



**UNIVERSITÉ DE STRASBOURG**



**ÉCOLE DOCTORALE Sciences Humaines et Sociales – Perspectives Européennes**

**Laboratoire Interuniversitaire des Sciences de l'Éducation et de la Communication (LISEC)  
UR 2310 – Équipe Activité, Travail et Identité Professionnelle (ATIP)**

**THÈSE** présentée par :

**Mathieu LORENZO**

soutenue le : **20 septembre 2021**

pour obtenir le grade de : **Docteur de l'Université de Strasbourg**

Discipline/ Spécialité : Sciences de l'Éducation et de la Formation

**Développer le raisonnement clinique, la réflexivité  
et le professionnalisme des futurs médecins  
généralistes**

**THÈSE dirigée par :**

**Pr TRIBY Emmanuel**

Sciences de l'éducation, Université de Strasbourg

**Pr FERNANDEZ Nicolas**

Professeur de clinique, Université de Montréal

**RAPPORTEURS :**

**Pr AUDETAT VOIROL Marie-Claude**

Sciences biomédicales, Université de Genève

**Pr GAGNAYRE Rémi**

Sciences de l'éducation, Université Sorbonne Paris nord

**AUTRES MEMBRES DU JURY :**

**Pr DURAT Laurence**

Sciences de l'éducation, Université de Haute Alsace

**Pr MORIN Marie-Paule**

Professeure de clinique, Université de Montréal

---

## **Remerciements**

### **Au Professeur Emmanuel Triby**

Merci d'avoir accepté de m'accompagner depuis le début dans ce travail de thèse. Les différents séminaires en Sciences de l'Éducation ainsi que nos échanges lors de nos entrevues m'ont permis d'enrichir mon approche conceptuelle et méthodologique de la recherche en Pédagogie des Sciences de la Santé. Malgré les difficultés, les inévitables retards et un parcours semé d'embûches, vous m'avez aidé et montré un soutien sans faille.

Certaines de vos thématiques telles que la didactique professionnelle et l'ingénierie de la formation et des compétences seront à coup sûr des axes de travail importants pour la suite de ma carrière universitaire. Merci pour tout cela.

### **Au Professeur Nicolas Fernandez**

Merci Nicolas de t'être proposé pour cette codirection dans des circonstances loin d'être idéales. Tu m'as apporté cette bienveillance légendaire de nos amis du Québec ainsi que ton expertise de pointe en Pédagogie des Sciences de la Santé. Merci pour ton engagement envers l'éducation médicale et notamment au sein de la communauté francophone.

Merci également pour ton accueil lors de mon séjour à Montréal et pour toutes les rencontres que tu m'as permises. Merci pour ton soutien au projet Hello-Clinical. Sans toi, ni cette thèse, ni le soutien du programme Samuel de Champlain, ni même mon grand intérêt pour les formations par concordance n'aurait existé.

### **À la Professeure Marie-Claude Audétat-Voirol**

Merci de me faire l'honneur de rapporter sur cette thèse. Vos travaux sur la supervision du raisonnement clinique sont de véritables références internationales. Merci pour votre engagement dans la communauté francophone en Éducation Médicale.

### **Au Professeur Rémi Gagnayre**

Merci de me faire l'honneur de sur rapporter cette thèse. Merci de vos critiques et remarques constructives lors des différents comités de suivi de thèse auxquels vous avez participé depuis le début de ce travail. Votre expertise en Pédagogie des Sciences de la Santé me fut fort utile.

**À la Professeure Laurence Durat**

Merci de me faire l'honneur de siéger dans le jury de cette thèse.

**À la Professeure Marie-Paule Morin**

Merci de me faire l'honneur de siéger dans le jury de cette thèse.

**Au Professeur Thierry Pelaccia**

Merci pour tous les échanges et débats autour des questions de pédagogie en Sciences de la Santé. Merci de m'avoir fait confiance au sein du CFRPS. Merci de ton aide précieuse en méthodologie de la recherche et en valorisation des travaux universitaires.

Ce travail n'aurait pas été possible sans toi, laisse-moi t'offrir quelques litres de coca zéro avec beaucoup de glaçons pour t'en remercier.

**À Divya**

Merci d'être là, de m'accompagner depuis toutes ces années. Ce travail n'aurait pas été possible sans ton soutien et ton aide. J'espère que ta relecture ne te donnera pas envie de mourir à cause d'un excès de fautes d'orthographe. Je progresse et j'ai d'ailleurs cette fois moi-même édité la page de garde de la thèse.

**À Naomi**

Merci d'être là, d'être un tel bonheur. Merci pour tes questionnements sur l'andragogie et tes suggestions pédagogiques.

**À ma famille**

À maman, papa, Charlotte : merci pour votre soutien et toutes les joies des dernières 35 années. Merci pour la curiosité intellectuelle qui a toujours été la vôtre et qui m'a permis de réaliser ce travail

de thèse. À mamie qui est tellement fière de ce travail. C'est un bonheur de te côtoyer. Merci à tous pour tous les bons moments.

### **À mes collègues de la maison de santé**

Sans vous je ne pourrais pas mener une telle carrière de clinicien et enseignant/chercheur. Travailler avec vous est une chance inouïe que je mesure davantage chaque jour.

