a pas été possible dans la configuration actuelle de nos ECOS, car les observateurs d'une même station varient d'un circuit à l'autre mais les étudiants aussi. C'est certainement une des limites de notre dispositif. Une autre limite bien documentée des listes de vérification binaire est qu'elles ne mesurent que les habiletés observables et procédurales, et favoriseraient une collecte exhaustive et non orientée des données [11,13,15,30]. C'est pour cela que nous préconisons d'apporter le plus grand soin dans la standardisation de leur rédaction, en incluant un nombre limité d'items par grille, focalisés sur les points clés de la prise en charge du patient, avec des instructions claires pour les observateurs, d'autant que la durée de nos stations est courte (sept minutes). Le respect de ces règles de rédaction est indispensable pour que ces grilles puissent mesurer réellement ce qu'elles doivent mesurer [20]. De plus, les listes binaires ne sont pas adaptées pour l'évaluation des habiletés organisationnelles et relationnelles qui est subjective, nécessitant une échelle qualitative en cinq points, que nous avons systématiquement intégrée pour ces deux items.

Le fait que les mêmes circuits soient administrés matin et après-midi a pu avoir un impact sur la reproductibilité des scores en favorisant la communication entre les étudiants ayant passé les ECOS le matin et ceux le passant l'après-midi. Nous n'avons pas été en mesure de vérifier l'existence d'une différence significative entre les scores obtenus le matin avec ceux obtenus l'après-midi. Les données publiées sur l'effet de la

communication entre les étudiants passant le même ECOS dans des circuits différents n'ont pas démontré un impact significatif sur les performances [31]. Néanmoins, pour diminuer l'impact éventuel d'une communication entre étudiants, nous avons apporté des mesures correctrices telles que l'introduction de stations sans patients simulés avec réalisation d'habiletés techniques de base et l'administration de l'ECOS 2020 en deux journées afin que tous les étudiants passent l'ECOS de façon simultanée. Une autre possibilité serait d'intégrer au sein des circuits ECOS qui sont administrés de façon non simultanée, des stations de contenus différents mais traitant des mêmes thèmes et avec des objectifs d'évaluation identiques. Compte-tenu de son impact probable sur la reproductibilité des scores et la cohérence interne de l'examen, ce dispositif devra être préalablement testé à large échelle car à ce jour, il n'est pas documenté dans la littérature.

Un autre élément de discussion concerne le nombre de stations. Ce point illustre le dilemme entre le concept de spécificité de contenu (la performance à une station n'est pas prédictive de la performance à une autre station de contenu différent et impose un nombre significatif de cas) [10,11,15,19] et les limites liées à la faisabilité de l'examen en termes de logistique. Un nombre d'au moins 20 stations par circuit était le minimum recommandé dans la littérature internationale pour atteindre un coefficient de fidélité acceptable [16,28,29]. Un certain nombre de données actuelles font état de circuit ECOS avec moins de stations, qui conservent cependant des scores stables et des coefficients de fidélité corrects. Ainsi, Heal *et al.* rapportent que l'ensemble des facultés australiennes qui ont introduit les ECOS dans leurs cursus depuis plus de 20 ans développent actuellement des ECOS de huit à 16 stations, de cinq à 10 minutes chacune, avec des données psychométriques probantes [8]. Ces données suggèrent qu'un ECOS comportant au minimum huit à 10 stations permet de produire un examen valide et des scores stables [8,17,18], sous réserve que l'ECOS soit intégré à un processus d'évaluation multidimensionnel utilisant de façon complémentaire d'autres formats évaluatifs [11,12], ce qui sera le cas lors de l'examen national de fin de second cycle [3,25]. Hormis les années où nous avons dû faire face à une augmentation importante du nombre d'étudiants liée à l'évolution du *numerus clausus*, nous avons toujours pu proposer huit stations par ECOS et actuellement 10 stations pour la dernière session. De plus, le format d'ECOS national retenu par la réforme en cours sera un circuit de 10 stations, administré simultanément sur le territoire national [25]. Cet ECOS devrait être fractionné en deux sessions de cinq stations chacune, administrées sur deux jours différents, soit le même modèle que celui que nous avons expérimenté dans notre UFR Santé.

