

Comment assurer l'équité d'une évaluation classante par les examens cliniques objectifs et structurés (ECOS) ? Enquête d'opinion auprès des externes sur les conditions d'acceptabilité de la réforme du deuxième cycle des études médicales en France

How to ensure the fairness of a grading evaluation by objective and structured clinical examinations (OSCE)? Opinion survey among medical students on the conditions of acceptability of the reform of the second cycle of medical studies in France

Carine ZUMSTEIN^{1,2,*} , Anne-Laure PHILIPPON^{3,4}, et Mathieu LORENZO^{1,5}

¹ Département de médecine générale et de formation territoriale, Faculté de médecine, maïeutique et science de la santé, Université de Strasbourg, Strasbourg, France

² Unité de simulation européenne en santé (UNISIMES), Faculté de médecine, maïeutique et science de la santé, Université de Strasbourg, Strasbourg, France

³ Sorbonne Université, Faculté de médecine, GRC Biofast, Paris, France

⁴ Service d'accueil des Urgences, Hôpital Pitié-Salpêtrière, Assistance Publique–Hôpitaux de Paris, Paris, France

⁵ Centre de formation et de recherche en pédagogie des sciences de la santé (CFRPS), Faculté de médecine, maïeutique et science de la santé, Université de Strasbourg, Strasbourg, France

Manuscrit reçu le 29 novembre 2021 ; commentaires éditoriaux formulés aux auteurs le 14 et le 29 novembre 2022 ; accepté pour publication le 1 décembre 2022

Résumé – Contexte : Les examens cliniques objectifs et structurés (ECOS) sont un outil d'évaluation de performances professionnelles en milieu simulé. Ils ont été intégrés à la réforme du deuxième cycle des études médicales en France afin d'évaluer les compétences des étudiants tout en leur garantissant l'équité et l'égalité des chances. Les ECOS favorisent l'objectivité de l'évaluation par leur forte standardisation mais de nombreuses variables sont à considérer. Afin de favoriser l'acceptabilité du dispositif, il est primordial de définir avec les étudiants les modalités de standardisation nécessaires et réalisables. **Objectif :** Explorer l'opinion des étudiants sur l'équité et la faisabilité des ECOS pour leur examen national. **Méthodes :** Nous avons utilisé une méthode qualitative basée sur des entretiens individuels semi-structurés avec des étudiants en deuxième cycle des études médicales. Les entretiens ont été codés en double aveugle et analysés de manière inductive. **Résultats :** Selon les étudiants, l'équité des ECOS repose principalement sur la construction de la grille d'évaluation, le choix des sujets, l'objectivité des évaluateurs et des patients standardisés et la formation préalable à l'examen. Ils précisent l'importance de la non-connaissance et l'absence de hiérarchie avec les évaluateurs et patients standardisés, et demandent une adaptation de la formation théorique et pratique centrée sur le patient. **Conclusion :** Selon les étudiants, l'équité des ECOS conditionne leur acceptabilité et repose sur le niveau de standardisation de tous les éléments les constituant. Au-delà de cette évaluation, la formation reste indispensable et devrait aboutir à un nouvel alignement pédagogique favorisant l'apprentissage de compétences cliniques au service des patients.

Mots clés : ECOS, évaluation de compétences, équité, standardisation, acceptabilité

Abstract. Background: Objective structured clinical examinations (OSCEs) are a tool for assessing professional performance in a simulated environment. They were integrated into the reform of the second cycle of medical studies in France in order to evaluate students' skills while guaranteeing equity and equal opportunity. OSCEs promote the objectivity of the evaluation by their high degree of standardization, but

*Correspondance et offprints : Carine ZUMSTEIN, Département de médecine générale et de formation territoriale, Faculté de médecine, maïeutique et sciences de la santé, 4 rue Kirschleger, 67000 Strasbourg, France.
Mailto : c.zumstein@unistra.fr.

many variables must be considered. In order to promote the acceptability of the system, it is essential to define with the students the necessary and feasible standardization methods. **Objective:** To explore students' opinions on the fairness and feasibility of OSCEs for their national exam. **Methods:** We used a qualitative method based on individual semi-structured interviews with graduate medical students. The interviews were double-blind coded and analyzed inductively. **Results:** According to the students, the fairness of OSCEs is mainly based on the construction of the evaluation grid, the choice of subjects, the objectivity of the evaluators and the standardized patients, and the pre-examination training. They point out the importance of non-knowledge and lack of hierarchy with the assessors and standardized patients, and ask for an adaptation of the theoretical and practical training focused on the patient. **Conclusion:** According to the students, the fairness of OSCEs conditions their acceptability and depends on the level of standardization of all the elements that make them up. Beyond this evaluation, training remains essential and should lead to a new pedagogical alignment that promotes the learning of clinical skills in the service of patients.

Keywords: OSCE, competence assessment, fairness, standardization, acceptability

Introduction

Les examens cliniques objectifs et structurés (ECOS) sont un outil d'évaluation de performances professionnelles en milieu simulé [1] qui, à côté d'autres sources d'information, peut légitimement contribuer à l'évaluation des compétences cliniques [2]. Les étudiants passent dans un circuit de plusieurs stations chronométrées, au sein desquelles ils effectuent des tâches spécifiques face à un patient simulé et standardisé selon les objectifs d'évaluation de la station [3,4]. Ils sont évalués par des observateurs à l'aide de grilles d'évaluation standardisées [1].

Contexte et problématique

En France, l'arrêté du 2 septembre 2020 a consacré l'intégration des ECOS à la réforme du deuxième cycle des études médicales (R2C) afin de « vérifier les compétences acquises par les étudiants et notamment leur capacité à développer un raisonnement clinique et à résoudre des problèmes » [5]. Ils prennent place dans le nouveau dispositif qui en fin de second cycle remplace les anciennes épreuves classantes nationales (ECN), pour organiser la procédure nationale d'affectation des étudiants dans les différentes filières de formation spécialisées. Alors que les ECN ne comportaient que des tests cognitifs (dossiers cliniques progressifs, questions à choix multiples et épreuve de lecture critique d'article scientifique), le dispositif introduit par la R2C organise une affectation des étudiants en troisième cycle sur la base d'un algorithme d'appariement entre les choix préférentiels des étudiants et leurs résultats à différentes épreuves. Celles-ci comportent un examen dématérialisé national (EDN) de connaissances, dont les résultats participent pour 60 % à la note globale, des ECOS pour 30 % de la note globale et la valorisation d'un parcours professionnel spécifique pour 10 % de la note globale. L'évaluation des connaissances par l'EDN différencie les connaissances de rang A, qui doivent être maîtrisées par tous les étudiants à leur entrée en troisième cycle, et des connaissances de rang B, spécifiques à chaque filière de spécialisation.

