



**HAL**  
open science

# Corrélation entre rang de classement aux ECNi de 2018 et performance des internes de second semestre de phase socle

Typhaine Guerard

## ► To cite this version:

Typhaine Guerard. Corrélation entre rang de classement aux ECNi de 2018 et performance des internes de second semestre de phase socle. Médecine humaine et pathologie. 2020. dumas-02992981

**HAL Id: dumas-02992981**

**<https://dumas.ccsd.cnrs.fr/dumas-02992981>**

Submitted on 6 Nov 2020

**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.



UNIVERSITÉ DE NICE SOPHIA ANTIPOLIS

FACULTÉ DE MÉDECINE

**Corrélation entre rang de classement aux ECNi de 2018 et  
performance des internes de second semestre de phase socle.**

**THÈSE DE DOCTORAT EN MÉDECINE**

en vue de l'obtention du diplôme d'Etat

par **GUERARD Typhaine**

née le 27/12/1991 à Vernon

Présentée et soutenue publiquement à Nice le 12 Octobre 2020

Président du jury: Monsieur le Professeur Fournier Jean-Paul

Assesseurs : Monsieur le Professeur Cluzeau Thomas

Monsieur le Professeur Staccini Pascal

Monsieur le Docteur Ganassi Alain

Liste des enseignants au 1er septembre 2020 à la Faculté de Médecine de Nice

**Doyen**

**Pr. BAQUÉ Patrick**

**Vice-doyens**

**Pédagogie  
Recherche  
Etudiants**

**Pr. ALUNNI Véronique  
Pr. DELLAMONICA Jean  
M. JOUAN Robin**

**Chargé de mission projet Campus**

**Pr. PAQUIS Philippe**

Conservateur de la bibliothèque

Mme AMSELLE Danièle

Directrice administrative des services

Mme CALLEA Isabelle

Doyens Honoraires

M. RAMPAL Patrick  
M. BENCHIMOL Daniel

Liste des enseignants au 1er septembre 2020 à la Faculté de Médecine de Nice

PROFESSEURS CLASSE EXCEPTIONNELLE

M.	BAQUÉ Patrick	Anatomie - Chirurgie Générale (42.01)
M.	BERNARDIN Gilles	Réanimation Médicale (48.02)
Mme	BLANC-PEDEUTOUR Florence	Cancérologie – Génétique (47.02)
M.	BOILEAU Pascal	Chirurgie Orthopédique et Traumatologique (50.02)
M.	DARCOURT Jacques	Biophysique et Médecine Nucléaire (43.01)
M.	DRICI Milou-Daniel	Pharmacologie Clinique (48.03)
M.	ESNAULT Vincent	Néphrologie (52-03)
M.	GILSON Éric	Biologie Cellulaire (44.03)
M.	GUGENHEIM Jean	Chirurgie Digestive (52.02)
M.	HASSEN KHODJA Reda	Chirurgie Vasculaire (51.04)
M.	HÉBUTERNE Xavier	Nutrition (44.04)
M.	HOFMAN Paul	Anatomie et Cytologie Pathologiques (42.03)
Mme	ICHAÏ Carole	Anesthésiologie et Réanimation Chirurgicale (48.01)
M.	LACOUR Jean-Philippe	Dermato-Vénéréologie (50.03)
M.	LEFTHERIOTIS Georges	Chirurgie vasculaire ; médecine vasculaire (51.04)
M.	MARQUETTE Charles-Hugo	Pneumologie (51.01)
M.	MARTY Pierre	Parasitologie et Mycologie (45.02)
M.	MICHIELS Jean-François	Anatomie et Cytologie Pathologiques (42.03)
M.	MOUNIER Nicolas	Cancérologie, Radiothérapie (47.02)
M.	MOUROUX Jérôme	Chirurgie Thoracique et Cardiovasculaire (51.03)
M.	PADOVANI Bernard	Radiologie et Imagerie Médicale (43.02)
Mme	PAQUIS Véronique	Génétique (47.04)
M.	PAQUIS Philippe	Neurochirurgie (49.02)
M.	PRADIER Christian	Épidémiologie, Économie de la Santé et Prévention (46.01)
M.	QUATREHOMME Gérald	Médecine Légale et Droit de la Santé (46.03)
M.	RAUCOULES-AIMÉ Marc	Anesthésie et Réanimation Chirurgicale (48.01)
M.	ROBERT Philippe	Psychiatrie d'Adultes (49.03)
M.	SCHNEIDER Stéphane	Nutrition (44.04)
M.	TRAN Albert	Hépatogastro-entérologie (52.01)

Liste des enseignants au 1er septembre 2020 à la Faculté de Médecine de Nice

PROFESSEURS PREMIERE CLASSE

Mme	ASKENAZY-GITTARD Florence	Pédopsychiatrie (49.04)
M.	BARRANGER Emmanuel	Gynécologie Obstétrique (54.03)
M.	BÉRARD Étienne	Pédiatrie (54.01)
M.	BONGAIN André	Gynécologie-Obstétrique (54.03)
Mme	BREUIL Véronique	Rhumatologie (50.01)
M.	CASTILLO Laurent	O.R.L. (55.01)
M.	CHEVALLIER Patrick	Radiologie et Imagerie Médicale (43.02)
M.	DE PERETTI Fernand	Anatomie-Chirurgie Orthopédique (42.01)
M.	FERRARI Émile	Cardiologie (51.02)
M.	FERRERO Jean-Marc	Cancérologie ; Radiothérapie (47.02)
M.	FONTAINE Denys	Neurochirurgie (49.02)
M.	GUÉRIN Olivier	Méd. In ; Gériatrie (53.01)
M.	HANNOUN-LEVI Jean-Michel	Cancérologie ; Radiothérapie (47.02)
M	JEAN BAPTISTE Elixène	Chirurgie vasculaire (51.04)
M.	LEVRAUT Jacques	Médecine d'urgence (48.05)
M.	LONJON Michel	Neurochirurgie (49.02)
M.	PASSERON Thierry	Dermato-Vénérologie (50-03)
M.	PICHE Thierry	Gastro-entérologie (52.01)
Mme	RAYNAUD Dominique	Hématologie (47.01)
M.	ROUX Christian	rhumatologie (50.01)
M.	ROSENTHAL Éric	Médecine Interne (53.01)
M.	STACCINI Pascal	Biostatistiques et Informatique Médicale (46.04)
M.	THOMAS Pierre	Neurologie (49.01)
M.	TROJANI Christophe	Chirurgie Orthopédique et Traumatologique (50.02)

Liste des enseignants au 1er septembre 2020 à la Faculté de Médecine de Nice

PROFESSEURS DEUXIEME CLASSE

Mme	ALUNNI Véronique	Médecine Légale et Droit de la Santé (46.03)
M.	ANTY Rodolphe	Gastro-entérologie (52.01)
M.	BAHADORAN Philippe	Cytologie et Histologie (42.02)
Mme	BAILLIF Stéphanie	Ophthalmologie (55.02)
Mme	BANNWARTH Sylvie	Génétique (47.04)
M.	BENIZRI Emmanuel	Chirurgie Générale (53.02)
M.	BENOIT Michel	Psychiatrie (49.03)
M.	BERTHET Jean-Philippe	Chirurgie Thoracique (51-03)
M.	BOZEC Alexandre	ORL- Cancérologie (47.02)
M.	BREAUD Jean	Chirurgie Infantile (54-02)
M.	BRONSARD Nicolas	Anatomie Chirurgie Orthopédique et Traumatologique (42.01)
Mme	BUREL-VANDENBOS Fanny	Anatomie et Cytologie pathologiques (42.03)
M.	CHEVALIER Nicolas	Endocrinologie, Diabète et Maladies Métaboliques (54.04)
Mme	CHINETTI Giulia	Biochimie-Biologie Moléculaire (44.01)
M.	CLUZEAU Thomas	Hématologie (47.01)
M.	DELLAMONICA Jean	réanimation médicale (48.02)
M.	DELOTTE Jérôme	Gynécologie-obstétrique (54.03)
Mme	ESTRAN-POMARES Christelle	Parasitologie et mycologie (45.02)
M	FAVRE Guillaume	Néphrologie (44-02)
M.	FOURNIER Jean-Paul	Thérapeutique (48-04)
Mme	GIORDANENGO Valérie	Bactériologie-Virologie (45.01)
Mme	GIOVANNINI-CHAMI Lisa	Pédiatrie (54.01)
M.	IANNELLI Antonio	Chirurgie Digestive (52.02)
M.	ILIE Marius	Anatomie et Cytologie pathologiques (42.03)
M.	ORBAN Jean-Christophe	Anesthésiologie-réanimation ; Médecine d'urgence (48.01)
M.	ROHRLICH Pierre	Pédiatrie (54.01)
M.	RUIMY Raymond	Bactériologie-virologie (45.01)
Mme	SACCONI Sabrina	Neurologie (49.01)
Mme	SEITZ-POLSKI barbara	Immunologie (47.03)
M.	VANBIERVLIET Geoffroy	Gastro-entérologie (52.01)

Liste des enseignants au 1er septembre 2020 à la Faculté de Médecine de Nice

MAITRES DE CONFÉRENCES DES UNIVERSITÉS - PRATICIENS HOSPITALIERS

M.	AMBROSETTI Damien	Cytologie et Histologie (42.02)
Mme	BERNARD-POMIER Ghislaine	Immunologie (47.03)
M.	CAMUZARD Olivier	Chirurgie Plastique (50-04)
Mme	CONTENTI-LIPRANDI Julie	Médecine d'urgence ( 48-04)
M.	DOGLIO Alain	Bactériologie-Virologie (45.01)
M.	DOYEN Jérôme	Radiothérapie (47.02)
M.	FOSSE Thierry	Bactériologie-Virologie-Hygiène (45.01)
M.	GARRAFFO Rodolphe	Pharmacologie Fondamentale (48.03)
Mme	HINAULT Charlotte	Biochimie et biologie moléculaire (44.01)
M.	HUMBERT Olivier	Biophysique et Médecine Nucléaire (43.01)
Mme	LAMY Brigitte	Bactériologie-virologie ( 45.01)
Mme	LONG-MIRA Elodie	Cytologie et Histologie (42.02)
M.	LOTTE Romain	Bactériologie-virologie ; Hygiène hospitalière (45.01)
Mme	MAGNIÉ Marie-Noëlle	Physiologie (44.02)
M.	MASSALOU Damien	Chirurgie Viscérale ( 52-02)
Mme	MOCERI Pamela	Cardiologie (51.02)
M.	MONTAUDIE Henri	Dermatologie (50.03)
Mme	MUSSO-LASSALLE Sandra	Anatomie et Cytologie pathologiques (42.03)
M.	NAÏMI Mourad	Biochimie et Biologie moléculaire (44.01)
M.	SAVOLDELLI Charles	Chirurgie maxillo-faciale et stomatologie (55.03)
M.	SQUARA Fabien	Cardiologie (51.02)
M.	TESTA Jean	Épidémiologie Économie de la Santé et Prévention (46.01)
Mme	THUMMLER Susanne	Pédopsychiatrie ( 49-04)
M.	TOULON Pierre	Hématologie et Transfusion (47.01)
M.	TRAN Antoine	Pédiatrie (54.01)



Liste des enseignants au 1er septembre 2020 à la Faculté de Médecine de Nice

PROFESSEUR DES UNIVERSITÉS

M. DARMON David Médecine Générale (53.03)

MAITRE DE CONFÉRENCES DES UNIVERSITÉS

Mme GROS Auriane Orthophonie (69)

