



HAL
open science

Association entre les performances cliniques des étudiants et leur réussite aux Épreuves Classantes Nationales informatisées : une étude de cohorte rétrospective monocentrique

L. Azoyan, Y. Lombardi, M.C. Renaud, A. Duguet, S. Georgan-Lavialle, F. Cohen-Aubart, G. Ibanez, O. Steichen

► To cite this version:

L. Azoyan, Y. Lombardi, M.C. Renaud, A. Duguet, S. Georgan-Lavialle, et al.. Association entre les performances cliniques des étudiants et leur réussite aux Épreuves Classantes Nationales informatisées : une étude de cohorte rétrospective monocentrique. *La Revue de Médecine Interne*, 2023, 44 (1), pp.5-11. 10.1016/j.revmed.2022.07.004 . hal-04070824

HAL Id: hal-04070824

<https://hal.science/hal-04070824>

Submitted on 16 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Titre : Association entre les performances cliniques des étudiants et leur réussite aux Épreuves Classantes Nationales informatisées : une étude de cohorte rétrospective monocentrique.

Auteurs :

L. Azoyan ¹, Y. Lombardi ², M.C. Renaud ³, A. Duguet ³, S. Georgin-Lavialle ¹, F. Cohen-Aubart ⁴, G. Ibanez ⁵, O. Steichen ¹

¹ Médecine interne, Hôpital Tenon, Paris

² Soins Intensifs Néphrologiques et Rein Aigu, Hôpital Tenon, Paris

³ Faculté de médecine, Sorbonne université, Paris

⁴ Médecine interne 2, Hôpital Pitié-Salpêtrière, Paris

⁵ Département de médecine générale, Sorbonne université, Paris

Mots clés : enseignement médical, externe, examen physique, recueil de l'anamnèse

Keywords : medical education, undergraduate, physical examination, history taking

Abstract :

Introduction. Les épreuves classantes nationales informatisées (ECNi) s'appuient actuellement sur une évaluation des connaissances des étudiants et non des compétences cliniques, pourtant déterminantes dans la performance professionnelle. Pour évaluer l'impact potentiel de cette absence de considération, nous avons évalué la concordance entre les rangs des étudiants aux épreuves facultaires cliniques et leur rang aux ECNi.

Méthodes. Nous avons inclus les étudiants de 6e année de médecine à Sorbonne Université des promotions 2017 à 2021 et récupéré leurs notes et rangs aux épreuves facultaires théoriques et cliniques et aux ECNi. Les épreuves théoriques sont au même format que les ECNi. Les épreuves cliniques comprennent une évaluation sémiologique au lit du malade et une évaluation des compétences relationnelles lors de deux entretiens cliniques simulés. Nous avons étudié la concordance entre les épreuves facultaires et les ECNi par le coefficient de corrélation intraclasse (CCI).

Résultats. Nous avons inclus 1806 étudiants. La concordance entre le rang aux épreuves facultaires théoriques et le rang aux ECNi est bonne (CCI 0,83 ; IC 95% 0,81-0,85) alors que la concordance entre le rang aux épreuves facultaires cliniques et le rang aux ECNi est médiocre (CCI 0,13 ; IC 95% 0,09-0,17).

Conclusion. L'accord entre le rang de classement aux ECNi et les compétences cliniques évaluées par une épreuve facultaire spécifique est actuellement très faible. Incorporer une évaluation pratique dans le classement national, comme le prévoit la réforme du deuxième cycle, motivera l'ensemble des étudiants à investir des efforts dans l'acquisition des compétences cliniques et valorisera les plus performants dans cette dimension.

English abstract :

Introduction. Before attending residency, 6th-year French medical students must validate a final examination including a practical clinical test in their faculty. However, the national ranking that determines their future specialty and region solely relies on a computerized knowledge test. Our goal was to investigate the association between the final faculty test and the national ranking test.

Methods. In our faculty, the final examination includes a computerized theoretical test (similar to the national one) and a practical test: a standardized evaluation of semiology skills at the bedside and a standardized assessment of relational skills with role plays. The agreements between the national test and faculty computerized and practical tests were analyzed by intraclass correlation coefficients (ICC).

Results. Data from 1806 students who underwent the three examinations from 2017 to 2021 were analyzed. There was a good agreement between the ranks in the faculty and national computerized

tests: ICC 0.83 (95% CI 0.81-0.85). By contrast, the agreement between the ranks in the faculty practical test and the national computerized test was poor: ICC 0.13 (95% CI 0.08-0.17). Results were stable over the years.

Conclusion. The agreement between the ranking of the current national test and the clinical skills assessed by a specific faculty test is minimal. This could relate to a true independence or to different levels of motivation to perform well. Indeed, the result of the national test is the most important one as it determines their career. Incorporating a clinical assessment into the national ranking test will motivate students to acquire clinical skills and value those who perform well this practical dimension.