La procédure d'affectation des étudiants se base alors sur autant de classements que de filières de spécialisation et reste fortement liée à une évaluation normative des performances des étudiants. L'examen ECOS organisé à l'échelle nationale en fin de troisième année du diplôme de formation approfondie en sciences médicales (DFASM3 – sixième année d'études) est donc l'un des déterminants majeurs du choix ultérieur de la spécialité et du lieu de poursuite des études pour chaque étudiant en médecine ; il conditionne de ce fait le champ de sa pratique professionnelle future.

Selon l'Association nationale des étudiants en médecine de France (ANEMF), l'addition des ECOS dans le dispositif d'évaluation et de classement de fin de sixième année de médecine vise à intégrer l'évaluation des compétences cliniques dans l'examen tout en garantissant « l'équité » et « l'égalité des chances entre étudiants » pour leurs choix de postes de troisième cycle [6]. L'ANEMF indique corollairement que l'introduction des ECOS était préférable à la prise en compte des évaluations de stage au motif que ces dernières ne conféraient « pas assez de garanties d'équité entre les étudiants » [6].

L'équité est un principe impliquant l'appréciation juste et le respect absolu de ce qui est dû à chacun [7]. L'égalité des chances, quant à elle, est un principe stipulant selon Rawls qu'à « capacités » et « talents » égaux, les individus devraient avoir les mêmes perspectives de succès [8]. En matière d'évaluation, elle se définit au regard de la capacité de juger sans biais ni discrimination [9], en visant une objectivité maximale alors que le jugement humain est par nature subjectif [10].

La standardisation de l'évaluation est souvent perçue comme une piste privilégiée pour réduire cette subjectivité car elle implique que tous les étudiants répondent aux mêmes questions, reçoivent les mêmes instructions claires et précises, qu'aucun ne profite d'avantages par rapport aux autres et que le système de correction s'applique uniformément pour tous les étudiants [11]. Ces quatre conditions peuvent être satisfaites dans le cadre des ECOS, assurant ainsi leur standardisation en réduisant les variables de cotation autres que les performances des étudiants [1].

Cette exigence de rigueur s'inscrit dans une perspective docimologique de l'évaluation, selon laquelle les observations qui permettent de documenter et évaluer les apprentissages doivent résulter d'une mesure. Les outils d'évaluation y sont en conséquence étudiés selon les différentes qualités métrologiques d'un bon instrument de mesure (validité, fidélité). Plus spécifiquement, la fidélité inter-juges (qui est l'une des exigences de l'objectivité) se réfère directement aux notions de standardisation et d'équité, puisqu'elle désigne le « degré de concordance entre le jugement porté par des examinateurs différents sur ce qui constitue la bonne réponse » [12,13]. Au regard de cette perspective docimologique, la formation des évaluateurs [1,12–14] et la standardisation des outils [1,15] permettent ainsi d'augmenter l'objectivité de l'évaluation [1,2].

De nombreux auteurs considèrent que l'obsession de l'objectivité, centrale dans la perspective docimologique, impose souvent des contraintes excessives qui, en contrepartie, peuvent avoir des conséquences réductrices quant à la qualité des apprentissages qui sont évalués. Plusieurs analyses ont ainsi argumenté que les jugements évaluatifs effectués à partir de mesures « objectives » ne sont finalement pas plus fiables que des jugements construits sur des données « subjectives » [10,16–19]. La standardisation des dispositifs d'évaluation n'est ainsi pas intrinsèquement nécessaire, y compris dans le cadre d'une évaluation certificative, pour autant que des conditions appropriées visant à assurer la crédibilité et la fiabilité du jugement évaluatif soient mises en place, dans une perspective d'évaluation à la fois plus qualitative et plus systémique, telle que la propose par exemple l'approche programmatique [18] qui est notamment à privilégier lorsque l'on revendique une évaluation de compétences [20,21]. Ce n'est ainsi pas la finalité certificative du dispositif d'évaluation des étudiants en fin de sixième année de médecine en France qui contraint à satisfaire l'exigence métrologique d'objectivité et de standardisation, mais le choix politique et contextuel des autorités administratives françaises de faire reposer la procédure d'affectation des étudiants dans chacune des filières spécialisées sur des classements nationaux, à partir des résultats d'une évaluation normative.

L'organisation des ECOS nécessite d'importants moyens humains et matériels [1,4]. Leur faisabilité dépend d'aspects logistiques tels que la disponibilité des salles, le coût du matériel, le nombre d'évaluateurs et de patients standardisés disponibles et le temps de préparation et de correction [1,12,14,15]. De nombreuses variables peuvent dès lors altérer les conditions visant l'objectivité et la standardisation du dispositif : des évaluateurs mal formés, des patients simulés insuffisamment standardisés, des sujets et des grilles d'évaluations non adaptés peuvent réduire la qualité psychométrique des ECOS [14]. Pour autant, la réussite de l'implantation des ECOS dans la R2C est primordiale pour les étudiants compte tenu des enjeux personnels qui lui sont associés, comme cela a été rappelé précédemment.

Les résultats d'enquêtes publiées dans la littérature quant au vécu d'étudiants de leurs expériences d'ECOS indiquent qu'ils perçoivent cet outil comme une préparation utile à leur pratique clinique [22,23] et à l'amélioration de leurs compétences communicationnelles [24], qui offre la possibilité d'évaluer en même temps un large éventail de compétences et disciplines [23–26]. Certains relevaient l'importance de la rétroaction et de l'opinion de leurs pairs à propos de leur performance dans leur processus d'apprentissage [22].

Les ECOS sont cependant vécus comme anxiogènes [23–25,27], notamment par peur de l'inconnu [27], par des étudiants qui signalaient un manque de préparation à cet outil d'évaluation [25,28]. Leur vécu des sessions d'ECOS est aussi conditionné par l'organisation globale du dispositif : ils mentionnaient notamment le nombre de stations et leur durée [24,28].

Selon les étudiants, les ECOS permettent une évaluation équitable [25,26], fiable et valide [23], mais ils s'inquiètent de la variabilité inter-évaluateur lors de l'usage de cet outil d'évaluation [24]. L'acceptabilité de ce dispositif reste faible [23] et la majorité des étudiants expriment préférer des examens écrits composés de questions à choix multiples plutôt que des ECOS [23].