PROFESSEURS AGRÉGÉS

Mme LANDI Rebecca Anglais

PRATICIEN HOSPITALIER UNIVERSITAIRE

M. DURAND Matthieu Urologie (52.04)

M. SICARD Antoine Néphrologie (52-03)

PROFESSEURS ASSOCIÉS

M. GARDON Gilles Médecine Générale (53.03)

Mme MONNIER Brigitte Médecine Générale (53.03)

MAITRES DE CONFÉRENCES ASSOCIÉS

Mme CASTA Céline Médecine Générale (53.03)

M. GASPERINI Fabrice Médecine Générale (53.03)

M. HOGU Nicolas Médecine Générale (53.03)



Liste des enseignants au 1er septembre 2020 à la Faculté de Médecine de Nice

Constitution du jury en qualité de 4ème membre

Professeurs Honoraires

M. AMIEL Jean	M. GÉRARD Jean-Pierre
M. ALBERTINI Marc	M. GIBELIN Pierre
M. BALAS Daniel	M. GILLET Jean-Yves
M. BATT Michel	M. GRELLIER Patrick
M. BLAIVE Bruno	M. GRIMAUD Dominique
M. BOQUET Patrice	M. HOFLIGER Philippe
M. BOURGEON André	M. JOURDAN Jacques
M. BOUTTÉ Patrick	M. LAMBERT Jean-Claude
M. BRUNETON Jean-Noël	M. LAZDUNSKI Michel
Mme BUSSIERE Françoise	M. LEFEBVRE Jean-Claude
M. CAMOUS Jean-Pierre	M. LE FICHOUX Yves
M. CANIVET Bertrand	Mme LEBRETON Elisabeth
M. CASSUTO Jill-patrice	M. MARIANI Roger
M. CHATEL Marcel	M. MASSEYEFF René
M. COUSSEMENT Alain	M. MATTEI Mathieu
Mme CRENESSE Dominique	M. MOUIEL Jean
M. DARCOURT Guy	Mme MYQUEL Martine
M. DELLAMONICA Pierre	M. ORTONNE Jean-Paul
M. DELMONT Jean	M. PRINGUEY Dominique
M. DEMARD François	M. SANTINI Joseph
M. DESNUELLE Claude	M. SAUTRON Jean Baptiste
M. DOLISI Claude	M. SCHNEIDER Maurice
Mme EULLER-ZIEGLER Liana	M. THYSS Antoine
M. FENICHEL Patrick	M. TOUBOL Jacques
M. FUZIBET Jean-Gabriel	M. TRAN Dinh Khiem
M. FRANCO Alain	M. VAN OBBERGHEN Emmanuel
M. FREYCHET Pierre	
M. GASTAUD Pierre	

M.C.U. Honoraires

M. ARNOLD Jacques	M. GIUDICELLI Jean
M. BASTERIS Bernard	M. MAGNÉ Jacques
M. BENOLIEL José	Mme MEMRAN Nadine
Mlle CHICHMANIAN Rose-Marie	M. MENGUAL Raymond
Mme DONZEAU Michèle	M. PHILIP Patrick
M. EMILIOZZI Roméo	M. POIRÉE Jean-Claude
M. FRANKEN Philippe	Mme ROURE Marie-Claire
M. GASTAUD Marcel	

Liste des enseignants au 1er septembre 2020 à la Faculté de Médecine de Nice

PROFESSEURS CONVENTIONNÉS DE L'UNIVERSITÉ

M.	BERTRAND François	Médecine Interne
M.	BROCKER Patrice	Médecine Interne Option Gériatrie
M.	CHEVALLIER Daniel	Urologie
Mme	FOURNIER-MEHOUAS Manuella	Médecine Physique et Réadaptation
M.	JAMBOU Patrick	Coordination prélèvements d'organes
M.	LEBOEUF Mathieu	gynécologie- obstétrique
Mme	NADEAU Geneviève	uro-gynécologie
M.	ODIN Guillaume	Chirurgie maxilo-faciale
M.	PEYRADE Frédéric	Onco-Hématologie
M.	PICCARD Bertrand	Psychiatrie
M.	QUARANTA Jean-François	Santé Publique

# RESUME

## INTRODUCTION

Les données sont discordantes sur le caractère prédictif des examens facultaires et/ou terminaux sur la performance clinique des futurs internes. Les Epreuves Classantes Nationales informatisées (ECNi) sont passées en ligne par tous les étudiants français en fin d'études et conditionnent leur internat (choix de spécialité et CHU) au rang de classement. Elles comportent 18 dossiers cliniques progressifs (DCP), 120 questions isolées de type QCM (QI) et l'analyse critique de deux articles scientifiques (LCA). Il n'existe aucune donnée sur leur éventuel caractère prédictif de la performance clinique des futurs internes. Notre objectif était de chercher une possible corrélation entre le rang de classement aux ECNi, les scores aux examens facultaires et la performance clinique des internes en phase socle.

## MATERIEL ET METHODES

L'étude porte sur la promotion 2018 des internes en poste au CHU de Nice et ceux formés à Nice mais étant en poste dans d'autres CHU, en spécialités cliniques. Avec leur accord, nous avons évalué leur performance lors du second semestre de leur phase socle au moyen d'un questionnaire reprenant neuf des treize compétences de l'*Entrustable Professional Activities* (EPA) selon la méthodologie utilisée par Lee en 2018 : recueil d'anamnèse, examen physique, présentation orale du cas, connaissances théoriques, jugement clinique, relations avec le patient, professionnalisme et évaluation globale. Chaque compétence a été évaluée au moyen d'une échelle de Likert avec sept points. Le score global a été comparé au rang de classement ECNi. Pour tous les étudiants, nous avons comparé le score global aux différents composants de l'ECNi : dossiers cliniques progressifs, questions isolées et lecture critique d'article scientifique. Nous avons également comparé, pour les étudiants formés à Nice les scores d'examens transversaux (Examen Cumulatif Transversal – ECT) de fin d'année en DFASM 1 et DFASM 2 et des trois examens transversaux de DFASM 3. Le test de Pearson a été utilisé avec la significativité définie par  $p < 0,05$ .

## RESULTATS

Les données complètes étaient disponibles pour 57,4% des étudiants sélectionnés. Il existait une corrélation significative entre le rang de classement et le score de performance globale. Pour les étudiants niçois, le score à l'ECT de DFASM 2 était significativement corrélé au score global de performance, de même que la somme des scores à l'ECT de DFASM 1 et DFASM 2 (le score à l'ECT de DFASM 1 n'était pas corrélé significativement). Deux des trois examens transversaux de DFAM3 étaient significativement corrélés au score global de performance. Pour les composantes de l'ECNi : Les scores aux DCP et aux QI étaient significativement corrélés au score global de performance, au contraire de la LCA.

## CONCLUSION ET DISCUSSION

Pour l'année 2018, les ECNi étaient prédictives de la performance clinique des futurs internes, de même que, partiellement, les examens facultaires transversaux de DFASM 2 à DFASM 3.

Notre étude comporte néanmoins plusieurs limites : elle est mono-centrique et pas forcément extrapolable. Nous ne disposons des données que pour 57,4% des internes concernés. Nous n'avons pas mesuré les treize compétences de l'EPA, mais seulement neuf d'entre elles, en suivant une méthodologie similaire à celle de Lee.

# ABSTRACT

## INTRODUCTION

The data are inconsistent on the predictive value of the medical school assessments on the clinical performance of future residents. The ECNi is taken online by all French students at the end of their studies and condition their residency (choice of their specialty and University Hospital based on their ranking). This assessment includes 18 progressive clinical records (DCP), 120 isolated multiple choice questions (QI) and the critical analysis of two scientific articles (LCA). There is no data on its potential predictive value on the clinical performance of future residents. Our objective was to look for a possible correlation between ECNi rankings, faculty assessments scores and the clinical performance of residents in their first year.

## MATERIALS AND METHODS

The study concerns the 2018 promotion of residents working at the Nice University Hospital and those trained in Nice but working in other University Hospitals, in clinical specialties. With their agreement, we evaluated their performance during the second half of their first year of residency by means of a questionnaire covering nine of the thirteen competencies of the Entrustable Professional Activities (EPA) according to the methodology used by Lee in 2018: history taking, physical examination, oral case presentation, theoretical knowledge, clinical judgment, patient relations, professionalism and overall assessment. Each competency was assessed using a 7-point Likert scale. The overall score was compared to the ECNi ranking. For all students, the overall score was compared to the different components of the ECNi: progressive clinical records, isolated questions and critical reading of scientific articles. We also compared, for the students trained in Nice, the scores of end-of-term exams (cumulative cross-curricular examination - ECT) in DFASM 1 and DFASM 2 and of the three cross-curricular exams of DFASM 3. The Pearson's test was used with significance defined by  $p < 0.05$ .

## RESULTS

Complete data was available for 57.4% of the selected students. There was a significant correlation between the rankings of the ECNi and the overall residency performance score. For students in Nice, the DFASM 2 ECT score was significantly correlated with the overall resident performance score, as was the sum of the DFASM 1 and DFASM 2 ECT scores (the DFASM 1 ECT score was not significantly correlated). Two of the three DFAM3 cross-curricular examinations were significantly correlated with the overall resident performance score. For the components of the ECNi : DCP and QI scores were significantly correlated with the overall resident performance score, unlike the LCA.

## CONCLUSION AND DISCUSSION

For 2018, the ECNi were predictive of the clinical performance of future residents, as were, in part, the DFASM 2 to DFASM 3 cross-curricular faculty exams.

However, our study has several limitations: it is mono-centric and cannot necessarily be extrapolated. We only have data for 57.4% of the selected residents. We did not measure the thirteen competencies of the EPA, but only nine of them, following a methodology similar to that of Lee.

## REMERCIEMENTS

Aux membres du jury:

Au **Professeur Fournier Jean Paul**, mon président de jury et directeur de thèse, merci de m'avoir accompagnée et soutenue tout au long de cette épreuve. Je vous prie de croire en ma profonde estime.

Au **Professeur Cluzeau Thomas**, je vous remercie d'avoir accepté de siéger au sein de mon jury, recevez toute ma gratitude.

Au **Professeur Staccini Pascal**, je vous remercie d'avoir répondu présent et pour l'intérêt que vous avez porté à mon travail.

Au **Docteur Ganassi Alain**, j'espère que tu sauras lire dans ces lignes l'expression de toute ma reconnaissance.

A ma famille et ma belle famille: merci pour votre soutien inconditionnel.

A mes amis: merci d'être là pour moi et de m'avoir supporté durant toutes ces épreuves.

A toutes les personnes rencontrées au cours de mon internat.

A mon amour et à ma fille: merci pour tout, de m'apporter le bonheur au quotidien, je vous aime tellement.