Mes plus grands remerciements pour toi Vincent, qui supporte depuis près de 10 ans l'adaptation de mon activité de soins en fonction de l'évolution de mon activité universitaire. Sans toi, rien dans mon parcours pédagogique n'aurait été possible.

Merci à Marion qui a dû endurer les errements de mes premiers mois de maître de stage. Ta présence est précieuse et c'est une joie de travailler avec toi au jour le jour. J'espère que la graine de la pédagogie qui sommeille en toi te permettra bientôt d'accueillir à ton tour des étudiants !

Merci à Sylvain de nous avoir fait confiance toi aussi depuis ton installation. Là aussi c'est une chance de t'avoir avec nous et je ne doute pas que tu franchisses bientôt le pas pédagogique qui t'amènera à devenir maître de stage et aussi peut-être un futur universitaire !

Merci aussi à Lyria d'avoir incarné plus que quiconque cet exercice coordonné pluriprofessionnel qui bénéficie chaque jour à nos patients. J'espère que ta carrière te comblera de bonheur.

Merci aussi à Christophe de ton expertise, d'être celui qui met en musique notre structure et de nous soulager de la gestion d'un nombre incalculable de tâches au quotidien.

Merci enfin à tous les autres, Joan, Karina, Simona, Lara, Sophie, Ouarda et Sophie de votre gentillesse, de votre professionnalisme et de tout simplement permettre le fonctionnement de notre structure.

### **À mes collègues du Département de Médecine Générale**

Merci de votre soutien, de votre gentillesse et votre fun. Cette équipe est vraiment exceptionnelle et je ne peux que souhaiter que cela dure le plus longtemps possible. Merci de m'avoir fait confiance dans des circonstances difficiles en 2014. Merci de m'avoir encore davantage confiance cette année pour diriger l'équipe. Merci en particulier à Philippe, sans qui, cette histoire n'aurait jamais démarré. Merci de ton appel et merci pour notre article sur le burnout.

Merci à tous les collègues à bouclette ou non de développer une véritable appétence pour la Pédagogie et de solides compétences en la matière. Rose est une rose est une rose est une rose.

### **Au Doyen Jean Sibilia**

Qui m'a fait confiance un soir de 2014, qui m'a soutenu pour devenir titulaire quelques années plus tard. Merci de ton soutien à la Médecine Générale. Merci de ton soutien au développement de la recherche en Soins Primaires. Merci de ton soutien au développement de la Pédagogie en Sciences de la Santé.

On avance ensemble !

### **À chloé**

Nous avons partagé tant de choses dans nos parcours universitaires. Merci pour toutes ces discussions parfois enflammées sur la formation et le fonctionnement de l'équipe. J'espère au plus vite pouvoir travailler à nouveau avec toi et ainsi bénéficier de ton expertise en recherche et en Pédagogie des Sciences de la Santé.

### **À Racha et Élodie**

Merci pour vos travaux en Pédagogie des Sciences de la Santé. Cette thèse n'aurait pas vu le jour sans vous. J'ai beaucoup appris (parfois à vos dépens) dans la supervision d'un travail de mémoire de recherche. Merci pour ces séminaires sur l'évaluation des apprentissages, tous ces modèles de Kane et autres croissants.

Nous mènerons et publierons à l'avenir bien d'autres travaux ensemble je l'espère.

### **À mes amis**

La vie ne serait pas belle sans vous.

## Table des matières

<b>Liste des tableaux</b>	<b>8</b>
<b>Liste des figures</b>	<b>9</b>
<b>Préambule</b>	<b>10</b>
<b>Posture épistémologique</b>	<b>10</b>
<b>la formation des médecins généralistes en France</b>	<b>11</b>
<b>Le profil souhaité des futurs médecins généralistes : les compétences attendues</b>	<b>16</b>
<b>Première partie : le raisonnement clinique</b>	<b>21</b>
<b>De la découverte de l'incertitude dans les soins au raisonnement clinique</b>	<b>21</b>
<b>Le raisonnement clinique</b>	<b>23</b>
<b>Évaluer le raisonnement clinique</b>	<b>29</b>
<b>Des outils et des méthodes pour évaluer le processus de raisonnement clinique</b>	<b>30</b>
<b>Un modèle universel du raisonnement clinique : le modèle MOT</b>	<b>31</b>
<b>Des outils pédagogiques visant à développer le raisonnement clinique et à illustrer l'incertitude dans les soins : les formations par concordance</b>	<b>45</b>
<b>Le test de concordance de script (TCS)</b>	<b>45</b>
<b>Un exemple de recherche sur le raisonnement clinique utilisant le TCS : le burnout affecte-t-il le raisonnement clinique ? Une étude observationnelle chez des internes de médecine générale</b>	<b>49</b>
Contexte de l'étude	50
Matériel et méthodes	50
Résultats	53
Discussion	54
Synthèse de l'étude	62
<b>Au-delà du TCS : les formations par concordance</b>	<b>63</b>
<b>Le développement de l'outil Hello-clinical : un outil permettant d'utiliser des patients virtuels pour évaluer le processus de raisonnement clinique</b>	<b>73</b>
<b>Un exemple de recherche sur le raisonnement clinique utilisant Hello-Clinical : lien entre expérience et prescriptions complémentaires en médecine générale. Une étude observationnelle utilisant des patients virtuels</b>	<b>79</b>
Contexte de l'étude	79
Matériel et méthodes	80
Résultats	81
Discussion	84
Synthèse	100
<b>Deuxième partie : la réflexivité comme moteur du développement du raisonnement clinique</b>	<b>102</b>
<b>De la métacognition à la réflexivité</b>	<b>102</b>