Objectifs

Dans le contexte rapporté de la R2C et au regard des données disponibles, il nous semble ainsi important d'étudier le vécu des sessions d'ECOS de la part d'étudiants français concernés, pour comprendre l'acceptabilité du dispositif et documenter d'éventuelles pistes d'amélioration, en lien avec une optimisation des conditions d'organisation lors du processus d'implantation, alors même que leur association représentative en argumente le caractère équitable.

L'objectif du présent travail était donc d'explorer l'opinion des étudiants sur l'équité et la faisabilité de l'organisation des ECOS, en tant que composante d'un examen national de fin de sixième année participant à la procédure d'affectation aux différentes filières de formation spécialisées de troisième cycle.

Méthodes

L'approche qualitative a été choisie en cohérence avec l'objectif de recueillir les opinions des étudiants dans le cadre d'une étude largement exploratoire. Le recours à des entretiens individuels semi-dirigés plutôt qu'à des groupes de discussion focalisée répondait à la préoccupation de favoriser la parole de chacun sans influence extérieure.

Ce travail a été construit selon les lignes directrices *Consolidated Criteria for Reporting Qualitative Research* (COREQ) [29]. L'enquêtrice principale en charge des entretiens avait une expérience limitée de la méthodologie qualitative mais a pu être guidée par les autres auteurs. Elle ne connaissait pas les participants et leur avait présenté ses objectifs et motivations par un courriel et une

notice d'information, en indiquant notamment son implication facultaire locale dans la mise en place des ECOS.

Échantillonnage

Les participants ont été recrutés parmi les étudiants inscrits de la troisième à la sixième année des études de médecine, en essayant d'obtenir un échantillon raisonné à variation maximale. Nous avons cherché à recruter des étudiants de différentes années d'étude, directement concernés par la réforme ou vivant l'actuel deuxième cycle à réformer, issus de différentes facultés de médecine utilisant ou non déjà des ECOS, et de différents niveaux d'implication au sein des associations étudiantes ou des instances universitaires.

Dans un premier temps, les étudiants ont été invités à participer *via* un courriel transmis par l'intermédiaire de leurs représentants facultaires et de leurs associations étudiantes locales et nationales, ainsi que par le réseau national de chercheurs en pédagogie des auteurs. Les étudiants intéressés étaient invités à renvoyer un courriel à l'auteur principal afin qu'il puisse s'assurer que son échantillon était diversifié et organiser le mode d'entretien.

Recueil des données

Les entretiens ont été conduits avec l'appui d'un guide d'entretien (Annexe A) visant à explorer les points de vue des étudiants. Ces entretiens ont été réalisés par l'auteur principal par téléphone ou en visioconférence, en toute confidentialité, et ont été enregistrés au moyen d'un dictaphone. Le guide d'entretien a été élaboré à partir de l'expérience de l'auteur principal dans l'organisation globale de sessions d'ECOS, enrichie par de nombreuses lectures d'articles et de formations en pédagogie médicale sur l'évaluation en sciences de la santé. Ce guide a été testé puis modifié après les trois premiers entretiens pour optimiser le recueil de données. Les entretiens se sont poursuivis jusqu'à obtenir une saturation suffisante des données, confirmée par la réalisation d'un entretien supplémentaire.

Neuf entretiens semi-structurés ont été réalisés entre le 1^{er} mars et le 23 juin 2021 et ont duré en moyenne 41 minutes (Tab. I). Trois participants initialement contactés n'ont pas donné suite. Les entretiens retranscrits n'ont pas été renvoyés aux participants. Un participant a complété son entretien *a posteriori* par renvoi de courriel.

Analyse des données

Les données enregistrées lors de chaque entretien ont été retranscrites à partir des enregistrements sur le logiciel Word pour former le verbatim de chaque entretien, mot à mot et fidèlement à l'enregistrement, en anonymisant les réponses, et en y ajoutant les notes de l'auteur prises après les entretiens. Ces retranscriptions n'ont pas été retournées aux participants.

Un double codage en aveugle a ensuite été effectué par deux auteurs (CZ, ALP), de façon longitudinale pour chaque entretien, puis transversale pour vérifier la cohérence des thèmes entre chaque entretien. L'arbre de codage a été étoffé au fil des codages sur Word et Excel. L'analyse des données a été faite de façon inductive par catégorisation et mise en relation du verbatim, selon une analyse thématique.

Résultats

L'analyse a retrouvé quatre thèmes principaux correspondant aux conditions d'équité des ECOS, selon les participants : 1) la construction de la grille d'évaluation ; 2) le choix des sujets ; 3) l'objectivité des évaluateurs et des patients standardisés et 4) la préparation à l'examen.

Les étudiants s'accordent sur la constance de la subjectivité d'un examen oral, même si tous ne l'acceptent pas, et mettent en avant l'importance de la neutralité de l'évaluation : « *y aura toujours un biais [...] c'est pas acceptable mais on peut pas s'en affranchir totalement* » (P8).

Certains demandent un cadre réglementaire pour assurer la transparence du dispositif : « *L'égalité des chances [...] c'est primordial, [...] ça doit être filmé pour contre-expertise [...] j pense que en fait c'est un droit, [...] un tirage au sort qui est fait sous huissier* » (P8), « *pouvoir prouver le... pourquoi est-ce que c'est fait comme ça* » (P1).

Des grilles détaillées pour réduire la subjectivité

L'un des points essentiels pour assurer la standardisation de l'évaluation est la construction de la grille : des termes ou gestes cruciaux doivent y figurer mais aussi des points attribués à l'évaluation de la relation médecin patient. Les étudiants insistent sur le détail de ces grilles pour réduire la subjectivité des évaluateurs : « *La première chose c'est la qualité des grilles* » (P1), « *il faut que tout soit détaillé parce que sinon effectivement c'est subjectif* » (P2).

Certains semblent favoriser les listes de vérification qui ne laissent pas de place à la subjectivité mais qui semblent peu discriminantes : « *le oui/non j pense que du coup c'est assez objectif, soit on l'a dit soit on l'a pas dit donc c'est assez simple à coter* » (P5), « *c'est pas très discriminant* » (P6) ; « *la nuance est importante* » (P8).

Le recours à des échelles de Likert semble plus neutre à condition qu'elles soient détaillées, ce qui permettrait d'augmenter la fidélité inter-juges et de favoriser leur acceptabilité : « *ce s'rait plus examinateur dépendant...* » (P5) ; « *acceptable oui euh... après servir au classement des étudiants euh... je saurais pas répondre [...] apporter le plus de précisions possible sur qu'est-ce que ça veut dire une échelle à 3, qu'est-ce que ça veut dire une échelle à 2* » (P3).