## LISTE DES ABREVIATIONS

CHU : Centre Hospitalo-Universitaire

CNG : Centre National de Gestion

DGP : Dossier Clinique Progressif

DFASM : Diplôme de Formation Approfondie en Sciences Médicales

ECN : Epreuves Classantes Nationales

ECNi : Epreuves Classantes Nationales Informatisées

ECOS : Examen Clinique Objectif et Structuré

ECT : Examen Cumulatif Transversal

EPA : Entrustable Professional Activities

KFP : Key Feature-Problems

LCA : Lecture Critique d'Article

QCM : Question à Choix Multiples

QI : Questions Isolées

QROC : Question à Réponse Ouverte et Courte

QRU : Question à Réponse Unique

TCEM : Troisième Cycle des Etudes Médicales

TCS : Test de Concordance de Script

USMLE : United States Medical Licensing Examination

# TABLE DES MATIÈRES

I.	INTRODUCTION .....	15
II.	MATERIELS ET METHODES .....	17
	A. Recrutement des étudiants : .....	17
	B. Etablissement du score EPA : .....	17
	C. Analyse des données : .....	18
	D. Analyse statistique : .....	19
III.	RESULTATS.....	20
	A. Population étudiée : .....	20
	B. Validation du score EPA : .....	22
	C. Relation performance/rang de classement : .....	22
	D. Relation performance/composants de l'ECNi : .....	24
	E. Relation performance/examens facultaires:.....	24
	F. Relation performance/docimologie des ECNi : .....	26
IV.	DISCUSSION .....	28
V.	CONCLUSION .....	35
VI.	REFERENCES BIBLIOGRAPHIQUES .....	36
VII.	ANNEXES.....	39



## I. INTRODUCTION

Il existe peu de données sur le caractère prédictif des examens terminaux nationaux et/ou facultaires sur la performance des futurs internes<sup>1-10</sup> et ces données sont discordantes. Le récent travail de Lee en 2018<sup>11</sup> est particulièrement intéressant. En effet, il montre que les examens facultaires et nationaux terminaux aux Etats-Unis, dont les formats sont communs avec ceux qui sont utilisés en France, sont effectivement prédictifs.

Actuellement, en France, on ne dispose d'aucune donnée sur ces deux points.

Depuis 2004, les études médicales en France sont clôturées par les ECN. Cette épreuve est passée simultanément par tous les étudiants en fin d'études (DFASM 3). A l'inverse des pays anglo-saxons ou européens qui utilisent une évaluation nationale terminale *certificative* (USA, Canada, Grande Bretagne, Allemagne, Suisse, Corée)<sup>12</sup>, on utilise en France une épreuve *classante*. Ce classement permet aux étudiants de choisir leur spécialité et le CHU où ils effectueront leur internat<sup>13</sup>.

Depuis 2016, les ECN sont désormais passés par voie électronique (ECNi). Leur durée est de deux jours et demi et ils se composent de trois types d'épreuves. La première épreuve est une analyse de dossiers cliniques progressifs répartie en trois blocs de six dossiers chacun (70 % de la note finale). La deuxième épreuve comporte 120 questions isolées à choix multiples (20 % de la note finale). Et enfin la dernière épreuve est un questionnaire fondé sur la lecture critique de deux articles scientifiques (10 % de la note finale)<sup>14</sup>.

A l'issue de ces épreuves, l'étudiant accède au troisième cycle des études de médecine, TCEM (internat). Depuis 2017, ce cycle est structuré en trois phases permettant à l'étudiant d'acquérir progressivement l'autonomie et les compétences nécessaires à sa spécialité. La phase 1 (phase socle), correspond à l'acquisition des connaissances de base de la spécialité et des compétences transversales nécessaires à l'exercice de la profession. La phase 2 (phase d'approfondissement) correspond à l'acquisition approfondie des connaissances et des compétences nécessaires à l'exercice de la spécialité suivie. Pour finir, la phase 3 (phase de consolidation) correspond à la consolidation de l'ensemble des connaissances et des compétences professionnelles nécessaires à l'exercice de la spécialité suivie<sup>15 13 16</sup>.

Les *Entrustable Professional Activities* (EPA) constituent une évaluation fiable et fidèle de l'acquisition des compétences que doivent maîtriser les étudiants au début de leur internat pour avoir une autonomie suffisante<sup>17 18 19 20</sup>.

Elles se composent de treize compétences : anamnèse et examen physique, hypothèses diagnostiques, prescription et interprétation des tests diagnostiques ou de dépistages courants, discussion des prescriptions, tenue du dossier médical, résumé oral d'une consultation, identification d'informations pertinentes à l'interrogatoire pour la prise en charge rationnelle du patient, qualité des transmissions, participation à une prise en charge inter-professionnelle, identification et prise en charge des situations d'urgences, obtention du consentement pour investigations et/ou procédures, pratique des procédures courantes, identification des échecs du système de santé, culture de sécurité et d'amélioration<sup>19 20</sup> (voir Annexe 1).

L'objectif de ce travail était de rechercher une éventuelle corrélation entre le rang de classement aux ECNi et les EPA et de préciser la(es)quelle(s) des trois modalités docimologiques (DCP, QI, LCA) des ECNi pouva(ien)t être corrélée(s). Enfin, nous prévoyions une approche similaire pour les examens académiques transversaux en DFASM 1, 2 et 3 administrés aux étudiants de la faculté de Médecine de Nice.

## **II. MATERIELS ET METHODES**

### **A. Recrutement des étudiants :**

Les étudiants éligibles sont ceux ayant passé les ECNi en juin 2018, étant au cours de notre étude en second semestre de leur phase socle en poste au CHU de Nice ou pour les internes formés à Nice, en stage dans d'autres CHU, dans une discipline clinique et en contact régulier avec des patients.

Les internes en cours de spécialisation en anatomie et cytologie pathologique, biologie médicale, génétique médicale, médecine nucléaire, imagerie ou santé publique n'étaient donc pas éligibles.

Pour la discipline santé au travail, seuls les internes en stage en service clinique étaient éligibles.

Les internes ont été contactés directement, via les internats des différents CHU et/ou les réseaux sociaux. Plusieurs relances ont été effectuées. Les internes n'étaient éligibles que s'ils donnaient leur accord pour participer à l'étude.

### **B. Etablissement du score EPA :**

Nous avons suivi la même méthodologie que Lee en utilisant son score composé de neuf compétences <sup>11</sup>, regroupant pour certaines rubriques plusieurs des composantes de la version initiale de l'EPA <sup>20</sup> : recueil de l'anamnèse, examen physique, présentation orale du cas, synthèse écrite, connaissances théoriques, jugement clinique, relation patient, professionnalisme, évaluation globale.

Le tableau comparatif entre les treize compétences de l'EPA et les neuf compétences de Lee est disponible en Annexe 1. Nous avons évalué chaque rubrique au moyen d'une échelle de Likert avec sept points (1: Très inférieur au niveau attendu, parmi les 14% moins bon. 2: Modérément inférieur au niveau attendu, parmi les 28% moins bon. 3: Légèrement inférieur au niveau attendu, parmi les 43% moins bon. 4: Atteint le niveau attendu, dans la moyenne. 5: Légèrement supérieur au niveau attendu, dans les 43% meilleurs. 6: Modérément supérieur au niveau attendu, dans les 28% meilleurs. 7: Très supérieur au niveau attendu, dans les 14% meilleurs).

Pour l'établissement du score EPA, chaque compétence a été affectée d'un coefficient de pondération (1 pour -3, 2 pour -2, 3 pour -1, etc...), permettant l'établissement d'un score numérique sur 63, secondairement ramené à 100.

Le fichier Excel pour le recueil des données est en Annexe 2.

Lee soulignait que son score n'avait pas été validé antérieurement <sup>11</sup>. Nous avons donc vérifié la reproductibilité des scores et la concordance inter-observateurs.

## C. Analyse des données :

Pour tous les étudiants, nous avons corrélé leur rang de classement aux ECNi avec leur performance en tant qu'interne (score EPA ramené à 100).

Nous avons également corrélé ce rang de classement avec les composantes du score EPA prises individuellement.

Plusieurs composantes complémentaires ont été regroupées et corrélées au rang de classement : anamnèse et examen physique, présentation orale et synthèse écrite, connaissances théoriques et jugement clinique, relation patient et professionnalisme.

Nous avons finalement comparé ces résultats entre les étudiants formés à Nice et ceux non formés à Nice.

Pour les étudiants niçois, nous disposons des notes des différentes composantes des ECNi grâce au CNG qui organise l'épreuve et transmet ces données à chacune des Facultés. De ce fait nous avons pu corrélérer le score EPA global à chaque type d'épreuve (DCP, QI et LCA).

Pour les DCP, nous avons corrélé le score EPA à chaque bloc de six DCP puis individuellement pour chaque DCP.

Lors de leur second cycle, les étudiants passent des examens transversaux (multi-disciplinaires) utilisant les mêmes formats que les ECN. En fin de DFASM 1 et en fin de DFASM 2, ils passent l'ECT, un examen composé de DCP et QI. En DFASM 3, ils passent trois épreuves réparties sur l'année ayant le même format.

Nous avons donc recherché une éventuelle corrélation entre chacune de ces épreuves, et chacune de leur composante et le score EPA défini ci-dessus.

Dans l'hypothèse où la qualité des questions pouvait conditionner la corrélation entre les notes aux épreuves et le score EPA, nous avons analysé quatre paramètres pour chaque épreuve.

Ces paramètres sont le nombre d'items différents traités par le DCP en se référant au programme de 2013<sup>22</sup>, le nombre de spécialités traitées par le DCP, la qualité des questions, qu'elles soient intégrées dans les DCP ou isolées (QI) et le niveau taxonomique des questions selon la taxonomie de Bloom révisée et adaptée aux études médicales.<sup>23</sup>

Concernant la qualité des questions, nous avons adapté les critères de Gierl<sup>24</sup> qui sont : vignette clinique et questions correctement rédigées, question explicite, réponse(s) correcte(s) non discutable(s), distracteurs plausibles.

Chacun de ces quatre paramètres a été affecté d'un coefficient 1 ; le score complet est donc de 4.

Pour ce qui est de la taxonomie de Bloom, le niveau 1 correspond aux « questions de cours », le niveau 2 à la physiopathologie et/ou à la reconnaissance de structures anatomiques et/ou histologiques sur un examen d'imagerie, le niveau 3 à une prise de décision (prescription d'examens complémentaires, diagnostic, prescription thérapeutique, ou choix d'une conduite à tenir), enfin le niveau 4 correspond au rationnel de prescription et/ou à l'identification de données pertinentes dans la vignette clinique. Pour ce critère, nous avons isolé les questions de niveau 1 et regroupé les questions de niveau 2 à 4.

## D. Analyse statistique :

Nous avons d'abord utilisé le score t de Student pour comparer les rangs de classement aux ECNi des étudiants ayant participé à notre étude et ceux n'ayant pas participé.

Nous avons ensuite validé le score EPA de Lee. Pour cela, deux seniors de l'unité de Médecine Générale d'Urgence du CHU de Nice ont évalué en aveugle quinze internes passés dans leur service. Les scores ont été comparés par un test de Mann et Whitney. Les scores attribués ont été corrélés par le test de Spearman. Pour finir, la concordance inter-observateurs a été vérifiée selon la méthode de Bland et Altman<sup>21</sup>.

La reproductibilité des résultats a été vérifiée par le score de Cronbach pour l'ensemble des étudiants ainsi que pour les étudiants niçois et non niçois.

La corrélation entre rang de classement et score EPA a été étudiée par le test de Pearson.

Ce test a également été utilisé pour rechercher une éventuelle corrélation entre chaque composant des ECNi (DCP, QI et LCA), pour chacun des trois blocs de six DCP, et chaque DCP.

La même analyse a été menée pour les ECT.

La comparaison de la qualité des questions (score de Gierl) a été comparée entre les trois blocs de DCP des ECNi grâce à une analyse de variance.

La proportion de questions classées 1 dans la taxonomie de Bloom a été comparée d'un bloc à l'autre grâce au test du Chi2 avec correction de Yates.

La reproductibilité des différentes notes aux ECNi a été vérifiée par le score de Cronbach.

Dans tous les cas, la significativité a été définie par **p < 0,05**.

