

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

An analysis of the content validity and formal quality of the isolated questions of the French National Ranking Examination (ECNi) for the academic year 2015–2016

Yves-Marie PERS^{1,*}, Clémentine MARAIS¹, Philippe LAMBERT², Claude JEANDEL³, Gérald CHANQUES⁴, Ilan SZWARC⁵, Jean-Luc FAILLIE⁶, Olivier MATHIEU⁶, et Moglie LE QUINTREC⁵

¹ Unité d'immunologie clinique et de thérapeutique ostéoarticulaire, Département de rhumatologie, Hôpital Lapeyronie, CHU de Montpellier, Université de Montpellier, Montpellier, France

² Département de médecine générale, Faculté de médecine de Montpellier-Nîmes, Montpellier, France

³ Pôle de gérontologie, Centre de prévention et de traitement des maladies du vieillissement, CHU de Montpellier, Université de Montpellier, Montpellier, France

⁴ Unité de réanimation et de transplantation, Département d'anesthésie-réanimation B, Hôpital Saint-Eloi, CHU de Montpellier, Université de Montpellier, Montpellier, France

⁵ Service de néphrologie et transplantation, Hôpital Lapeyronie, CHU de Montpellier, Université de Montpellier, Montpellier, France

⁶ Département de pharmacologie médicale et toxicologie, CHU de Montpellier, Université de Montpellier, Montpellier, France

Manuscrit soumis pour publication le 21 mai 2017 ; commentaires éditoriaux formulés aux auteurs le 20 octobre 2018 et le 7 octobre 2019 ; accepté pour publication le 9 octobre 2019

Résumé - Contexte : Les premières épreuves classantes nationales informatisées (ECNi) clôturant le deuxième cycle des études médicales en France ont eu lieu en juin 2016 et ont comporté des dossiers cliniques progressifs (DCP), des questions isolées (QI), ainsi qu'une épreuve de lecture critique d'articles (LCA). Les QI représentaient une nouveauté de cet examen puisque les DCP et la LCA étaient utilisés depuis 2004. **But :** Analyse de la validité du contenu et de la qualité formelle des QI des premières épreuves ECNi. **Méthodes :** Un questionnaire contenant les critères docimologiques à recueillir a été défini et validé au préalable. Les 120 QI de l'ECNi 2016 ont été analysées par un binôme indépendant. Les discordances entre les évaluateurs ont été interprétées. Toutes les analyses ont été colligées sur tableur Excel puis les statistiques ont été réalisées sur Graphpad Prism version 6 et Stata v12. **Résultats :** Les rédacteurs des QI ont majoritairement respecté les consignes de forme. Les niveaux d'habiletés intellectuelles requis concernaient majoritairement les processus de mémorisation, de simple récupération de l'information et de compréhension (82 % des QI). Dans 7 à 23 % des QI, l'utilisation de distracteurs était jugée inappropriée. Une prédominance de QI portait sur les disciplines médicales (72,5 %). L'unité d'enseignement (« Circulation – Métabolismes ») était la plus représentée (26 %). Les discordances étaient faibles (6,2 % des évaluations). **Conclusion :** Il s'agit de la première analyse de la validité du contenu et de la qualité formelle des QI des ECNi. Leur format de rédaction a été largement respecté par les auteurs. Elles ont peu fait appel à des processus de raisonnement complexe. Il semblerait utile de favoriser les courtes vignettes cliniques pour améliorer cet aspect.

Mots clés : ECNi, analyse de contenu, qualité formelle, questions isolées

Abstract. Context: The first French electronic National Ranking Examination (ECNi), occurring at the end of the undergraduate medical studies took place for the first time in June 2016. Assessments included progressive clinical cases scenarios (DCPs), isolated questions (IQ) and critical appraisal of scientific articles (LCA). The IQs represented the novelty of this evaluation since the DCPs and LCA were already in use since

*Correspondance et offprints : Yves-Marie PERS. Unité d'immunologie clinique et de thérapeutique ostéoarticulaire, Département de rhumatologie, Hôpital Lapeyronie, CHU Montpellier, 371, avenue du doyen Gaston Giraud, 34295 Montpellier, France. Mailto: ympers2000@yahoo.fr.

2004. **Aim:** To analyse the content validity and formal quality of the IQ content for the first ECNi tests. **Methods:** A survey stating quality criteria to be collected was defined and beforehand validated. The 120 IQ of the ECNi were analysed by an independent pair. Discrepancies between the two evaluators were interpreted. All the analyses were compiled on an Excel spreadsheet and then statistics were performed on Graphpad Prism version 6 and Stata v12. **Results:** The majority of IQ writers complied with the formal instructions. The required intellectual skills mainly consist of information retrieval and knowledge (82% of IQ). In 7 to 23% of IQ, the use of distractors was considered inappropriate. A predominance of IQ was found in non-surgical disciplines (72.5%). The Teaching Unit (“Circulation – Metabolism”) was the most represented (26%). Discrepancies were low between the two assessors (6.2% of the evaluation criteria). **Conclusion:** This is the first analysis of content validity and formal quality of ECNi IQ tests. The writing format was widely respected by the authors. IQ have made little use of complex reasoning processes. It would seem useful to promote short clinical vignettes to improve this feature.

Keywords: ECNi, content validity, formal quality, isolated questions

Introduction

Dans le contexte français, les épreuves classantes nationales (ECN), introduites en 2004 à l’issue des deux premiers cycles d’études médicales (soit l’équivalent de ce que d’autres pays nomment formation pré-graduée), ont constitué un changement majeur dans l’évaluation sommative finale des apprentissages. Le concours d’internat, autrefois limité aux étudiants souhaitant accéder à une filière de formation spécialisée de troisième cycle hors médecine générale, s’est imposé dès lors à l’ensemble des étudiants en médecine pour obtenir un poste d’interne. En fonction de son classement à ces épreuves, l’étudiant doit choisir à la fois la filière de formation, le centre hospitalo-universitaire de rattachement et la faculté de formation, et donc sa ville d’affectation. Le nombre de places dans chaque filière est limité ce qui en fait un concours classant. Les ECN comportaient à l’époque neuf dossiers cliniques, explorés par des questions rédactionnelles à réponses ouvertes et courtes, avec un temps d’épreuve moyen d’une heure par dossier et, depuis 2009, une épreuve de lecture critique d’article (LCA). Une double correction était réalisée, ce qui nécessitait la mobilisation de près de 400 enseignants sur une durée de plus de trois semaines sur un même site, entraînant un coût organisationnel important. Depuis 2004, l’effectif d’étudiants a considérablement augmenté et la discrimination des étudiants en fonction des notes est devenue très faible : en 2012, 7658 étudiants ont été classés dans un intervalle de seulement 350 points entiers [1]. De plus, concernant l’aspect pédagogique, ces ECN ne semblaient pas explorer le programme de manière assez exhaustive. Ainsi, en 2012, seuls 110 items étaient abordés sur les 360 au programme [1]. Des critiques du format de l’épreuve, favorisant l’excès de « bachotage » (néologisme français désignant un travail de révision intensif et bref, centré exclusivement sur la performance à l’examen, au détriment d’un apprentissage signifiant et en profondeur), avaient également été émises.