Certains étudiants apprécieraient des échelles d'activité fiable pour évaluer les stations de façon globale, mais elles semblent peu objectives et peu acceptables : « *l'inconnu dans l'évaluation [...] on n'aime pas trop* » (P4).

Tableau I. Caractéristiques des participants et des entretiens.

	Sexe	Âge	Année d'études	Région d'études	Implication associative ou facultaire	Expérience antérieure d'ECOS	Durée entretien (min)
P1	Féminin	21	3 ^e -4 ^e	Normandie	Forte implication	oui	39
P2	Féminin	24	6 ^e	Bourgogne-Franche Comté	Faible implication	oui	37
P3	Masculin	23	5 ^e	Grand Est	Pas d'implication	oui	33
P4	Masculin	21	4 ^e	Bourgogne-Franche Comté	Faible implication	oui	43
P5	Féminin	23	6 ^e	Hauts-de-France	Faible implication	oui	26
P6	Féminin	24	6 ^e	Nouvelle Aquitaine	Faible implication	oui	43
P7	Féminin	23	6 ^e	Auvergne-Rhône-Alpes	Faible implication	oui	35
P8	Masculin	26	6 ^e	Grand Est	Faible implication	non	66
P9	Masculin	23	5 ^e	Grand Est	Pas d'implication	non	47

ECOS : Examens cliniques objectifs et structurés

Accepter la variabilité des sujets : entre validité et fidélité des ECOS

Le choix de sujets qui seraient, respectivement, identiques au plan national ou différents selon chaque site facultaire fait débat : « *un sujet partout l'même, ce s'rait mieux* » (P9) ; « *je vois pas l problème si on a pas les mêmes cas partout* » (P6).

Ils seraient tirés au sort dans un pool national de sujets issus du programme et répartis en différents types selon les tâches demandées, sur lesquels chaque étudiant serait évalué pour garantir la validité de contenu de l'outil : « *un pool national de sujets que chaque faculté tire au sort* » (P4), « *évaluant les mêmes domaines d'apprentissage et les mêmes compétences* » (P1).

La fidélité de l'outil nécessiterait que les sujets soient similaires et de même niveau de difficulté, ou identiques avec un examen simultané sans risque de fuite de sujet : « *on s'rendait compte qu'y en avaient qui étaient tombé sur des trucs quand même un peu plus pointus que d'autres* » (P6).

Enfin, la validité écologique semble remise en question pour certains sujets hospitaliers ne reflétant pas l'authenticité du métier : « *on est quand même pas dans un box à l'hôpital* » (P2).

L'objectivité des évaluateurs et patients standardisés en question

Le recrutement des évaluateurs doit reposer sur leur volontariat et leur qualité d'enseignant, universitaire ou non, ou même non-médecin. L'objectivité de l'évaluation devrait pouvoir être assurée par la formation des évaluateurs, par le recours à deux évaluateurs par station, à des évaluateurs d'autres spécialités que celle de la station et par l'absence de lien de connaissance entre les étudiants et les évaluateurs. Les étudiants proposent ainsi des échanges d'évaluateurs entre facultés mais reconnaissent la difficulté d'une telle organisation : « *que tous les examinateurs aient une formation commune de toutes les facs* » (P2) ; « *faire fonctionner ça en binôme moi j'pense que ça pourrait être pas mal [...] c'est p't-être plus*

objectif » (P6) ; « *j'me suis retrouvée avec un grand chef d'anatomie qui est aussi le père d'un de mes amis, ça m'a juste super stressée et je l'ai, "fin ce sujet complètement complètement raté"* » (P5).

Cette évaluation devrait s'accompagner d'une rétroaction : « *vu qu'on a pas d'debriefing après, si c'que j'ai dit c'est n'importe quoi, au final personne viendra jamais m'dire que c'était n'importe quoi* » (P5) ; « *faut qu'ça ait plus un aspect pédagogique* » (P6).

Le recrutement des patients standardisés est envisagé de façon très ouverte : patients, acteurs, externes, internes, chefs de cliniques, professeurs des universités étaient régulièrement cités. Selon les participants, le vrai patient a l'avantage de sa faible maîtrise du discours médical et de sa capacité à fournir une meilleure prestation de simulation de sa pathologie mais il risque de se perdre dans ses consignes, réduisant sa standardisation. L'acteur n'a pas de connaissances médicales mais doit être formé pour savoir quelles informations donner aux étudiants. Recruter des internes pourrait rassurer les externes, alors qu'un médecin pourrait être moins investi dans son rôle et indiquer par son comportement non verbal que ce n'est pas la réponse attendue. Le patient standardisé doit être volontaire, intègre, en sécurité psychologique et inconnu des étudiants voire sans lien hiérarchique : « *est-ce que le vrai patient peut pas se perdre dans un flot d'informations et que ce soit pas assez standardisé pour les ECOS* » (P4) ; « *prendre des patients qui ont déjà présenté la maladie, qui partiellement la simuleront beaucoup mieux* » (P8) ; « *y aurait aussi un peu d'gêne, surtout d'examiner quelqu'un qui hiérarchiquement est supérieur à nous* » (P9) ; « *personne n'a déshabillé le patient [...] tout le monde a dit oui mais si c'était un vrai patient dans un box je l'aurais fait spontanément* » (P2). Le rôle du patient doit être scrupuleusement standardisé pour augmenter sa neutralité et favoriser l'authenticité des stations, sans rentrer dans les extrêmes, d'où la possibilité d'une certaine liberté de comportement verbal ou non verbal : « *faut leur apprendre la scène quoi, pour que ce soit plus réaliste possible* » (P3) ; « *ça pourra pas être parfait, on veut l'standardiser au maximum* » (P4) ; « *au final il l'aidait [...] et moi il m'a pas dit ça [...] on perd l'équité. [...] ils seront*

peut-être un peu plus sévères avec ceux de fin de matinée parce qu'ils ont vu pleins d'étudiants passer et ils se comportent pas de la même manière [...] après il faut trouver un juste milieu » (P2).

Un alignement pédagogique nécessaire

L'examen de fin de sixième année de médecine est un enjeu majeur pour les étudiants : *« c'est quelque chose qui va conditionner une vie »* (P8).