Manuscrit

1. Introduction

Depuis 2015, les étudiants en 3^e année du Diplôme de Formation Approfondie en Sciences Médicales (DFASM3) passent les épreuves classantes nationales informatisées (ECNi) [1]. Ces épreuves visent une meilleure discrimination des étudiants sans modification du modèle docimologique, reposant toujours sur une restitution de connaissances ou leur application directe [2]. Les données de la littérature concernant l'association entre la performance des étudiants en médecine aux examens théoriques et leur performance aux épreuves cliniques sont hétérogènes tant les contextes éducatifs et les enjeux des examens sont différents [3–5]. En France, une étude portant sur un nombre limité d'étudiants avant 2015 ne montrait pas d'association entre leur savoir-faire en sémiologie clinique et leur performance aux épreuves classantes nationales manuscrites [6].

La réforme du deuxième cycle des études médicales de 2013 a imposé une évaluation facultaire des compétences cliniques pour valider l'entrée en troisième cycle [7–9]. Le format d'évaluation des compétences cliniques (examen de patients, épreuves de simulation) est laissé à la libre appréciation des facultés. Avec la réforme en cours pour la mise en place des épreuves dématérialisées nationales (EDN), le classement national des étudiants intégrera la performance à des examens cliniques objectifs structurés (ECOS) et la prise en compte du parcours personnel [10]. La question se pose de savoir dans quelle mesure l'adjonction de cette dimension pratique peut modifier le classement des étudiants et leur manière de se préparer.

Nous avons évalué la concordance entre le rang des étudiants aux épreuves facultaires cliniques et leur rang aux ECNi afin de déterminer dans quelle mesure le classement aux ECNi est concordant avec un classement sur les performances cliniques.

2. Méthodes

Schéma expérimental et participants

Cette étude observationnelle rétrospective monocentrique a inclus les étudiants de DFASM3 de Sorbonne Université issus de 5 promotions successives de 2017 à 2021, ayant passé la même année les épreuves facultaires de fin de second cycle, théoriques et cliniques, et les ECNi.

Épreuves

Les ECNi se déroulent en fin de DFASM3 et comprennent 18 dossiers progressifs de 15 questions, 120 questions isolées et 2 épreuves de lecture critique d'article de 15 questions pour un total de 10 800 points. Chaque question comporte 5 propositions de réponse, à choix unique ou multiple. La note totale a été rapportée sur 20.

Les différentes épreuves facultaires ont lieu au sein de la faculté 4 mois avant les ECNi. Les épreuves théoriques présentent un contenu proche des ECNi avec une taille réduite : 10 dossiers de 15 questions, 60 questions isolées pour une note totale rapportée sur 20.

Les épreuves cliniques comprennent 2 parties : une évaluation standardisée des compétences en sémiologie clinique au lit du malade et une évaluation standardisée des compétences relationnelles et communicationnelles lors d'une épreuve d'entretiens simulés, dérivée du modèle des mini entretiens multiples [11]. Les étudiants sont évalués par un jury composés de 2 médecins séniors (PU ou MCU-PH et CCA). Chaque jury évalue 3 étudiants en 4 heures (débriefing compris). La composition des jurys diffère d'une année à l'autre.

L'examen sémiologique est réalisé au sein des services avec des patients hospitalisés ayant donné leur consentement. En 2020, les étudiants s'étaient examinés les uns les autres compte tenu de l'épidémie de COVID-19 et certains jurys ont fait de même en 2021. Les étudiants sont évalués sur la bonne réalisation d'un item d'entretien (par exemple caractériser une dyspnée) et de 3 items d'examen physique (par exemple chercher une hépatomégalie et une splénomégalie) provenant d'une grille fournie aux examinateurs (d'autres exemples sont fournis dans l'**annexe 1**). Le dernier item sémiologique est laissé à la libre appréciation du jury. La performance à ces 5 items comptent pour 15 points et l'attitude globale au lit du malade est notée sur 5 points. L'évaluation porte uniquement sur la performance dans le recueil des signes et non sur leur interprétation.

L'épreuve d'entretiens simulés comporte deux jeux de rôles dans 3 familles de situations de communication avec les patients ou leurs proches : l'adhésion thérapeutique, le refus de soins et les annonces difficiles. Un exemple est donné dans l'**annexe 2**. Chaque étudiant se voit attribuer une

paire de situations différentes des autres étudiants de sa session. Il reçoit un énoncé par situation et dispose de 20 minutes pour préparer les deux entretiens. Toutes les connaissances médicales nécessaires sont rappelées dans l'énoncé. Les entretiens durent entre 5 et 7 minutes chacun. Un des membres du jury joue l'interlocuteur de l'étudiant. Chaque jeu de rôle permet d'évaluer certaines compétences attendues : communicateur, scientifique, responsable au plan éthique, acteur de santé publique, réflexif et collaborateur. Chaque compétence concernée est notée sur 2 ou 3 points moyennés sur les deux entretiens et le jury ajoute jusqu'à 4 points pour l'attitude globale pour un total de 20 points.

