

Tests de concordance de scripts en évaluation sommative en internat de médecine générale à Rouen : fidélité inter-juge et réflexions sur l'élaboration du test

Aliénor Broutin Moussette

► To cite this version:

Aliénor Broutin Moussette. Tests de concordance de scripts en évaluation sommative en internat de médecine générale à Rouen : fidélité inter-juge et réflexions sur l'élaboration du test. Médecine humaine et pathologie. 2016. <dumas-01464604>

HAL Id: dumas-01464604

<https://dumas.ccsd.cnrs.fr/dumas-01464604>

Submitted on 10 Feb 2017

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.

FACULTÉ MIXTE DE MÉDECINE ET DE PHARMACIE DE ROUEN

ANNÉE 2016

N°

**THÈSE POUR LE
DOCTORAT EN MÉDECINE**

(Diplôme d'État)

PAR

MOUSSETTE Aliénor

NÉE LE 08/08/1984 À ROUEN

PRÉSENTÉE ET SOUTENUE PUBLIQUEMENT LE 01/12/2016

**TESTS DE CONCORDANCE DE SCRIPTS EN ÉVALUATION
SOMMATIVE EN INTERNAT DE MÉDECINE GÉNÉRALE À
ROUEN : FIDÉLITE INTER-JUGE ET RÉFLEXIONS SUR
L'ÉLABORATION DU TEST**

PRÉSIDENT DU JURY : Pr Jean-Loup HERMIL

MEMBRES DU JURY : Pr Louis SIBERT, Dr Elisabeth MAUVIARD

DIRECTRICE DE THÈSE : Dr Yveline SEVRIN

Par délibération en date du 3 mars 1967, la faculté a arrêté que les opinions émises dans les dissertations qui lui seront présentées doivent être considérées comme propres à leurs auteurs et qu'elle n'entend leur donner aucune approbation ni improbation.

ANNEE UNIVERSITAIRE 2016 - 2017
U.F.R. DE MEDECINE ET DE-PHARMACIE DE ROUEN

DOYEN : **Professeur Pierre FREGER**

ASSESEURS : **Professeur Michel GUERBET**
Professeur Benoit VEBER
Professeur Pascal JOLY
Professeur Stéphane MARRET

I - MEDECINE

PROFESSEURS DES UNIVERSITES – PRATICIENS HOSPITALIERS

Mr Frédéric ANSELME	HCN	Cardiologie
Mme Isabelle AUQUIT AUCKBUR	HCN	Chirurgie plastique
Mr Fabrice BAUER	HCN	Cardiologie
Mme Soumeya BEKRI	HCN	Biochimie et biologie moléculaire
Mr Ygal BENHAMOU	HCN	Médecine interne
Mr Jacques BENICHOU	HCN	Bio statistiques et informatique médicale
Mr Olivier BOYER	UFR	Immunologie
Mr François CARON	HCN	Maladies infectieuses et tropicales
Mr Philippe CHASSAGNE (<i>détachement</i>)	HCN	Médecine interne (gériatrie) – Détachement
Mr Vincent COMPERE	HCN	Anesthésiologie et réanimation chirurgicale
Mr Jean-Nicolas CORNU	HCN	Urologie
Mr Antoine CUVELIER	HB	Pneumologie
Mr Pierre CZERNICHOW (<i>surnombre</i>)	HCH	Epidémiologie, économie de la santé
Mr Jean-Nicolas DACHER	HCN	Radiologie et imagerie médicale
Mr Stéfan DARMONI	HCN	Informatique médicale et techniques de communication
Mr Pierre DECHELOTTE	HCN	Nutrition
Mr Stéphane DERREY	HCN	Neurochirurgie
Mr Frédéric DI FIORE	CB	Cancérologie
Mr Fabien DOGUET	HCN	Chirurgie Cardio Vasculaire

Mr Jean DOUCET	SJ	Thérapeutique - Médecine interne et gériatrie
Mr Bernard DUBRAY	CB	Radiothérapie
Mr Philippe DUCROTTE	HCN	Hépatogastro-entérologie
Mr Frank DUJARDIN	HCN	Chirurgie orthopédique - Traumatologie
Mr Fabrice DUPARC	HCN	Anatomie - Chirurgie orthopédique et traumatologique
Mr Eric DURAND	HCN	Cardiologie
Mr Bertrand DUREUIL	HCN	Anesthésiologie et réanimation chirurgicale
Mme Hélène ELTCHANINOFF	HCN	Cardiologie
Mr Thierry FREBOURG	UFR	Génétique
Mr Pierre FREGER	HCN	Anatomie - Neurochirurgie
Mr Jean François GEHANNO	HCN	Médecine et santé au travail
Mr Emmanuel GERARDIN	HCN	Imagerie médicale
Mme Priscille GERARDIN	HCN	Pédopsychiatrie
Mr Michel GODIN (<i>surnombre</i>)	HB	Néphrologie
M. Guillaume GOURCEROL	HCN	Physiologie
Mr Dominique GUERROT	HCN	Néphrologie
Mr Olivier GULLIN	HCN	Psychiatrie Adultes
Mr Didier HANNEQUIN	HCN	Neurologie
Mr Fabrice JARDIN	CB	Hématologie
Mr Luc-Marie JOLY	HCN	Médecine d'urgence
Mr Pascal JOLY	HCN	Dermato – Vénérologie
Mme Bouchra LAMIA	Havre	Pneumologie
Mme Annie LAQUERRIERE	HCN	Anatomie et cytologie pathologiques
Mr Vincent LAUDENBACH	HCN	Anesthésie et réanimation chirurgicale
Mr Joël LECHEVALLIER	HCN	Chirurgie infantile
Mr Hervé LEFEBVRE	HB	Endocrinologie et maladies métaboliques
Mr Thierry LEQUERRE	HB	Rhumatologie
Mme Anne-Marie LEROI	HCN	Physiologie
Mr Hervé LEVESQUE	HB	Médecine interne
Mme Agnès LIARD-ZMUDA	HCN	Chirurgie Infantile
Mr Pierre Yves LITZLER	HCN	Chirurgie cardiaque
Mr Bertrand MACE	HCN	Histologie, embryologie, cytogénétique
M. David MALTETE	HCN	Neurologie
Mr Christophe MARGUET	HCN	Pédiatrie
Mme Isabelle MARIE	HB	Médecine interne
Mr Jean-Paul MARIE	HCN	Oto-rhino-laryngologie
Mr Loïc MARPEAU	HCN	Gynécologie - Obstétrique
Mr Stéphane MARRET	HCN	Pédiatrie

Mme Véronique MERLE	HCN	Epidémiologie
Mr Pierre MICHEL	HCN	Hépatogastro-entérologie
M. Benoit MISSET	HCN	Réanimation Médicale
Mr Jean-François MUIR (<i>sumombre</i>)	HB	Pneumologie
Mr Marc MURAINÉ	HCN	Ophthalmologie
Mr Philippe MUSETTE	HCN	Dermatologie - Vénérologie
Mr Christophe PEILLON	HCN	Chirurgie générale
Mr Christian PFISTER	HCN	Urologie
Mr Jean-Christophe PLANTIER	HCN	Bactériologie - Virologie
Mr Didier PLISSONNIER	HCN	Chirurgie vasculaire
Mr Gaëtan PREVOST	HCN	Endocrinologie
Mr Jean-Christophe RICHARD (<i>détachement</i>)	HCN	Réanimation médicale - Médecine d'urgence
Mr Vincent RICHARD	UFR	Pharmacologie
Mme Nathalie RIVES	HCN	Biologie du développement et de la reproduction
Mr Horace ROMAN	HCN	Gynécologie - Obstétrique
Mr Jean-Christophe SABOURIN	HCN	Anatomie - Pathologie
Mr Guillaume SAVOYE	HCN	Hépatogastrologie
Mme Céline SAVOYE-COLLET	HCN	Imagerie médicale
Mme Pascale SCHNEIDER	HCN	Pédiatrie
Mr Michel SCOTTE	HCN	Chirurgie digestive
Mme Fabienne TAMION	HCN	Thérapeutique
Mr Luc THIBERVILLE	HCN	Pneumologie
Mr Christian THUILLEZ (<i>sumombre</i>)	HB	Pharmacologie
Mr Hervé TILLY	CB	Hématologie et transfusion
M. Gilles TOURNEL	HCN	Médecine Légale
Mr Olivier TROST	HCN	Chirurgie Maxillo-Faciale
Mr Jean-Jacques TUECH	HCN	Chirurgie digestive
Mr Jean-Pierre VANNIER (<i>sumombre</i>)	HCN	Pédiatrie génétique
Mr Benoît VEBER	HCN	Anesthésiologie - Réanimation chirurgicale
Mr Pierre VERA	CB	Biophysique et traitement de l'image
Mr Eric VERIN	HB	Service Santé Réadaptation
Mr Eric VERSPYCK	HCN	Gynécologie obstétrique
Mr Olivier VITTECOQ	HB	Rhumatologie
Mr Jacques WEBER	HCN	Physiologie