Les épreuves classantes nationales (ECN) actuelles et leur préparation restent très théoriques alors que la réforme propose une évaluation de compétences en lien direct avec leur futur métier, mais le manque de formation théorique et pratique adaptée rend l'épreuve inéquitable : *« C'est pas dans le bon sens qu'est l'entonnoir d'apprentissage »* (P1) ; *« c'est super qu'on soit enfin évalués en tant qu'étudiants et futurs médecins, et pas qu'en tant que machine à QCM [(Question à Choix Multiple)] »* (P2) ; *« des ECOS à l'ECN l'année prochaine y aurait un problème d'équité »* (P3) ; *« nous sommes beaucoup formés « sur le tas » et non uniformément à ces choses pendant les stages »* (P7) ; *« on commencerait à former les médecins ne serait-ce qu'à évaluer les étudiants, déjà rendre obligatoire les évaluations de fin d'étage [...] ne pas avoir de retour j'trouve que c'est pas du tout constructif »* (P6).

La nouvelle modalité d'évaluation permettrait d'adapter les enseignements et les apprentissages et de placer le patient au centre du dispositif : *« ça permet vraiment de redonner du sens et du concret à ce qu'on apprend »* (P1) ; *« le but des ECOS c'est pas forcément pour les ECN futurs, c'est un peu pour toute la formation, [...] y a un autre enjeu aussi sur la façon dont seront soignés les patients »* (P6).

Discussion

Nos résultats révèlent les nombreuses inquiétudes des étudiants qui portent à la fois sur la faisabilité des ECOS dans le cadre de la R2C et sur les conditions de mise en œuvre qui pourraient menacer l'équité du dispositif. Ces inquiétudes font écho à plusieurs constats déjà disponibles dans la littérature, résultant d'enquêtes auprès des étudiants, et elles peuvent être mises en lien avec des réflexions générales concernant la problématique de l'évaluation certificative finale des étudiants en médecine.

Elles sont cependant à interpréter aussi au regard des particularités du contexte français, caractérisé notamment par la dimension concurrentielle qu'induit la procédure d'affectation en troisième cycle, organisée sur la base d'un système de classement des étudiants. Ainsi, l'acceptabilité de l'implantation des ECOS, telle qu'elle est en cours, laisse sans doute la place à la possibilité de compromis, à condition que ces derniers n'altèrent pas l'exigence d'équité, telle qu'elle est perçue par les étudiants au regard des enjeux personnels et professionnels qui sont leurs.

Les points de vigilance exprimés par les étudiants concernent notamment le niveau de standardisation requis pour chacune des composantes opérationnelles de

l'organisation d'un ECOS. Au-delà du dispositif d'évaluation lui-même, ils concernent aussi l'ensemble de la formation théorique et pratique. En mettant en exergue la nécessité que le curriculum réponde aux exigences d'un alignement pédagogique, les étudiants questionnent la cohérence globale du curriculum dans la perspective d'une préparation à leur futur métier, en suggérant que les ECOS devraient d'abord être un moyen au service de leurs apprentissages avant d'être une fin.

L'équité avant la faisabilité : une question de standardisation ?

Les étudiants interrogés dans l'étude estiment que l'équité des ECOS implique une standardisation plus ou moins forte à la fois des sujets des stations, des grilles d'évaluation, des évaluateurs et des rôles joués par les patients standardisés, résultats en accord avec la littérature [1,10,15–19]. Sans doute imprégnés de l'approche positiviste qu'ils ont connue tout au long de leurs études, il leur semble difficile de s'affranchir de la standardisation comme unique moyen d'assurer l'équité entre tous, notamment du fait que leur future carrière professionnelle repose sur un dispositif de classements des étudiants. Pour autant, le niveau de standardisation reste encore à définir, en cherchant à concilier ce qui pourrait favoriser l'acceptabilité du dispositif, tout en ne l'enfermant pas dans des contraintes qui ne seraient pas nécessaires. Un travail de communication approfondi entre toutes les parties prenantes semble en tout état de cause nécessaire pour permettre une meilleure acceptabilité du dispositif auprès de ces étudiants.

Sur un premier plan, le choix entre, respectivement, des sujets nationaux uniformisés, option qui compliquerait potentiellement la standardisation des évaluateurs et des patients simulés, ou des sujets locaux, alternative qui nécessiterait une harmonisation des types de sujets et de leur niveau de difficulté entre les différents centres d'examen [4], reste une question ouverte.

Sur un deuxième plan, celui des différents formats de grille d'évaluation, les étudiants semblent considérer qu'ils se complètent et sont à utiliser en fonction de ce qu'on veut évaluer. De fait, certaines données de la littérature indiquent que l'utilisation d'une liste de vérification, plutôt qu'une échelle globale, favorise une évaluation plus objective [14,30], sauf dans le domaine affectif et communicationnel [31] car la mesure reste très réductrice [11]. Cependant, la construction et la validation d'une liste de vérification ne dispensent pas d'un jugement subjectif [30]. L'échelle d'évaluation globale peut être utilisée pour de multiples tâches évaluées simultanément, où elle permet une évaluation plus nuancée des étudiants, voire un repérage de comportements potentiellement dangereux par rapport aux pratiques recommandées [30].

Concernant l'utilisation des grilles, la littérature fait état d'un large consensus quant au fait que les évaluateurs doivent y être formés [1,13,14,32]. Les étudiants rapportent notamment des effets d'ordre de correction, de séquence, de halo et d'instabilité du correcteur [33] qui

influent sur les notes obtenues. Nos propres résultats sont convergents avec ces constats. Ils illustrent la conviction de la part des étudiants que la formation des évaluateurs, visant une compréhension similaire de chaque item de la grille d'évaluation, est une exigence incontournable, dans l'optique d'une standardisation maximale pour éviter ces facteurs de variations.

Il en va de même quant à la nécessité d'une forte standardisation des rôles à jouer par les patients, qui doivent être détaillés et faire l'objet d'un entraînement préalable, condition indispensable pour qu'ils sachent quand et comment délivrer chaque information verbale ou non verbale en fonction de chaque étudiant [32].

À l'instar de ce qui est rapporté dans la littérature, les étudiants sont par ailleurs sensibles à l'absence de lien de connaissance et de hiérarchie avec l'évaluateur et le patient standardisé, condition visant à ne pas perturber la performance des étudiants et à assurer la neutralité de l'évaluateur et du patient standardisé. Un évaluateur qui connaît l'étudiant pourrait en effet modifier son score en fonction de l'étudiant, de son niveau social ou par effet d'inertie par rapport aux scores des évaluations antérieures du même étudiant [33]. Un patient standardisé pourrait par son jeu modifier la standardisation de l'outil pour favoriser ou défavoriser certains étudiants.