Données

Cette étude a été autorisée par le conseil des études de la faculté de médecine Sorbonne Université. La scolarité nous a fourni les données anonymisées des étudiants comprenant leur âge, leurs notes aux épreuves facultaires, leur jury et leurs notes aux ECNi.

Analyse statistique

Pour obtenir des classements uniformes, les résultats aux épreuves facultaires et aux ECNi ont chacun été transformés en rang au sein de la promotion avant les analyses de concordance par le coefficient de corrélation intraclasse (CCI). Le CCI quantifie l'accord observé entre deux mesures d'un phénomène, au-delà du simple accord attendu par l'effet du hasard. La forme du CCI utilisée ici est celle à deux facteurs à effets aléatoires, évaluateur unique et concordance absolue [12,13]. Nous avons fait le choix d'utiliser le CCI pour privilégier la concordance des rangs de classement en fonction de différentes modalités d'examens plutôt que le coefficient de corrélation de Pearson qui calcule dans quelle mesure ces deux variables tendent à changer ensemble. L'interprétation est la suivante : concordance faible pour un CCI inférieur à 0,5, moyenne entre 0,5 et 0,75, bonne entre 0,75 et 0,9, excellente au-dessus de 0,9. Les intervalles de confiance ont été estimés par rééchantillonnage (bootstrapping).

La capacité du rang aux épreuves facultaires à prédire le rang aux ECNi a été évalué par une décomposition de variance obtenue dans un modèle de régression linéaire multiple avec le rang intrapromotion aux ECNi comme variable à expliquer et les rangs aux épreuves facultaires comme variables explicatives. L'hypothèse de normalité des résidus a été vérifiée graphiquement. Un modèle mixte comportant en outre l'année de promotion et le jury en variables d'ajustement à effets aléatoires a permis de vérifier que les estimations restaient très proches lorsque ces deux effets étaient pris en compte.

Pour estimer l'impact du jury sur les notes obtenues aux épreuves facultaires cliniques, nous avons estimé la part de variance attribuable au jury rapporté à la variance totale des notes entre les étudiants au sein de modèles à effets mixtes prenant la note comme variable à expliquer et le jury comme variable explicative à effet aléatoire.

Les analyses ont été faites avec le logiciel R (4.0.2) à l'aide des packages irr et lme4 [14–16].

3. Résultats

Participants et notes obtenues aux différentes épreuves

Parmi les 1996 étudiants de DFASM3 au sein des promotions 2017 à 2021, 1806 avaient passé la même année les épreuves facultaires et les ECNi. L'âge moyen des étudiants était de 25,2 ans, les épreuves facultaires cliniques étaient très bien réussies avec une moyenne de 16,4 pour l'examen clinique et 15,7 pour les jeux de rôle. Les épreuves théoriques étaient également bien réussies avec une moyenne de 13,2 pour l'épreuve facultaire et 13,5 pour les ECNi. Les notes moyennes étaient stables d'une année à l'autre (**Tableau 1**).

Concordance entre performance aux épreuves facultaires et aux ECNi

Les rangs intrapromotion des étudiants aux ECNi selon leur rang intrapromotion aux épreuves facultaires théoriques et cliniques sont représentés sur la **figure 1**. La concordance entre le rang intrapromotion aux ECNi et le rang intrapromotion aux épreuves facultaires théoriques est bonne (CCI 0,83, IC95% 0,81-0,86) ; elle est très faible avec les épreuves facultaires cliniques (CCI 0,13, IC95 % 0,08-0,17). Les résultats sont similaires après différenciation des rangs aux épreuves de sémiologie au lit du malade et d'entretiens simulés. Ils sont stables sur les 5 années (**Tableau 2**).

Dans le modèle de régression linéaire multiple, le rang aux épreuves facultaires théoriques explique 68,7% de la variabilité du rang intrapromotion aux ECNi alors que le rang à l'examen clinique et aux jeux de rôles en expliquent moins de 0,1% de la variabilité.

Variance attribuable au jury

La variance attribuable au jury en fonction des années est représentée dans la **figure 2**. Elle est de 45,1% (IC 95% 39,8-49,1) en 2017, 45,3% (IC 95% 40,1-49,1) en 2018, 27,1% (IC 95% 19,2-32,7) en 2019, 31,0% (IC 95% 24,0 - 36,1) en 2020 et 37,1% (30,9- 41,8) en 2021.