La décision d'établir le score de passage à 10 sur 20 a été un choix arbitraire. Compte-tenu que les étudiants qui passaient l'ECOS avaient validé préalablement l'ensemble des autres unités d'enseignements, que l'objectif pédagogique de l'ECOS était plus d'identifier les étudiants en

difficulté dans leurs habiletés cliniques que de décider le passage dans l'année suivante, cette approche a été maintenue. L'augmentation récente du score de passage à 12 sur 20 nous a semblé justifiée eu égard à l'investissement logistique et à l'augmentation du nombre de stations par ECOS. Il faut reconnaître que si la problématique du score de passage représente une des limites docimologiques de notre travail, elle n'est cependant pas spécifique à l'ECOS. Elle rejoint la problématique des dispositifs généralement mis en place dans notre milieu pour mesurer l'atteinte des objectifs pédagogiques par les examens facultaires.

Par ailleurs, un certain nombre d'arguments plaident pour la pertinence du contenu pédagogique de nos ECOS. La planification est un des éléments clés [12,15,17-20]. Nous nous sommes ainsi assuré que pour chaque session, l'échantillonnage des cas cliniques choisis soit bien représentatif du programme de second cycle. De même, nous avons pris le plus grand soin à ce que le mode de présentation (urgence, consultation), les caractéristiques du patients (âge, sexe) soient équitablement répartis au sein des stations, que les principales habiletés composant la compétence clinique soient harmonieusement réparties au sein des stations, que les tâches demandées soient explicites et ne mesurent pas plus d'une à trois habiletés cliniques par station. Au sein des habiletés cliniques, il nous a semblé important d'accorder un plus grand poids aux habiletés de prise en charge thérapeutique qu'aux capacités d'interrogatoire et d'examen physique pour éviter d'accorder trop d'importance à la sémiologie plutôt qu'à la prise en charge du patient (Tab. III et IV).

Nous avons pu ainsi bâtir une banque de cas ECOS couvrant l'ensemble des disciplines et des points clés du programme, possiblement adaptables et réutilisables (Tab. II). Le niveau de fidélité des scores obtenus au sein des différents circuits et aux différentes habiletés est relativement peu élevé (Tab. V). Ceci est le reflet d'un certain nombre de biais qu'il est important de bien identifier : a) le nombre relativement restreint de stations dans nos premières sessions d'ECOS ; b) une hétérogénéité dans les objectifs d'évaluation décrits dans les stations, que nous nous sommes efforcé de corriger progressivement au fil des ans (*cf.* Tab. III) ; c) la difficulté d'évaluer de façon objective et standardisée les habiletés relationnelles et organisationnelles ; d) un manque de standardisation de la rédaction des grilles ; e) un manque de standardisation de la formation des patients simulés et surtout des observateurs. Ce dernier point peut représenter un challenge dans nos structures hospitalo-universitaires et est un point d'amélioration certain de notre dispositif. Néanmoins, compte tenu de notre manque d'expérience en l'absence de directive nationale, et comparativement aux données de la littérature établissant qu'une valeur de coefficient Alpha de 0,7 à 0,8 est adéquate pour un examen sanctionnant à fort enjeux, nos résultats méritent considération par rapport aux données publiées concernant la cohérence interne des ECOS [12,14-16,28,29].

Par ailleurs, nous n'avons pas pu identifier des arguments en faveur des qualités de discrimination des ECOS mais la distribution des scores obtenus par les étudiants montre que nos ECOS peuvent répartir les étudiants selon une courbe gaussienne normale et donc établir un classement. La tendance à la normalisation des coefficients de distribution des scores des trois dernières sessions suggère d'ailleurs que notre dispositif d'établissement des scores s'est amélioré progressivement, avec l'augmentation du nombre de stations, l'amélioration de la formation des patients simulés et des observateurs (Fig. 4 et Tab. VI). La valeur le plus souvent négative du coefficient d'asymétrie de Skewness suggère que les étudiants les moins performants dans les habiletés cliniques sont ceux qui obtiennent les moins bons scores aux ECOS. L'identification des étudiants en difficulté est un objectif majeur pour un examen facultaire et ces résultats nous confortent sur la pertinence des ECOS. Par ailleurs, il est documenté que les étudiants peu performants aux ECOS à un niveau pré-gradué seront plus souvent en difficulté lors de leur pratique clinique au cours de leur internat [32]. Ces données devraient inciter à proposer un suivi pédagogique personnalisé aux étudiants les moins performants aux ECOS.