<b>Accompagner le développement de la réflexivité via des écrits réflexifs-----</b>	<b>106</b>
<b>Évaluer la réflexivité des traces écrites d'apprentissages : la traduction de la grille REFLECT-----</b>	<b>109</b>
Contexte de l'étude-----	109
Matériel et méthodes-----	111
Résultats-----	112
Discussion-----	112
Synthèse-----	126
<b>Le lien entre les formations par concordance et la réflexivité-----</b>	<b>127</b>
Contexte de l'étude-----	127
Matériel et méthodes-----	128
Résultats-----	128
Discussion-----	129
Synthèse-----	155
<b>Troisième partie : développer le professionnalisme en médecine générale-----</b>	<b>157</b>
<b>Professionnalisme, professionnels et professionnalisation-----</b>	<b>157</b>
<b>Comment accompagner le développement du professionnalisme ?-----</b>	<b>161</b>
<b>Enrichir la rétroaction : inclure des patients dans les panels de référence des formations par concordance de jugement-----</b>	<b>162</b>
Contexte de l'étude-----	163
Matériels et méthodes-----	163
Résultats-----	164
Discussion-----	168
Synthèse-----	193
<b>Synthèse et perspectives-----</b>	<b>194</b>
<b>Bibliographie-----</b>	<b>197</b>
<b>Annexes-----</b>	<b>213</b>
<b>Annexe n°1 : indicateurs des niveaux de développement des compétences en médecine générale</b>	<b>213</b>
<b>Annexe n°2 : Les problèmes à éléments clés-----</b>	<b>220</b>
<b>Annexe n°3 : Qu'est-ce que le test de concordance de script ?-----</b>	<b>225</b>

## LISTE DES TABLEAUX

Tableau 1 : représentations des connaissances selon la modélisation par objets typés.....	32
Tableau 2 : exemple de questions de TCS.....	47
Tableau 3: seuils des dimensions du burnout du MBI-HSS dans l'approche conventionnelle.....	51
Tableau 4 : exemple de questions de TCS employées dans l'étude.....	52
Tableau 5 : Répartition des scores des dimensions du burnout dans l'approche conventionnelle.....	53
Tableau 6 : seuils des dimensions du burnout du MBI-HSS dans l'approche " épuisement +1 ".....	53
Tableau 7 : comparaison des scores moyens du SCT sur la sévérité du burnout dans l'approche conventionnelle.....	54
Tableau 8 : analyse univariée des caractéristiques des groupes.....	83
Tableau 9 : grille nationale d'évaluation des RSCA.....	109
Tableau 10 : grille REFLECT dans sa version originale.....	110
Tableau 11 : détails des répartitions des réponses des panels à la FpC-jugement.....	168

## LISTE DES FIGURES

Figure 1 : synthèse des liens entre les différents travaux réalisés.....	19
Figure 2 : Représentation du double processus selon Croskerry (en anglais).....	27
Figure 3: modèle principal du processus de raisonnement clinique selon la MOT.....	34
Figure 4: détail de la procédure "identifier les indices dès les premiers instants".....	35
Figure 5 : composants et spécialisations du concept de "contexte".....	36
Figure 6 : Attributs et composants du concept de "patient".....	37
Figure 7 : sous-procédures et intrants du concept de "déterminer les objectifs de la rencontre".....	38
Figure 8 : sous-procédures et intrants de la procédure « catégoriser pour décider de l'action ».....	40
Figure 9 : sous-procédures et intrants de la procédure « mettre en œuvre les actions appropriées ».....	42
Figure 10: sous-procédures et intrants de la procédure « évaluer les résultats ».....	43
Figure 11 : sous-procédures, intrants et spécialisations de la procédure « utiliser des stratégies de recours ».....	44
Figure 12 : exemple de vignette de FpC-raisonnement.....	64
Figure 13 : exemple de questions d'une vignette de fpc-jugement utilisant le format "A".....	66
Figure 14 : exemple de questions d'une vignette de fpc-jugement utilisant le format "B".....	68
Figure 15 : exemple de feedback de FpC-jugement concernant les questions présentées dans le "format A".....	70
Figure 16 : deuxième exemple de feedback de FpC-jugement concernant les questions présentées dans le "format B".....	71
Figure 17 : exemple de synthèse d'un cas de fpc-jugement.....	72
Figure 18 : capture d'écran de la page d'accueil d'un scénario du logiciel Hello-Clinical.....	75
Figure 19 : capture d'écran de l'affiche du dossier médical simulé.....	76
Figure 20: boucle de l'apprentissage expérientiel selon Kolb.....	103
Figure 21 : l'apprentissage réflexif selon Boud, Keogh et Walker.....	104
Figure 22 : les composantes de la réflexivité selon Nguyen et al.....	106
Figure 23 : marguerite des compétences des médecins généralistes.....	158