4. Discussion

Résultats principaux

Nous avons observé une très bonne concordance entre les rangs aux épreuves facultaires théoriques et aux ECNi mais une concordance très faible entre les rangs aux épreuves facultaires cliniques et aux ECNi. Le poids du jury dans la note et donc le rang aux épreuves facultaires cliniques est important notamment lors des premières années de mise en place.

Forces et limites

Notre étude analyse de façon rigoureuse les résultats obtenus par un grand nombre d'étudiants à des épreuves dont les modalités sont bien définies. Toutefois, des limites découlent de son caractère monocentrique. En particulier, la signification des notes obtenues aux épreuves cliniques organisées dans notre faculté est incertaine, ce qui complique l'interprétation et l'extrapolation des résultats.

Ces épreuves cliniques ont une bonne validité de construction. En effet, les items de sémiologie et les points de performance ont été définis avec les département d'enseignement de chaque spécialité. Les jeux de rôle s'inspirent des mini-entrevues multiples [11]. La grille d'évaluation est conforme aux attendus d'apprentissage fixés pour nos étudiants par le département d'enseignement de relation de soins et conforme avec la réforme du deuxième cycle des études médicales de 2013. Les vignettes sont relues par un groupe de travail de ce département. Toutefois, la cohérence interne et la reproductibilité des épreuves n'a pas été formellement évaluée, et leur capacité à prédire la performance professionnelle future encore moins.

Contrairement aux ECOS qui peuvent mobiliser des connaissances théoriques, notre épreuve clinique est construite pour porter uniquement sur les compétences génériques. L'épreuve de sémiologie évalue la bonne réalisation des gestes sémiologiques mais pas l'interprétation du résultat ni le raisonnement clinique. L'épreuve d'entretiens simulés fournit les connaissances médicales nécessaires dans les énoncés. Toutefois, la performance clinique est connue pour être dépendante du contexte et on peut imaginer que des étudiants, par exemple, mobilisent moins bien leurs compétences cliniques dans une situation qui leur est peu familière, même si les connaissances nécessaires leur sont fournies.

Les études faites sur les ECOS et les mini entrevues multiples soulignent l'importance d'un nombre suffisant de stations diversifiées avec des jurys différents pour obtenir des évaluations reproductibles [11,17]. Dans notre épreuve, chaque étudiant est évalué par un jury unique sur cinq items d'examen clinique et deux jeux de rôle, ce qui la rend sensible à un effet jury, mis en évidence par nos analyses.

Interprétation

Plusieurs études ont souligné le niveau insuffisant des performances sémiologiques d'étudiants en médecine au cours ou en fin de second cycle [6,18–20]. Les erreurs fréquemment mises en évidence étaient une mauvaise application de la technique d'examen, des oublis et une mauvaise interprétation des données recueillies par l'examen. Cela contraste avec les bonnes notes obtenues dans nos épreuves, qui sont plus ciblées. En effet, elles ne portent que sur la bonne réalisation des gestes d'examen clinique, pas sur leur interprétation ; elles ne portent pas sur un examen clinique complet mais sur un nombre limité d'items spécifiquement demandés aux étudiants.

La très bonne concordance entre les rangs aux épreuves facultaires théoriques et aux ECNi montre la bonne reproductibilité de ce type d'épreuve de connaissances. La très faible concordance entre les rangs aux épreuves cliniques et aux ECNi peut avoir plusieurs explications. Il existe indiscutablement un effet jury dont nous discuterons plus loin de l'impact sur l'association du rang aux épreuves cliniques avec le rang aux ECNi. On peut faire l'hypothèse, peu probable, d'une dissociation complète entre le potentiel d'acquisition des connaissances et le potentiel d'acquisition des compétences cliniques, qui relèveraient de mécanismes et de dispositions personnelles différentes. On peut surtout suspecter une motivation différente des étudiants à acquérir les connaissances et les compétences cliniques, dans la mesure où ces dernières ne sont pas encore directement prises en compte dans le classement national qui détermine les possibilités de formation en troisième cycle.

L'association entre réussite aux épreuves cliniques et théoriques est variable dans la littérature. Certaines études mentionnent une corrélation faible voire négative. Aux États-Unis, une étude rapporte un coefficient de corrélation de 0,39 [21], une autre entre 0,00 et 0,14 [3]. En France, une étude trouve un coefficient de corrélation à -0,13 [6] et une autre à 0,18 [22]. Au Japon enfin, une étude trouve un coefficient de corrélation à 0,00 [5]. D'autres études avancent des coefficients de corrélation plus importants : 0,68 en Iran [23], 0,66 en Arabie Saoudite [4], 0,41 au Nigéria [24] et 0,45 au Koweït [25]. Cette variabilité peut avoir plusieurs sources : des contextes éducatifs différents selon les pays, des promotions d'âge différent, l'hétérogénéité de contenu et de format des examens théoriques et pratiques, ainsi que l'enjeu des examens. Le contenu des examens n'est pas systématiquement décrit. Les ECOS comportant des questions théoriques ont une meilleure corrélation avec les examens écrits [23,24]. Toutefois, une étude réalisée en France sur la mise en place d'un format réduit d'ECOS dans une autre faculté montre une association avec les résultats aux épreuves théoriques aussi faible que dans notre étude [22].