### III. RESULTATS

#### A. Population étudiée :

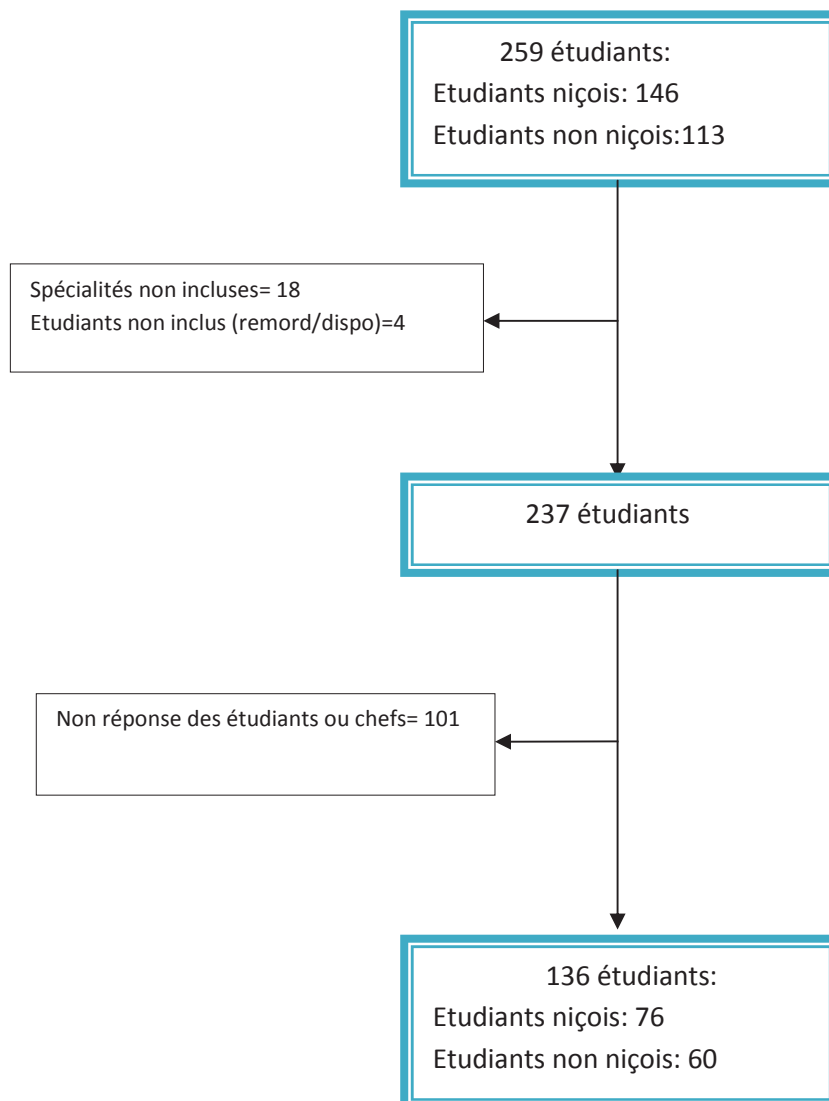
Comme illustré dans le diagramme de flux, 259 étudiants de la promotion 2018 ont été recrutés au départ (146 étudiants ayant fait leur cursus à Nice et 113 ayant fait leur cursus en dehors de Nice).

Sur ces 259 étudiants, 22 n'ont pas été inclus, amenant notre échantillon à 237 étudiants.

Sur ces 237 étudiants, seuls 136 chefs de service ont répondu à notre questionnaire après de multiples relances, soit un taux de participation de **57,4%**.

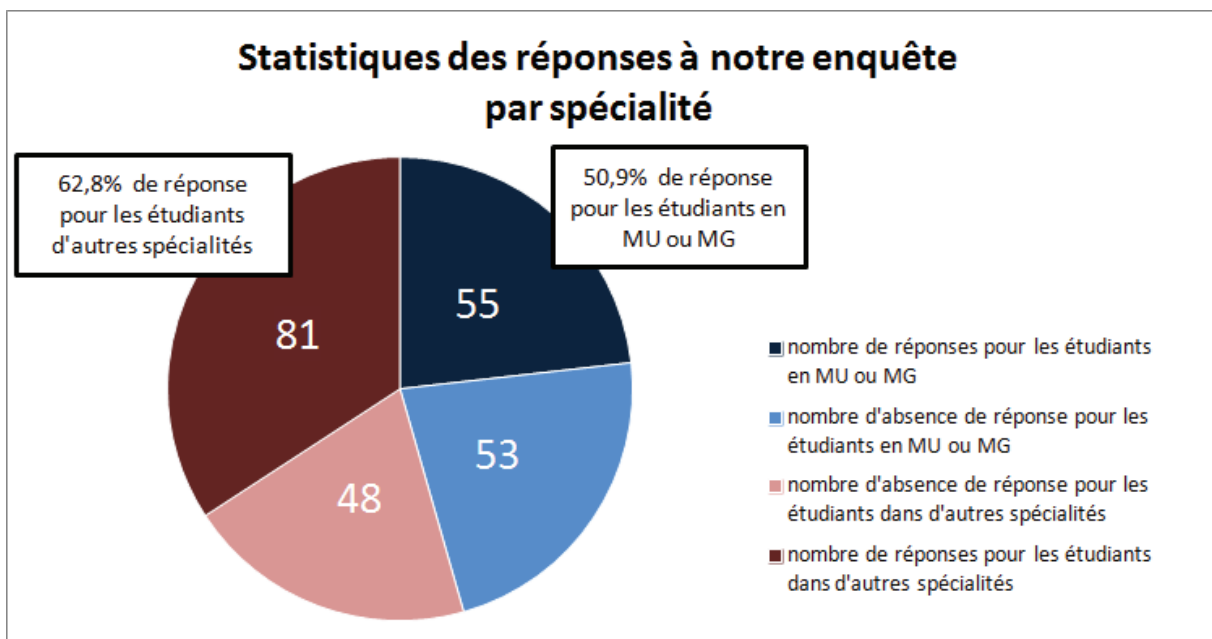
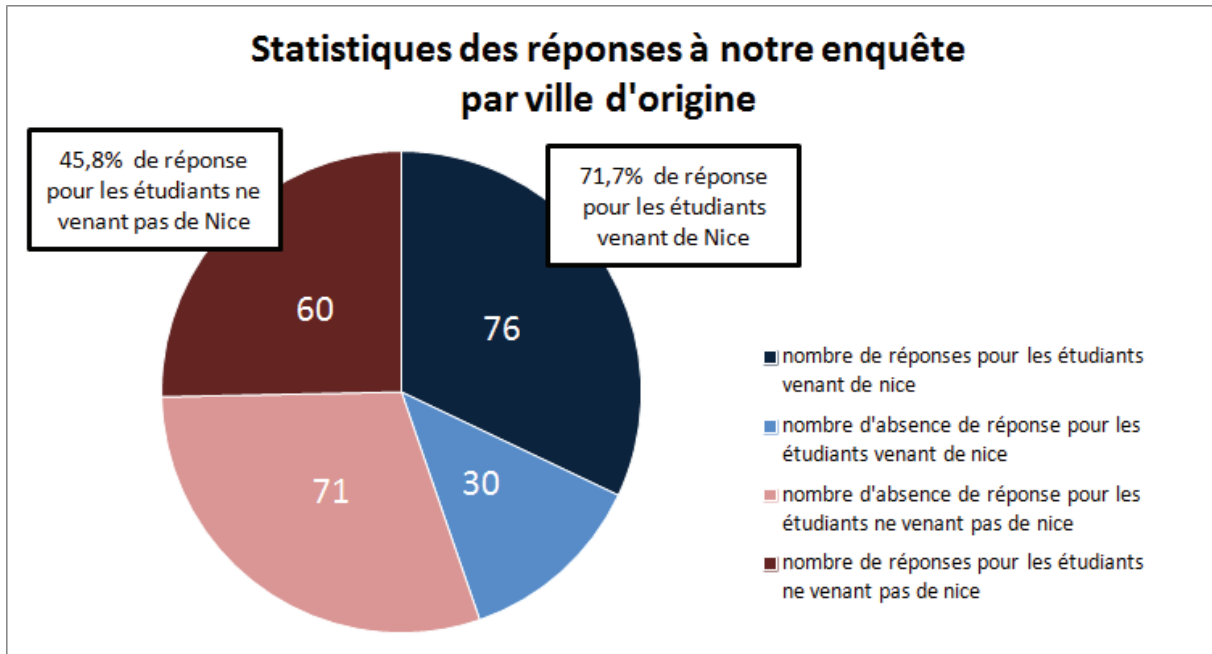
Cependant, un test de Student montre qu'il n'y a pas de différence significative sur les rangs de classement aux ECNi entre le groupe d'étudiant dont nous disposons des données et ceux qui n'ont pas participé ( $p=0,05$ ).

**Figure 1 : Diagramme de flux de la population étudiée**



De plus, nous avons remarqué, que le taux de réponse à notre enquête diffère en fonction de la ville d'origine et de la spécialité de l'étudiant.

**Figure 2 : Statistiques de réponses à notre étude par ville d'origine et par spécialité**





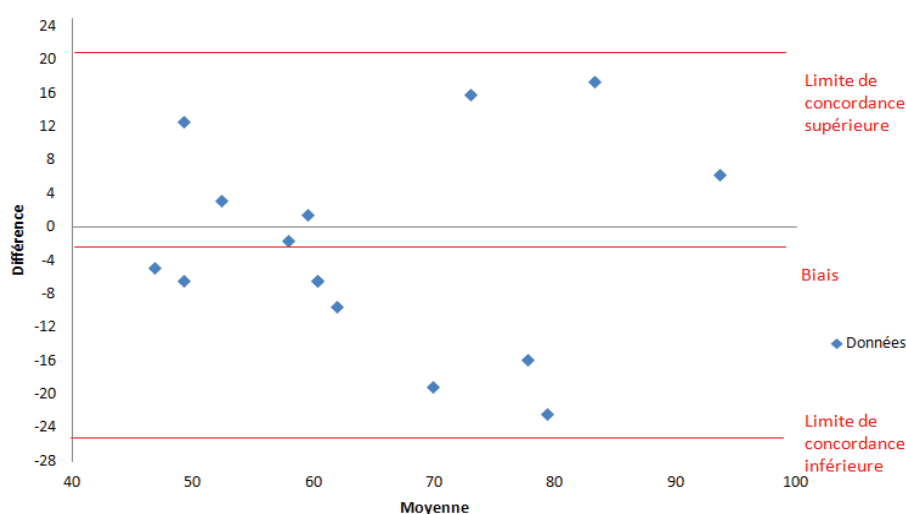
## B. Validation du score EPA :

Quinze internes ont été évalués en aveugle par deux seniors de l'unité de MGU de Pasteur 2.

Les deux observateurs ont tous les deux donné en moyenne des résultats comparables et ont fourni des notes avec un écart type comparable également ( $63,81 \pm 15,35$  et  $66,14 \pm 15,10$ ).

L'analyse selon Bland et Altman montre que les deux observations sont très peu biaisées l'une par rapport à l'autre (-2,3%), et qu'il y a une variance modérée entre elles.

**Figure 3 : Concordance inter-observateurs**



Enfin, le test de Pearson montre que les scores attribués sont très significativement corrélés ( $r = 0,686$ ,  $p=0,0047$ ).

Le score de Cronbach était de **0,9575** pour l'ensemble du groupe, de **0,9569** pour les étudiants niçois et de **0,9574** pour les étudiants non-niçois, ce qui montre une excellente cohérence interne.

## C. Relation performance/rang de classement :

Les données étaient donc disponibles pour **136 étudiants**, 76 originaires de Nice et 60 originaires d'autres Facultés.

Les étudiants formés à Nice étaient significativement mieux classés ( $3620 \pm 2491$  vs  $4774 \pm 1999$ ,  $p = 0,006$ ).