MAITRES DE CONFERENCES DES UNIVERSITES – PRATICIENS HOSPITALIERS

Mme Noëlle BARBIER-FREBOURG	HCN	Bactériologie – Virologie
Mme Carole BRASSE LAGNEL	HCN	Biochimie
Mme Valérie BRIDOUX HUYBRECHTS	HCN	Chirurgie Vasculaire
Mr Gérard BUCHONNET	HCN	Hématologie
Mme Mireille CASTANET	HCN	Pédiatrie
Mme Nathalie CHASTAN	HCN	Neurophysiologie
Mme Sophie CLAEYSSENS	HCN	Biochimie et biologie moléculaire
Mr Moïse COEFFIER	HCN	Nutrition
Mr Manuel ETIENNE	HCN	Maladies infectieuses et tropicales
Mr Serge JACQUOT	UFR	Immunologie
Mr Joël LADNER	HCN	Epidémiologie, économie de la santé
Mr Jean-Baptiste LATOUCHE	UFR	Biologie cellulaire
Mr Thomas MOUREZ	HCN	Virologie
Mme Muriel QUILLARD	HCN	Biochimie et biologie moléculaire
Mme Laëtitia ROLLIN	HCN	Médecine du Travail
Mr Mathieu SALAUN	HCN	Pneumologie
Mme Pascale SAUGIER-VEBER	HCN	Génétique
Mme Anne-Claire TOBENAS-DUJARDIN	HCN	Anatomie
Mr David WALLON	HCN	Neurologie

PROFESSEUR AGREGE OU CERTIFIE

Mr Thierry WABLE	UFR	Communication
-------------------------	-----	---------------

II - PHARMACIE

PROFESSEURS

Mr Thierry BESSON	Chimie Thérapeutique
Mr Jean-Jacques BONNET	Pharmacologie
Mr Roland CAPRON (PU-PH)	Biophysique
Mr Jean COSTENTIN (Professeur émérite)	Pharmacologie
Mme Isabelle DUBUS	Biochimie
Mr Loïc FAVENNEC (PU-PH)	Parasitologie
Mr Jean Pierre GOULLE (Professeur émérite)	Toxicologie
Mr Michel GUERBET	Toxicologie
Mme Isabelle LEROUX - NICOLLET	Physiologie
Mme Christelle MONTEIL	Toxicologie
Mme Martine PESTEL-CARON (PU-PH)	Microbiologie
Mme Elisabeth SEGUIN	Pharmacognosie
Mr Rémi VARIN (PU-PH)	Pharmacie clinique
Mr Jean-Marie VAUGEUIS	Pharmacologie
Mr Philippe VERITE	Chimie analytique

MAITRES DE CONFERENCES

Mme Cécile BARBOT	Chimie Générale et Minérale
Mr Jérémy BELLIEN (MCU-PH)	Pharmacologie
Mr Frédéric BOUNOURE	Pharmacie Galénique
Mr Abdeslam CHAGRAOUI	Physiologie
Mme Camille CHARBONNIER	Statistiques
Mme Marie Catherine CONCE-CHEMTOB	Législation pharmaceutique et économie de la santé
Mme Elizabeth CHOSSON	Botanique
Mme Cécile CORBIERE	Biochimie
Mr Eric DITTMAR	Biophysique
Mme Nathalie DOURMAP	Pharmacologie
Mme Isabelle DUBUC	Pharmacologie
Mme Dominique DUTERTE- BOUCHER	Pharmacologie

Mr Abdelhakim ELOMRI	Pharmacognosie
Mr François ESTOUR	Chimie Organique
Mr Gilles GARGALA (MCU-PH)	Parasitologie
Mme Najla GHARBI	Chimie analytique
Mme Marie-Laure GROULT	Botanique
Mr Hervé HUE	Biophysique et mathématiques
Mme Laetitia LE GOFF	Parasitologie - Immunologie
Mme Hong LU	Biologie
Mme Marine MALLETER	Biologie Cellulaire
Mme Sabine MENAGER	Chimie organique
Mme Tiphaine ROGEZ-FLORENT	Chimie analytique
Mr Mohamed SKIBA	Pharmacie galénique
Mme Malika SKIBA	Pharmacie galénique
Mme Christine THARASSE	Chimie thérapeutique
Mr Frédéric ZIEGLER	Biochimie

PROFESSEURS ASSOCIES

Mme Cécile GUERARD-DETUNCQ	Pharmacie officinale
Mr Jean-François HOUIVET	Pharmacie officinale

PROFESSEUR CERTIFIE

Mme Mathilde GUERIN	Anglais
----------------------------	---------

ASSISTANT HOSPITALO-UNIVERSITAIRE

Mme Sandrine DAHYOT	Bactériologie
----------------------------	---------------

ATTACHES TEMPORAIRES D'ENSEIGNEMENT ET DE RECHERCHE

Mme Hanane GASMI	Galénique
Mme Benedetta CORNELIO	Chimie organique

LISTE DES RESPONSABLES DES DISCIPLINES PHARMACEUTIQUES

Mme Cécile BARBOT	Chimie Générale et minérale
Mr Thierry BESSON	Chimie thérapeutique
Mr Roland CAPRON	Biophysique
Mme Marie-Catherine CONCE-CHEMTOB	Législation et économie de la santé
Mme Elisabeth CHOSSON	Botanique
Mr Jean-Jacques BONNET	Pharmacodynamie
Mme Isabelle DUBUS	Biochimie
Mr Loïc FAVENNEC	Parasitologie
Mr Michel GUERBET	Toxicologie
Mr François ESTOUR	Chimie organique
Mme Isabelle LEROUX-NICOLLET	Physiologie
Mme Martine PESTEL-CARON	Microbiologie
Mme Elisabeth SEGUIN	Pharmacognosie
Mr Mohamed SKIBA	Pharmacie galénique
Mr Rémi VARIN	Pharmacie clinique
Mr Philippe VERITE	Chimie analytique

III – MEDECINE GENERALE

PROFESSEUR

Mr Jean-Loup **HERMIL** UFR Médecine générale

PROFESSEURS ASSOCIES A MI-TEMPS

Mr Emmanuel **LEFEBVRE** UFR Médecine Générale

Mme Elisabeth **MAUVIARD** UFR Médecine générale

Mr Philippe **NGUYEN THANH** UFR Médecine générale

MAITRE DE CONFERENCES ASSOCIE A MI-TEMPS

Mr Pascal **BOULET** UFR Médecine générale

Mr Emmanuel **HAZARD** UFR Médecine Générale

Mme Lucile **PELLERIN** UFR Médecine générale

Mme Yveline **SEVRIN** UFR Médecine générale

Mme Marie Thérèse **THUEUX** UFR Médecine générale

ENSEIGNANTS MONO-APPARTENANTS

PROFESSEURS

Mr Serguei FETISSOV (med)	Physiologie (ADEN)
Mr Paul MULDER (phar)	Sciences du Médicament
Mme Su RUAN (med)	Génie Informatique

MAITRES DE CONFERENCES

Mr Sahil ADRIOUCH (med)	Biochimie et biologie moléculaire (Unité Inserm 905)
Mme Gaëlle BOUGEARD-DENOYELLE (med)	Biochimie et biologie moléculaire (UMR 1079)
Mme Carine CLEREN (med)	Neurosciences (Néovasc)
M. Sylvain FRAINEAU (phar)	Physiologie (Inserm U 1096)
Mme Pascaline GAILDRAT (med)	Génétique moléculaire humaine (UMR 1079)
Mr Nicolas GUEROUT (med)	Chirurgie Expérimentale
Mme Rachel LETELLIER (med)	Physiologie
Mme Christine RONDANINO (med)	Physiologie de la reproduction
Mr Antoine OUVRARD-PASCAUD (med)	Physiologie (Unité Inserm 1076)
Mr Frédéric PASQUET	Sciences du langage, orthophonie
Mme Isabelle TOURNIER (med)	Biochimie (UMR 1079)

CHEF DES SERVICES ADMINISTRATIFS : Mme Véronique DELAFONTAINE

HCN - Hôpital Charles Nicolle

HB - Hôpital de BOIS GUILLAUME

CB - Centre Henri Becquerel

CHS - Centre Hospitalier Spécialisé du Rouvray

CRMPR - Centre Régional de Médecine Physique et de Réadaptation

SJ – Saint Julien Rouen

REMERCIEMENTS :

A Monsieur le Professeur Jean-Loup HERMIL,

Merci de me faire l'honneur de présider le jury de cette thèse. Merci également pour le travail que vous accomplissez au quotidien dans l'accompagnement des internes de médecine générale au travers de leur cursus.