## PRÉAMBULE

### POSTURE ÉPISTÉMOLOGIQUE

Ce travail s'inscrit dans le champ disciplinaire général des Sciences de l'Éducation, plus précisément de la Pédagogie des Sciences de la Santé. Bien que n'ayant pas de véritable reconnaissance universitaire en France (sous forme notamment de filière universitaire et de postes dédiés), nous allons prendre le parti comme Pelaccia et Triby de qualifier la Pédagogie des Sciences de la Santé de "discipline" (1,2). Cette discipline est également appelée "Pédagogie Médicale" ou "Éducation Médicale" dans la littérature - cette dernière dénomination étant la traduction de l'anglais "*medical education*". Je choisirai de favoriser dans ce tapuscrit la dénomination de Pédagogie en Sciences de la Santé pour désigner cette discipline. En effet, si comme les terminologies "Éducation Médicale" et "Pédagogie Médicale" le laissent supposer, cette discipline est née dans les Facultés de Médecine (2), elle s'intéresse à présent à la formation de l'ensemble des professionnels de santé.

L'éducation médicale est définie en anglais sur Wikipédia (il n'existe pas au jour de l'écriture de cette thèse d'article en français sur les termes de "Pédagogie Médicale" ou "Éducation Médicale") comme "*l'éducation liée à la pratique de la médecine, y compris la formation initiale pour devenir médecin (les premier et deuxième cycles en Faculté de Médecine et les stages) et la formation supplémentaire par la suite (troisième cycle, bourse de recherche et formation médicale continue)*. [...] *L'éducation médicale est également le domaine académique didactique de la formation des médecins à tous les niveaux, y compris la formation médicale initiale, de troisième cycle et continue*" (3). Il s'agit selon ce même article d'un domaine "*actif de recherche en éducation*" (3).

La Pédagogie des Sciences de la Santé est inscrite depuis plus de 20 ans dans le mouvement de l'éducation fondée sur les preuves (4). Le but de ce mouvement est de s'éloigner d'une Éducation fondée sur des "opinions" pour tendre vers des pratiques pédagogiques reposant sur des données scientifiques (comme pour le mouvement de la Médecine fondée sur des preuves – *evidence-based medicine* – dans les soins) (4). Pour Albert *et al.*, la recherche en Pédagogie des Sciences de la Santé penche fortement vers le pôle "*production pour les utilisateurs*" (par opposition au pôle "*production pour les producteurs*") (5). Il s'agit comme l'écrivent Pelaccia et Triby de "*résoudre les problèmes soulevés par les utilisateurs de terrain des*

*connaissances produites par la discipline*" (2). Ces mêmes auteurs appellent cependant à diversifier les productions de la discipline pour favoriser son développement (2,6).

Dans cette thèse, la majorité des travaux pourrait se classer plutôt à proximité du pôle "*production pour les utilisateurs*" d'Albert (5).

Les travaux présentés dans ce tapuscrit empruntent à deux paradigmes principaux de recherche en Pédagogie des Sciences de la Santé : le post-positivisme et l'interprétativisme (7). Dans le post-positivisme, la posture épistémologique est que la "*connaissance objective du monde n'est pas nécessairement complètement accessible et le but est d'établir une vérité probable*" selon Bunniss et Kelly (7). Concernant l'interprétativisme, la posture épistémologique est que "*la connaissance est subjective et qu'il existe de multiples et diverses interprétations de la réalité*" (7).

Le post-positivisme découle relativement naturellement de la culture dominante de la recherche en Médecine et de la formation des médecins en France. Il s'agit également d'une posture globalement dominante en Pédagogie des Sciences de la Santé (2). Cependant, j'ai également été influencé par l'interprétativisme dans ma formation en recherche en Médecine Générale ainsi que dans ma formation en Pédagogie des Sciences de la Santé.

En tant que formateur au sein d'une Faculté de Médecine, les questions principales qui sous-tendent mon activité de formation sont : comment mieux former nos étudiants afin qu'ils deviennent de meilleurs soignants ? Comment nos choix pédagogiques peuvent-ils avoir un impact sur la qualité des soins prodigués par les étudiants ?

Puisque mon activité actuelle principale de formation concerne les futurs médecins généralistes, je propose de présenter les grandes lignes de la formation des médecins généralistes en France, de détailler le profil de formation attendu avant d'explicitier les problématiques ayant conduit à la réalisation des différents travaux de recherche présentés dans ce document.

## LA FORMATION DES MÉDECINS GÉNÉRALISTES EN FRANCE

La formation actuelle des médecins généralistes en France s'effectue au sein d'une Faculté de Médecine. La durée minimale de formation est de neuf années réparties en trois cycles successifs. Je

commencerais par décrire l'état de la formation des médecins généralistes à la période à laquelle je fus étudiant moi-même (2003-2012) pour décrire ensuite la formation proposée à la Faculté de Médecine de Strasbourg en 2021 à l'issue des évolutions récentes. Si les grandes lignes de la formation des médecins sont définies au niveau national et sont donc communes à toutes les Facultés de Médecine de France, certains choix pédagogiques locaux font varier les types de stages proposés, leurs durées, l'alternance avec la formation théorique et les modalités d'évaluation des apprentissages. Il n'est donc pas possible de brosser un portrait unique de la formation des médecins généralistes en France en 2021 mais seulement d'en esquisser les principaux contours.