Ces éléments ont conduit les deux ministères de tutelle (enseignement supérieur et recherche ; affaires sociales et santé) à proposer une réforme technologique, utilisant des tablettes numériques comme interface

pour l’évaluation des étudiants avec correction automatisée. Les objectifs de cette réforme étaient d’améliorer la fonction discriminante des épreuves, de simplifier la logistique, de diversifier les épreuves et d’harmoniser les chances de réussite des étudiants. Le 20 juillet 2015, le Ministère de l’enseignement supérieur et de la recherche a validé ce projet et a publié un arrêté fixant les nouvelles modalités d’organisation et de déroulement des premières ECN informatisées (ECNi) [2].

La première session des ECNi a eu lieu du 20 au 22 juin 2016 et a comporté un ensemble d’épreuves répondant à trois formats, coté sur un barème total de 10 000 points [3] : les dossiers cliniques progressifs (DCP), les questions isolées (QI) et la lecture critique d’articles (LCA). L’intention affichée de ces épreuves était de tester les connaissances factuelles, la capacité à résoudre des problèmes cliniques simples et les problèmes complexes en situation d’incertitude. Les modalités du déroulement des ECNi ont été réparties de la façon suivante :

- 18 DCP : partant d’un énoncé initial, 15 questions à choix multiple (QCM) de cinq propositions chacune, apparaissent au fur et à mesure de la validation des QCM par l’étudiant, ajoutant de nouvelles données au cas clinique mais dont le raisonnement est censé être indépendant de la réponse précédemment donnée par l’étudiant. Les points attribués aux réponses à ces questions comptaient pour 70 % du score final soit 389 points par DCP. Les étudiants ont examiné six DCP par demi-journée de trois heures ;
- 120 QI, constituées de QCM à type de questions à réponse unique parmi cinq propositions, complètement indépendantes les unes des autres, couvrant l’ensemble des items du programme (en théorie ceux qui n’ont pas été abordés dans les DCP) en une demi-journée. Les points attribués aux réponses aux QI représentaient 20 % du score final soit 17 points par QI ;
- LCA : deux articles scientifiques (en français en 2016, et en anglais par la suite) avec 15 QCM indépendants portant sur la méthodologie, l’analyse critique des résultats ainsi que la qualité des articles. L’ensemble représentait 10 % du score final soit 33 points par QCM et s’est déroulé en une demi-journée.

Afin de mettre en place ces nouvelles modalités et détecter rapidement les éventuelles failles, des épreuves « tests » nationales ont eu lieu du 7 au 10 décembre 2015 et du 7 au 10 mars 2016. Ces épreuves « tests » ont donné lieu à des retours positifs d'expérience malgré des difficultés techniques [4].

En décembre 2004, le conseil scientifique de la section médecine du Centre national des concours d'internat (CNCI) a défini plusieurs consignes sur la rédaction des QCM pour les ECNi [5]. À notre connaissance, il n'existe pas de recommandations aux auteurs plus récentes relatives au format des ECNi. Les consignes pour les enseignants rédacteurs étaient les suivantes :

- éviter les questions de mémorisation pure ;
- diversifier au maximum les questions en faisant apparaître l'approche pluridisciplinaire de la pathologie en cause ;
- poser des questions suscitant une réflexion ;
- ne pas négliger dans les questions la gestion de l'incertitude ;
- veiller à ne pas poser des questions sur les objectifs libellés dans le texte ;
- dans une démarche diagnostique, demander à l'étudiant de hiérarchiser les données issues de l'examen et d'établir une stratégie d'exploration paraclinique.

En raison de ces changements de modalités d'évaluation prévus pour l'ECNi 2016, chaque faculté a dû s'adapter en appliquant (progressivement et inégalement) ces modalités d'évaluation aux enseignements et aux examens afin de permettre aux étudiants, jusqu'alors habitués aux dossiers rédactionnels, de s'entraîner. Pour cela depuis novembre 2013, une base docimologique nationale d'entraînement a été créée pour l'ensemble des universités françaises : le Système informatique distribué d'évaluation en santé (SIDES). Sur cette plateforme SIDES, le contenu (forme et fond) des DCP, des QI et des LCA doit respecter certaines recommandations afin de préparer au mieux les étudiants aux conditions de l'examen final. Ainsi, SIDES a émis des consignes rédactionnelles, référencées ci-dessous et recommandées aux enseignants lors de leur formation à l'utilisation de la plateforme [6] :

- la question mesure un objectif important d'apprentissage et se rapporte au référentiel ;
- l'énoncé est constitué d'une phrase, d'une courte histoire clinique ou d'une iconographie ;
- l'énoncé présente un seul problème à résoudre ;
- l'énoncé utilise un langage simple et clair, sans ambiguïtés ni jugement de valeur ;
- l'énoncé est formulé, autant que possible, à la forme affirmative ou interrogative ;
- les propositions sont au nombre de cinq ;
- la bonne réponse n'est pas plus longue, ni plus explicite, ni mieux construite ;
- les propositions sont homogènes dans leur contenu, leur forme et leur structure grammaticale ;
- les propositions ne sont pas synonymes, ne se chevauchent pas ;
- tous les distracteurs sont plausibles ;

- les propositions exactes et éventuellement celles indispensables ou inacceptables sont précisées ;
- les propositions suivent les conseils typographiques : les propositions sont identifiées par des majuscules suivies d'un point. Pas de majuscule au début de chaque proposition, pas de ponctuation après les propositions.

Les QI ont été introduites en 2016 en complément des DCP et des dossiers de LCA déjà présents au concours depuis 2004. Toutes les questions (DCP, LCA ou QI) de l'ECNi sont préparées en amont par des groupes d'experts de spécialité. Par la suite, les sujets sont tirés au sort à partir de cette banque confidentielle, gérée par le CNCI et indépendante du système SIDES. Au sein des facultés, les examens sous forme de QI ont été mis en application de façon variable. Certains étudiants peuvent ainsi être favorisés ou inversement défavorisés selon la fréquence respective des QI dans les examens facultaires, avec pour conséquence une préparation hétérogène et davantage d'incertitude des étudiants aux ECNi.