La subjectivité du jury joue nécessairement un rôle dans l'évaluation des compétences, avec des effets de contraste (l'évaluateur aura tendance à noter plus favorablement un second étudiant si le premier avait un niveau insuffisant), des effets de halo (la note de l'évaluateur ira dans le sens de sa

première impression) et une tendance personnelle à l'indulgence ou à la sévérité [26–28]. Il est admis que la cible pour la variabilité des notes attribuable à l'effet jury dans une épreuve de compétence (station d'ECOS, mini-entrevue) est inférieure à 30% [29,30]. Les valeurs observées pour nos épreuves sont supérieures à cette cible en 2017, 2018 et 2021. L'importance de l'effet jury dans nos épreuves peut en grande partie être expliquée par le fait que chaque étudiant est évalué par un unique jury. Les valeurs observées particulièrement importantes en 2017 et 2018 correspondent aux années de mise en place de ce format d'épreuve dans notre faculté. En 2019 et 2020, en revanche, la variabilité liée au jury se situe dans la cible, probablement par un effet d'acculturation des hospitalo-universitaires titulaires et de clarification des grilles d'évaluation. Toutefois, le CCI entre les rangs aux épreuves facultaires cliniques et aux ECNi n'est pas meilleur en 2019 et en 2020, ce qui suggère que l'effet jury n'est pas la principale explication de la très faible concordance entre l'évaluation de la performance clinique et celle des connaissances. En 2021, la variabilité augmente à nouveau, soulignant l'importance de maintenir des efforts pour assurer une bonne reproductibilité de l'évaluation pour ce type d'épreuve pratique.

Perspectives

La réforme du 2^e cycle et des modalités d'accès au 3^e cycle des études médicales valorisera les étudiants qui s'investissent dans l'apprentissage de la sémiologie et des compétences cliniques de manière générale. Nous supposons que les étudiants investiront alors plus d'efforts dans l'apprentissage de ces dimensions au cours du second cycle des études médicales. Les modalités doivent limiter l'effet jury pour que l'examen soit juste. Les mesures principales sont un nombre suffisant de stations et de jurys différents, la construction rigoureuse des stations et l'évaluation standardisée des étudiants par exemple via l'utilisation d'enregistrement vidéo [17,29,31]. La réforme est également l'occasion de mettre en place des innovations pédagogiques pour enseigner les compétences cliniques, que ce soit au sein ou en dehors des stages : mise en place de feedback formalisé [32, 33], de tutorat par des étudiants plus avancés [34] ou encore de référentiels vidéo pour l'apprentissage des techniques d'examen clinique [35]. Ces différentes solutions seront adaptées aux contraintes techniques dans chaque faculté. Leur efficacité devra être évaluée formellement. Il sera ensuite intéressant de réévaluer l'association entre la performance aux examens nationaux et la performance aux épreuves cliniques et théoriques après la mise en place des EDN.

5. Conclusion

L'accord entre le rang de classement aux ECNi et les compétences cliniques évaluées par une épreuve facultaire spécifique est actuellement très faible. L'incorporation d'une épreuve pratique

dans le futur classement national des étudiants motivera l'ensemble des étudiants à investir des efforts dans l'acquisition des compétences cliniques et valorisera les plus performants dans cette dimension.