A Monsieur le Professeur Louis SIBERT,

Merci d'avoir accepté de juger mon travail et de faire partie de mon jury de thèse. Merci pour les différentes études auxquelles vous avez participé concernant les TCS et qui m'ont valu quelques heures de lecture.

A Madame le Dr Yveline SEVRIN,

Merci d'avoir accepté d'être ma directrice de thèse, de m'avoir guidée tout au long de ce travail. Merci de ta disponibilité et de ton investissement, du travail que tu accomplis au sein du département de médecine générale. Merci de m'avoir fait confiance au cabinet et de m'avoir confié pour quelques semaines tes patients. Merci de m'avoir appris à travers ce remplacement la méticulosité du suivi des patients.

A Madame le Dr Elisabeth MAUVIARD,

Merci d'avoir accepté de faire partie de mon jury de thèse, de m'avoir conseillée et recadrée parfois au fil de mon travail. Merci pour le travail que vous effectuez au sein du département de médecine générale. Merci pour les pauses café et les conseils de prise en charge des patients au cabinet.

A Monsieur le Dr Philippe NGUYEN-THANH,

Merci d'avoir accepté d'assister à ma soutenance et d'apporter votre expérience sur le sujet concerné par cette thèse.

A la mémoire de Monsieur le Dr Francis ROUSSEL,

Ses travaux sur les TCS à Rouen ont constitué le point de départ de mon étude. J'aurais aimé bénéficier davantage de ses connaissances et de son expertise. J'espère qu'il aurait approuvé mon travail.

À ma famille,

À mes parents, qui m'ont toujours poussée dans les études sans pour autant m'opprimer et en me laissant profiter de ma jeunesse. Merci d'avoir su trouver cet équilibre, et de m'avoir fait confiance. Merci d'être toujours présents pour moi et pour vos petits-enfants, je vous aime.

À ma ptite sœur, je suis fière de toi, et même si on ne se voit pas assez à mon goût je suis heureuse de retrouver à chaque fois que je suis à tes côtés cette complicité qui m'est chère. Je t'aime. À mon beauf, merci de prendre soin d'elle et de la rendre heureuse.

À mon homme, mon mari, mon complice, le père de mes enfants. Merci pour ton soutien indéfectible, merci de partager mes joies et mes tristesses, tu es le pilier de ma vie quotidienne, l'amour de ma vie, sans toi je ne serai rien.

À mes enfants, à Gabriel mon grand garçon d'amour je vais enfin être un « vrai docteur » !, à Clémentine ma petite fleur qui s'épanouit chaque jour sous nos yeux émerveillés. Il n'y a pas de mots pour décrire combien je vous aime.

À ma belle famille, vous m'avez tout de suite acceptée au sein de votre famille et je ne l'oublierai jamais. Merci pour votre accueil et pour l'atmosphère paisible qui règne dans votre foyer. Merci aussi d'être toujours présents pour vos petits-enfants.

À mamie Mérina, tu es loin mais je pense à toi,

À la mémoire des êtres chers que j'ai perdus : papi Jean, papi Roger, mamie Laurette, je pense toujours à vous.

À mes amis,

À mes compagnons de galère devenus mes amis:

Aux vieilles meufs, Alice Céline Fanny et Marie-Anne, mon gang, mon « crew », ma bouffée d'oxygène, je suis heureuse de vous avoir rencontrées. Merci pour les barres de rire mais aussi pour avoir su partager les moments de doute ou de tristesse. À nos années d'externat, puis d'internat, puis aux suivantes et à toutes celles qu'on va encore passer ensemble. À nos soirées passées à se déhancher, à nos vacances, à nos vieux mecs aussi ! Merci à eux de nous supporter, de partager (parfois malgré eux) notre amitié qui prend parfois de la place, en tout cas qui a une grande place dans ma vie.

Aux vieux mecs, Thibault et Greg, merci pour les soirées passées ensemble, merci pour les souvenirs de ~~vacances~~ stage et principalement celui où on s'est le plus ennuyé au CHU de Pointe-à-Pitre (c'était affreux !).

À Charlotte, merci pour les moments partagés avec toi depuis l'externat, en stage ou en rempla, au boulot, en formation, ou autour d'un café à papoter. Merci pour avoir été ma pote de galère pour la thèse, merci pour ta disponibilité et ton écoute, ton aide m'a été précieuse. Merci pour ton amitié qui a grandi progressivement tout au long de ces années. Et merci aussi à ton homme pour ses conseils d'expert !!!

À Élise, merci pour les moments passés en stage (les barres de rire de la gynéco...qu'est-ce qu'on a pu cravacher !!!), puis pour m'avoir permis de découvrir ton cabinet qui va bientôt devenir aussi le mien grâce à toi. Tu vas devoir me supporter encore quelque temps... Merci pour ton amitié, et pour la belle personne que tu es. Merci aussi à ton homme, quand on entend que tout le monde se gausse on sait qu'il est là...va falloir optimiser l'insonorisation de la salle de pause!

À Clémousse, merci d'avoir fait de nos 2 P1 des années plus que supportables, merci pour les journées BU et les soirées, pour m'avoir hébergé un nombre incalculable de fois... le sol de ta salle de bains m'a paru très confortable.

À Nico, Béber et toute la clique

À tous mes cointernes avec qui j'ai partagé de bons moments : Étienne, Stéphanie, les 2 Caro, Pauline, Joumi... j'en passe et des meilleurs !

À mes potos :

À Tonton Pirate, merci pour les souvenirs partagés depuis plus de 15 ans maintenant, pour les soirées de maboul et les conneries innombrables qu'on a pu faire ensemble... Merci pour ton originalité et ta désobéissance, ton système D, ta constante soif d'apprendre et d'évoluer, pour ton combat de chaque jour. Tu es en quelque sorte le frère que je n'ai pas eu!

À Virg', ma « supercondriaque », merci pour ton amitié, d'abord au lycée puis après ton come-back... Tu as bien fait de me recontacter. Merci pour les sorties au resto qui font du bien, tes histoires farfelues et nos discussions de gonzesses.

À Céline A., merci pour ton amitié, grâce à notre rencontre je ne regrette pas d'avoir « perdu » un an en fac de sciences. Merci pour ta générosité et pour m'avoir introduit dans tes soirées moules, puis dans la bande de guedins : Dr Suzette, Céline G. et tous les autres. Merci à vous pour les soirées, la bonne humeur, et pour m'avoir demandé un nombre incalculable de fois : alors t'en es où dans ta thèse ???

À ma Cômète intergalactique, tu es sûrement le potto que je connais depuis le plus longtemps, déjà au CP tu...non je ne dirai rien, toi-même tu sais ;-) Tellement de souvenirs avec toi, merci pour tout ça. Tu me manques. Tu feras bientôt, j'en suis persuadée, un super papa.

À mes aînés,

Merci aux Dr Lemarchand, Robert et Villamaux pour mon premier stage prat.

Merci aux Dr Casaubieilh, Clavier et Odoux pour mon SASPAS.

J'ai beaucoup appris à vos côtés.

Merci à la fine équipe de Lyons-la-Forêt : Dr Collard, Le Roy et Louis-Juste ainsi qu'à Aline, Simone, et tous les autres. Je suis vraiment contente de me lancer dans cette aventure avec vous.

À tous ceux qui m'ont demandé 1000 fois où j'en étais dans mes études, dans ma thèse, quand j'allais être enfin docteur et à tous ceux qui n'y ont jamais rien compris : c'est TER-MI-NÉ !!!!

TABLE DES MATIÈRES

RÉSUMÉ	2
ABSTRACT	3
INTRODUCTION	4
Le raisonnement clinique et la théorie des scripts	4
Les principes du TCS	4
Format du TCS	5
Le processus d'établissement des scores et du score de passage	5
Forces et faiblesses du TCS	6
BUTS DE L'ÉTUDE	6
MÉTHODES	7
Fidélité inter-juge	7
Analyse d'items	8
Experts déviants	8
Réponses faites au hasard et stratégies de réponse	8
RÉSULTATS	8
Fidélité inter-juge	8
Analyse des items	10
Experts déviants	10
Réponses faites au hasard et stratégies de réponses	10
DISCUSSION	11
Fidélité inter-juge	11
Analyse d'items	11
Experts déviants	12
Réponses faites au hasard et stratégies de réponse	13
CONCLUSION	13
BIBLIOGRAPHIE	14

RÉSUMÉ :

Contexte : Le test de concordance de scripts (TCS) est un outil d'évaluation répandu. Peu d'études sont disponibles en évaluation sommative. Certains débats persistent concernant l'étude insuffisante de la fidélité inter-juge, la trop grande confiance accordée au coefficient alpha, la nécessité ou non d'exclure les experts déviants et l'efficacité de certaines stratégies de réponse.