Notre travail avait pour objectif de réaliser une analyse de la validité du contenu et de la qualité formelle des QI des premières épreuves ECNi afin d'évaluer si le cahier des charges prédéfini avait été respecté.

Méthodes

Pour les ECNi de juin 2016, il y a eu 15 DCP de 15 questions et 3 DCP de 14 questions, soit 267 questions, et 120 QI ainsi que deux dossiers de LCA comportant 15 questions. Les dossiers des ECNi ont été récupérés *via* deux sources étudiantes : un blog indépendant intitulé « Medshake » [7] et un blog en lien avec l'Association nationale des étudiants en médecine de France (ANEMF) [8].

Nous avons réalisé une analyse de la validité et de la pertinence des QI sur la forme et le fond en se référant aux consignes publiées par le SIDES et le CNCI [5,6]. Les critères référencés en annexe ont été séparément évalués pour chaque QI selon un choix de réponse binaire (OUI / NON) ou un choix multiple selon les cas (Annexe A). La liste des critères d'évaluation (R) était numérotée de un à 18 (Annexe A). L'étude des niveaux d'habiletés cognitives s'est basée sur la taxonomie de Bloom révisée [9]. Nous avons classé le niveau cognitif sollicité pour pouvoir répondre aux QI en trois catégories (Annexe A) : 1) reconnaissance et compréhension, soit le niveau 1 « connaissance » et 2 « compréhension » du domaine cognitif de la taxonomie de Bloom révisée ; 2) application, soit le niveau 3 « application » du domaine cognitif de la taxonomie de Bloom révisée ; 3) analyse, soit les niveaux 4 « analyse », 5 « synthèse » et 6 « évaluation » du domaine cognitif de la taxonomie de Bloom révisée [9].

L'évaluation des 120 QI de l'ECNi 2016 a été réalisée séparément par deux évaluateurs spécialistes (néphrologie et rhumatologie-thérapeutique). Chaque évaluateur procédait à une analyse individuelle sans connaître les résultats de l'analyse de l'autre évaluateur. Les discordances dans les réponses du binôme ont été analysées pour

Tableau I. Résultats de l'analyse de la validité du contenu et de la qualité formelle par le binôme.

	Evaluateur 1	Evaluateur 2	Bilan après concertation
R1: Objectif apprentissage ?	111 (92,5 %)	120 (100 %)	
R2: Courte histoire clinique ?	20 (16,7 %)	25 (20,8 %)	
R3: Iconographie ?	9 (7,5 %)	7 (5,8 %)	
R4: Un seul problème ?	120 (100 %)	120 (100 %)	
R5: Langage simple ?	120 (100 %)	118 (98,3 %)	
R6: Formulation des QI	120 (100 %)	117 (97,5 %)	
R7: Propositions de 5 ?	120 (100 %)	120 (100 %)	
R8: Bonne réponse normale ?	120 (100 %)	120 (100 %)	
R9: Homogénéité des propositions ?	112 (93,3 %)	120 (100 %)	
R10: Propositions synonymes ?	120 (100 %)	118 (98,3 %)	
R11: Distracteurs ?	112 (93,3 %)	88 (73,3 %)	
R12: Niveau d'habileté intellectuelle			
Reconnaissance / compréhension	98 (81,7 %)	99 (82,5 %)	
Application des connaissances	22 (18,3 %)	20 (16,7 %)	
Analyse	0	1 (0 %)	
R13: Mono / pluri-disciplinaire	7 (5,8 %)	15 (12,5 %)	13 (10,8 %)
R14: Groupe disciplinaire			
Spécialités non-chirurgicales	86 (71,7 %)	105 (87,5 %)	87 (72,5 %)
Spécialités chirurgicales	9 (7,5 %)	4 (3,3 %)	15 (12,5 %)
Thérapeutique	13 (10,8 %)	10 (8,3 %)	13 (10,8 %)
Santé publique	10 (8,3 %)	3 (2,5 %)	10 (8,3 %)
R16: Nombre d'items > 1 ?	14 (11,7 %)	14 (11,7 %)	20 (16,7 %)
R18: Hors programme ?	5 (4,2 %)	5 (4,2 %)	5 (4,2 %)

QI: questions isolées.

chaque critère d'évaluation (R) et ont été calculées de façon binaire pour obtenir un pourcentage de discordance et une déviation standard.

Les deux examinateurs devaient se concerter pour chaque critère d'évaluation (R) lorsque les discordances sur les 120 QI étaient supérieures à 20 %. Une relecture conjointe des QI a été effectuée afin d'accorder les réponses et les homogénéiser.

La répartition des QI sur les 32 disciplines suivantes a été répertoriée (santé publique, obstétrique, pédopsychiatrie, pédiatrie, infectiologie, endocrinologie, rhumatologie, oncologie, oto-rhino-laryngologie (ORL), ophtalmologie, néphrologie, nutrition, hématologie, gynécologie, neurologie, orthopédie, hépato-gastro-entérologie, addictologie, urologie, cardiologie, psychiatrie, dermatologie, pneumologie, allergologie, chirurgie cardiaque et thoracique, réanimation, chirurgie digestive, médecine physique et réadaptation, neurochirurgie, génétique, médecine légale et médecine du travail). Certaines QI pouvaient faire intervenir plusieurs spécialités différentes.

Toutes les analyses ont été colligées sur tableur Excel puis les statistiques ont été réalisées sur Graphpad Prism version 6 et Stata v12.

Résultats

Description de l'analyse de la validité du contenu et de la qualité formelle

Les 120 QI de l'ECNi ont été analysées par deux évaluateurs indépendants, initialement sans concertation. Les résultats sont résumés dans le [tableau I](#).

Les rédacteurs ont respecté, pour la majorité des QI, les consignes formulées pour la rédaction. Cinq des 120 questions ne faisaient pas partie du référentiel et ont été considérées hors programme (questions N° 17, 18, 19, 95, 106). La majorité des questions comportait des distracteurs plausibles et adaptés (80 %), ainsi qu'une forme grammaticale ou un contenu homogène.

Les processus cognitifs requis pour pouvoir répondre à ces QI étaient très souvent basés sur la simple connaissance ou sur leur application. Une seule question a fait appel à une réelle analyse ([Tab. I](#)). L'utilisation de courtes histoires cliniques ou d'iconographie était peu souvent retrouvée (< 20 %). Sur les 22 questions ayant fait appel à une application de connaissance ou à une analyse, 11 ont utilisé une courte présentation clinique et cinq ont utilisé une iconographie.