Les scores EPA des étudiants niçois étaient légèrement supérieurs, sans que la différence soit significative ( $77,00 \pm 14,41$  vs  $73,20 \pm 17,30$ ,  $p = 0,17$ ).

Globalement, pour les 136 étudiants, il existait une corrélation faible mais significative entre performance et rang de classement ( $r=-0,184$ ,  $p = 0,03$ ), presque significative pour les étudiants niçois ( $p = 0,059$ ) et non significative pour les étudiants non niçois ( $p = 0,457$ ).

Il n'y avait aucune différence significative entre les scores rubrique par rubrique obtenus par les étudiants niçois et les étudiants non niçois.

Une corrélation faible mais significative entre la performance et le rang de classement était retrouvée pour quelques composantes du score EPA : recueil de l'anamnèse, recueil de l'anamnèse et examen physique, présentation orale du cas, synthèse écrite du cas, présentation orale et synthèse écrite du cas, jugement clinique.

Cette corrélation était retrouvée chez les étudiants niçois, mais pas chez les étudiants non niçois.

Il est difficile de dire si ces résultats relèvent des ECNi et/ou de leur apprentissage facultaire et/ou de leur apprentissage en stage.

**Tableau 1 : Corrélation entre le rang de classement aux ECNi et performance des internes**

Compétence(s)	Corrélation globale (p)	Corrélation Nice (p)	Corrélation hors Nice (p)	Scores Nice (/7 et 14)	Scores hors Nice (/7 et 14)
Recueil anamnèse	<b>0,040</b> ( $r=-0,179$ )	<b>0,037</b>	0,842	$5,30 \pm 1,17$	$5,06 \pm 1,32$
Examen physique	0,061	0,084	0,393	$5,10 \pm 1,22$	$5,06 \pm 1,32$
Anamnèse + Examen	<b>0,037</b> ( $r=-0,18$ )	<b>0,041</b> ( $r=-0,23$ )	0,576	$10,40 \pm 2,25$	$10,05 \pm 2,47$
Présentation orale	<b>0,007</b> ( $r=-0,226$ )	<b>0,008</b> ( $r=-0,295$ )	0,463	$5,35 \pm 1,15$	$5,00 \pm 1,47$
Synthèse écrite	0,059	0,056	0,896	$5,39 \pm 1,24$	$5,15 \pm 1,19$
Présentation + Synthèse	<b>0,0167</b> ( $r=-0,203$ )	<b>0,019</b> ( $r=-0,263$ )	0,628	$10,75 \pm 2,31$	$10,15 \pm 2,55$
Connaissances	0,143	0,136	0,778	$5,26 \pm 1,24$	$5,08 \pm 1,44$
Jugement	<b>0,045</b> ( $r=-0,17$ )	0,085	0,511	$5,34 \pm 1,30$	$5,07 \pm 1,36$
Connaissances + Jugement	0,071	0,095	0,628	$10,60 \pm 2,46$	$10,15 \pm 2,68$
Relation Patient	0,110	0,596	0,151	$5,67 \pm 1,05$	$5,40 \pm 1,33$
Professionnalisme	0,129	0,302	0,440	$5,63 \pm 1,01$	$5,28 \pm 1,68$
Relation + professionnalisme	0,095	0,399	0,258	$11,30 \pm 1,90$	$10,69 \pm 2,84$
Evaluation Globale	0,071	0,182	0,489	$5,44 \pm 1,07$	$5,05 \pm 1,41$

## D. Relation performance/composants de l'ECNi :

Ces données ne concernaient que les étudiants niçois. Elles étaient disponibles pour **76 étudiants**.

Le résultat du test de Cronbach sur les 21 parties des ECNi (18 DCP, 2 DCP de la LCA et QI) donne une cohérence interne correcte ( $\alpha=0,937$ ).

Il existait une corrélation faible mais significative entre performance et l'épreuve de DCP prise globalement (3 blocs de six DCP chacun, regroupant 256 questions) :  $r=0,264$ ,  $p = 0,0234$ .

Pour le bloc 1 (6 DCP – 84 questions) et le bloc 2 (6DCP – 88 questions), la corrélation était faible mais significative :  $r=0,247$ ,  $p = 0,0286$  pour le bloc 1 et  $r=0,272$ ,  $p = 0,0157$  pour le bloc 2 alors qu'elle ne l'était pas pour le bloc 3 (6 DCP – 84 questions) :  $p = 0,0696$ .

Pris individuellement, les DCP 1-2, 1-3, 2-1, 2-5 et 2-6 étaient faiblement mais significativement corrélés au score EPA.

Il existait également une corrélation faible mais significative avec l'épreuve de QI, regroupant 120 questions :  $r=0,316$ ,  $p = 0,0046$ .

Par contre, la corrélation était non significative pour la LCA ( $p = 0,0609$ ).

**Tableau 2 : Relation performance/composantes de l'ECNi**

Pour les 76 étudiants niçois	DCP bloc 1	DCP bloc 2	DCP bloc 3	QI	LCA
<b>Corrélation</b>	6 DCP (84 questions) $r=0,247$ , $p=0,0286$	6 DCP (88 questions) $r=0,272$ , $p=0,0157$	6 DCP (84 questions) $p=0,0696$	120 questions $r=0,316$ , $p=0,0046$	2 DCP (30 questions) $p=0,0609$
<b>Au total</b>	Pour les DCP $r=0,264$ , $p= 0,0234$				

## E. Relation performance/examens facultaires:

Ces données ne concernaient que les étudiants niçois. Elles étaient disponibles pour **62 étudiants** (DFASM 1 et 2) et **69 étudiants** (DFASM 3). La discordance des effectifs correspond à des étudiants ayant redoublé, changé de Faculté ou ayant choisi de différer le passage des ECNi.

N'ont été pris en compte que les examens trans-disciplinaires administrés en fin de DFASM1 et DFASM 2 (ECT), et les trois épreuves DFASM 3.

Pour les ECT en DFSAM 1 et 2, il existait une corrélation faible mais significative en DFASM 2 ( $r=0,285$ ,  $p = \mathbf{0,0223}$ ), et lorsque l'on associe les ECT en DFASM 1 et DFASM 2 ( $r=0,281$ ,  $p = \mathbf{0,0241}$ ) (DFASM 1 ( $p = 0,0702$ )).

L'ECT de DFASM 1 était composé de quatre DCP et 40 QI. Les QI étaient faiblement mais significativement corrélées ( $r= 0,261$ ,  $p = \mathbf{0,0366}$ ). L'ECT de DFASM 2 était composé de six DCP. Si l'on regroupait les DCP des deux ECT (10 au total), la corrélation était significative :  $r=0,602$ ,  $p = \mathbf{0,028}$ . Pris individuellement, le DCP 4 de l'ECT DFASM 1 et les DCP 1 et 6 de l'ECT DFASM 2 étaient significativement corrélés au score EPA.

Pour les examens de DFASM 3 : la corrélation était faible mais significative pour les trois examens :  $r= 0,316$ ,  $p = \mathbf{0,0168}$ . Elle était faible mais significative pour l'examen 1 ( $r= 0,342$ ,  $p = \mathbf{0,0034}$ ), l'examen 3 ( $r=0,275$ ,  $p = \mathbf{0,0199}$ ), mais pas significative pour l'examen 2 ( $p = 0,099$ ). Les examens 1 et 3 comportaient six DCP chacun, l'examen 2 comportait deux DCP et 60 QI. Les QI étaient faiblement mais significativement corrélées ( $r=0,237$ ,  $p = \mathbf{0,0465}$ ). Si l'on regroupe tous les DCP (14 en tout), la corrélation était faible mais significative :  $r=0,318$ ,  $p = \mathbf{0,0067}$ .

Pris individuellement, les DCP 1-4, 1-5, 3-6 étaient significativement corrélés au score EPA.

**Tableau 3: Relation performance/examens facultaires (ECT)**

	ECT DFASM 1	ECT DFASM 2
<b>Corrélation</b>	4 DCP: $p= 0,0702$ 40 QI: $r= 0,261$ , $p=\mathbf{0,0366}$	6 DCP: $r=0,285$ , $p= \mathbf{0,0223}$
<b>Au total</b>	Association ECT DFASM 1 et DFASM 2: $r=0,281$ , $p= \mathbf{0,0241}$ Association des DCP des deux épreuves soit 10 DCP: $r=0,602$ $p= \mathbf{0,028}$	

DFASM 3	Examen 1	Examen 2	Examen 3
<b>Corrélation</b>	6 DCP: $r= 0,342$ , $p=\mathbf{0,0034}$	2 DCP: $p=0,099$ 60 QI: $r=0,237$ , $p=\mathbf{0,0465}$	6 DCP: $r=0,275$ , $p=\mathbf{0,0199}$
<b>Au total</b>	Association des trois examens: $r= 0,316$ , $p=\mathbf{0,0168}$ Association des DCP des trois examens soit 14 DCP: $r=0,318$ , $p=\mathbf{0,0067}$		

## F. Relation performance/docimologie des ECNi :

L'analyse a porté sur les 18 DCP qui sont répartis en trois blocs de six. Nous avons étudié plusieurs paramètres qui sont les suivants :

- Tout d'abord, le nombre total de questions dans chaque DCP était similaire dans les trois blocs, il existait une tendance à l'augmentation du nombre d'item traités ainsi que du nombre moyen de spécialités abordées par DCP selon que les blocs étaient significativement corrélés au score EPA (blocs 1 et 2) ou non (bloc 3).

La différence est significative seulement pour le nombre moyen d'items du programme traité par bloc de six DCP ( $p = 0,0012$ ).

**Tableau 4 : Répartition du nombre d'item, de spécialité et du nombre total de questions pour chaque DCP des trois blocs**

	Dossier	Nombre d'items du programme	Nombre de spécialités	Nombre total de questions
<b>Bloc 1</b>	DCP 1-1	9	7	14
	DCP 1-2	9	5	14
	DCP 1-3	5	5	14
	DCP 1-4	8	4	14
	DCP 1-5	6	4	14
	DCP 1-6	8	1	14
	Moyenne Ecart type	8 2	4 2	14 0
<b>Bloc 2</b>	DCP 2-1	2	2	14
	DCP 2-2	4	2	15
	DCP 2-3	2	1	15
	DCP 2-4	3	1	15
	DCP 2-5	6	6	13
	DCP 2-6	5	5	14
	Moyenne Ecart type	4 2	3 2	14 1
<b>Bloc 3</b>	DCP 3-1	2	1	13
	DCP 3-2	5	4	15
	DCP 3-4	1	1	15
	DCP 3-3	5	2	15
	DCP 3-5	2	1	15
	DCP 3-6	5	4	15
	Moyenne Ecart type	3 2	2 1	15 1

- Ensuite, selon le score de Gierl, on observe une tendance similaire : le score moyen est plus élevé dans les deux blocs significativement corrélés à la performance ; la différence est significative ( $p < 0,001$ ).

**Tableau 5: qualité des questions mesurée selon le score de Gierl .**

	<b>Bloc 1</b>	<b>Bloc 2</b>	<b>Bloc 3</b>
<b>Moyenne</b>	2,77	2,66	2,29
<b>Ecart type</b>	0,83	0,80	0,53

- Pour finir, la taxonomie de Bloom : la proportion de questions de type 1 (« questions de cours ») était respectivement de 25%, 31% et 38% pour les blocs 1 à 3. Les différences n'étaient pas significatives. Elle était de 58% pour les 120 QI.

En définitive, plus que les aspects quantitatifs, ce sont les aspects qualitatifs qui déterminent la corrélation entre les notes aux DCP et la performance du futur interne.