L'accès aux études de médecine était historiquement conditionné par la réussite après obtention du baccalauréat d'un concours d'admission très sélectif ayant lieu en fin de première année (Première Année Commune des Études de Santé – PACES ; anciennement nommée "P1"). Ma formation initiale en Médecine a débuté en 2003 à la Faculté de Médecine de la Pitié-Salpêtrière à Paris (actuellement Sorbonne Université). La première année était principalement dédiée à l'acquisition superficielle de connaissances sur l'histoire de la médecine, la physique, les mathématiques et les sciences fondamentales de la vie. Le concours permettait à environ un étudiant sur sept d'accéder à la deuxième année. L'évaluation des apprentissages consistait majoritairement en des questions à choix multiples (QCM). Ces QCM exploraient principalement les connaissances déclaratives, du "savoir par cœur".

Dès la deuxième année, la formation prenait la forme d'une alternance entre les stages à l'hôpital et les enseignements théoriques à la faculté. Le contenu des enseignements était alors davantage professionnalisant : il établissait des liens entre les savoirs fondamentaux et la pratique de la médecine.

Le deuxième cycle suivait ce modèle d'alternance entre la troisième et la sixième année. Il visait à acquérir les connaissances de base nécessaires pour raisonner face à un problème clinique mais aussi à préparer le concours national de fin de sixième année : les Épreuves Classantes Nationales (ECN). Les ECN n'étant pas des examens, il n'y avait pas de score minimal exigé pour intégrer le troisième cycle : il était déterminé par le résultat de chacun aux ECN. Les étudiants (entre 6000 et 8000 par an) choisissaient, en fonction de leur classement, leur subdivision (faculté de rattachement) et leur spécialité médicale ou chirurgicale pour le troisième cycle. Le nombre de postes à pourvoir par discipline et par subdivision était fixé par un arrêté ministériel annuel. Dans le deuxième cycle, les facultés les mieux organisées

proposaient au maximum trois mois de stage en médecine générale. Beaucoup de facultés ne proposaient aucun stage dans cette spécialité. Pourtant, la moitié des étudiants de deuxième cycle deviendraient médecins généralistes. Seule une demi-journée sur les six années du premier et deuxième cycle portait sur la médecine générale et était animée par des médecins généralistes.

Le troisième cycle (nommé dans le langage courant "internat" ou "DES" pour diplôme d'études spécialisées) visait à former les étudiants à l'exercice de leur spécialité. Dans le cas de la spécialité médecine générale, trois années de formation étaient nécessaires. La formation se déroulait quasi-exclusivement en stage par périodes de six mois. A Strasbourg, où j'ai réalisé mon troisième cycle, la majorité des étudiants (internes) ne réalisaient que six mois de stage en médecine générale. Les plus chanceux bénéficiaient de deux stages (soit une année complète). Sur l'ensemble des études médicales, les futurs médecins généralistes les plus chanceux bénéficiaient donc de 15 mois de stage dans leur spécialité sur six à sept ans de stages. La grande majorité des étudiants n'y passait que six mois. La formation hors stage en troisième cycle se composait de quelques demi-journées par semestre d'enseignements obligatoires (une dizaine par an). Ces formations prenaient des formes variées : cours magistraux animés par les enseignants ou les pairs, groupes d'analyse de pratique, analyse critique d'articles. Un séminaire obligatoire d'une journée par an avait lieu : l'accueil des nouveaux internes, une formation sur la thèse et la démarche qualité en médecine générale, une session de préparation aux examens de fin de formation. En toute fin d'internat (en fin de neuvième année des études médicales), deux examens certificatifs visaient à vérifier que tous les étudiants soient aptes à exercer la médecine générale : des tests de concordance de script (TCS) et une série de jeux de rôle standardisés appelés des "examens cliniques objectifs structurés" (ECOS). Les TCS étaient utilisés pour documenter la qualité du raisonnement clinique et les ECOS des compétences relationnelles, certains aspects du raisonnement clinique et du savoir-faire comme des manœuvres cliniques. La validation des stages étant quasi-systématique même en cas de carences professionnelles graves, certains étudiants se retrouvaient identifiés comme en difficulté seulement le dernier mois de leur dernière année de formation sans réelles possibilités de remédiation. L'obtention du diplôme était enfin conditionnée par la production et la soutenance d'un mémoire de DES de médecine générale ainsi qu'une thèse d'exercice.

Plusieurs réformes d'ampleur ont profondément transformé les études médicales ces dernières années sans en modifier la durée. Depuis la rentrée 2020, la réforme du premier cycle des études médicales a introduit deux nouvelles voies d'accès principales aux études de médecine au lieu du concours de première année : le parcours spécifique "accès santé" (PASS) et les licences avec option "accès santé" (L.AS) (8). Une partie des candidats accéderont en deuxième année de médecine en fonction de leurs notes aux examens et une autre partie sera sélectionnée sur une série d'entretiens avec les candidats. Le premier objectif ici est d'éviter aux étudiants qui échoueront à intégrer une Faculté de Médecine de "perdre" une ou deux années. Les années de PASS ou L.AS serviront à valider une licence. Le deuxième objectif de cette réforme est de diversifier le profil des futurs médecins. Les oraux de sélection doivent notamment évaluer les capacités relationnelles des étudiants.