Les questions ont majoritairement porté sur des spécialités non-chirurgicales ([Tab. I](#)). Indépendamment, certaines QI faisaient référence à l'enseignement de thérapeutique et de santé publique (respectivement 10,8 et 8,3 %). Les QI se rapportaient le plus souvent à un seul item du programme de l'ECNi. L'unité d'enseignement (UE) 8 « circulation – métabolismes » était la plus représentée avec 25,8 % des questions de l'épreuve ECNi ([Tab. II](#)).

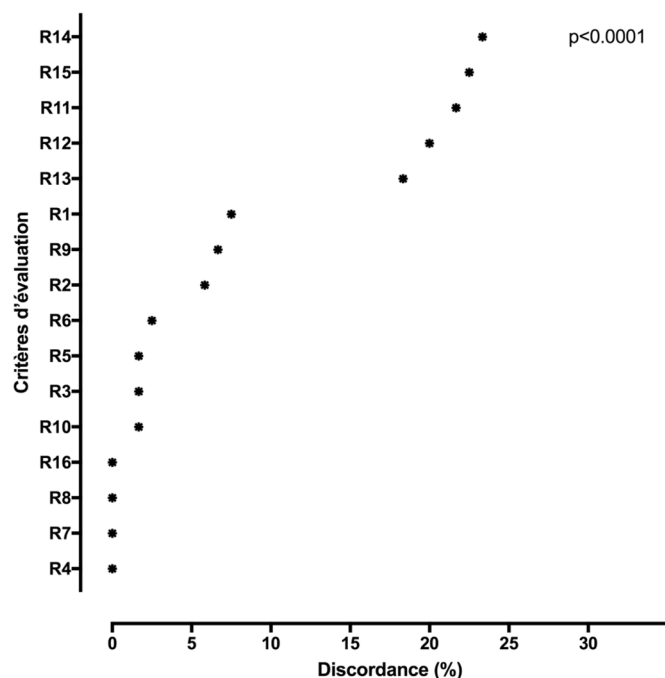
Analyse des discordances entre évaluateurs

La discordance médiane était de 6,2 % [IC95% ; 0–12,5]. La discordance en fonction des QI était variable et supérieure à 10 % pour 44 (36 %) des QI ([Fig. 1](#)). La

Tableau II. Répartition des questions isolées en fonction des unités d'enseignement du programme des épreuves classantes nationales.

	UE1	UE2	UE3	UE4	UE5	UE6	UE7	UE8	UE9	UE10	UE11
Nb QI	9	9	13	15	7	9	10	31	13	7	6
%	7,5	7,5	10,8	12,5	5,8	7,5	8,3	25,8	10,8	5,8	5,0

QI : questions isolées ; ECN : épreuves classantes nationales ; UE : unité d'enseignements ; UE1 : apprentissage de l'exercice médical et de la coopération interprofessionnelle ; UE2 : de la conception à la naissance – pathologie de la femme – hérédité – enfant – adolescent ; UE3 : maturation – vulnérabilité – santé mentale – conduites addictives ; UE4 : perception – système nerveux – revêtement cutané ; UE5 : handicap – vieillissement – dépendance – douleur – soins palliatifs – accompagnement ; UE6 : maladies transmissibles – risques sanitaires – santé au travail ; UE7 : inflammation – immunopathologie – poumon – sang ; UE8 : circulation – métabolismes ; UE9 : cancérologie – onco-hématologie ; UE10 : le bon usage du médicament et des thérapeutiques non-médicamenteuses ; UE11 : urgences et défaillances viscérales aiguës.

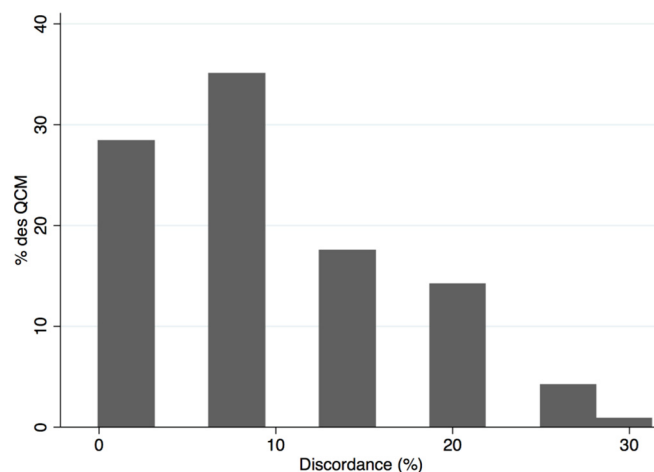
**Fig. 1.** Discordance en pourcentage en fonction des différents items (R) d'évaluation.

distribution des discordances ne suivait pas une loi normale. Le pourcentage de discordance était pour plus de 60 % des questions inférieur à 10 % (Fig. 2). Pour 5 % des QI, le pourcentage de discordance était supérieur à 30 %.

Les discordances supérieures à 20 % entre les deux examinateurs ont été retrouvées pour les items 11 à 15 (Fig. 1). Les discordances entre les deux évaluateurs portaient principalement sur l'évaluation des distracteurs (adaptés ou non) et la classification des spécialités. Pour les autres points d'évaluations, les résultats des évaluations des deux auteurs étaient sensiblement identiques. Les discordances étaient variables selon les spécialités et étaient supérieures à 10 % pour 13 des 24 spécialités.

Analyse après correction des discordances

Après l'évaluation des discordances entre les deux enseignants, les critères R13, R14, R15 et R16, ont été harmonisés car jugés correspondant aux items les plus

**Fig. 2.** Variabilité de la discordance d'appréciation entre les évaluateurs en fonction des 120 questions isolées.

discordants. Une prédominance de QI portant sur les spécialités non-chirurgicales (72,5 %) a été observée (Tab. I). Au total, 16,7 % des QI abordaient plus d'un item du programme de l'ECNi. Pour répondre aux QI, les étudiants devaient faire appel à plusieurs disciplines dans 10,8 % des cas. La synthèse de la correction du classement des UE et des disciplines sur les QI discordantes a été résumée dans le tableau III.

Trente et une question portaient sur l'UE8 (Tab. II). Plus de 10 questions concernaient respectivement l'UE3 (maturation – vulnérabilité – santé mentale – conduites addictives), l'UE4 (perception – système nerveux – revêtement cutané) et l'UE9 (cancérologie – onco-hématologie).

Sur les 32 disciplines, neuf n'ont pas donné lieu à des QI : l'allergologie, la chirurgie cardiaque et thoracique, la réanimation, la chirurgie digestive, la médecine physique et réadaptation, la neurochirurgie, la génétique, la médecine légale et la médecine du travail.