Buts: Apprécier la fidélité inter-juge des TCS utilisés en évaluation sommative en DES de médecine générale à Rouen et discuter de l'impact de certains critères d'élaboration du test sur le classement des étudiants.

Méthodes : Une analyse quantitative rétrospective des résultats aux TCS en DES de médecine générale à Rouen entre 2011 et 2016 a été effectuée. La fidélité inter-juge a été évaluée via le calcul de la corrélation intra-classe (ICC). L'impact sur le classement des étudiants a été évalué après amélioration du coefficient alpha, exclusion des experts déviants, éviction des réponses extrêmes, et utilisation des stratégies de réponses au hasard.

Résultats : La fidélité inter-juge était excellente ($ICC=0.984$). L'amélioration de la cohérence interne du test modifiait le classement des étudiants mais pas l'exclusion des experts déviants. Les stratégies d'éviction des réponses extrêmes et de réponse au hasard se révélaient parfois efficaces.

Conclusion : Le panel d'experts utilisé à Rouen en DES de médecine générale semble quantitativement et qualitativement adapté. Certaines améliorations concernant la construction des items pourraient être réalisées. Une méfiance devrait être accordée concernant l'interprétation du coefficient alpha. D'autres études sont nécessaires afin d'estimer quelles dimensions du raisonnement clinique sont mesurées par le TCS.

MOTS-CLÉS :

test de concordance de script, raisonnement clinique, médecine générale, évaluation sommative

ABSTRACT:

Context: The script concordance test (SCT) is a common assessment tool. Few studies are available in summative assessment. Some debates continue regarding the insufficient study of inter-panel reliability, excess of confidence considering alpha coefficient, deviant experts exclusion and the effectiveness of some response strategies.

Aims: To assess the inter-panel reliability of the SCT used for summative assessment in family medicine residency program at Rouen Medical University, and to discuss the impact of some specific test development criteria on students' classification.

Methods: A retrospective analysis of quantitative results to SCT in family medicine residency program at the University of Rouen between 2011 and 2016 was performed. The inter-panel reliability was assessed by calculating the intraclass correlation coefficient (ICC). The impact on students' classification was evaluated after improving the alpha coefficient, excluding deviant experts, avoiding extreme responses, and use guessing.

Results: The inter-panel reliability was excellent ($ICC = 0.984$). Internal consistency improvement changed students' classification but not deviant experts' exclusion. Strategies of extreme responses eviction and guessing were sometimes effective.

Conclusion: The expert panel used in family medicine residency program at Rouen seems to be quantitatively and qualitatively adequate. Some improvements in the construction of the items could be realized. Alpha coefficient should be interpreted with distance and care. Further studies are necessary to estimate which dimensions of clinical reasoning are measured by the SCT.

KEYWORDS:

script concordance test, clinical reasoning, family practice, summative assessment

INTRODUCTION:

Le raisonnement clinique et la théorie des scripts :

Selon Bernard Charlin (1), la compétence clinique nécessaire à tout professionnel de santé comprend 3 composantes :

- Les connaissances théoriques
- Les habiletés (habiletés à recueillir les données cliniques: savoir mener un bon interrogatoire, habiletés techniques : réaliser des gestes, habiletés humaines etc...)
- Le raisonnement clinique

Le raisonnement clinique est l'activité intellectuelle qui sélectionne, synthétise les informations obtenues à partir d'une situation clinique spécifique, qui l'intègre aux connaissances et expériences antérieures et l'utilise pour prendre des décisions de diagnostic et de prise en charge du patient (1).

L'acquisition d'un bon raisonnement clinique est un objectif primordial de la formation médicale puisque deux tiers des erreurs diagnostiques en médecine seraient consécutives à des erreurs de raisonnement et de prise de décision (2).

De nombreuses études en psychologie cognitive ont été menées ces dernières décennies afin d'explorer les fondements du processus de raisonnement clinique et différents modèles ont été décrits (3), dont la théorie des scripts. Les scripts en médecine sont les réseaux de connaissances élaborés par le clinicien, spécifiquement organisés pour être efficaces dans une tâche clinique, ce qui permet au clinicien d'acquérir des automatismes concernant la prise de décision. Les chemins ou « scripts » empruntés sont mis à jour et améliorés tout au long de l'expérience du clinicien (4).

Alors que les connaissances théoriques peuvent être facilement évaluées par des examens comme le questionnaire à choix multiples (QCM), l'évaluation du raisonnement clinique s'avère moins aisée et de nombreux formats de test ont été proposés (1,5). En se basant sur la théorie cognitive des scripts, un nouvel outil d'évaluation du raisonnement clinique a été créé à la fin des années 90: le test de concordance de scripts (TCS) (6).

Les principes du TCS :

Le but du TCS est l'analyse d'une étape au sein d'un processus de raisonnement et non le résultat de ce processus. Il permet de mesurer le degré d'organisation des connaissances, vérifie si les connaissances sont élaborées (c'est-à-dire organisées pour agir efficacement dans le contexte clinique), plutôt que la simple présence de ces connaissances.

Par ailleurs, le TCS est un outil d'évaluation dont l'optique est centrée sur la pratique clinique. Or, comme le rapporte Schön (7), les professionnels de santé sont confrontés dans leur pratique à deux types de problèmes:

- Les problèmes simples, bien structurés : l'ensemble des données nécessaires est accessible d'emblée et la solution résulte d'une simple application des connaissances, les professionnels s'accordent en général sur la solution à apporter.
- Les problèmes complexes, mal structurés : les données du problème ne sont pas toutes disponibles d'emblée, la solution n'est pas univoque et nécessite une réflexion mobilisant des connaissances diverses. Les professionnels sont souvent en désaccord

sur la solution à apporter, même s'ils parviennent généralement aux mêmes diagnostics et à des décisions similaires en matière de prise en charge des patients.

Les problèmes auxquels les professionnels de santé sont le plus souvent confrontés dans leur pratique quotidienne sont complexes. Le TCS explore le processus de raisonnement clinique mis en œuvre en contexte d'incertitude dans des situations cliniques complexes, ambiguës et donc au plus proche de la pratique quotidienne.

Format du TCS :

Les questions d'un TCS se présentent sous forme de vignettes qui correspondent à des situations courantes rencontrées en pratique clinique. Chaque vignette est volontairement courte car ceci participe à créer de l'incertitude; les données sont volontairement ambiguës ou des informations sont manquantes. Chaque cas de TCS doit être un problème même pour un médecin expérimenté.

On présente au candidat une hypothèse initiale de diagnostic, d'investigation, de traitement ou de pronostic. Le candidat est invité à évaluer l'impact d'une nouvelle information sur cette hypothèse initiale par le choix d'un des cinq échelons de l'échelle de Likert (l'hypothèse initiale est-elle éliminée, moins probable, inchangée, plus probable ou certaine?).

Le tableau 1 présente un exemple d'item du TCS.

Une patiente de 65 ans est adressée aux urgences pour une suspicion d'embolie pulmonaire.				
Hypothèse initiale	Nouvelle information	Conséquence sur l'hypothèse initiale		
Si vous pensiez faire...	Et qu'alors vous trouvez...	L'hypothèse initiale est...		
Un dosage des D-dimères	Un cancer de l'ovaire en cours de traitement	A	B	C D E

A: éliminée; B: moins probable; C: inchangée; D: plus probable; E: certaine

Tableau 1: exemple d'un item de TCS (Charlin et al., 2005)

Le processus d'établissement des scores et du score de passage:

Le TCS est un format de test original puisqu'il n'y a pas qu'une seule bonne réponse à un item.

On fait une évaluation du degré de concordance entre les réponses de l'étudiant et celles d'un panel de référence constitué de médecins expérimentés (anonymes et experts de la spécialité). Les réponses des experts sont utilisées pour bâtir la grille de correction, qui constitue la référence pour déterminer le score du candidat, par une méthode de pondération des choix de réponse (les points sont attribués proportionnellement à la répartition de la réponse des experts à la question).

La réponse de chaque membre du panel de référence est ainsi prise en compte, la méthode permet de prendre en compte la variabilité trouvée habituellement entre médecins expérimentés lorsqu'ils répondent à des questions complexes.

On établit enfin un score de passage en deçà duquel un étudiant est recalé à l'examen (habituellement 2 écarts-type en dessous de la moyenne du panel de référence) (8).

Forces et faiblesses du TCS :

Le TCS est maintenant un outil assez répandu et son utilisation a fait l'objet de recommandations et de guides de bonne pratique à l'usage des enseignants en professions de santé (9–11). Il peut être utilisé sous forme papier ou en ligne (12). Il peut être utilisé aussi bien en évaluation formative (13) qu'en évaluation sommative (9), même si comme nous allons le voir sa validité pour son utilisation lors d'examens à enjeu élevé est débattue (14).