Conflit d'intérêt : aucun

Financement : aucun

Remerciements : aucun

Références

- [1] Arrêté du 20 juillet 2015 relatif à l'organisation des épreuves classantes nationales anonymes donnant accès au troisième cycle des études médicales - Légifrance n.d. <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000030916532/2020-10-19/> (accessed October 19, 2020).
- [2] Rivière E, Quinton A, Dehail P. Analyse de la discrimination des notes aux ECN informatisées de 2016. *Rev Médecine Interne* 2019;40:286–90. <https://doi.org/10.1016/j.revmed.2018.10.386>.
- [3] Dong T, Saguil A, Artino A, Gilliland W, Waechter D, Lopreato J, et al. Relationship Between OSCE Scores and Other Typical Medical School Performance Indicators: A 5-Year Cohort Study. *Mil Med* 2012;177:44–6. <https://doi.org/10.7205/MILMED-D-12-00237>.
- [4] Al-osail AM, Al-osail EM, Al-ghamdi MA, Al-hawas AM, Al-bahussain AS, Aldajani AA. Correlations between the Objective Structured Clinical Examination Score and Written Examinations in Undergraduates. *IJLSR* 2014;2:13.
- [5] Komasa N, Terasaki F, Nakano T, Kawata R. Relationships between objective structured clinical examination, computer-based testing, and clinical clerkship performance in Japanese medical students. *PLoS ONE* 2020;15. <https://doi.org/10.1371/journal.pone.0230792>.
- [6] Steichen O, Georgin-Lavialle S, Gâteau G, Ranque B. Évaluation du savoir-faire en sémiologie clinique des étudiants en fin de deuxième cycle des études médicales. *Rev Médecine Interne* 2015;36:312–8. <https://doi.org/10.1016/j.revmed.2014.10.003>.
- [7] Arrêté du 8 avril 2013 relatif au régime des études en vue du premier et du deuxième cycle des études médicales - Légifrance n.d. <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000027343762/2020-10-19/> (accessed October 19, 2020).
- [8] 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. <https://doi.org/10.3109/0142159X.2013.818634>.
- [9] 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. <https://doi.org/10.3109/0142159X.2013.818635>.
- [10] Arrêté du 21 décembre 2021 relatif à l'organisation des épreuves nationales donnant accès au troisième cycle des études de médecine. n.d.
- [11] Eva KW, Rosenfeld J, Reiter HI, Norman GR. An admissions OSCE: the multiple mini-interview. *Med Educ* 2004;38:314–26. <https://doi.org/10.1046/j.1365-2923.2004.01776.x>.
- [12] Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 2016;15:155–63. <https://doi.org/10.1016/j.jcm.2016.02.012>.
- [13] Bruton A, Conway JH, Holgate ST. Reliability: What is it, and how is it measured? *Physiotherapy* 2000;86:94–9. [https://doi.org/10.1016/S0031-9406\(05\)61211-4](https://doi.org/10.1016/S0031-9406(05)61211-4).

- [14] R Core Team (2020). R: The R Project for Statistical Computing n.d. <https://www.r-project.org/> (accessed October 20, 2020).
- [15] Bates D, Maechler M, Bolker [aut B, cre, Walker S, Christensen RHB, et al. lme4: Linear Mixed-Effects Models using “Eigen” and S4. 2020.
- [16] Gamer M, Lemon J, Singh IFP. irr: Various Coefficients of Interrater Reliability and Agreement. 2019.
- [17] Roberts C, Newble D, Jolly B, Reed M, Hampton K. Assuring the quality of high-stakes undergraduate assessments of clinical competence. *Med Teach* 2006;28:535–43. <https://doi.org/10.1080/01421590600711187>.
- [18] Ramani S, Ring BN, Lowe R, Hunter D. A pilot study assessing knowledge of clinical signs and physical examination skills in incoming medicine residents. *J Grad Med Educ* 2010;2:232–5. <https://doi.org/10.4300/JGME-D-09-00107.1>.
- [19] Li Y, Li N, Han Q, He S, Bae RS, Liu Z, et al. Performance of Physical Examination Skills in Medical Students during Diagnostic Medicine Course in a University Hospital of Northwest China. *PLOS ONE* 2014;9:e109294. <https://doi.org/10.1371/journal.pone.0109294>.
- [20] Haring CM, Cools BM, van der Meer JW, Postma CT. Student performance of the general physical examination in internal medicine: an observational study. *BMC Med Educ* 2014;14:73. <https://doi.org/10.1186/1472-6920-14-73>.
- [21] Simon SR, Bui A, Day S, Berti D, Volkan K. The relationship between second-year medical students’ OSCE scores and USMLE Step 2 scores. *J Eval Clin Pract* 2007;13:901–5. <https://doi.org/10.1111/j.1365-2753.2006.00768.x>.
- [22] Matet A, Fournel L, Gaillard F, Amar L, Arlet J-B, Baron S, et al. Impact of integrating objective structured clinical examination into academic student assessment: Large-scale experience in a French medical school. *PloS One* 2021;16:e0245439. <https://doi.org/10.1371/journal.pone.0245439>.
- [23] Eftekhari H, Labaf A, Anvari P, Jamali A, Sheybaee-Moghaddam F. Association of the pre-internship objective structured clinical examination in final year medical students with comprehensive written examinations. *Med Educ Online* 2012;17:15958. <https://doi.org/10.3402/meo.v17i0.15958>.
- [24] Tijani K, Giwa S, Abiola A, Adesanya A, Nwawolo C, Hassan J. A comparison of the objective structured clinical examination and the traditional oral clinical examination in a Nigerian university. *J West Afr Coll Surg* 2017;7:59–72.
- [25] Al Rushood M, Al-Eisa A. Factors predicting students’ performance in the final pediatrics OSCE. *PloS One* 2020;15:e0236484. <https://doi.org/10.1371/journal.pone.0236484>.
- [26] Yeates P, Moreau M, Eva K. Are Examiners’ Judgments in OSCE-Style Assessments Influenced by Contrast Effects? *Acad Med* 2015;90:975–80. <https://doi.org/10.1097/ACM.0000000000000650>.
- [27] Fuller R, Homer M, Pell G, Hallam J. Managing extremes of assessor judgment within the OSCE. *Med Teach* 2016;39:1–9. <https://doi.org/10.1080/0142159X.2016.1230189>.
- [28] Chong L, Taylor S, Haywood M, Adelstein B-A, Shulruf B. Examiner seniority and experience are associated with bias when scoring communication, but not examination, skills in objective structured clinical examinations in Australia. *J Educ Eval Health Prof* 2018;15:17. <https://doi.org/10.3352/jeehp.2018.15.17>.
- [29] Pell G, Fuller R, Homer M, Roberts T. How to measure the quality of the OSCE: A review of metrics – AMEE guide no. 49. *Med Teach* 2010;32:802–11. <https://doi.org/10.3109/0142159X.2010.507716>.
- [30] Yazbeck Karam V, Park YS, Tekian A, Youssef N. Evaluating the validity evidence of an OSCE: results from a new medical school. *BMC Med Educ* 2018;18:313. <https://doi.org/10.1186/s12909-018-1421-x>.
- [31] Bellier A, Chaffanjon P, Krupat E, Francois P, Labarère J. Cross-cultural adaptation of the 4-Habits Coding Scheme into French to assess physician communication skills. *PLOS ONE* 2020;15:e0230672. <https://doi.org/10.1371/journal.pone.0230672>.