Les disciplines les plus représentées (> 5 %) pour cette épreuve ECNi 2016 (Fig. 3) étaient respectivement la santé publique (7,5 %), la pédiatrie (6,7 %), l'infectiologie (6,7 %), l'endocrinologie (8,3 %), la rhumatologie (6,7 %) l'oncologie (9,2 %), l'ophtalmologie (5,8 %), l'hématologie (8,3 %), la neurologie (6,7 %), l'urologie (6,7 %), la cardiologie (8,3 %) et la psychiatrie (5 %).

Tableau III. Correction des discordances entre évaluateurs retrouvées sur le classement des unités d'enseignement et des disciplines (spécialité(s) concernée(s) ; unités d'enseignement du programme des épreuves classantes nationales (item)).

Questions	Évaluateur 1	Évaluateur 2	Synthèse
N° 7 Diabète gestationnel	Pédiatrie + Endocrinologie ; UE8 (252)	Obstétrique ; UE2 (23)	Endocrinologie + Obstétrique ; UE2 (23) + UE8 (252)
N° 8 Développement psy moteur normal	Pédopsychiatrie ; UE3 (53)	Pédiatrie ; UE3 (53)	Pédiatrie ; UE3 (53)
N° 17 Parois orbite	Hors programme	Ophthalmologie	Hors programme
N° 18 Uvule	Hors programme	ORL ; UE4 (146)	Hors programme
N° 19 Améloblastome	ORL	Hors programme	Hors programme
N° 20 Calcémie para néoplasique	Oncologie + Rhumatologie	Oncologie	Oncologie + Rhumatologie
N° 44 Aménorrhée I	Pédiatrie + Endocrinologie ; UE2 (40)	Gynécologie ; UE2 (34)	Endocrinologie + Gynécologie ; UE2 (34,40)
N° 50 Compression nerf périph	Neurologie ; UE4 (90)	Neurologie + Orthopédie ; UE4 (94)	Neurologie + Orthopédie ; UE4 (90,94)
N° 56 Zollinger Ellison	Gastro-entérologie ; UE8 (269)	Gastro-entérologie + Oncologie ; UE8 (300)	Gastro-entérologie + Oncologie ; UE8 (269,300)
N° 57 Dépistage selles	Gastro-entérologie ; UE9 (298)	Gastro-entérologie + Oncologie ; UE9 (287)	Gastro-entérologie + Oncologie ; UE9 (287)
N° 68 Urétrite	Urologie ; UE6 (157)	Urologie + Infectieux ; UE6 (158)	Urologie + Infectieux ; UE6 (157,158)
N° 72 Schizophrénie	Psychiatrie ; UE3 (61)	Psychiatrie + Thérapeutique ; UE3 (61)	Psychiatrie ; UE3 (61)
N° 93 Hypergammaglobulinémie	Hématologie ; UE7 (208)	Hématologie + Autoimmunité ; UE7 (188)	Hématologie ; UE7 (188)
N° 94 Plaquenil	Rhumatologie ; UE7 (190)	Thérapeutique ; UE10 (319)	Rhumatologie ; UE7 (190)
N° 95 Syndrome d'activation macrophagique	Pédiatrie ; UE8 (272)	Hématologie ; UE7 (?)	Hors programme
N° 100 Rubéole congénitale	Obstétrique ; UE2 (26)	Infectieux ; UE6 (160)	Obstétrique + Infectieux ; UE2 (26) ; UE6 (160)
N° 103 Mononucléose	Hématologie ; UE7 (213)	Infectieux ; UE6 (160)	Hématologie ; UE7 (213)
N° 105 Rx Thorax	Pneumologie ; UE4 (86)	Pneumologie + ORL ; UE4 (86) ; UE8 (199)	Pneumologie + ORL ; UE4 (86) ; UE8 (199)
N° 106 EFR	Pneumologie ; UE7 (184, 205)	Pneumologie ; UE7 (204)	Hors programme
N° 118 Inhibiteur calcique	Cardiologie + Thérapeutique ; UE8 (221) ; UE10 (326)	Cardiologie ; UE8 (234)	Cardiologie + Thérapeutique ; UE8 (221,234) ; UE10 (326)

ECN : épreuves classantes nationales ; UE : unité d'enseignements ; EFR : explorations fonctionnelles respiratoires ; ORL : oto-rhino-laryngologie.

Discussion

Il s'agit de la première analyse du contenu et de la qualité formelle des QI des ECNi. Ces résultats sont particulièrement intéressants à considérer alors que les ECNi sont actuellement remises en cause par les organismes de tutelle et qu'une réforme est engagée pour leur évolution. Nous avons souhaité analyser les qualités des QI des épreuves ECNi, et examiner leur adéquation avec les objectifs annoncés de cet examen. Nous pouvons retenir un respect général des consignes de rédaction par les auteurs sur la forme. Cependant, nous avons observé que plusieurs critères de rédaction émis par la section médecine du CNCI n'étaient pas respectés (transversalité des QI, absence de raisonnement complexe, choix de certains distracteurs).

Dans l'ensemble, les QI ont majoritairement porté sur les spécialités non-chirurgicales. La thérapeutique représentait plus de 10 % des questions et les items de santé publique 8 % des questions. Notre étude a révélé une inégalité de répartition des QI avec une surreprésentation de l'UE 8 « circulation – métabolismes ». De même, certaines disciplines n'ont pas donné lieu à des QI comme l'allergologie, la chirurgie cardiaque, la réanimation, la chirurgie digestive, la médecine physique et réadaptation, la neurochirurgie, la génétique, la médecine légale et la médecine du travail. Du fait des modalités de l'examen et du tirage au sort des QI, il était attendu d'observer un déséquilibre entre les disciplines. Il est intéressant de noter que parmi les 18 DCP et les deux articles de LCA des épreuves ECNi 2016, seules trois de ces disciplines n'ont pas donné lieu à des questions (médecine