Pris dans leur ensemble, les discours des étudiants recueillis dans notre enquête peuvent s'interpréter en lien avec la nécessité de réfléchir aux qualités psychométriques de l'outil d'évaluation choisi, afin d'optimiser son acceptabilité [12]. Il est admis que les ECOS peuvent satisfaire un haut niveau de validité [1,13], notamment de validité écologique [1,4,12,32] mais aussi de validité de contenu [32], parce qu'ils explorent le niveau taxonomique de la « démonstration » de la pyramide de Miller grâce au recours à la simulation.

Le choix des sujets explorés dans chacune des stations est déterminant pour garantir la fidélité ou la reproductibilité de la mesure que permet l'outil et donc l'équité du classement qu'elle permet d'établir [1,12,13]. Les questions liées au choix de sujets nationaux ou de sujets locaux ont déjà été évoquées. Les étudiants évoquent la possibilité de classer les sujets par types et d'être systématiquement exposés à tous les types de stations pour être évalués sur l'ensemble des compétences nécessaires aux médecins.

La formation des évaluateurs [1,12–14] et la standardisation des outils d'évaluation [1,15], cités par les étudiants, conditionnent l'objectivité en lien avec la fidélité inter-juges des ECOS. Mais lorsqu'on vise l'évaluation globale de performances lors de situations complexes, la validité du jugement évaluatif « subjectif » peut être tout autant assurée, moyennant le recours à une variété suffisante des cas et à une multiplicité des évaluateurs [11–13]. En l'espèce, si les étudiants ne mentionnent pas le nombre de stations, ils insistent bien sur la nécessité d'avoir plusieurs évaluateurs par station, ce qui peut néanmoins réduire la fidélité de l'outil [12].

C'est souligner à nouveau les tensions complexes qui émergent à l'interface des deux tendances lourdes de la démarche évaluative, celle qui met en avant la rigueur et

l'objectivité de la mesure et des indicateurs quantitatifs à partir de l'observation de tâches standardisées (approche docimologique) et celle qui privilégie la pertinence du jugement et des indicateurs qualitatifs recueillis à partir de l'observation et l'analyse de tâches authentiques (approche programmatique) [10–20]. Ces mêmes tensions sont à l'œuvre dans les orientations et les dispositions de la R2C qui, d'un côté, souhaite favoriser l'évolution des dispositifs curriculaires vers une approche par compétences – ce qui justifierait de privilégier une approche programmatique – mais qui, de l'autre, ne rompt pas avec l'impératif d'un classement des étudiants – ce qui contraint à satisfaire les exigences de la psychométrie.

Il faut donc comprendre que les conditions de l'équité sont multidimensionnelles et ne se résument pas à des chiffres ou des qualités psychométriques que pourraient facilement garantir la standardisation [34]. Les étudiants savent qu'une part de subjectivité persiste dans tout jugement évaluatif. À ce titre, ce n'est sans doute pas tant la « subjectivité » qu'ils récuse que le caractère arbitraire et opaque des décisions qui les concernent, dont ils pourraient craindre d'être les victimes.

Un alignement pédagogique nécessaire

Une telle perspective soutient dès lors l'exigence d'alignement pédagogique que les discours des étudiants révèlent. Au-delà de leur usage au moment de l'évaluation certificative, les ECOS peuvent fournir un impact positif sur les apprentissages [4]. Sous certaines conditions, la simulation, grâce à la reconstitution raisonnablement réaliste de situations que permettent les ECOS, peut favoriser l'apprentissage et l'acquisition de capacités opératoires au-delà des seules connaissances, les unes et les autres constituant des ressources pour le développement de compétences [1,11]. Les étudiants indiquent vouloir changer leurs méthodes d'apprentissage pour être mieux préparés à soigner les patients, notamment par un meilleur investissement en stage qui prime pour certains sur les révisions théoriques de leurs cours. Dans une telle perspective, ils conçoivent les ECOS comme une modalité d'évaluation cohérente et pertinente, avant même que se pose la question de leur acceptabilité comme épreuve concourant à leur classement.

Les enseignants doivent tout autant être mobilisés pour favoriser l'alignement pédagogique et la cohérence entre la formation et le dispositif d'évaluation choisi car cela conditionne les modalités d'apprentissage [11]. Les étudiants souhaitent que les évaluations deviennent systématiques dans les stages, réalisées par des encadrants formés à l'évaluation, avec des entraînements réguliers [11] pour se familiariser à l'outil [6]. Ces entraînements ou évaluations formatives doivent selon eux s'accompagner d'une rétroaction de niveau adapté, dans le cadre de débriefings récurrents, afin que puissent s'ajuster, respectivement, les stratégies d'apprentissage des étudiants et les interventions pédagogiques des enseignants [4,11].

L'enjeu de l'intégration des ECOS dans le deuxième cycle ne devrait donc pas s'arrêter aux questions que soulèvent les nécessités d'une évaluation certificative et classante, mais il devrait intégrer la perspective d'une opportunité à saisir pour reconsidérer les curriculums de formation des futurs médecins.

Forces et limites de l'étude

La triangulation des chercheurs au cours du codage et de l'analyse est un élément au bénéfice de la crédibilité de nos résultats. La neutralité de l'analyse a été renforcée par un double codage en aveugle, réalisé en alternance avec le recueil de données, en faisant émerger les thèmes à partir des données. La saturation des données a été obtenue après huit entretiens et confirmée par un entretien supplémentaire.

La transférabilité de nos résultats est favorisée par le caractère multicentrique de notre étude intéressant six régions françaises, par la qualité de notre échantillonnage à variation maximale en termes d'âge, de sexe, d'année d'études, sélectionné parmi des étudiants volontaires et motivés pour exprimer leurs opinions. Notre échantillon n'est cependant peut-être pas totalement représentatif des étudiants en deuxième cycle de médecine car seuls deux d'entre eux n'avaient pas d'expérience préalable d'ECOS.

En raison du caractère exploratoire de notre méthode, nous n'avons utilisé qu'un seul outil de recueil des données. Une triangulation méthodologique pourrait compléter ce travail, *via* la réalisation de groupes de discussion focalisée pour confronter les opinions des étudiants et favoriser leurs interactions sociales, afin de construire collectivement des conditions d'acceptabilité des ECOS. Celles-ci pourraient ultérieurement faire l'objet d'une enquête quantitative pour recueillir un maximum d'opinions et d'observations directes d'évaluations par les ECOS, avec un retour immédiat de la part des étudiants observés pour affiner les niveaux de standardisation nécessaires pour garantir l'équité du dispositif. Un recueil complémentaire des opinions des enseignants engagés dans la réforme pourrait également enrichir notre réflexion.