- [32] Martinsen SSS, Espeland T, Berg EAR, Samstad E, Lillebo B, Slørdahl TS. Examining the educational impact of the mini-CEX: a randomised controlled study. *BMC Med Educ* 2021;21:228. <https://doi.org/10.1186/s12909-021-02670-3>.
- [33] Singhal A, Yadav AK, Anup H, Anjali B, Singh S. To evaluate the effect of small-group demonstration-based training module in improving clinical skills in undergraduate medical students. *Med J Armed Forces India* 2022;78:32–5. <https://doi.org/10.1016/j.mjafi.2020.01.003>.
- [34] Gripay B, André T, De Laval M, Peneau B, Secourgeon A, Lerolle N, et al. Benefits of semiology taught using near-peer tutoring are sustainable. *BMC Med Educ* 2022;22:26. <https://doi.org/10.1186/s12909-021-03086-9>.
- [35] Bedside Medicine Exam Skills. *Stanf Med* 25 n.d. <https://stanfordmedicine25.stanford.edu/> (accessed May 3, 2022).

Tableaux :

Année	Nombre	Age	Épreuve facultaire théorique	Épreuve facultaire clinique			ECNi
				Moyenne	Examen clinique	Jeux de rôle	
Total	1806	25,2 (2,5)	13,2 (1,4)	16,0 (1,8)	16,4 (2,1)	15,7 (2,3)	13,5 (1,4)
2017	355	25,0 (2,5)	12,9 (1,3)	15,9 (1,9)	16,1 (2,3)	15,8 (2,3)	13,3 (1,3)
2018	351	24,9 (2,3)	12,9 (1,3)	16,0 (1,9)	16,4 (2,2)	15,7 (2,4)	13,1 (1,4)
2019	363	25,0 (2,5)	13,5 (1,4)	16,0 (1,8)	16,3 (2,0)	15,7 (2,2)	13,2 (1,2)
2020	372	25,4 (2,1)	13,5 (1,3)	16,1 (2,3)	16,3 (2,1)	15,9 (2,3)	13,8 (1,3)
2021	365	25,6 (2,8)	13,5 (1,4)	16,1 (2,2)	16,6 (2,0)	15,6 (2,2)	14,2 (1,5)

Tableau 1

Année	Épreuve facultaire théorique	Épreuve facultaire clinique		
		Moyenne	Examen clinique	Jeux de rôle

Total	0,83 (0,81 - 0,85)	0,13 (0,08 - 0,17)	0,13 (0,09 - 0,17)	0,08 (0,03 - 0,12)
2017	0,88 (0,85 - 0,90)	0,16 (0,06 - 0,26)	0,11 (0,00 - 0,21)	0,15 (0,05 - 0,25)
2018	0,84 (0,81 - 0,87)	0,12 (0,02 - 0,22)	0,12 (0,02 - 0,22)	0,08 (-0,02 - 0,18)
2019	0,81 (0,76 - 0,86)	0,06 (-0,04 - 0,16)	0,05 (-0,05 - 0,15)	0,04 (-0,06 - 0,14)
2020	0,82 (0,76 - 0,86)	0,14 (0,04 - 0,24)	0,19 (0,10 - 0,29)	0,04 (-0,05 - 0,15)
2021	0,80 (0,74 - 0,86)	0,15 (0,05 - 0,25)	0,18 (0,07 - 0,27)	0,08 (-0,02 - 0,19)

Tableau 2

Légendes des tableaux et figures :

Tableau 1 : caractéristiques des étudiants et notes rapportées sur 20 aux différentes épreuves. Les valeurs moyennes sont données avec leur écarts types entre parenthèses.