Des résultats similaires étaient obtenus pour les examens facultaires.

## IV. DISCUSSION

Nous mettons en évidence une corrélation significative entre le rang de classement aux ECNi et le score EPA de performance des internes formés à la Faculté de Médecine de Nice et/ou en choix au CHU de Nice lors de leur phase socle. Cette corrélation est néanmoins faible.

Intuitivement, nous pensions que la corrélation fonctionnerait avec les éléments du score explorant des connaissances théoriques, ce que nous n'avons pas confirmé.

Les deux formats d'épreuves corrélées à ce résultat sont les DCP et les QI, alors que l'épreuve de LCA n'est pas significativement corrélée.

L'analyse de la composition des DCP et des paramètres des questions les composant (QCM) montre que le nombre moyen d'items du programme traités par DCP et la qualité des QCM sont vraisemblablement responsables de ces résultats.

Des résultats similaires sont retrouvés sur les examens académiques transversaux administrés à la faculté de Médecine de Nice en DFASM 2 et DFASM 3.

Le caractère prédictif ou non des examens terminaux, nationaux ou non, et académiques est un enjeu majeur pour les autorités et les responsables de programmes académiques<sup>11 19 25</sup>.

Les données de la littérature sont peu importantes et parfois contradictoires.

Alors que dans les articles de Lee<sup>11</sup>, Blacklow<sup>24</sup>, Hojat<sup>1</sup>, Markert<sup>2</sup>, Dong<sup>3</sup>, Sharma<sup>10</sup>, on retrouve une corrélation positive entre les évaluations académiques et/ou terminales et la performance durant le résidanat, Smith<sup>4</sup>, Borowitz<sup>5</sup> et Kahn<sup>6</sup> ne retrouvent quant à eux aucune corrélation.

Dans sa revue de la littérature, Harfman<sup>7</sup> ne retrouve aucune corrélation avec les différentes évaluations pré-graduées prises individuellement.

A l'inverse, Hamdy<sup>8</sup> identifie les épreuves terminales théoriques (Step 2 CK) et pratiques (Step 2 CS – composée de stations d'Examen Clinique Objectif et Structuré – ECOS) comme prédictives de la performance au cours du résidanat (Step 3).

Lee<sup>11</sup> met en évidence une corrélation significative entre la performance des résidents évaluée par le score EPA que nous avons utilisé et le Step 2 CK, les huit stations ECOS et l'examen de médecine interne du *National Board of Medical Examiners* (NBME).

Très récemment, Sharma<sup>10</sup> corrèle le Step 2 CK et la performance multimodale (validé tout au cours de l'année sur un large éventail de compétences) durant le résidanat. En analyse multi-variée, c'est le seul paramètre prédictif.

Nous avons suivi une méthodologie similaire à celle de Lee<sup>11</sup>, en adaptant sa méthode aux examens couramment utilisés dans les Facultés françaises. Le score EPA qu'elle utilise n'a cependant pas été



validé. Comme vu dans le paragraphe II-B, nous utilisons neuf composantes sur les treize que comporte initialement l'EPA<sup>20 19</sup>. La plupart sont regroupées dans le score de Lee, ce qui ne pose pas de problème<sup>26</sup>. Par contre, trois compétences n'y apparaissent pas explicitement : identification et gestion des situations d'urgences (EPA n°10), obtention du consentement (EPA n°12) et identification des échecs du système de santé, culture de la sécurité et d'amélioration (EPA n°13).<sup>20 19</sup>. (voir Annexe 1)

Les composantes n°10 et 13 ne sont pas forcément partagées par toutes les spécialités représentées dans notre population.

Avant d'utiliser ce score EPA, nous l'avons validé en documentant la concordance inter-observateur, en vérifiant l'attribution de scores non significativement différents et significativement corrélés.

L'utilisation de ce score expose au risque d'effet de halo où lorsqu'un examinateur doit noter plusieurs aspects d'une performance, l'attribution d'une bonne note dans un de ces aspects entraîne une surévaluation dans les autres aspects<sup>27</sup>. Ce risque peut être limité en évaluant une compétence à la fois ou en utilisant plusieurs examinateurs<sup>27 28</sup>. Pulito montre que l'évaluation d'une seule compétence est un bon prédicteur de la note finale pour toutes les compétences<sup>28</sup>. Enfin, en matière d'évaluation des compétences, il peut exister une différence significative des notes obtenues en fonction du genre de l'observateur et de l'observé<sup>29</sup>.

Malgré le faible effectif des internes testés et les limites méthodologiques de l'utilisation d'un score de compétences, la cohérence interne du test est parfaite avec un score de Cronbach à 0,957, ce qui fournit une approximation satisfaisante de sa reproductibilité<sup>30</sup>.

Initialement on aurait pu penser que seul les internes les mieux classés avaient accepté de participer à notre étude et de ce fait autorisé leurs responsables de stage à communiquer leur score.

Cependant, le test de Student sur le rang aux ECNi des étudiants ayant accepté de participer à notre étude et de ceux n'ayant pas répondu montre que ces deux populations sont statistiquement proches, tant au niveau de la moyenne que de l'écart type. Nous avons donc pu infirmer cette hypothèse.

La cohérence interne des ECNi a aussi pu être vérifiée grâce à un test de Cronbach qui donne un coefficient alpha de 0,937, indiquant une excellente cohérence interne<sup>30</sup>.

La très récente revue de la littérature de Daniel montre que les QCM, regroupés ou non, représentent un format permettant la mesure du raisonnement clinique, notamment pour l'identification du diagnostic, la prise en charge ou le traitement, la collecte d'informations, et les hypothèses diagnostiques<sup>31</sup>.

Les DCP n'ont pas d'équivalent dans la littérature anglo-saxonne. Le format qui s'en rapproche le plus est le *Modified Essay Question*. Ce format focalise surtout sur les questions de mémorisation pure. Plusieurs articles prouvent que la reproductibilité des scores et la concordance inter-observateurs sont médiocres. Ce format est donc unanimement abandonné depuis 2010<sup>32 33</sup>.

On ne dispose pratiquement d'aucune donnée sur les DCP ou plus généralement sur les ECNi<sup>34 35 36</sup>. Rivière et Pers ont analysé la première épreuve, déroulée en 2016. Leurs conclusions sont similaires : les QCM, qu'ils soient isolés ou regroupés en DCP respectent les consignes de rédaction, sont pertinents, et mesurent majoritairement des aspects de mémorisation pure<sup>35 36</sup>.

Rivière compare DCP et QI et note que les DCP fonctionnent à un niveau cognitif plus élevé que les QI (55 vs 38%,  $p = 0,003$ ). Il montre également que les dernières questions des DCP sont significativement moins pertinentes et propose de limiter les DCP à 10 questions tout en augmentant leur nombre pour garder un nombre global de questions constant<sup>35</sup>.

On note toutefois que ces auteurs n'utilisent pas la taxonomie de Bloom de la même manière.

Ainsi Rivière utilise trois catégories : mémoire seule (niveau 1 de la taxonomie de Bloom), réflexion simple (niveaux 2 à 3) et raisonnement complexe (niveaux 4 à 6)<sup>35</sup>.

Cette appropriation est aussi variable dans la littérature anglo-saxonne.

D'autre part, Palmer<sup>37</sup> en distingue trois niveaux : mémorisation pure, compréhension-application et résolution de problème.

Quoiqu'il en soit, la proportion de questions de mémorisation pure varie entre 50% et 90% pour ces trois auteurs<sup>35 36 37</sup>.

Nos résultats sont concordants avec ces approches. Nous confirmons l'importance du nombre d'items du programme comme élément prédictif de la performance, ce qui revient à « découper » les DCP en questions séquentielles, c'est à dire à relier plusieurs items à une même vignette clinique.

Cette approche est conforme au concept de spécificité de cas, battu en brèche par Norman<sup>38</sup>.

En utilisant une approche par généralisabilité, il montre que deux à trois items rattachés à une vignette clinique permettent de garantir la meilleure reproductibilité des scores (fidélité).

Nous confirmons que les questions des DCP (et des QI) de cette épreuve étaient plutôt correctement rédigées (au moins en termes d'adéquation avec les recommandations fournies aux auteurs).

Le format le plus recommandé par la littérature anglo-saxonne est le modèle *one best answer*<sup>39</sup> où toutes les options sont vraisemblables, mais où la réponse correcte est non discutable (encore plus « vraisemblable »).

Ce format était très minoritairement utilisé lors des ECNi de 2018, de même que l'iconographie.

Qu'il s'agisse des DCP ou des QI, le taux de QCM de type *one best answer*, comportant une vignette était très faible (5% dans chaque groupe), de même que le taux de QCM comportant de l'imagerie (10% pour les DCP et 12,5% pour les QI). Ces constats étaient déjà effectués en 2016<sup>35 36</sup>.

La qualité de rédaction des questions est un élément crucial de la qualité d'un examen, et des décisions qui en découlent<sup>40</sup>. C'est très précisément ce que nous retrouvons ici en comparant les valeurs du score de qualité des questions des blocs de DCP significativement corrélés au score EPA ou non (respectivement :  $2,77 \pm 0,83$ ,  $2,66 \pm 0,80$  pour les blocs 1 et 2 et  $2,29 \pm 0,53$  pour le bloc 3). Compte-tenu de la variabilité de l'application de la taxonomie de Bloom aux QCM, nous avons retenu comme critère de comparaison uniquement la proportion de questions de mémorisation pure (niveau 1), commune aux trois références mentionnées : elle est plus basse en 2018, au moins pour les DCP (25% pour le bloc 1, 31% pour le bloc 2 et 38% pour le bloc 3) et sensiblement constante pour les QI<sup>35 36</sup>. Nous mettons en évidence pour les DCP une tendance non significative à la relation inverse entre proportion de QCM de mémorisation pure et score EPA.

Pour de multiples raisons, comme suggéré par Pugh dans son article de 2019, les QCM sont et resteront probablement longtemps, le format quantitativement le plus important de ces examens terminaux<sup>41</sup>.

La préparation de QCM de qualité est difficile, consommatrice de temps et coûteuse.

L'épreuve que nous avons analysée a « consommé » 378 QCM. On trouve de très nombreuses recommandations pour leur préparation dont l'une des plus séduisante est « l'approche cognitive » décrite par Pugh en 2019<sup>41</sup> qui propose d'adapter un processus d'intelligence artificielle pour la génération automatique de questions.

Cette approche permet de « cibler » l'utilisation des connaissances plutôt que les connaissances factuelles. La génération automatique de questions permet de préparer énormément de questions à partir du même thème, indiscernables des QCM rédigés « artisanalement » par des experts<sup>24 42</sup>.

L'enjeu est doublement important car les QCM sont régulièrement mal rédigés, y compris dans les revues les plus prestigieuses<sup>43</sup>. Fort heureusement, l'apprentissage permet d'améliorer significativement la qualité des questions<sup>44</sup>. D'après Tarrant, l'enjeu est important, car des QCM mal rédigés débouchent sur des décisions inadéquates<sup>40</sup>, avantageant notamment les candidats les plus faibles, d'où une perte de discrimination de l'épreuve<sup>40</sup>. Ce point est crucial ici où les étudiants ne sont pas *certifiés*, mais *classés*. Ce qui peut expliquer les résultats de l'étude de Rivière montrant que les ECNi de 2016 n'étaient pas plus discriminantes que celles de 2015 (dernière édition papier)<sup>34</sup>. Il proposait alors d'élargir la typologie des formats utilisés, avec notamment l'introduction de QCM de type *one-best answer*, d'augmenter le nombre de QI et de diminuer la longueur des DCP en augmentant le nombre pour rester à un nombre global de questions constant.