La réforme du deuxième cycle s'applique depuis la rentrée 2020 et tente d'inscrire davantage les études de médecine dans une approche par compétences (8,9). L'alternance entre stage et formation théorique n'est pas remise en question. Un stage entre six et seize semaines en médecine générale est obligatoire pour tous les étudiants (9). L'objectif principal du deuxième cycle reste l'acquisition de connaissances avec un examen national en sixième année. Il devrait être exigé un score minimal de 80% concernant un ensemble de connaissances jugées "indispensables à tout médecin" pour pouvoir accéder au troisième cycle. Sept compétences feront l'objet d'une évaluation dans les stages via un portfolio tout au long du deuxième cycle et des ECOS devront être organisés en fin de sixième année pour certifier ces compétences. Le programme se réclame de la conception de la compétence de Jacques Tardif : « *un savoir-agir complexe reposant sur la mobilisation et la combinaison efficace d'une variété de ressources internes et externes au sein d'une famille de situations* » (10). Les compétences visées en deuxième cycle sont nommées (i) clinicien, (ii) communicateur, (iii) coopérateur, (iv) acteur de santé publique, (v) praticien réflexif, (vi) scientifique, (vii) responsable aux plans éthique et déontologique (8). Ces compétences sont largement inspirées du référentiel de compétences CANMEDS en vigueur pour les médecins du Canada. On peut noter que ces compétences correspondent davantage à des "rôles" qu'à des compétences selon la définition de Tardif (11).

Une procédure complexe de "matching" inspirée des systèmes étasuniens et allemands doit remplacer le système du classement national pour permettre le choix de la spécialité et de la subdivision d'accueil en troisième cycle. Les étudiants candidateront dans les spécialités et les subdivisions de leurs

choix, et seront sélectionnés en fonction de leurs scores aux examens nationaux mais également en fonction du contenu de leur portfolio, sur des critères de motivation et d'un projet professionnel. Au jour de la rédaction de ce document, les arbitrages nécessaires n'ont pas encore été réalisés au niveau national.

La réforme du troisième cycle des études médicales est entrée en vigueur à la rentrée 2017 (12). Je vais en détailler les principales modalités pour le DES de médecine générale. Deux phases de formation au sein du troisième cycle sont distinguées : une phase "socle" d'une année et une phase "d'approfondissement" de deux années. A noter que pour les DES plus longs que celui de médecine générale, une troisième phase nommée "de consolidation" correspondant à un nouveau statut de "docteur junior" prend place sur les quatrième voire parfois cinquième et sixième années. Une validation de chaque phase est nécessaire pour passer à la suivante introduisant des "points d'étape" bienvenus dans le DES de médecine générale. Les étudiants doivent développer six compétences spécifiques à la médecine générale. Celles-ci se nomment (i) premier recours/urgence, (ii) relation/communication/approche centrée patient, (iii) approche globale/prise en compte de la complexité, (iv) éducation/prévention/santé individuelle et communautaire, (v) continuité/suivi/coordination des soins autour du patient et (vi) professionnalisme.

Pour chaque compétence, trois niveaux de développement sont décrits : novice, intermédiaire et compétent. Pour chaque niveau, diverses grilles utilisant des descripteurs de niveaux de compétences sont utilisées pour documenter le niveau de développement. Il est exigé un certain niveau minimal de développement en fin de phase socle pour accéder à la phase d'approfondissement ainsi qu'en fin de phase d'approfondissement pour valider le DES.

La maquette des stages est devenue extrêmement rigide avec prescription de l'exacte nature des stages à réaliser selon la phase. Les futurs médecins généralistes doivent à présent effectuer six mois de stage en médecine générale ainsi que six mois dans un service d'accueil des urgences en phase socle. Ils doivent ensuite réaliser six mois de stage dans un service hospitalier de "médecine adulte" en deuxième année et refaire six mois de stage en médecine générale avec une large autonomie en troisième année. Ils doivent réaliser six mois de stage "santé de l'enfant" et six mois "santé de la femme" indifféremment en deuxième ou troisième année. Ces deux derniers stages peuvent s'effectuer exclusivement en milieu hospitalier, en ambulatoire dans des structures spécialisées type protection maternelle et infantile (PMI)

ou planning familial, ou auprès de médecins généralistes ayant une activité fortement orientée sur ces thématiques. Cette maquette permet d'offrir au minimum un an de stage en troisième cycle en médecine générale aux futurs médecins généralistes. Au maximum, deux années de stage en médecine générale peuvent être effectuées sur les trois années de formation.

La formation hors-stage comprend la participation à une dizaine de groupe d'échange et d'analyse de pratique ainsi que des formations optionnelles à choisir parmi une offre qui en comprend une trentaine. Les étudiants tiennent un portfolio qui contient diverses traces d'apprentissage : des récits narratifs réflexifs, un journal de bord, des rapports de formation, des rapports de supervision en stage. Un système de tutorat est mis en place pour les trois années du DES avec un accompagnement individuel.