Le TCS a démontré sa validité de construit (c'est-à-dire que les cliniciens expérimentés réussissent mieux le TCS que les étudiants) ainsi que sa fidélité (cohérence interne mesurée par le coefficient alpha de Cronbach) au travers de son application dans de nombreuses spécialités médicales ou paramédicales (15,16). Sa validité prédictive a également été démontrée (les étudiants qui réussissent le TCS à un instant donné réussiront de nouveau si le TCS est réadministré plus tard dans leur cursus) (17). Les TCS sont relativement faciles à créer, à administrer et à corriger (correction automatisée possible) (11). L'acceptabilité du format de ce test par les étudiants est bonne (18).

En revanche, de nombreux points sont le sujet de débats dans la littérature:

- Depuis ses débuts la détermination d'un score de passage est le problème du TCS (9). Cependant la détermination d'un score de réussite est toujours arbitraire pour n'importe quelle méthode d'évaluation. Une étude a comparé différentes méthodes de détermination du score de passage (19) mais n'a pas retrouvé de différences entre elles.
- En 2013, Lineberry et al.(14) pointent du doigt le fait que la plupart des études évaluant la fidélité du TCS n'utilisent comme critère d'appréciation que le coefficient alpha de Cronbach et qu'elles mettent de côté les autres critères de fidélité comme notamment la fidélité inter-juge (stabilité des réponses des experts selon les jurys).
- Lineberry et al. suggèrent également que la validité de contenu du test peut être remise en question par le fait que la méthode d'établissement des scores prend en compte la variabilité de réponse de tous les experts, y compris ceux qui pourraient être considérés comme déviants.
- Enfin, Lineberry et al. mettent en avant le processus par lequel un étudiant qui éviterait les réponses situées aux deux extrêmes de l'échelle de Likert (réponses A ou E) ou qui choisirait tout le temps la réponse située au milieu de l'échelle (réponse C) augmenterait son score de façon non négligeable.

BUTS DE L'ÉTUDE :

Le TCS est utilisé en fin d'internat de médecine générale à Rouen à but sommatif depuis de nombreuses années. Sa réussite fait partie des conditions de validation du diplôme d'étude spécialisé (DES) de médecine générale. Or il n'existe que très peu d'études dans la littérature concernant les TCS en évaluation sommative (20). Nous avons décidé d'utiliser les résultats aux TCS à Rouen sur plusieurs années afin d'étudier les points suivants :

- La fidélité inter-juge concernant 21 items communs aux TCS ayant été administrés de 2011 à 2013.
- L'utilité de l'analyse d'items et de l'amélioration a posteriori du coefficient alpha (en éliminant du test les items ayant une faible corrélation par rapport au score total du test, ce que nous nommerons par la suite la corrélation item/examen) : cette manœuvre provoque-t-elle une modification des scores des étudiants et par conséquent de leur classement ?
- L'intérêt d'exclure les experts déviants : est-ce que cela modifie le classement des étudiants ?
- Le problème des réponses faites au hasard et du style de réponse des candidats : un candidat qui répondrait au hasard ou qui éviterait volontairement les réponses modales extrêmes sur l'échelle de Likert améliorerait-il de façon significative sa note et donc son classement?

MÉTHODES :

Les résultats des internes de médecine générale et des experts aux TCS à Rouen en DES de médecine générale entre 2011 et 2016 ont été analysés de façon rétrospective.

L'examen était composé chaque année de 90 questions et durait 90 minutes. Les experts ont systématiquement répondu à l'examen en même temps que les internes dans le même amphithéâtre afin d'être exactement dans les mêmes conditions. Les experts choisissaient pour chaque question la réponse qui leur semblait être la meilleure ainsi que la réponse qui leur semblait être la plus mauvaise; un point négatif était alors accordé à cet échelon dans la clef de notation, si bien que la note finale de l'étudiant à un item (et même à l'examen) pouvait être négative.

Fidélité inter-juge :

La corrélation intra-classe (ICC) a été calculée entre les points accordés par les experts à chaque échelon de l'échelle de Likert concernant les 21 questions communes servies aux TCS de 2011 à 2013 (il n'y avait pas de questions en commun les autres années). L'ICC est utilisée pour mesurer le degré de consensus entre 2 (ou plus) évaluateurs ou méthodes d'évaluation. Un $ICC > 0.7$ traduit un accord fort et un $ICC > 0.8$ traduit un accord presque parfait.

Afin de mieux apprécier l'aspect des réponses du panel d'experts selon les années, un graphique à colonnes a été construit pour chaque question avec en abscisse les échelons de l'échelle de Likert et en ordonnée la note accordée à cet échelon.

Puis une simulation a été effectuée ainsi qu'une comparaison des notes et du classement qu'auraient eu les étudiants de la session 2013 sur ces 21 questions en fonction des jurys entre 2011 et 2013.

L'ICC a été calculée entre les pourcentages d'étudiants ayant choisi chacun des échelons de l'échelle de Likert chaque année sur ces mêmes questions communes entre 2011 et 2013, afin d'observer si cette corrélation était globalement comparable à celle des experts.

Analyse d'items :

Le coefficient alpha de Cronbach a été calculé pour les épreuves de 2011 à 2016 (sauf 2014, année pour laquelle nous n'avons pas pu obtenir toutes les données nécessaires au calcul). Un coefficient alpha est considéré acceptable s'il est supérieur à 0.7 et plus il se rapproche de 1, plus le construit mesuré est homogène.

Pour les 21 questions communes de 2011 à 2013 la corrélation item/examen (sur le nombre total de questions à l'examen) concernant chaque item chaque année a été calculée afin d'observer s'il existait une stabilité globale des corrélations en fonction des années.

Pour les sessions 2013, 2015 et 2016 la corrélation item/examen concernant chaque item a été calculée (les données nécessaires au calcul n'étaient pas disponibles pour les autres années). Une simulation des notes des étudiants a été effectuée ainsi qu'une comparaison de leur classement virtuel en fonction de l'élimination ou non des items à moins bonne corrélation (corrélation négative ou <0.1). Ceci nous a permis d'apprécier si l'amélioration a posteriori du coefficient alpha avait un quelconque effet sur le classement des étudiants et sur leur élimination ou non à l'examen.

Experts déviants :

Pour l'année 2013 il a été effectué un calcul des notes qu'auraient eu les étudiants en ayant au préalable éliminé les réponses des experts déviants de la clef de notation (experts dont la note était inférieure à -1DS par rapport à la moyenne des notes obtenues par les experts). Ceci nous a permis d'apprécier si l'exclusion des experts déviants avait un effet sur le classement des étudiants et sur leur élimination ou non à l'examen. Les données nécessaires au calcul n'étaient pas disponibles pour les autres années.

Réponses faites au hasard et stratégies de réponse :

Pour les sessions 2013, 2015 et 2016 il a été effectué un calcul de la note et du nouveau classement qu'aurait eu l'étudiant classé en dernière position s'il avait systématiquement pour chaque question évité de répondre en choisissant les deux échelons extrêmes de l'échelle de Likert (A et E).

Pour ces mêmes sessions 2013, 2015 et 2016 il a ensuite été effectué un calcul de la note et du classement qu'aurait obtenu un étudiant virtuel qui aurait répondu systématiquement B. Le même calcul a été effectué concernant les échelons C et D.

Les données nécessaires à ces calculs n'étaient pas disponibles pour les autres années.

RÉSULTATS :

Les étudiants étaient respectivement au nombre de 82, 74, 73, 78, 70 et 106 aux sessions des TCS de 2011 à 2016. Les experts étaient respectivement au nombre de 32, 32, 34, 26, 30 et 29 aux mêmes sessions.

Fidélité inter-juge :

La corrélation intra-classe entre les notes des experts aux 21 questions communes de 2011 à 2013 était de 0.984. Au vu de cette forte corrélation, il en découle un aspect graphique des réponses du panel visuellement comparable d'une année sur l'autre comme on peut le voir

dans la figure 1 (sauf pour 2 questions pour lesquelles l'homogénéité peut être discutable, cf figure 2).

La simulation effectuée en prenant les réponses des étudiants de la session 2013 sur les 21 questions communes et en calculant leur note et classement en fonction des notes attribuées par les différents jurys de 2011, 2012 et 2013 donnaient un classement similaire d'une année sur l'autre : les 5 derniers étudiants étaient systématiquement les mêmes.

La corrélation intra-classe entre les pourcentages des effectifs d'étudiants ayant répondu A, B, C, D ou E en fonction des 3 années était de 0.97. Un seul item avait une corrélation un peu plus faible de 0.74 (cf. figure 3), alors que pour tous les autres items la corrélation était supérieure à 0.9 ; Cet item fait partie des 2 items dont l'aspect graphique des réponses du panel d'expert paraissait moins homogène sur 3 années.