Notre étude présente plusieurs limites. La plus importante est l'insuffisance du taux de réponses (57,4%) en dépit de plusieurs relances. Les multiples relances ont débouché sur des réponses tardives qui ne permettent pas d'exclure un éventuel biais de mémorisation.

L'analyse des spécialités des responsables de stage montre que ce sont les urgentistes et les médecins généralistes qui ont eu le taux de participation le plus faible.

L'étude est monocentrique et ne concerne qu'une promotion d'internes, regroupant des internes en poste au CHU de Nice et des internes formés à Nice, mais en stage dans d'autres CHU. Les internes formés à Nice étaient significativement mieux classés que leurs collègues. Ce point rend compte d'un phénomène observé depuis plusieurs années : le CHU de Nice a un classement moyen (15<sup>ème</sup> sur 28)<sup>45</sup> et les étudiants niçois les mieux classés quittent Nice et sont remplacés par des étudiants moins bien classés. Par contre, il n'y avait pas de différence significative en termes de performance ( $48,51 \pm 9,08$  vs  $46,12 \pm 10,9$ ,  $p = \text{NS}$ ).

Ces limites obèrent la généralisation de nos résultats.

Le score EPA que nous avons utilisé est directement adapté de celui que Lee<sup>11</sup> a utilisé et en conserve les limites (notamment le passage de treize à neuf compétences). Comme tous les scores d'évaluation de performances, il possède ses limites intrinsèques discutées plus haut.

Tous les internes ont été évalués au second semestre de leur phase socle. On ne peut exclure que l'on ait pu surestimer leur performance (mesure de compétences acquises durant le premier semestre d'internat en phase socle), ce qui n'implique pas forcément que la corrélation avec le rang de classement soit impactée. En effet, tous les internes ayant été évalués de la même façon, l'impact est sans doute négligeable.

De plus, Le score de qualité des QCM que nous avons utilisé est inspiré de celui de Gierl<sup>24</sup>. Ce score comporte initialement huit rubriques, et nous en avons utilisé quatre. Nous nous sommes inspirés de l'approche de Pugh<sup>42</sup> qui est partie du même score et a éliminé et/ou reformulé les rubriques qu'elle avait jugé inadaptées. Ce score a été adopté par le groupe chargé d'établir le cahier des charges de la plateforme qui administrera l'épreuve remplaçant les ECNi à partir de 2023.

Les scores ont été attribués par un seul observateur, au contraire des études de Gierl (21) et Pugh<sup>42</sup>. L'analyse de Pers sur les questions des ECNi 2016 montre un taux de discordance inter-observateurs faible, inférieur à 10 %<sup>36</sup>. Il n'est donc pas certain que nos mesures soient très entachées d'erreurs. En analysant les composantes du scores EPA, nous pensions observer une corrélation significative pour les rubriques « connaissance », « jugement » et « connaissances et jugement », intuitivement explorées par les ECNi. Nous ne retrouvons pas cette corrélation (sauf pour les internes formés à Nice pour la rubrique « jugement »).

Par contre, les compétences pour lesquelles nous retrouvons une corrélation significative ou presque significative (« anamnèse », « examen physique », « anamnèse + examen physique », « présentation orale », « synthèse écrite » et « présentation orale + synthèse écrite ») relèvent plutôt de l'apprentissage en stage. A condition d'une organisation rigoureuse, l'apprentissage en stage peut

être extrêmement efficace<sup>46</sup>. On ne peut exclure que les étudiants les mieux classés aux ECNi aient aussi bénéficié des stages les plus formateurs.

L'arrêté du 8 avril 2013, stipule que "le deuxième cycle des études médicales a pour objectif de permettre aux étudiants d'acquérir les compétences nécessaires à leur formation du troisième cycle et non seulement un apprentissage des connaissances théoriques."<sup>47</sup>

Or, les ECNi, dans leur format actuel, ne mesurent majoritairement que des connaissances théoriques. Des pans entiers des compétences à acquérir lors de ce second cycle ne sont donc pas évalués (communication, professionnalisme, fonctionnement interdisciplinaire, etc...)<sup>34</sup>.

Comme l'a fait remarquer le groupe de travail interministériel sur les ECNi : « les épreuves qui constituent les ECN actuelles ont un retentissement majeur sur le travail de préparation que les étudiants y consacrent [...] Ces épreuves favorisent à l'excès un « bachotage » axé sur une mémorisation factuelle »<sup>48</sup>. Ce mode de fonctionnement éloigne les étudiants de l'apprentissage du raisonnement clinique car trop attachés au « bachotage » de connaissances. Or posséder un savoir n'est pas posséder un savoir-faire, un élève « empli de savoirs » peut se trouver incapable de les réactiver pour agir en situation réelle<sup>49</sup>

Les autorités de tutelle vont dans ce sens : Le Ministère de la Santé : « La procédure des ECN est jugée rigide et inadaptée aux enjeux de la formation des médecins : l'objectif des étudiants et des formateurs n'est plus de former les étudiants à l'exercice de la médecine ou des fonctions d'interne, mais à obtenir de bons résultats aux ECN ». Le Pr Dubois-Randé, précédent président de la Conférence des Doyens, aux manettes de la vaste réforme des études de santé qui réclamait la suppression des ECN : « Il faut sortir du « tout ECN » classant - qui crée un bachotage entre étudiants - en revalorisant les stages et les compétences, qui seront évalués pour tous les étudiants y compris ceux souhaitant entrer dans le système de soin français. Aujourd'hui encore, trop d'enseignants sont focalisés sur l'acquisition des connaissances et non pas des compétences.»<sup>50</sup>

Une meilleure adéquation entre l'apprentissage et leur évaluation en vue de l'acquisition et de la vérification de l'acquisition des compétences essentielles est donc indispensable.

C'est l'objet de la réforme du second cycle des études médicales qui concerne les étudiants entrant en DFASM 1 à la rentrée 2021.

Cette approche implique une modification des programmes<sup>25</sup>. Plusieurs études suggèrent l'intérêt de stages de remise à niveau entre la fin des études et le début du résidanat<sup>51 52</sup>, notamment pour les disciplines très « techniques » comme la chirurgie.

C'est le principe des *boot camp*. Cette approche est d'autant plus logique qu'au début de leur résidanat les résidents estiment maîtriser moins de 50 % des compétences qu'ils devraient maîtriser à ce stade<sup>53 54</sup>.

C'est la logique de la réforme du second cycle qui fait du DFASM 3 une année pré-professionnalisante. La réforme des modalités d'examens, tant facultaires que terminaux constitue une approche complémentaire.

Plusieurs études confirment le caractère prédictif des ECOS concernant l'acquisition des compétences à maîtriser par les futurs résidents<sup>8 11 10</sup>.

Cette approche est d'autant plus séduisante que les ECOS ont un impact significatif, poussant les étudiants vers des apprentissages actifs, tels que les examens cliniques de patients ou l'apprentissage technique dans un centre de simulation<sup>55</sup>.

En France, la réforme du second cycle sera en place en 2023. Elle comportera trois parties: 40 % du score total sera basé sur des ECOS afin d'évaluer les compétences de l'étudiant, 50% du score final portera sur ses connaissances et enfin 10% du score total sur son parcours professionnel.

Le nouveau mode d'évaluation des connaissances devra dans la mesure du possible, mesurer le raisonnement clinique, être docimologiquement largement documenté et être utilisable à grande échelle par voie électronique.

Plusieurs formats de questions répondent à ce cahier des charges : QCM à contexte riche (vignette clinique) de type *one-best answer*, Test de Concordance de Scripts (TCS), *Key-Feature Problems* (KFP) (qui comprendront QCM, QRU et QROC)<sup>31 56 57</sup>.

TCS et KFP permettent également d'appréhender le raisonnement en contexte d'incertitude<sup>58</sup> ce qui constitue un objectif majeur de l'apprentissage des années à venir<sup>59</sup>.

Le format de QCM décrit par Sam classant les options de réponses par vraisemblance pourrait répondre également à cet objectif<sup>60</sup>.

Il est indispensable d'utiliser plusieurs formats, chacun évaluant des facettes différentes et complémentaires<sup>61</sup>.

Ce serait une erreur de sous-estimer la pertinence des QCM dans la mesure du raisonnement clinique. Plusieurs études utilisant le raisonnement à voix haute montrent que les étudiants les plus performants mobilisent les mêmes voies de raisonnement que les experts lors de la résolution de QCM à contexte riche<sup>62 63</sup>.

## **V. CONCLUSION**

En dépit de limites méthodologiques évidentes, nous montrons que le rang de classement aux ECNi pouvait être corrélé au niveau de performance attendu, mesuré en première année d'internat (phase socle). Ce résultat repose sur deux des trois épreuves des ECNi : les DCP et les QI. Il est conditionné par la qualité des QCM utilisés et l'aptitude des DCP à balayer plusieurs items du programme.

A condition d'évoluer vers des formats plus exigeants d'un point de vue docimologique, et plus orientés vers la mesure du raisonnement clinique plutôt que de connaissances factuelles, les futures ECN pourraient atteindre également cet objectif.

## VI. REFERENCES BIBLIOGRAPHIQUES

1. Hojat, M., Gonnella, J. S., Veloski, J. J. & Erdmann, J. B. Is the glass half full or half empty? A reexamination of the associations between assessment measures during medical school and clinical competence after graduation. *Acad. Med. J. Assoc. Am. Med. Coll.* 68, S69-76 (1993).
2. Markert, R. J. The relationship of academic measures in medical school to performance after graduation. *Acad. Med.* 68, S31 (1993).
3. Dong, T. *et al.* Validity evidence for medical school OSCEs: associations with USMLE® step assessments. *Teach. Learn. Med.* 26, 379–386 (2014).
4. Smith, S. R. Correlations between graduates' performances as first-year residents and their performances as medical students. *Acad. Med. J. Assoc. Am. Med. Coll.* 68, 633–634 (1993).
5. Borowitz, S. M., Saulsbury, F. T. & Wilson, W. G. Information collected during the residency match process does not predict clinical performance. *Arch. Pediatr. Adolesc. Med.* 154, 256–260 (2000).
6. Kahn, M. J., Merrill, W. W., Anderson, D. S. & Szerlip, H. M. Residency program director evaluations do not correlate with performance on a required 4th-year objective structured clinical examination. *Teach. Learn. Med.* 13, 9–12 (2001).
7. Harfmann, K. L. & Zirwas, M. J. Can performance in medical school predict performance in residency? A compilation and review of correlative studies. *J. Am. Acad. Dermatol.* 65, 1010-1022.e2 (2011).
8. Hamdy, H. *et al.* BEME systematic review: predictive values of measurements obtained in medical schools and future performance in medical practice. *Med. Teach.* 28, 103–116 (2006).
9. Blacklow, R. S., Goepf, C. E. & Hojat, M. Further psychometric evaluations of a class-ranking model as a predictor of graduates' clinical competence in the first year of residency. *Acad. Med. J. Assoc. Am. Med. Coll.* 68, 295–297 (1993).
10. Sharma, A. *et al.* USMLE Step 2 CK: Best Predictor of Multimodal Performance in an Internal Medicine Residency. *J. Grad. Med. Educ.* 11, 412–419 (2019).
11. Lee, M. & Vermillion, M. Comparative values of medical school assessments in the prediction of internship performance. *Med. Teach.* 40, 1287–1292 (2018).
12. An International Association For Medical Education - AMEE.  
<https://amee.org/conferences/amee-past-conferences/amee-2018>.
13. ECNi\_2017\_VDT\_SI.pdf.
14. Arrêté du 20 juillet 2015 relatif à l'organisation des épreuves classantes nationales anonymes donnant accès au troisième cycle des études médicales. JORF n°0168 du 23 juillet 2015
15. Arrêté du 12 avril 2017 portant organisation du troisième cycle des études de médecine. JORF n°0089 du 14 avril 2017
16. Décret n° 2016-1597 du 25 novembre 2016 relatif à l'organisation du troisième cycle des études de médecine et modifiant le code de l'éducation. 2016-1597 (2016). JORF n°0276 du 27 novembre 2016
17. ten Cate, O. Entrustability of professional activities and competency-based training. *Med. Educ.* 39, 1176–1177 (2005).
18. ten Cate, O. & Scheele, F. Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Acad. Med. J. Assoc. Am. Med. Coll.* 82, 542–547 (2007).
19. Cutrer, W. B., Russell, R. G., Davidson, M. & Lomis, K. D. Assessing medical student performance of Entrustable Professional Activities: A mixed methods comparison of Co-Activity and Supervisory Scales. *Med. Teach.* 42, 325–332 (2020).
20. 484778-epa13toolkit.pdf.
21. Bland, J. M. & Altman, D. G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet Lond. Engl.* 1, 307–310 (1986).