La validation des différentes phases s'effectue en documentant devant un jury l'atteinte du niveau de compétences atteint à l'aide du portfolio.

La formation actuelle des médecins généralistes en France semble ainsi avoir nettement progressé ces dernières années en termes de qualité et de cohérence pédagogique. Il reste pourtant de nombreux défis lorsque l'on s'y penche de plus près.

#### LE PROFIL SOUHAITÉ DES FUTURS MÉDECINS GÉNÉRALISTES : LES COMPÉTENCES ATTENDUES

Dans l'arrêté de 2017 fixant les modalités pédagogiques du DES de médecine générale (12), les six compétences transversales citées au chapitre précédent prescrivent le profil attendu des futurs médecins généralistes. Ces compétences sont issues d'un travail de consensus d'experts de la discipline publié en 2013 par Compagnon *et al.* (13). La création de ce référentiel n'a pas été précédée d'un travail rigoureux de l'analyse de l'activité du médecin généraliste. Alors que les intitulés de ces compétences sont relativement obscurs à la première lecture, les définitions fournies par Compagnon *et al.* permettent de mieux cerner ce profil attendu des futurs généralistes (13) :

**Premier recours, urgence** : gérer avec la personne les problèmes de santé indifférenciés, non sélectionnés, programmés ou non, selon les données actuelles de la science, le contexte et les possibilités de la personne, quels que soient son âge, son sexe, ou toute autre caractéristique, en organisant une accessibilité (proximité, disponibilité, coût) optimale.

**Relation, communication, approche centrée patient** : construire une relation avec le patient, son entourage, les différents intervenants de santé ainsi que les institutionnels, en utilisant dans les différents contextes les habiletés communicationnelles adéquates, dans l'intérêt des patients.

**Approche globale, prise en compte de la complexité** : mettre en œuvre une démarche décisionnelle centrée patient selon un modèle global de santé (EBM, Engel, etc.) quel que soit le type de recours aux soins dans l'exercice de la médecine générale.

**Éducation, prévention, santé individuelle et communautaire** : accompagner « le » patient dans une démarche autonome visant à maintenir et améliorer sa santé, prévenir les maladies, les blessures et les problèmes psychosociaux dans le respect de son propre cheminement, et donc à intégrer et à articuler dans sa pratique l'éducation et la prévention.

**Continuité, suivi, coordination des soins autour du patient** : Capacité à assurer la continuité des soins et la coordination des problèmes de santé du patient engagé dans une relation de suivi et d'accompagnement.

**Professionalisme** : assurer l'engagement envers la société et à répondre à ses attentes, à développer une activité professionnelle en privilégiant le bien-être des personnes par une pratique éthique et déontologique, à améliorer ses compétences par une pratique réflexive dans le cadre de la médecine fondée sur des faits probants, à assumer la responsabilité des décisions prises avec le patient.

A partir de ce travail de définition des compétences à développer pour exercer la médecine générale, un second travail de consensus d'experts non formalisé par la même équipe a défini trois niveaux de développement pour chacune de ces six compétences : novice, intermédiaire et compétent (14). Pour chacun de ces niveaux, une liste d'indicateurs explicite ce qui le caractérise. Ces indicateurs permettent, encore plus finement, de cerner le profil attendu en fin de formation chez les étudiants (internes) en DES de médecine générale. Les tableaux en annexe n°1 présentent ces différents indicateurs selon le niveau et selon la compétence d'après le travail de Compagnon *et al.* (14).

Ces indicateurs sont considérés par certains auteurs davantage comme des "descripteurs" des différents niveaux de compétences (15). Un travail supplémentaire pourrait consister à lister des éléments directement observables en supervision ou dans les traces d'apprentissages écrites qui rendraient compte du développement des compétences (15).

Les noms des compétences retenues pour le référentiel de médecine générale interpellent par rapport aux termes habituellement employés dans les professions de santé (16–18). La compétence "premier recours, urgences" se réfère à la gestion des problèmes de santé des patients en médecine générale. Il s'agit en réalité de développer le raisonnement clinique des futurs généralistes (16). Si le professionnalisme fait partie du profil habituellement exigé chez les professionnels de santé, sa caractérisation comme compétence est sujette à débat (18–20). Le concept de réflexivité est mentionné dans la définition de la compétence "professionnalisme" mais n'est pas ensuite explicite dans les indicateurs de niveaux de développement. On devine entre les lignes son rôle, en particulier dans l'indicateur "améliore ses compétences" du niveau "compétent". La réflexivité est une compétence individualisée dans le deuxième cycle des études médicales sous le terme "praticien réflexif" (9). Il s'agit d'une caractéristique importante du profil attendu des futurs médecins généralistes (17).

L'article 5 de l'arrêté du 21 avril 2017 relatif au troisième cycle des études de médecine stipule que "*L'utilisation de méthodes pédagogiques innovantes dans le cadre d'une approche par compétences et adaptées aux caractéristiques des étudiants concernés est encouragée.*" (21). Il est précisé dans le même article qu'est notamment privilégiée l'utilisation de méthodes "*d'enseignement à distance asynchrone utilisant des modalités différées d'apprentissage ; d'évaluation et d'échange d'information*" (21). L'annexe du même arrêté concernant le DES de Médecine Générale prescrit l'utilisation d'un portfolio avec une soutenance argumentée de son contenu afin d'évaluer les compétences des étudiants (21).