Nous n'avons pas trouvé de corrélation entre les scores et classements aux ECNi et les performances aux ECOS des étudiants des promotions 2018 et 2019 (Fig. 5). Cela confirme que la compétence clinique est un construit multidimensionnel et dépasse largement les connaissances théoriques, qui sont, à ce jour, les seuls éléments de la compétence clinique mesurés par les ECNi. Ce constat ponctuel est un argument supplémentaire pour une nécessaire diversification docimologique telle qu'elle est souhaitée par la réforme du deuxième cycle [1,3,25].

Conclusion

Au total, dans notre expérience, l'ECOS s'est avéré utilisable à large échelle en tant qu'examen facultaire validant. Il permet de mesurer des éléments essentiels de la compétence clinique, son impact éducatif potentiel est majeur. Il ne faut cependant pas perdre de vue que l'ECOS ne propose qu'une observation de la performance en contexte simulé [17,18]. Des composants majeurs de la compétence clinique sont susceptibles de « passer sous les radars » de l'ECOS, tels que les mécanismes cognitifs mis en jeu par les étudiants dans l'élaboration de leurs stratégies de raisonnement clinique, dans leurs interactions avec les patients, leurs enseignants, leurs pairs, les autres professionnels de santé. Finalement, la principale limite de l'ECOS est de proposer une approche morcelée de la compétence clinique et le format de l'examen ne permet pas d'appréhender les capacités de prise en charge globale d'un patient. Comme tous les outils d'évaluation, il a ses forces et ses faiblesses et il faut être conscient qu'un processus d'évaluation intégrant le concept d'approche par compétences impose l'emploi de multiples formats d'évaluation de façon complémentaire, tel que la réforme

du second cycle semble l'envisager [1,3]. Il a néanmoins l'immense mérite d'évaluer les habiletés cliniques de façon homogène et standardisée. Son intégration officielle au sein du cursus des études médicales en France devrait entraîner un profond changement dans les méthodes d'apprentissage des habiletés cliniques. Un enseignement contextualisé avec des sessions d'entraînement avec patients simulés serait justifié, d'autant qu'il est documenté que ce type d'entraînement a un impact positif sur les performances aux ECOS et est actuellement réclamé, à juste titre, par les étudiants [33,34]. Les modifications du cursus des études médicales prévues dans la réforme vont d'ailleurs intégrer ce point avec la mise en place d'ECOS d'entraînement annuels dès le début du second cycle [25]. À ce titre, l'entraînement aux habiletés cliniques procédurales et aux interactions avec des patients simulés devrait inciter à optimiser le rapprochement entre UFR et centres de simulation. Le coût concernant le recrutement et l'entraînement des patients simulés standardisés doit s'intégrer dans la réflexion actuellement en cours sur le financement de la simulation en santé pour la formation initiale.

Enfin, pour asseoir l'implantation des ECOS à grande échelle en France, au vu de notre expérience monocentrique, l'effort collectif devra porter sur tous les éléments qui concourent à la standardisation et à l'objectivité des ECOS, et déterminer une stratégie consensuelle concernant la rédaction des synopsis des stations et des grilles d'observation, ainsi que la formation des observateurs. De plus, il va être nécessaire d'approfondir les analyses statistiques des données issues des ECOS au niveau national afin de les étayer rapidement par des données probantes plus scientifiquement fondées.

Contributions

Louis Sibert a participé à la conception du protocole de recherche, au recueil de données, à l'interprétation des résultats, à l'écriture du manuscrit, ainsi qu'à sa correction et relecture. Pascale Schneider a participé à la conception du protocole de recherche, au recueil de données, à l'interprétation des résultats, à la correction et à la relecture du manuscrit. Agnès Liard-Zmuda a participé à la conception du protocole de recherche, au recueil de données, à l'interprétation des résultats, à la correction et à la relecture du manuscrit. Jean-François Menard a participé à la conception du protocole de recherche, au recueil de données et à l'analyse statistique. Isabelle Auquit-Auckbur a participé au recueil des données, à la correction et à la relecture du manuscrit. Benoît Veber a participé à la conception du protocole de recherche, au recueil de données, à l'interprétation des résultats, à la correction et à la relecture du manuscrit.