Conclusion

L'enquête rapportée dans ce travail montre qu'aux yeux des étudiants, l'équité d'une évaluation classante par les ECOS est d'abord conditionnée par son niveau de standardisation, qui pourrait faire l'objet d'aménagements définis avec eux, pour favoriser l'acceptabilité du dispositif. Cette standardisation concerne tous les éléments opérationnels constituant les ECOS. Mais au-delà de l'évaluation, la réforme du deuxième cycle est un tremplin pour mettre en place un nouvel alignement pédagogique favorisant l'apprentissage de compétences cliniques au service des patients. Des travaux complémentaires permettront de confronter les opinions des étudiants et d'aboutir au meilleur compromis entre faisabilité et équité.

Contributions

Carine Zumstein a participé à la conception du protocole de recherche, au recueil des données, à l'analyse des données, à l'interprétation des résultats et à l'écriture du manuscrit. Anne-Laure Philippon et Mathieu Lorenzo ont participé à l'analyse des données, à l'interprétation des résultats et ont contribué à l'écriture du manuscrit.

Ce travail a été élaboré dans le cadre d'un mémoire de recherche pour l'obtention du diplôme de master de pédagogie en sciences de la santé à l'Université de Strasbourg.

Approbation éthique

Ce travail a fait l'objet d'une déclaration à la Commission nationale de l'informatique et des libertés (CNIL) en date du 26 février 2021 et d'un avis favorable préalable du Comité pour l'intégrité et l'éthique de la recherche en éducation des professions de la santé de la Société internationale francophone d'éducation médicale (SIFEM) en date du 10 mars 2021. Tous les participants ont donné leur accord oral et écrit.

Liens d'intérêts

Aucun auteur ne déclare de conflit d'intérêts en lien avec le contenu de cet article. L'auteure principale est cependant impliquée dans le dispositif facultaire local d'implantation des ECOS à l'Université de Strasbourg.

Annexe A: Guide d'entretien

- 1) Accueil du participant et confirmation orale du consentement à participer à l'étude.
- 2) As-tu déjà expérimenté des ECOS dans ta faculté ? Si oui, comment les choses se sont passées ? Si non, comment imaginerais-tu que ça puisse se passer ?
Je te propose maintenant de rentrer dans le détail du dispositif, en partant de ton expérience/idée, puis d'essayer d'élargir les choses pour les imaginer au niveau national.
- 3) Comment envisages-tu que les sujets des stations soient choisis dans chaque faculté de France ?
 - Comment pourrait-on envisager le choix des stations de telle sorte à ce que ce soit acceptable pour tous les étudiants dans l'idée du matching ?
 - Comment imagines-tu l'organisation pratique du dispositif ? Comment penses-tu que ce soit faisable en termes de salles, de nombre d'évaluateurs et de patients standardisés ? Comment pourrait-on l'organiser ?
- 4) L'évaluation des stations se fait au moyen d'une grille d'évaluation interprétée par un évaluateur. Comment envisages-tu que l'évaluation des stations puisse se faire partout en France ?

- Comment faudrait-il construire les grilles d'évaluation selon toi ? Quels seraient les points de vigilance à avoir ? La grille peut se construire avec des items à cocher, des échelles plus globales ou des échelles d'activités fiables, que penses-tu de ces différentes possibilités ?
- Une évaluation d'un examen oral réalisée par un individu me semble plutôt subjective, même avec une grille d'évaluation interprétée par chaque évaluateur. Pourtant c'est ce qui est fait dans les ECOS. Selon toi, comment pourrait-on uniformiser l'interprétation des différents évaluateurs ?
- Et si tu te retrouvais face à un évaluateur que tu avais déjà rencontré, comment réagiras-tu ? Dans quelle mesure serait-ce acceptable ?

5) Comment envisages-tu que les patients standardisés soient recrutés et formés partout en France ?

- Les facultés sont actuellement très hétérogènes dans leur programme d'ECOS et de patients standardisés. Certains débutent à peine, d'autres ont déjà une expérience avancée mais chacun fait les choses différemment : comment uniformiser nos pratiques pour que ce soit acceptable pour tous les étudiants dans l'idée du matching ?
- Quels seraient pour toi les points de vigilance à avoir concernant le recrutement et la formation des patients standardisés ?
- Et si tu te retrouvais face à un patient standardisé que tu avais déjà rencontré, comment réagiras-tu ? Dans quelle mesure serait-ce acceptable ?

6) Finalement, après toutes ces réflexions, comment envisages-tu en pratique l'intégration des ECOS dans la R2C dans toutes les facultés de France ?

- Relance : Lors de l'élaboration de la R2C, les syndicats ont avancé un argument d'« équité » pour intégrer les ECOS dans le matching. Comment fais-tu le lien entre l'équité qui est demandée et le dispositif tel que nous venons d'en discuter ?
- À quel point penses-tu qu'un tel dispositif puisse être acceptable dans le matching s'il est mis en place de cette façon ?
- Quel serait le bon compromis entre une standardisation nécessaire des ECOS partout en France et sa faisabilité pratique ?

7) Par rapport à ces ECOS, comment envisagerais-tu la formation pendant le 2^e cycle ?

- Changerais-tu ta façon d'apprendre, et si oui comment ?
- Quels éléments devrions-nous changer dans la formation actuelle du 2^e cycle ?
- Quels éléments te sembleraient nécessaires dans la formation pour t'aider à préparer ces ECOS ?

8) Nous avons abordé beaucoup de choses au cours de cet entretien, et je te remercie déjà énormément pour ta participation. Il me reste encore quelques petites questions d'ordre démographiques, mais avant, est-ce que tu souhaiterais rajouter quelque chose à tout ce que tu as

dit ? Ou est-ce qu'il y a quelque chose dont tu aimerais parler mais que nous n'avons pas abordé pendant cet entretien ?

9) Profil du participant :

- Quel est ton âge ?
- En quelle année de médecine es-tu ?
- Dans quelle Faculté de Médecine es-tu ?
- Es-tu adhérent à une association ou une corporation d'étudiants en médecine et si oui quelle est ton implication dedans ?
- Es-tu un représentant élu des étudiants en médecine dans ta Faculté ?