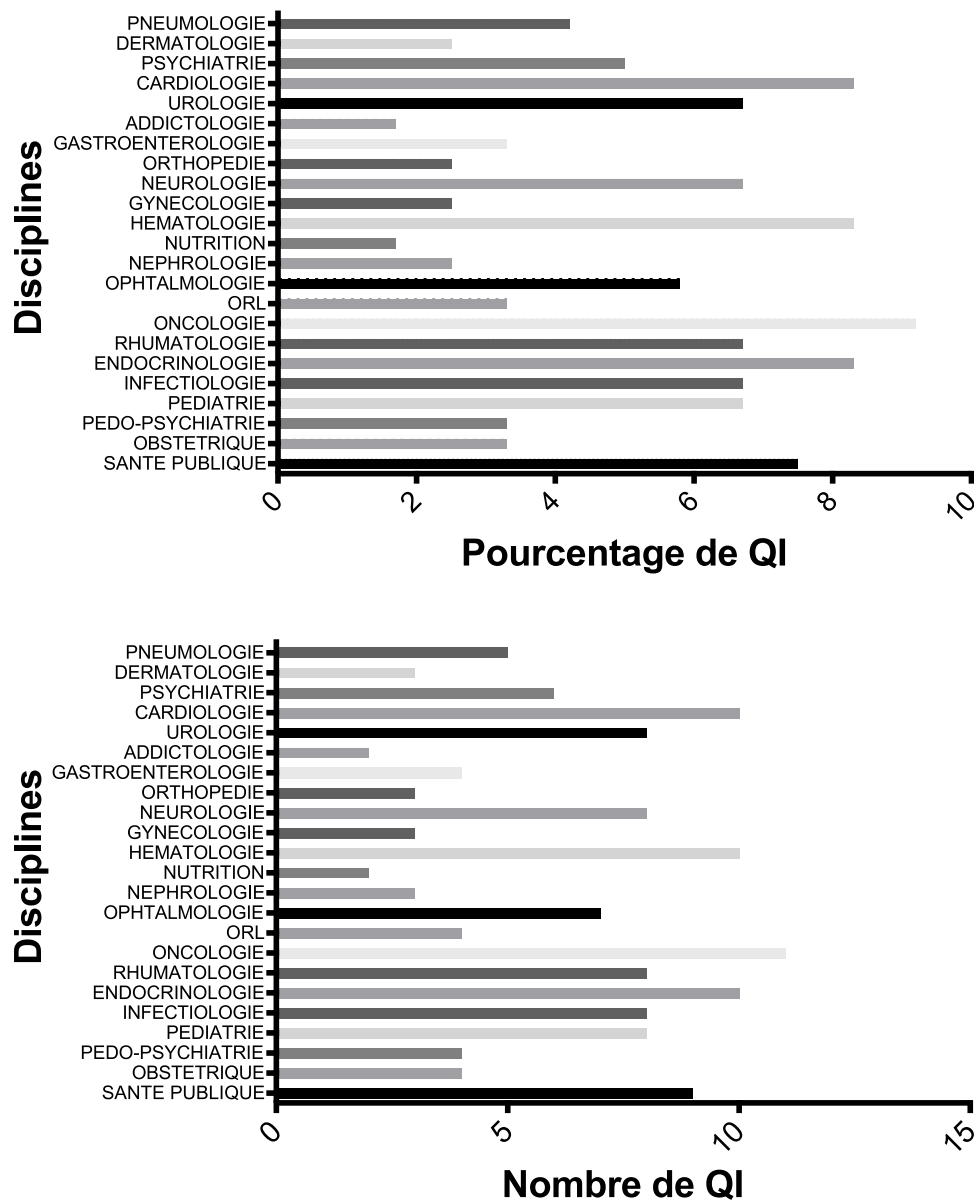


Fig. 3. Répartition des questions isolées (QI) par discipline.

Distribution des 120 QI selon les disciplines (exprimée soit en chiffre absolu, soit en pourcentage du nombre de QI). Les disciplines suivantes n'étaient pas concernées par des QI en 2015–2016 : allergologie, chirurgie cardiaque et thoracique, réanimation, chirurgie digestive, médecine physique et réadaptation, neurochirurgie, génétique, médecine légale, médecine du travail.

du travail, allergologie, chirurgie digestive). Comme attendu et même si l'on peut le regretter, nous n'avons pas observé de transversalité au sein des QI puisque plus de 80 % des QI correspondaient à une discipline unique.

L'évaluation des niveaux d'habiletés cognitives requis pour répondre aux QI, même si celles-ci sont parfois difficiles à catégoriser, était importante à évaluer. Sont-ils représentatifs des objectifs attendus des étudiants en fin de sixième année? À l'aide de la classification révisée de Bloom [9], nous avons séparé en trois catégories la mobilisation cognitive requise pour répondre aux questions : reconnaissance et compréhension ; application ; analyse (Annexe A). Nous avons retrouvé une prédomi-

nance de l'utilisation de la reconnaissance ou de la compréhension simple (80 %) alors qu'environ 20 % des QI faisaient appel à l'application de ces connaissances. Aucun des QI ne faisait vraiment appel à une capacité d'analyse selon notre définition. Cet item a cependant été source de discordances de plus de 20 % entre les deux évaluateurs. Cet aspect avait été préalablement remarqué par les auteurs de l'Université de Bordeaux sur l'analyse du contenu des ECNi 2016 [10]. Il s'agissait d'une analyse documentaire des DCP et des QI, plus focalisée sur le fond, et réalisée par un expert de chaque discipline. Ils ont analysé les questions selon quatre critères : la pertinence des questions (pertinentes / discutables / non pertinentes),

leurs difficultés (facile / difficile), le niveau d'étude requis (deuxième cycle / troisième cycle spécialisé) et le niveau cognitif pour y répondre (mémoire seule / réflexion simple / mémorisation complexe). Leur étude a utilisé une définition du niveau cognitif assez proche de la nôtre. Elle a mis en évidence que l'utilisation du simple rappel de connaissances déclaratives était observé dans 62 % des cas et que 38 % des QI faisaient appel à une réflexion simple. Comparativement, l'appel à la « mémoire pure » était moins souvent utilisé au cours des DCP (45 %). Ce système d'évaluation récompense un apprentissage théorique plutôt que pratique et n'incite pas les étudiants à s'investir dans leurs stages hospitaliers. En parallèle, l'apprentissage au lit du patient par le compagnonnage dans les premiers gestes comme dans les relations humaines se réduit. Ainsi, il nous paraît indispensable de réintroduire une démarche diagnostique et un raisonnement clinique dans le cursus de l'étudiant en médecine. Le contenu des QI et des DCP devrait contenir plus de questions de réflexion, de raisonnement complexe et possiblement de prise de décision clinique. Un examen pratique de mise en situation (comme le certificat de compétence clinique, envisagé par la nouvelle réforme des études de médecine en France à l'image de ce qui se fait dans les pays anglo-saxons depuis plusieurs années) pourrait être utile pour une prise en charge médicale efficiente, pragmatique, transversale et factuelle des patients.