22. Études médicales. *Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation* [//www.enseignementsup-recherche.gouv.fr/pid20536/bulletin-officiel.html](http://www.enseignementsup-recherche.gouv.fr/pid20536/bulletin-officiel.html).
23. Krathwohl, D. R. A Revision of Bloom's Taxonomy: An Overview. *Theory Pract.* 41, 212–218 (2002).
24. Gierl, M. J. & Lai, H. Evaluating the quality of medical multiple-choice items created with automated processes. *Med. Educ.* 47, 726–733 (2013).
25. Lyss-Lerman, P. *et al.* What training is needed in the fourth year of medical school? Views of residency program directors. *Acad. Med. J. Assoc. Am. Med. Coll.* 84, 823–829 (2009).
26. Englander, R. *et al.* Toward Defining the Foundation of the MD Degree: Core Entrustable Professional Activities for Entering Residency. *Acad. Med. J. Assoc. Am. Med. Coll.* 91, 1352–1358 (2016).
27. Margolis, M. J. *et al.* Use of the mini-clinical evaluation exercise to rate examinee performance on a multiple-station clinical skills examination: a validity study. *Acad. Med. J. Assoc. Am. Med. Coll.* 81, 556–60 (2006).
28. Pulito, A. R., Donnelly, M. B. & Plymale, M. Factors in faculty evaluation of medical students' performance. *Med. Educ.* 41, 667–675 (2007).
29. Riese, A., Rappaport, L., Alverson, B., Park, S. & Rockney, R. M. Clinical Performance Evaluations of Third-Year Medical Students and Association With Student and Evaluator Gender. *Acad. Med. J. Assoc. Am. Med. Coll.* 92, 835–840 (2017).
30. Bertrand R & Blais JG. *Modèles de mesure. L'apport de la théorie des réponses aux items.* (Presses de l'Université de Québec, 2004).
31. Daniel, M. *et al.* Clinical Reasoning Assessment Methods: A Scoping Review and Practical Guidance. *Acad. Med. J. Assoc. Am. Med. Coll.* 94, 902–912 (2019).
32. Feletti, G. I. Reliability and validity studies on modified essay questions. *J. Med. Educ.* 55, 933–941 (1980).
33. Palmer, E. J., Duggan, P., Devitt, P. G. & Russell, R. The modified essay question: its exit from the exit examination? *Med. Teach.* 32, e300–307 (2010).
34. Rivière, E., Quinton, A. & Dehail, P. Analyse de la discrimination des notes aux ECN informatisées de 2016. *Rev. Médecine Interne* 40, 286–290 (2019).
35. Rivière, E. *et al.* Analyse docimologique des premières Épreuves classantes nationales informatisées (ECNi) de 2016 : propositions d'amélioration. *Rev. Médecine Interne* 40, 47–51 (2019).
36. Pers, Y.-M. *et al.* Analyse de la validité du contenu et de la qualité formelle des questions isolées des épreuves classantes nationales informatisées (ECNi) de l'année universitaire 2015–2016. *Pédagogie Médicale* 19, 171–180 (2018).
37. Palmer, E. J. & Devitt, P. G. Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? Research paper. *BMC Med. Educ.* 7, 49 (2007).
38. Norman, G., Bordage, G., Page, G. & Keane, D. How specific is case specificity? *Med. Educ.* 40, 618–623 (2006).
39. Case SM & Swanson DB. *Constructing written test questions for the basic and clinical sciences, 3rd edition.* (The National Board of Medical Examiners, 2002).
40. Tarrant, M. & Ware, J. Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments. *Med. Educ.* 42, 198–206 (2008).
41. Pugh, D., Champlain, A. D. & Touchie, C. Plus ça change, plus c'est pareil: Making a continued case for the use of MCQs in medical education. *Med. Teach.* 41, 569–577 (2019).
42. Pugh, D., De Champlain, A., Gierl, M., Lai, H. & Touchie, C. Can automated item generation be used to develop high quality MCQs that assess application of knowledge? *Res. Pract. Technol. Enhanc. Learn.* 15, 12 (2020).
43. Stagnaro-Green, A. S. & Downing, S. M. Use of flawed multiple-choice items by the New England Journal of Medicine for continuing medical education. *Med. Teach.* 28, 566–568 (2006).

44. Jozefowicz, R. F. *et al.* The quality of in-house medical school examinations. *Acad. Med. J. Assoc. Am. Med. Coll.* 77, 156–161 (2002).
45. Classement des CHU. *What's Up Doc* <https://www.whatsupdoc-lemag.fr/gros-dossier/classement-des-chu> (2019).
46. Norcini, J. & Burch, V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Med. Teach.* 29, 855–871 (2007).
47. *Arrêté du 8 avril 2013 relatif au régime des études en vue du premier et du deuxième cycle des études médicales.* JORF n°0095 du 23 avril 2013
48. Drummond, D. *et al.* [Newly formed French residents in pediatrics are not well prepared for conducting pediatric resuscitation after medical school]. *Arch. Pediatr. Organe Off. Soc. Francaise Pediatr.* 23, 150–158 (2016).
49. Compagnon, L. *et al.* Définitions et descriptions des compétences en médecine générale. *Vo U M E* 8.
50. Cousin, S. Fin des ECN : entrée en vigueur en 2019. *Remede.org* <http://www.remede.org/documents/fin-des-ecn-entree-en-vigueur-en-2019.html>.
51. Esterl, R. M., Henzi, D. L. & Cohn, S. M. Senior medical student 'Boot Camp': can result in increased self-confidence before starting surgery internships. *Curr. Surg.* 63, 264–268 (2006).
52. Lindeman, B. M., Sacks, B. C. & Lipsett, P. A. Graduating Students' and Surgery Program Directors' Views of the Association of American Medical Colleges Core Entrustable Professional Activities for Entering Residency: Where are the Gaps? *J. Surg. Educ.* 72, e184–e192 (2015).
53. Santen, S. A. *et al.* How Competent Are Emergency Medicine Interns for Level 1 Milestones: Who Is Responsible? *Acad. Emerg. Med.* 20, 736–739 (2013).
54. Wancata, L. M., Morgan, H., Sandhu, G., Santen, S. & Hughes, D. T. Using the ACMGE Milestones as a Handover Tool From Medical School to Surgery Residency. *J. Surg. Educ.* 74, 519–529 (2017).
55. 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.* 19, 146 (2019).
56. Sibert L & Fournier JP. *ECNi – Epreuve TCS Test de concordance de script. Le guide méthodologique.* (Maloine, 2015).
57. Hrynychak, P., Takahashi, S. G. & Nayer, M. Key-feature questions for assessment of clinical reasoning: a literature review. *Med. Educ.* 48, 870–883 (2014).
58. Cooke, S. & Lemay, J.-F. Transforming Medical Assessment: Integrating Uncertainty Into the Evaluation of Clinical Reasoning in Medical Education. *Acad. Med. J. Assoc. Am. Med. Coll.* 92, 746–751 (2017).
59. Gheihman, G., Johnson, M. & Simpkin, A. L. Twelve tips for thriving in the face of clinical uncertainty. *Med. Teach.* 42, 493–499 (2020).
60. Sam, A. H. *et al.* Clinical prioritisation questions: A novel assessment tool to encourage tolerance of uncertainty? *Med. Teach.* 42, 416–421 (2020).
61. van der Vleuten, C. P. M. & Schuwirth, L. W. T. Assessing professional competence: from methods to programmes. *Med. Educ.* 39, 309–317 (2005).
62. Heist, B. S., Gonzalo, J. D., Durning, S., Torre, D. & Elnicki, D. M. Exploring Clinical Reasoning Strategies and Test-Taking Behaviors During Clinical Vignette Style Multiple-Choice Examinations: A Mixed Methods Study. *J. Grad. Med. Educ.* 6, 709–714 (2014).
63. Surry, L. T., Torre, D. & Durning, S. J. Exploring examinee behaviours as validity evidence for multiple-choice question examinations. *Med. Educ.* 51, 1075–1085 (2017).

## VII. ANNEXES

### **Annexe 1 : Comparaison entre les treize compétences de l'EPA et les neuf compétences du framework de Lee**

<b>Core Entrustable Performance Activities</b>	<b>Lee's Framwork competencies</b>
EPA 1: Gather a History and Perform a Physical Examination	History taking
EPA 2: Prioritize a Differential Diagnosis Following a Clinical Encounter	Physical examination
EPA 3: Recommend and Interpret Common Diagnostic and Screening Tests	Oral case presentation
EPA 4: Enter and Discuss Orders and Prescriptions	Case write-up,
EPA 5: Document a Clinical Encounter in the Patient Record	Fund of knowledge
EPA 6: Provide an Oral Presentation of a Clinical Encounter	Clinical judgment
EPA 7: Form Clinical Questions and Retrieve Evidence to Advance Patient Care	Physician–patient interaction
EPA 8: Give or Receive a Patient Handover to Transition Care Responsibility	Professional attitudes and behaviors
EPA 9: Collaborate as a Member of an Interprofessional Team	Overall rotation performance
EPA 10: Recognize a Patient Requiring Urgent or Emergent Care and Initiate Evaluation and Management	
EPA 11: Obtain Informed Consent for Tests and/or Procedures	
EPA 12: Perform General Procedures of a Physician	
EPA 13: Identify System Failures and Contribute to a Culture of Safety and Improvement	

## Annexe 2: Fichier servant à l'évaluation des internes par leur chef de service

Compétence évaluée	1: Très inférieur au niveau attendu (parmi les 14% moins bons)	2: Modérément inférieur au niveau attendu (parmi les 28% moins bons)	3: Légèrement inférieur au niveau attendu (parmi les 43% moins bons)	4: Atteint le niveau attendu (dans la moyenne)	5: Légèrement supérieur au niveau attendu (parmi les 19% meilleurs)	6: Modérément supérieur au niveau attendu (parmi les 28% meilleurs)	7: Très supérieur au niveau attendu (parmi les 14% meilleurs)	Nombre de croix (1 max par ligne)
Recueil de l'anamnèse				X				1
Examen physique				X				1
Présentation orale du cas				X				1
Synthèse écrite du cas				X				1
Connaissances théoriques				X				1
Jugement clinique				X				1
Relation avec le patient				X				1
Attitudes et comportements professionnels				X				1
Évaluation globale				X				1
Totaux	0	0	0	0	36	0	0	0
Note Globale sur 63	36							
%	57,14							