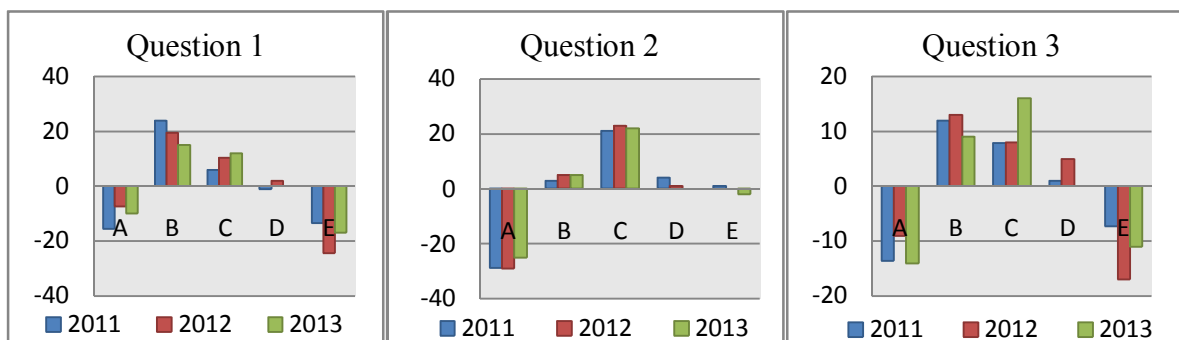


Figure 1: échantillon d'items dont l'aspect des réponses du panel est superposable d'une année sur l'autre. En ordonnée, le nombre de points accordés par les experts à chaque réponse A, B, C, D, E.

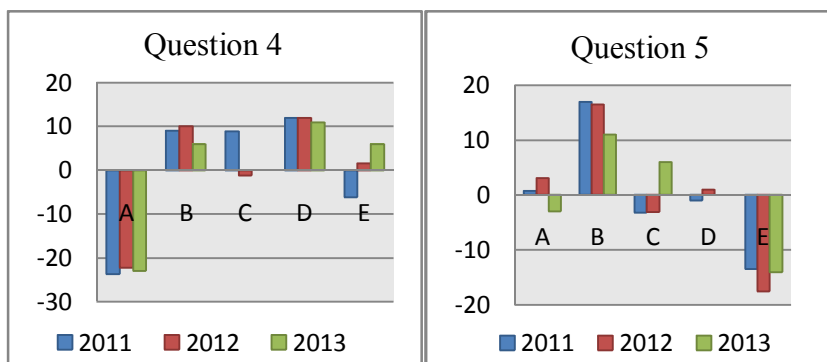


Figure 2: deux items pour lesquels l'aspect graphique des réponses du panel n'est pas tout à fait homogène d'une année sur l'autre. En ordonnée, le nombre de points accordés par les experts à chaque réponse A, B, C, D, E.

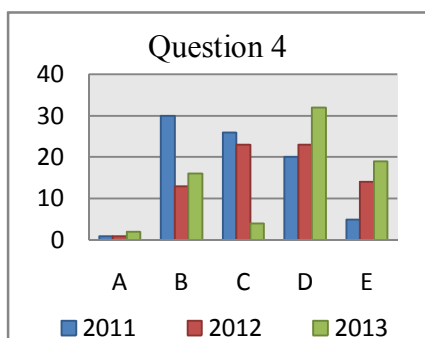


Figure 3: aspect graphique des réponses des étudiants, concernant l'item dont la corrélation intra-classe des étudiants est mauvaise. En ordonnée, le pourcentage d'étudiants ayant choisi chaque réponse A, B, C, D, E.

Analyse des items :

Le coefficient alpha de Cronbach était respectivement égal à 0.71, 0.73 et 0.84 de 2011 à 2013. En 2015, le Cronbach était égal à 0.66 et en 2016, à 0.77.

Concernant la comparaison des corrélations item/examen des 21 questions (en prenant le score total aux 90 questions chaque année) : les corrélations n'étaient pas tout à fait stables d'une année sur l'autre. 4 items obtenaient une fois sur les 3 ans une corrélation négative et 1 item obtenait une fois une corrélation inférieure à 0.1, alors que les autres années les corrélations de ces items étaient bonnes (>0.1) voire très bonnes (>0.4).

Après avoir éliminé les items à corrélation inférieure à 0.1 pour les examens de 2013, 2015 et 2016, il a été conservé respectivement 82, 62 et 67 items pour 2013, 2015 et 2016. Les coefficients alpha de Cronbach en résultant étaient respectivement égaux à 0.86 (amélioration +0.02), 0.77 (+0.11) et 0.83 (+0.06) pour 2013, 2015 et 2016. Les classements virtuels en résultant ne montraient pas de différence pour la session 2013. Les étudiants éliminés (dont la note était inférieure à 2 écarts-typé par rapport à la moyenne des étudiants) auraient été les mêmes. Pour la session 2015, 1 étudiant supplémentaire aurait été éliminé. Pour la session 2016, à l'inverse, 1 étudiant n'aurait pas été éliminé alors qu'il l'avait été lors de l'examen comprenant 90 questions.

Experts déviants :

Après élimination des réponses des experts déviants (dont la note était inférieure à -1 DS par rapport à la moyenne du panel) de la clé de notation pour la session 2013, il restait 23 experts sur 29 et les étudiants éliminés auraient été les mêmes.

Réponses faites au hasard et stratégies de réponses :

Après avoir éliminé les réponses modales extrêmes de l'étudiant arrivé en dernière position du classement pour les sessions 2013, 2015 et 2016 (nous avons transformé leurs réponses A en réponses B et leurs réponses E en réponse D), nous avons obtenu les résultats suivants :

- En 2013, l'examen comportait 8 items à réponse modale extrême (soit 9% des items). L'étudiant arrivé dernier avait répondu de façon extrême à 39 items (soit 43%). Adopter la stratégie consistant à éviter les réponses A et E lui aurait permis de passer de la 73^{ème} à la 61^{ème} place du classement et ainsi il n'aurait pas été éliminé.
- En 2015, l'examen comportait 18 items à réponse modale extrême (soit 20% des items). L'étudiant arrivé dernier avait répondu de façon extrême à 29 items (soit 32%). Adopter la stratégie consistant à éviter les réponses A et E ne lui aurait pas permis de modifier son classement et il aurait été éliminé de toute façon.
- En 2016, l'examen comportait 7 items à réponse modale extrême (soit 8% des items). L'étudiant arrivé dernier avait répondu de façon extrême à 43 items (soit 48%). Adopter la stratégie consistant à éviter les réponses A et E lui aurait permis de passer de la 106^{ème} à la 40^{ème} place et ainsi il n'aurait pas été éliminé.

Concernant les réponses faites au hasard en 2013, 2015 et 2016:

- Un étudiant qui aurait répondu systématiquement B à toutes les questions du TCS n'aurait été éliminé à aucune session sur les 3.
- Un étudiant ayant répondu C à toutes les questions n'aurait pas été éliminé en 2013 mais il l'aurait été en 2015 et 2016.
- Un étudiant ayant répondu D aurait été éliminé à chaque fois.

DISCUSSION :

Cette étude a pour originalité d'étudier la fidélité inter-juge en évaluation sommative entre des panels recrutés à l'intérieur de la même spécialité et dont l'origine géographique est similaire, ce qui n'avait pas encore été rapporté à notre connaissance dans la littérature. Nous allons désormais discuter point par point nos résultats :

Fidélité inter-juge :

La corrélation intra-classe des experts entre 2011 et 2013 s'est révélée excellente. Changer de jury n'a pas modifié le classement des étudiants de la session 2013 concernant les 21 questions étudiées. Les étudiants ont répondu également de façon homogène à 20 questions sur les 21 questions étudiées en 2011, 2012 et 2013.

Cependant, il faut rappeler que nous n'avons pu calculer la fidélité inter-juge que pour 21 questions. Il aurait été plus probant de pouvoir réaliser cette étude sur l'ensemble de l'examen qui comprend 90 questions. Par ailleurs, les 21 questions étudiées avaient été sélectionnées de nouveau en 2012 et 2013 en raison d'une bonne corrélation item/examen retrouvée en 2011 ; ceci pourrait créer un biais de sélection des questions étudiées.

De plus, parmi les 58 experts ayant assuré la composition des panels entre 2011 et 2013, 13 experts étaient communs aux 3 années. 12 experts ont participé 2 années et 33 experts n'ont participé qu'une seule année. Ceci amène indiscutablement un biais ; puisque certains experts sont communs alors ils vont probablement répondre de la même façon aux questions posées.