Tableau 2 : coefficient de corrélation intraclasse entre le rang aux ECNi au sein de la promotion et le rang aux différentes épreuves facultaires. Intervalles de confiance à 95% sont entre parenthèses.

Figure 1 : rangs intrapromotion aux ECNi en fonction des rangs intrapromotion aux épreuves facultaires théoriques (A) et cliniques (B). La pente de régression linéaire est représentée sur chacun des graphiques.

Figure 2 : pourcentage de variance attribuable au jury en fonction des années. Les barres d'erreurs représentent les intervalles de confiance. L'objectif de limite supérieure de variance attribuable à l'effet jury (30%) est représenté par la ligne horizontale.

Annexes :

Sémiologie cardiovasculaire	
Faire l'interrogatoire d'une douleur thoracique	Type : oppression, piquûre, brûlure... Localisation et irradiations Durée et mode d'installation, rythme et périodicité, épisodes antérieurs Intensité, facteurs modulant la douleur (déclenchant, aggravant, améliorant), évolution depuis le début Signes fonctionnels associés (cardiaque, respiratoires, digestifs) Antécédents et facteurs de risque cardiovasculaire et respiratoires, traitements
Chercher une hypotension orthostatique	Pression artérielle couchée depuis 10 minutes et debout immédiatement, puis à 1 et 3 minutes (diminution de 20 mmHg pour la PAS, 10 pour la PAD) ; noter la variation de FC (attendue > 20/min si hypotension non neurogène)
Auscouter les 4 principaux foyers cardiaque et les zones d'irradiation, décrire un souffle	Temps : proto/méso/télé/holo- -systolique/diastolique Timbre Localisation de l'intensité maximale : Aortique : 2e espace intercostal droit, parasternal Pulmonaire : 2e espace intercostal gauche, parasternal Tricuspide : 4e espace intercostal gauche, zone xyphoïdienne Mitral : 5e espace intercostal gauche, ligne médioclaviculaire Irradiation : carotides, axillaire Intensité : cotation sur 6
Chercher un syndrome d'insuffisance ventriculaire gauche	Orthopnée et dyspnée d'effort Tachycardie à l'auscultation cardiaque (+/- galop) Crépitants symétriques, prédominant aux bases Hypoperfusion périphérique (extrémités froides, cyanose des ongles, temps de recoloration cutanée > 3 secondes) Signes de choc (marbrures, hypotension)
Chercher un syndrome d'insuffisance ventriculaire droite	Hépatalgies d'effort Jugulaire inspectée au bon endroit, malade proclive à 45° RHJ cherché en maintenant la pression pendant au moins 10 secondes, respiration non bloquée Œdèmes de surcharge avec signe du godet cherché en déclive (cheville si assis, lombes si allongé)
Palper les pouls périphériques	Pédieux, tibiaux postérieurs (derrière la malléole interne), poplités, fémoraux, radiaux, temporaux

Annexe 1 : exemple d'items d'entretien et d'examen physique de sémiologie cardiovasculaire.

Vous êtes dermatologue. Vous recevez en consultation d'urgence une patiente de 30 ans suivie pour une dermatite atopique sévère, actuellement en poussée. Elle était traitée depuis 9 mois par dupilumab (anticorps monoclonal antirécepteur de l'interleukine 4). La patiente a arrêté ce traitement depuis environ 2 mois, pourtant efficace et bien toléré. Elle présente une poussée de dermatite atopique touchant plus de 50% de surface corporelle dont le visage et les mains, sans signe de surinfection.

Vous savez que la dermatite atopique est une maladie chronique cutanée débutant généralement dans l'enfance, avec des poussées fortement liées au stress. Le dupilumab est un traitement systémique approuvé dans cette indication, en injections sous-cutanées auto-administrées par le patient tous les 15 jours. Les effets indésirables les plus fréquents sont les réactions au site d'injection et les infections (conjonctivite, blépharite, herpès buccal).

Vous entamez la discussion en cherchant à comprendre les raisons qui ont conduit la patiente à interrompre le traitement, alors que la dermatite atopique était bien contrôlée et que cette maladie chronique, que la patiente connaît bien, nécessite un traitement au long cours pour contrôler la survenue de poussées cutanées pouvant se compliquer de surinfections.

Annexe 2 : exemple de situation proposée lors de l'épreuve de jeux de rôle.