Liens d'intérêt

Aucun auteur ne déclare de conflit d'intérêts en lien avec le contenu de cet article.

Approbation éthique

S'agissant d'une analyse rétrospective de données anonymes et déjà diffusées, ce travail n'a pas été soumis à un comité de protection des personnes mais il a été approuvé par le comité pédagogique de l'Unité de formation et recherche Santé de l'Université de Rouen.

Remerciements

Nous remercions Madame Nikki Sabourin pour la lecture et la correction du résumé en anglais.

Références

1. Loi n°2019-774 du 24 juillet 2019 relative à l'organisation et à la transformation du système de santé. Journal Officiel n°0172 du 26 juillet 2019. 2019 [On-line]. Disponible sur <https://www.legifrance.gouv.fr/eli/loi/2019/7/24/2019-774/jo/texte>.
2. Nguyen DQ, Blais JG. 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;8:232-5.
3. Arrêté du 2 septembre 2020 portant modification de diverses dispositions relatives au régime des études en vue du premier et du deuxième cycle des études médicales et à l'organisation des épreuves classantes nationales. Journal Officiel n°0221 du 10 septembre 2020. 2020 [On-line]. Disponible sur <https://www.legifrance.gouv.fr/eli/jo/2021/11/7/0260>.
4. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *BMJ* 1975;1:447-51.
5. Patrício MF, Julião M, Fareleira F, Carneiro AV. Is the OSCE a feasible tool to assess competencies in undergraduate medical education? *Med Teach* 2013;35:503-14.
6. Cömert M, Zill JM, Christalle E, Dirmaier J, Härter M, Scholl I. Assessing communication skills of medical students in Objective Structured Clinical Examinations (OSCE). A systematic review of rating scales. *PLoS One* 2016;11:e01527.
7. Müller S, Koch I, Settmacher U, Dahmen U. How the introduction of OSCEs has affected the time students spend studying: Results of a nationwide study. *BMC Med Educ* 2019;19:146.
8. Heal C, D'Souza K, Banks J, Malau-Aduli BS, Turner R, Smith J *et al.* A snapshot of current Objective Structured Clinical Examination (OSCE) practice at Australian medical schools. *Med Teach* 2019;41:441-7.
9. Kakadia R, Chen E, Ohyama H. Implementing an online OSCE during the COVID-19 pandemic. *J Dent Educ* 2020;1-3.
10. Newble DI. Eight years'experience with a structured clinical examination. *Med Educ* 1988;22:200-4.