La méthode que nous avons retenue, avec une analyse réalisée en binôme, nous a permis de mettre en évidence des discordances importantes sur certains critères d'évaluation, notamment la pertinence des distracteurs (R11), le niveau cognitif requis pour répondre à la QI (R12), l'approche mono ou pluridisciplinaire (R13), le groupe disciplinaire exigé pour répondre à la QI (R14) et l'UE permettant de répondre à la QI (R15). Ces critères d'évaluation (R11 à R15) étaient plus ambigus et subjectifs et d'autre part étaient liés entre eux, ce qui peut expliquer les discordances plus importantes. Ces discordances mettent en évidence le caractère subjectif de l'interprétation des QI, probablement du fait de la transversalité du système par UE qui peut entraîner des redondances d'informations pour les pathologies au croisement de plusieurs spécialités. Cette hiérarchisation par UE pourrait également être responsable des cinq questions jugées hors programme. Nous avons analysé les causes de ces discordances, et corrigé les items les plus discordants permettant d'homogénéiser les réponses. La question de la pertinence des distracteurs (R11) est complexe. Dans l'étude bordelaise [10], l'analyse des QI/DCP a été réalisée par des experts de chaque discipline ce qui peut augmenter la pertinence, et la précision de l'évaluation car ils connaissaient les prérequis demandés de leur spécialité pour répondre aux questions. Ils ont observé que 30 % des distracteurs étaient litigieux et non crédibles dans l'ensemble des questions intégrant DCP et QI. L'étude ne séparait pas les résultats des QI/DCP et ne nous a pas permis de comparer nos résultats. Ce critère d'évaluation reste également aléatoire comme l'attestent les discordances rapportées par les deux

évaluateurs dans notre étude (respectivement 7 et 27 % de distracteurs non pertinents). Le choix de « bons » distracteurs est difficile. Ils doivent être plausibles mais incontestablement faux. Ils devraient également présenter la même probabilité d'être choisis afin d'éviter les réponses par élimination. Nous n'avions pas, au préalable, appréhendé ce critère d'évaluation de façon concertée ce qui aurait été probablement nécessaire pour diminuer la subjectivité. Il est également possible que le jugement de la pertinence des distracteurs ait pu être erroné par méconnaissance de la question. Par conséquent, il nous paraît nécessaire que l'analyse de ce critère de jugement se fasse par plusieurs experts de la même spécialité et avec une concertation préalable à l'évaluation, bien qu'il existe des contraintes importantes à une telle organisation.

L'ancienne version des ECN au format écrit, malgré ses inconvénients, semblait davantage utile pour évaluer la démarche diagnostique de l'étudiant comme l'atteste l'étude de Sananes *et al.* [11] sur la validité et la conformité formelle du contenu des ECN de 2004 à 2011. Dans cette étude, ont été colligés et analysés 72 dossiers et 572 questions. Les auteurs n'ont retrouvé que 12 questions de mémorisation pure (soit 2 %) et une réelle réflexion de l'étudiant était nécessaire dans au moins 78 % des dossiers. De même, une véritable stratégie paraclinique était demandée dans 68 % des dossiers. De nombreux éléments de « savoir-faire », au sens de la définition formulée par le CNCI, étaient également évalués dans les ECN. Ainsi plus d'un tiers des dossiers nécessitait l'interprétation de résultats biologiques et plus d'un dossier sur deux celle d'une imagerie ou d'autres examens paracliniques, simulant une réelle démarche diagnostique. Par ailleurs, un tiers des dossiers évaluait la capacité des étudiants à définir des priorités et près de trois quarts exploraient certains aspects d'une prise en charge globale. Enfin, l'accent était mis sur l'information que le médecin devait être capable de donner à son patient dans différentes situations ainsi que sur certaines mesures médico-sociales. Les QI sont un système de tests rapides et faciles à analyser, adaptés pour un mode d'évaluation des connaissances, de la compréhension, ainsi que de l'application de ces connaissances, mais qui apparaît en réalité comme un frein au développement de la pensée et du raisonnement individuel [12]. On peut également leur reprocher une parcellisation des connaissances, une forme de conditionnement et une intervention du hasard. Notre étude en est l'illustration mais il existe des moyens pour éviter cette dérive. Comme cela a été observé dans notre étude, des courtes histoires cliniques et/ou des iconographies étaient plus souvent associées à un processus d'analyse. Cela suggère l'intérêt d'utiliser des QI contenant des vignettes cliniques ou des iconographies afin d'améliorer l'apprentissage au raisonnement clinique. L'utilisation d'outils numériques permettrait également de vérifier les capacités d'analyse de l'étudiant, ainsi que les tests de concordance de script qui prennent en compte l'incertitude. Enfin, les champs de compétence hors savoirs cognitifs tels que le savoir-faire et le savoir-être semblent difficilement évaluables par les QI. Des

questions techniques sur les différentes étapes d'un savoir-faire sont cependant possibles et devraient être promues dans les examens futurs.

Notre travail comporte certaines limites. Les QI des épreuves « tests » n'ont pas été analysées car nous n'avons pas pu y avoir accès. Ces QI sont également issus de la banque du CNCI et leur analyse aurait permis une vision plus exhaustive et approfondie, notamment sur la fréquence respective des disciplines. Il aurait été utile d'étendre notre méthode d'analyse au contenu des DCP et des LCA afin de les comparer aux QI, à l'exemple de l'étude Bordelaise [10]. Notre travail n'aborde pas la problématique de la valeur discriminative de cet examen pour la sélection des « bons » médecins et/ou de la pertinence du classement eu égard à la valeur supposée de l'étudiant. Compte tenu des résultats de notre étude, il nous semblerait intéressant de mettre en place avant les examens une évaluation systématique approfondie des QI. Celle-ci aurait pour objectif d'améliorer la pertinence des distracteurs et de favoriser le raisonnement. Il serait également utile que les enseignants soient formés et sensibilisés à la rédaction de QI comportant des vignettes cliniques ou des iconographies afin d'améliorer la qualité pédagogique de leurs QI. Nous pouvons citer quelques obstacles à ces propositions. En effet, la confidentialité des épreuves ne permet pas de faire circuler l'ensemble des QI à tous les experts et limite la possibilité d'évaluation préalable des QI. Il faut également souligner la difficulté de création d'un examen répondant à tous les critères prédéfinis, tant du point de vue de la réflexion que du coût et de la logistique.