Il faut également noter que la comparaison des jurys a été effectuée sur 3 années consécutives, ce qui aurait pu nous amener à observer davantage de réponses différentes d'une année sur l'autre en raison de l'évolution de la médecine et de ses recommandations. C'est justement l'explication que nous avançons concernant la question 5 pour laquelle les réponses du panel d'experts sont quasiment les mêmes en 2011 et 2012 (plutôt A et B) mais qui changent en 2013 (plutôt B et C) (cf. figure 2) ; cette question concernait le diabète du sujet âgé et de nouvelles recommandations ont été émises par la Haute Autorité de Santé (HAS) en 2013 concernant cette thématique. Par contre, les réponses des étudiants étaient les mêmes pour cette question d'une année sur l'autre.

Pour la question 4, les réponses différaient à la fois pour les étudiants mais également pour les experts. Il faut noter que cette question n'était pas exactement posée de la même façon les 3 années. En 2011, elle intervenait dans l'examen à la suite de 2 autres questions concernant le même cas, et en 2012 et 2013, elle était posée seule. Ceci suggère la difficulté d'assurer l'indépendance des questions posées à travers un cas comprenant plusieurs questions. Mais on a pu constater que les réponses des étudiants différaient tout de même entre 2012 et 2013, ce qui pourrait suggérer que cette question est finalement trop ambiguë et qu'elle devrait peut-être être éliminée.

Analyse d'items :

Les corrélations item/examen calculées pour les 21 questions communes ne sont pas toujours stables d'un examen à l'autre pour la même question. Ceci sous-entend que la corrélation item/examen ne devrait pas forcément être un critère de choix afin de déterminer si un item doit être conservé ou non. Par ailleurs, comme le soulignent Lubarsky et al.(11), « une corrélation faible peut refléter l'hétérogénéité des pratiques des membres du panel ou du domaine testé et pas forcément une défaillance de l'item lui-même ». De plus, ceci soulève une question plus profonde, déjà soulevée par Wilson et al. en 2014 (21) : quelle dimension du

raisonnement clinique le TCS mesure-t-il vraiment ? Et concernant notre étude est-ce que les TCS administrés en évaluation sommative à Rouen mesurent la même dimension du raisonnement clinique d'une année sur l'autre ?

Après avoir amélioré a posteriori le coefficient alpha en éliminant les items dont la corrélation item/examen était inférieure à 0.1, les épreuves de 2015 et 2016 ne disposent plus que de 62 et 67 items respectivement. Chaque année un étudiant a vu son sort changer quant à son élimination ou non à l'examen. En 2015, 4 étudiants auraient été éliminés au lieu de 3. Par contre, les 4 derniers étudiants sont toujours les mêmes avant ou après avoir retiré les items à corrélation < 0.1. En 2016, 4 étudiants auraient été éliminés au lieu de 5, mais de la même façon, les 5 derniers étudiants sont toujours les mêmes. L'épreuve de 2013 dont le coefficient alpha était déjà très bon conserve quant à elle 82 items sur 90 et les étudiants éliminés restent les mêmes. Ces résultats suggèrent que l'élimination des items à mauvaise corrélation pourrait modifier les résultats des étudiants. Ceci va à l'encontre des résultats de l'étude de P.Duggan et B.Charlin (20), mais il est important de noter que la méthode utilisée pour le calcul de la note de passage diffère dans notre étude : effectivement à Rouen la note de passage est fixée à -2DS par rapport à la moyenne des étudiants et non pas par rapport à la moyenne du panel d'experts comme il est plus coutumier de le voir dans la littérature. Cependant une étude réalisée en 2013 (19) n'a pas trouvé de différence significative entre les différentes méthodes d'établissement des scores de passage. Ceci serait à vérifier dans notre contexte d'étude. Notre résultat pourrait également s'expliquer par le plus petit nombre de questions par rapport à l'étude de P.Duggan et B.Charlin ; en 2015 et 2016 nos épreuves se sont vues amputer d'un bon nombre de questions. Il faudrait donc probablement prévoir un examen plus long (avec plus de questions) afin de pouvoir conserver un nombre de questions plus important après élimination a posteriori des items à mauvaise corrélation (comme le suggère l'étude de Dory et al.(22) : 100 items recommandés, avec élimination a posteriori en moyenne de 25% des questions pour obtenir 75 questions au total). Nos résultats suggèrent par ailleurs que les résultats des étudiants peuvent être modifiés quant à leur passage ou non à l'examen, cependant lors d'un examen classant sans score de passage comme ceci est fait à l'examen national classant (ENC) en France, il n'y aurait peut-être pas eu de réelle différence avant ou après élimination des items à corrélation < 0.1, puisque dans notre étude les derniers étudiants étaient de toute façon systématiquement les mêmes.

Ainsi, en conclusion sur ce point, la sélection des questions a posteriori via la corrélation item/examen n'apparaît pas être une méthode dépourvue de défaut. L'instabilité des corrélations item/examen pour une même question en fonction des sessions suggère que les TCS à Rouen ne mesurent pas forcément la même dimension du raisonnement clinique d'une année sur l'autre. De plus, il faudrait accroître le nombre de questions à l'examen pour pouvoir si besoin améliorer a posteriori la cohérence interne du test, ce qui semble modifier sensiblement le sort des étudiants. Enfin, l'attention devrait peut-être être plutôt portée sur la stratégie consistant à éliminer les items dont l'aspect des réponses du panel d'experts sur l'échelle de Likert est trop étalé (question trop ambiguë) ou au contraire trop concentré sur un seul échelon (question de cours).

Experts déviants :

Nos résultats semblent être en accord avec les données de la littérature (23). Il faut noter que nous n'avons pu tester l'élimination des experts déviants que sur une année (2013). Le nombre d'experts composant le panel utilisé à Rouen est assez élevé tous les ans par rapport à ce qui est retrouvé de façon courante dans la littérature. Par conséquent, la taille du panel restant après avoir éliminé les experts déviants était encore très correcte (il est recommandé

habituellement d'en recruter au moins 20 pour les examens à fort enjeu (24)) ce qui explique probablement que les résultats des étudiants étaient inchangés.

Réponses faites au hasard et stratégies de réponse :

Après avoir éliminé les réponses modales extrêmes du dernier candidat pour les sessions 2013, 2015 et 2016, nous avons étudié si ceci impactait son élimination ou non à l'examen. Il a été constaté que pour une session sur les 3 (2015), ceci n'avait pas d'impact. Ce résultat peut s'expliquer par le fait qu'en 2015, l'examen comportait plus de questions à réponse modale extrême que les autres années (20% contre respectivement 9 et 8% en 2013 et 2016). Ceci suggère qu'il faudrait un pourcentage minimum de questions à réponse modale extrême parmi l'ensemble des questions d'un TCS afin que la stratégie consistant à éviter les réponses modales extrêmes ne fonctionne pas. Cette idée a déjà émergé dans la littérature (25), en évoquant la nécessité d'un pourcentage de 50% de questions à réponses modales extrêmes sur l'ensemble d'un examen. Nos résultats amènent à penser que ce pourcentage est peut-être surévalué. Cependant dans notre étude l'interne en dernière position en 2015 avait répondu moins souvent de façon extrême (32% de réponses A et E) que les internes des 2 autres sessions étudiées (43% de réponses A et E pour le dernier étudiant en 2013 et 48% pour celui de 2016). Un pourcentage de questions à réponse modale extrême se situant entre 20 et 50% est donc probablement nécessaire afin d'éviter de réussir en optant pour la stratégie « sans réponse A et E ».

En optant pour un examen avec un minimum de questions à réponse modale extrême, il faudra cependant se méfier de l'impact sur la qualité des questions ; un item à réponse modale extrême a probablement une plus forte tendance à ressembler à une question de cours. Pour remédier à cela et comme le suggèrent Lubarsky et al. via leur guide de construction d'un TCS (11), il est préférable d'utiliser pour les échelons extrêmes une phrase qui n'est pas trop « catégorique ou sans équivoque » ; ceci permettrait d'encourager le choix des réponses extrêmes sans pour autant que la question soit trop théorique.

Par la suite, une simulation des réponses au hasard a été effectuée pour les 3 sessions 2013, 2015 et 2016. Répondre systématiquement B est une stratégie gagnante à chaque session, répondre D ne l'est jamais et répondre C est gagnant une fois sur 3 seulement. Ceci n'est pas en accord avec les résultats de l'étude de Lineberry et al (14) qui montraient que la réponse C récoltait le plus souvent la réponse modale et ainsi obtenait le plus de points. Nos résultats suggèrent néanmoins qu'il faudrait apporter une plus grande attention aux réponses modales attendues lors de la construction du TCS à Rouen afin que celles-ci soient équitablement réparties sur l'échelle de Likert et qu'aucune stratégie de réponse au hasard ne fonctionne.