10) Remerciements

Références

1. 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-e1446.
2. Harden RM. What is an OSCE? *Med Teach* 1988;10:19-22.
3. Lewis KL, Bohnert CA, Gammon WL, Hölzer H, Lyman L, Smith C, *et al.* The Association of Standardized Patient Educators (ASPE) Standards of Best Practice (SOBP). *Adv Simul* 2017;2.
4. Epstein RM. Assessment in medical education. *N Engl J Med* 2007;356:387-96.
5. 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. NOR : ESR2018628A. Journal officiel de la République Française 2020(0221) [On-line]. Disponible sur : <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000042320018/2020-09-17/>.
6. Association nationale des étudiants en médecine de France (ANEMF). Le certificat de compétence clinique aux épreuves classantes nationales. 2018 [On-line]. Disponible sur : https://nanopdf.com/download/contribution-ccc_pdf/.
7. Centre national de ressources textuelles et lexicales (CNRTL). Équité: définition. 2012 [On-line]. Disponible sur : <https://www.cnrtl.fr/lexicographie/%C3%A9quit%C3%A9>.
8. Maréchal JP. L'éthique économique de John Rawls. *L'Économie politique* 2003;1:94-112.
9. Harden RM, Lilley P, Patricio M. The definitive guide to the OSCE: The objective structured clinical examination as a performance assessment. Edinburgh, New York: Elsevier, 2016.
10. Valentine N, Durning S, Shanahan EM, Schuwirth L. Fairness in human judgement in assessment: a hermeneutic literature review and conceptual framework. *Adv Health Sci Educ Theory Pract* 2021;26:713-38.
11. Jouquan J. L'évaluation des apprentissages des étudiants en formation médicale initiale. *Pédagogie Médicale* 2002;3:38-52.
12. Bertrand C, Dory V, Pelaccia T, Durand E, Vaux J, Charlin B, *et al.* Choisir un outil d'évaluation. In: Pelaccia T (sous la direction de). Comment (mieux) former et évaluer les étudiants en médecine et en sciences de la santé ? Louvain-la-Neuve: De Boeck Supérieur, 2017:357-70.

13. Froger O. Comment évaluer les étudiants avec des patients standardisés en situation simulée ? In : Comment évaluer les apprentissages dans l'enseignement supérieur professionnalisant ? Louvain-la-Neuve : De Boeck Supérieur, 2017:115-31.
14. 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-e1463.
15. Loye N, Fontaine S. S'instrumenter pour évaluer. *Pédagogie Médicale* 2018;19:95-107.
16. Van der Vleuten CP, Norman GR, De Graaff E. Pitfalls in the pursuit of objectivity: issues of reliability. *Med Educ* 1991;25:110-8.
17. Norman GR, Van der Vleuten CPM, De Graaff E. Pitfalls in the pursuit of objectivity: issues of validity, efficiency and acceptability. *Med Educ* 1991;25:119-26.
18. Van der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to programmes. *Med Educ* 2005;39:309-17.
19. Heeneman S, de Jong LH, Dawson LJ, Wilkinson TJ, Ryan A, Tait GR, *et al.* Ottawa 2020 consensus statement for programmatic assessment – 1. Agreement on the principles. *Med Teach* 2021;43:1139-48.
20. Wiggins G. 27 characteristics of authentic assessment. TeachThought. 2018 [On-line]. Disponible sur : <https://www.teachthought.com/pedagogy/authentic-assessment/>.
21. Tardif J. L'évaluation des compétences. Documenter le parcours de développement. Montréal : Chenelière Education, 2006.
22. García-Mayor S, Quemada-González C, León-Campos Á, Kaknani-Uttumchandani S, Gutiérrez-Rodríguez L, Del Mar Carmona-Segovia A, *et al.* Nursing students' perceptions on the use of clinical simulation in psychiatric and mental health nursing by means of objective structured clinical examination (OSCE). *Nurse Educ Today* 2021;100:104866.
23. Labaf A, Eftekhari H, Majlesi F, Anvari P, Sheybae-Moghaddam F, Jan D, Jamali A. Students' concerns about the pre-internship objective structured clinical examination in medical education. *Educ Health (Abingdon)* 2014;27:188-92.
24. Nasir AA, Yusuf AS, Abdur-Rahman LO, Babalola OM, Adeyeye AA, Popoola AA, *et al.* Medical students' perception of objective structured clinical examination: a feedback for process improvement. *J Surg Educ* 2014;71:701-706.
25. Purver J, Neville P, Fowler E. Between fairness and fear – Dental undergraduates' attitudes towards objective structured clinical examinations. *Eur J Dent Educ* 2019;23:323-31.
26. Vanlint A, Tam KL, Yu S, Visvanathan R. Evaluation of the introduction of the OSCE to the fifth-year Geriatric Medicine Teaching Programme. *Australas J Ageing* 2016;35:285-8.
27. Duncumb M, Cleland J. Student perceptions of a sequential objective structured clinical examination. *J R Coll Physicians Edinb* 2019;49:245-9.
28. Ataro G, Worku S, Asaminew T. Experience and challenges of Objective Structured Clinical Examination (OSCE): Perspective of students and examiners in a Clinical Department of Ethiopian University. *Ethiop J Health Sci* 2020;30:417-26.
29. Gedda M. Traduction française des lignes directrices COREQ pour l'écriture et la lecture des rapports de recherche qualitative. *Kinésithérapie, La Revue* 2015;15:50-4.
30. Ilgen JS, Ma IWY, 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. Miller C. Improving and enhancing performance in the affective domain of nursing students: insights from the literature for clinical educators. *Contemp Nurse* 2010;35: 2-17.
32. Bertrand C, Pelaccia T, Durand E, Revaux F, Leveque M, Delacour C, *et al.* Construire et utiliser les outils d'évaluation choisis. In: Pelaccia T (sous la direction de). Comment (mieux) former et évaluer les étudiants en médecine et en sciences de la santé ? Louvain-la-Neuve : De Boeck Supérieur, 2017:371-96.
33. Leclercq D, Nicaise J, Demeuse M. Docimologie critique : des difficultés de noter des copies et d'attribuer des notes aux élèves. In : Introduction aux théories et aux méthodes de la mesure en sciences psychologiques et en sciences de l'éducation. Liège : Les éditions de l'Université de Liège, 2004:273-92.
34. Cling JP, Cogneau D, Loup J, Naudet JD, Razafindrakoto M, Roubaud F. L'égalité des chances : un nouveau défi pour le développement ? *L'Economie politique* 2006;2:21-40.

Citation de l'article : Zumstein C, Philippon A-L, Lorenzo M. Comment assurer l'équité d'une évaluation classante par les examens cliniques objectifs et structurés (ECOS) ? Enquête d'opinion auprès des externes sur les conditions d'acceptabilité de la réforme du deuxième cycle des études médicales en France. *Pédagogie Médicale*, <https://doi.org/10.1051/pmed/2022038>