11. Newble DI, Swanson DB. Psychometric characteristics of the objective structured clinical examination. *Med Educ* 1988;22:325-34.
12. Grand'Maison P, Brailovsky CA, Lescop J. Content validity of the Quebec licensing examination (OSCE). *Can Fam Physician* 1996;42:254-9.
13. Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med* 1998;73:993-7.
14. Brailovsky CA, Grand'Maison P. Using evidence to improve evaluation: A comprehensive psychometric assessment of a SP-based OSCE licensing examination. *Adv Health Sci Educ Theory Pract* 2000;5:207-19.
15. Newble D. Techniques for measuring clinical competence: Objective structured clinical examinations. *Med Educ* 2004;38:199-203.
16. Brailovsky CA, Grand'Maison P, Lescop J. A large-scale multicenter Objective Structured Clinical Examination for licensure. *Acad Med* 1992;67:S37-S39.
17. Khan KZ, Ramachandran S, Gaunt K, Pushkar P. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part I: An historical and theoretical perspective. *Med Teach* 2013;35:e1437-46.
18. Khan KZ, Gaunt K, Ramachandran S, Pushkar P. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part II: Organisation & administration. *Med Teach* 2013;35:e1447-63.
19. Swanson DB, van der Vleuten CP. Assessment of clinical skills with standardized patients: State of the art revisited. *Teach Learn Med* 2013;25 Suppl 1:S17-25.
20. Daniels VJ, Pugh D. Twelve tips for developing an OSCE that measures what you want? *Med Teach* 2018;40:1208-13.
21. Arrêté du 8 avril 2013 relatif au régime des études en vue du premier et du deuxième cycle des études médicales, Journal Officiel du 23 avril 2013, Bulletin officiel n°20 du Ministère de L'enseignement Supérieur, de la Recherche et de l'Innovation du 16 mai 2013. 2013 [On-line]. Disponible sur : https://www.enseignementsup-recherche.gouv.fr/pid20536/bulletin-officiel.html?cid_bo=71544&cbo=1.
22. Sibert L, Mairesse JP, Aulanier S, Olombel P, Becret F, Thiberville J *et al.* Introducing the objective structured clinical examination to a general practice residency programme: Results of a French pilot study. *Med Teach* 2001;23:383-8.
23. Sibert L, Grand'Maison P, Doucet J, Weber J, Grise P. Initial experience of an objective structured clinical examination in evaluating urology residents. *Eur Urol* 2000;37:621-7.
24. Sibert L, Grand'Maison P, Charlin B, Grise P. Développement d'un examen clinique objectif structuré pour évaluer les compétences des internes en urologie. *Pédagogie Médicale* 2000;1:33-9.
25. Réforme du 2^e cycle des études médicales. Vade-Mecum pour les référents enseignants et étudiants. Ministère de l'enseignement Supérieur, de la Recherche et de l'Innovation. 2021 [On-line]. Disponible sur : https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwiC0NTu0efzAhVMXhoKHUYMCAIQFnoECAYQAQ&url=https%3A%2F%2Fcache.media.enseignementsup-recherche.gouv.fr%2Ffile%2FReforme_des_etudes_de_Sante%2F16%2F5%2Fsante_vade-m_reforme_2ecycle_A4_03_bis_1400165.pdf&usg=AOvVaw0HWM5li93JNlinwEjZ5Vsc.
26. SIDES. Système Inter-Universitaire Dématérialisé d'Évaluation en Santé, Université Numérique en Santé et Sport. [On-line]. Disponible sur : <https://www.uness.fr/plateformemes-sides/sides>.
27. Ho AD, Yu CC. Descriptive statistics for modern test score distributions: Skewness, kurtosis, discreteness, and ceiling effect. *Educ Psychol Meas* 2015;75:365-88.
28. Reznick R, Smee S, Rothman A, Chalmers A, Swanson D, Dufresne L *et al.* An objective structured clinical examination for the licentiate: Report of the pilot project of the medical council of Canada. *Acad Med* 1992;67:487-94.
29. Brannick MT, Erol-Korkmaz HT, Prewett M. A systematic review of the reliability of objective structured clinical examination scores. *Med Educ* 2011;45:1181-9.
30. Ilgen JS, Ma IW, Hatala R, Cook DA. A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. *Med Educ* 2015;49:161-73.
31. Rutala PJ, Witzke DB, Leko EO, Fulginiti JV, Taylor PJ. Sharing of information by students in an objective structured clinical examination. *Arch Intern Med* 1991;151:541-4.
32. Terry R, Hing W, Orr R, Milne N. Do coursework summative assessments predict clinical performance? A systematic review. *BMC Med Educ* 2017;17:40.
33. Pugh D, Bhanji F, Cole G, Dupre J, Hatala R, Humphrey-Murto S *et al.* Do OSCE progress test scores predict performance in a national high-stakes examination? *Med Educ* 2016;50:351-8.
34. Nuzzo A, Tran-Dinh A, Courbebaisse M, Peyre H, Plaisance P, Matet A *et al.* Improved clinical communication OSCE scores after simulation-based training: Results of a comparative study. *PLoS One* 2020;15:e0238542.

Citation de l'article : Sibert L, Schneider P, Liard A, Lefevre-Scelles A, Menard J-F, Auquit-Auckbur I, Veber B. L'examen clinique objectif structuré (ECOS) comme examen validant des compétences cliniques des étudiants en médecine français : 13 ans d'expérience rouennaise. *Pédagogie Médicale*, 2022;23:69-84