CONCLUSION :

Le TCS utilisé à Rouen en évaluation sommative en DES de médecine générale a su démontrer des qualités : le recrutement du panel d'experts est productif puisque les experts sont nombreux à participer chaque année, et la fidélité inter-juge est excellente montrant une constance des réponses selon les jurys. Le fait d'éliminer les experts déviants ne change pas les résultats des étudiants et continuer à les inclure participe probablement à refléter la diversité des attitudes possibles face à un problème clinique complexe.

Concernant la construction du test, il faudrait probablement augmenter légèrement le nombre de questions afin de pouvoir en conserver suffisamment après élimination des items à mauvaise corrélation item/examen, puisque cette manœuvre a potentiellement un effet sur le score des étudiants et ainsi sur leur élimination ou non au test. L'aspect des réponses du panel pourrait être utilisé afin d'éliminer les questions de cours et les questions trop ambiguës. Il faudrait également veiller à ce que les pourcentages de réponses modales attendues aux items soient équilibrés sur les 5 échelons de l'échelle de Likert afin d'éviter que les stratégies de réponses au hasard ou d'éviction des réponses modales extrêmes soient payantes.

Enfin, après les observations faites dans cette étude concernant la variabilité de la corrélation item/examen pour une même question sur plusieurs sessions de tests, il faut se poser la question de ce que le TCS évalue vraiment et notamment quelles dimensions du raisonnement clinique il mesure.

BIBLIOGRAPHIE :

1. Charlin B, Bordage G, Van Der Vleuten C. L'évaluation du raisonnement clinique. *Pédagogie Médicale*. févr 2003;4(1):42-52.
2. Modi JN, Anshu, Gupta P, Singh T. Teaching and Assessing clinical reasoning skills. *Indian Pediatr*. 2015;52(9):787-794.
3. Monteiro SM, Norman G. Diagnostic Reasoning: Where We've Been, Where We're Going. *Teach Learn Med*. janv 2013;25(sup1):S26-32.
4. Lubarsky S, Dory V, Audétat M-C, Custers E, Charlin B. Using script theory to cultivate illness script formation and clinical reasoning in health professions education. *Can Med Educ J*. 2015;6(2):e61.
5. Van Bruggen L, Manrique-Van Woudenberg M, Spierenburg E, Vos J. Preferred question types for computer-based assessment of clinical reasoning: a literature study. *Perspect Med Educ*. nov 2012;1(4):162-71.
6. Charlin B, Brailovsky C, Leduc C, Blouin D. The diagnosis script questionnaire: a new tool to assess a specific dimension of clinical competence. *Adv Health Sci Educ*. 1998;3(1):51-58.
7. Schön DA. *The Reflective Practitioner: How Professionals Think In Action*. 1st edition; Basic Books; 1984. 388 p.
8. Charlin B, Gagnon R, Lubarsky S, Lambert C, Meterissian S, Chalk C, et al. Assessment in the Context of Uncertainty Using the Script Concordance Test: More Meaning for Scores. *Teach Learn Med*. 22 juin 2010;22(3):180-6.
9. Giet D, Massart V, Gagnon R, Charlin B. Le test de concordance de script en 20 questions. *Pédagogie Médicale*. févr 2013;14(1):39-48.

10. Fournier JP, Demeester A, Charlin B. Script concordance tests: guidelines for construction. *BMC Med Inform Decis Mak.* 2008;8(1):1.
11. Lubarsky S, Dory V, Duggan P, Gagnon R, Charlin B. Script concordance testing: From theory to practice: AMEE Guide No. 75. *Med Teach.* mars 2013;35(3):184-93.
12. Sibert L, Giorgi R, Dahamna B, Doucet J, Charlin B, Darmoni SJ. Is a web-based concordance test feasible to assess therapeutic decision-making skills in a French context? *Med Teach.* janv 2009;31(4):e162-8.
13. HAS. test_de_concordance_de_script_tcs_fiche_technique_2013_01_31.pdf [Internet]. Disponible sur: http://www.has-sante.fr/portail/upload/docs/application/pdf/2013-02/test_de_concordance_de_script_tcs_fiche_technique_2013_01_31.pdf
14. Lineberry M, Kreiter CD, Bordage G. Threats to validity in the use and interpretation of script concordance test scores. *Med Educ.* déc 2013;47(12):1175-83.
15. Iravani K, Amini M, Doostkam A, Dehbozorgian M. The validity and reliability of script concordance test in otolaryngology residency training. *J Adv Med Educ Prof.* 2016;4(2):93.
16. Côté S, Tanguay A, Gagnon R, Charlin B, Michaud C. Élaboration et validation d'un test de concordance de script pour évaluer le raisonnement clinique des infirmières en contexte de détérioration de la condition clinique. *Pédagogie Médicale.* févr 2014;15(1):7-20.
17. Brailovsky C, Charlin B, Beausoleil S, Cote S, Van der Vleuten C. Measurement of clinical reflective capacity early in training as a predictor of clinical reasoning performance at the end of residency: an experimental study on the script concordance test. *Med Educ.* 2001;35(5):430-436.
18. Hoff L, Bestawros A, Kassis J, Charlin B. Le test de concordance de script comme outil d'enseignement et d'apprentissage: un projet-pilote pour les étudiants de première année de médecine. *Pédagogie Médicale.* févr 2010;11(1):51-6.
19. Linn AMJ, Tonkin A, Duggan P. Standard setting of script concordance tests using an adapted Nedelsky approach. *Med Teach.* avr 2013;35(4):314-9.
20. Duggan P, Charlin B. Summative assessment of 5th year medical students' clinical reasoning by script concordance test: requirements and challenges. *BMC Med Educ.* 9 mai 2012;12:29.
21. Wilson AB, Pike GR, Humbert AJ. Preliminary Factor Analyses Raise Concerns about Script Concordance Test Utility: Exploring SCT constructs. *Med Sci Educ.* mars 2014;24(1):51-8.
22. Dory V, Gagnon R, Vanpee D, Charlin B. How to construct and implement script concordance tests: insights from a systematic review. *Med Educ.* juin 2012;46(6):552-63.

23. Gagnon R, Lubarsky S, Lambert C, Charlin B. Optimization of answer keys for script concordance testing: should we exclude deviant panelists, deviant responses, or neither? *Adv Health Sci Educ Theory Pract.* déc 2011;16(5):601-8.
24. Gagnon R, Charlin B, Coletti M, Sauve E, van der Vleuten C. Assessment in the context of uncertainty: how many members are needed on the panel of reference of a script concordance test? *Med Educ.* mars 2005;39(3):284-91.
25. See KC, Tan KL, Lim TK. The script concordance test for clinical reasoning: re-examining its utility and potential weakness. *Med Educ.* nov 2014;48(11):1069-77.

Aliénor Moussette

TESTS DE CONCORDANCE DE SCRIPTS EN ÉVALUATION SOMMATIVE EN INTERNAT DE MÉDECINE GÉNÉRALE À ROUEN : FIDÉLITE INTER-JUGE ET RÉFLEXIONS SUR L'ÉLABORATION DU TEST.

RÉSUMÉ

Contexte : Le test de concordance de scripts (TCS) est un outil d'évaluation répandu. Peu d'études sont disponibles en évaluation sommative. Certains débats persistent concernant l'étude insuffisante de la fidélité inter-juge, la trop grande confiance accordée au coefficient alpha, la nécessité ou non d'exclure les experts déviants et l'efficacité de certaines stratégies de réponse.

Buts: Apprécier la fidélité inter-juge des TCS utilisés en évaluation sommative en DES de médecine générale à Rouen et discuter de l'impact de certains critères d'élaboration du test sur le classement des étudiants.

Méthodes : Une analyse quantitative rétrospective des résultats aux TCS en DES de médecine générale à Rouen entre 2011 et 2016 a été effectuée. La fidélité inter-juge a été évaluée via le calcul de la corrélation intra-classe (ICC). L'impact sur le classement des étudiants a été évalué après amélioration du coefficient alpha, exclusion des experts déviants, éviction des réponses extrêmes, et utilisation des stratégies de réponses au hasard.

Résultats : La fidélité inter-juge était excellente ($ICC=0.984$). L'amélioration de la cohérence interne du test modifiait le classement des étudiants mais pas l'exclusion des experts déviants. Les stratégies d'éviction des réponses extrêmes et de réponse au hasard se révélaient parfois efficaces.

Conclusion : Le panel d'experts utilisé à Rouen en DES de médecine générale semble quantitativement et qualitativement adapté. Certaines améliorations concernant la construction des items pourraient être réalisées. Une méfiance devrait être accordée concernant l'interprétation du coefficient alpha. D'autres études sont nécessaires afin d'estimer quelles dimensions du raisonnement clinique sont mesurées par le TCS.

MOTS-CLÉS

test de concordance de script, raisonnement clinique, médecine générale, évaluation sommative