Conclusion

Nous avons effectué la première analyse du contenu et de la qualité formelle du système des QI complémentaire des DCP introduits aux ECNi en 2016. Cette analyse a été réalisée par un binôme indépendant. Il est ressorti de ce travail que les QI ne semblent pas permettre dans l'état actuel une évaluation optimale des connaissances de l'étudiant en fin de sixième année de médecine du fait du caractère limité à la restitution simple de connaissance. Les rédacteurs de QI devraient davantage respecter les consignes de la section médecine du CNCI. Les QI sont par contre formulées de manière adéquate. Certaines spécialités sont surreprésentées au détriment d'autres parfois absentes. Il semblerait utile de favoriser les courtes vignettes cliniques ainsi que l'iconographie pour améliorer la qualité de réflexion demandée aux étudiants. Les outils numériques ou les tests de concordance pourraient également élargir le champ des compétences évaluées.

Contributions

Yves-Marie Pers a participé à la conception du protocole, au recueil des données, à l'interprétation des résultats et à l'écriture du manuscrit. Clémentine Marais, Philippe Lambert et Claude Jeandel ont participé à

l'interprétation des résultats et à la relecture critique du manuscrit. Gérald Chanques a participé à la relecture critique du manuscrit. Ilan Szwarc a participé à l'analyse statistique, à l'interprétation des résultats et à la relecture critique du manuscrit. Jean-Luc Faillie a participé à la conception du protocole, à l'interprétation des résultats et à la relecture critique du manuscrit. Olivier Mathieu a participé à l'interprétation des résultats, relecture critique du manuscrit. Moglie Le Quintrec a participé au recueil des données, à l'interprétation des résultats, à l'analyse statistique et à l'écriture du manuscrit.

Liens d'intérêts

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

Approbation éthique

Non sollicitée.

Remerciements

Nous remercions pour leur soutien les différentes ressources du Diplôme d'université de pédagogie médicale de la Faculté de médecine de Montpellier-Nîmes.

Annexe A: Critères d'évaluation utilisés dans l'analyse du contenu et de la qualité formelle des QI des épreuves ECNi 2015–2016.

Réponse binaire sur la formulation des QI :

R1. La question mesure un objectif important d'apprentissage.

R2. L'énoncé est constitué d'une courte histoire clinique.

R3. L'énoncé est constitué d'une iconographie.

R4. L'énoncé présente un seul problème à résoudre.

R5. L'énoncé utilise un langage simple et clair, sans ambiguïtés ni jugement de valeur.

R6. L'énoncé est formulé à la forme affirmative ou interrogative.

R7. Les propositions sont au nombre de 5.

R8. La bonne réponse n'est pas plus longue, ni plus explicite, ni mieux construite.

R9. Les propositions sont homogènes dans leur contenu, leur forme et leur structure grammaticale.

R10. Les propositions ne sont pas synonymes, ne se chevauchent pas.

R11. Utilisation de distracteurs plausibles.

Réponse binaire ou à choix multiples sur le contenu des QI :

R12. Les niveaux d'habileté intellectuelle requis pour pouvoir répondre à la QI :

– reconnaissance et compréhension (définis par la faculté à puiser dans sa mémoire à long terme des données factuelles pour répondre à la QI ; établir des liens entre des connaissances et un nouveau fait) ;

- application (définie par le fait d'appliquer ses connaissances et sa compréhension à un exercice pratique en transférant une procédure apprise à une tâche) ;
- analyse (défini par la capacité à distinguer, organiser des idées afin de déterminer un choix).

R13. Approche monodisciplinaire ou pluridisciplinaire: la réponse à la question nécessite (ou non) des disciplines différentes.

R14. Groupe disciplinaire exigé pour répondre à la question: médicale, chirurgicale, thérapeutique, santé publique.

R15. Quelle UE permet de répondre à la question ? (Choix de 1 à 12).

R16. Le nombre d'items ECN abordés: 1 ou > 1.

R17. Nom de la discipline concernée par la QI.

R18. La question est hors programme car elle ne se réfère pas au référentiel.

Références

1. Jolly D, Lorette G, Ambrosi P, Dreyfuss D, Chaffanjon P, Le Jeune C *et al.* Résultats des épreuves classantes nationales (ECN) 2012. *Presse Med* 2013;42:1138-40.
2. Arrêté fixant les nouvelles modalités de l'ECN (version consolidée au 24 février 2016). 2016. Disponible sur https://www.legifrance.gouv.fr/affichTexte.do?sessionId=048B3DD8D62F3F9E4DF73C2C76208836.tpdila08v_2?cidTexte=JORFTEXT000030916532&dateTexte=20160224.
3. Le Jeune C. Les ECNi bilan de la 1^e épreuve les différences en 2017. DIU de pédagogie médicale 20 janvier 2017. 2017. Disponible sur http://www.chups.jussieu.fr/diupmweb/2017/23janv2017/ECN_bilan.pdf.
4. Roux D, Meyer G, Cymbalista F, Bouaziz J-D, Falgarone G, Tesniere A *et al.* Épreuves informatisées de type ECNi: retour d'expérience sur un test multiuniversité. *Rev Med Interne* 2016;37:166-172.
5. Conseil scientifique du CNCI. Section médecine. Conseils pour l'élaboration des dossiers de l'ECN. 2004. Disponible sur <http://www.cnci.univ-paris5.fr/medecine/ConseilAuteur.pdf>.
6. SIDES. Les épreuves classantes nationales informatisées. 2014. Disponible sur <http://side-sante.org/>.
7. Blog Medshake. ECNi. Les sujets 2016. 2016. Disponible sur <https://www.medshake.net/blog/a114/ecni-les-sujets-2016/>.
8. E-Carabin. ECNi. Les sujets 2016. 2016. Disponible sur <https://e-carabin.anemf.org/>.
9. Anderson LW, Krathwohl DR, Bloom BS. A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. New York, NY: Longman, 2001.
10. Rivière E, Quinton A, Neau D, Constans J, Vignes JR, Dehail P. Analyse docimologique des premières épreuves classantes nationales informatisées (ECNi) de 2016: propositions d'amélioration. *Rev Med Interne* 2019;40:47-51.
11. Sananes N, Langer B, Patris M, Pottecher T, Andrès E. Analyse de la validité de contenu et de la qualité formelle des épreuves classantes nationales administrées en France de 2004 à 2011. *Pédagogie Médicale* 2013;14:255-267.
12. Bouvy T, Warnier L. Évaluer les acquis des étudiants à l'aide de QCM. 2016. Disponible sur https://cdn.uclouvain.be/public/Exports%20reddot/ipm/documents/VADEMECUM_Mars_2016.pdf.

Citation de l'article : Pers Y.-M., Marais C., Lambert P., Jeandel C., Chanques G., Szwarc I., Faillie J.-L., Mathieu O., Le Quintrec M. 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* 2018;19:171-180