Mon travail de doctorat en sciences de l'éducation s'inscrit dans ce contexte de la formation des futurs médecins généralistes au sein du troisième cycle de médecine générale. J'ai débuté ma carrière de formateur dans un environnement où la plupart des outils et méthodes utilisés ne semblaient pas reposer sur un argumentaire pédagogique solide. Ce cadre de formation prescrit et finalement relativement rigide m'a interrogé en regard des pratiques pédagogiques locales en troisième cycle de Médecine Générale à la Faculté de Médecine de Strasbourg. Les modalités pédagogiques proposées aux étudiants favorisent-elles les apprentissages visés ? Comment accompagner au mieux le développement des compétences professionnelles des étudiants ? Comment l'évaluation des apprentissages proposée peut-elle favoriser le transfert des apprentissages de la salle de classe en stage ?

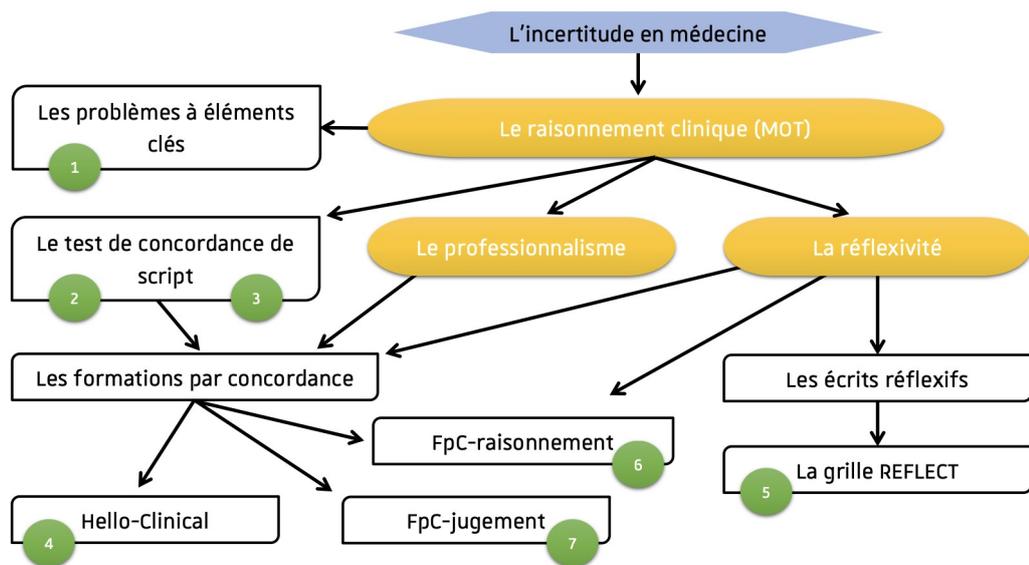
L'objectif global de ce travail était donc avant tout de contribuer à améliorer la formation des futurs médecins généralistes.

En lien avec ces questions, ce tapuscrit décrit mon cheminement intellectuel concernant trois des aspects critiques du profil attendu des futurs médecins généralistes : le raisonnement clinique, la réflexivité et le professionnalisme.

Concernant la forme, cette thèse est une thèse par articles. Il s'agit donc d'un ensemble de travaux différents autour de la question de la formation des médecins généralistes en France. Ce choix de la thèse par articles a été guidé principalement par les exigences universitaires en matière de carrière qui valorisent principalement la publication d'articles originaux. Ce format démontre ma capacité à créer des protocoles de recherche originaux et à valoriser ces travaux par la publication d'articles dans les revues de Médecine Générale et de Pédagogie des Sciences de la Santé.

Le présent tapuscrit tente de présenter l'ensemble des travaux réalisés de manière cohérente en retraçant cinq années de recherches en Pédagogie des Sciences de la Santé.

A partir de l'identifications de difficultés dans le développement de ces trois caractéristiques attendues, je propose une série de travaux ayant pour objectif d'améliorer la formation des étudiants de troisième cycle de médecine générale. Le schéma de la figure n°1 ci-dessous illustre ce cheminement et les liens entre les études réalisées. Les numéros dans les bulles vertes correspondent aux sept articles issus de la thèse.



**FIGURE 1 : SYNTHÈSE DES LIENS ENTRE LES DIFFÉRENTS TRAVAUX RÉALISÉS**

A partir de la découverte en stage de l'incertitude en Médecine, je me suis intéressé au processus de raisonnement clinique. Mes lectures m'ont fait découvrir deux outils d'évaluation du raisonnement clinique : les problèmes à éléments clés et le test de concordance de script. Je me suis intéressé au concept de professionnalisme en Pédagogie des Sciences de la Santé et à celui de réflexivité. J'ai travaillé sur l'utilisation de la méthode de la concordance appliquée à la formation des futurs médecins généralistes et j'ai travaillé à traduire et adapter culturellement un outil d'évaluation de la réflexivité des traces écrites d'apprentissage.

Cette thèse se divisera en trois parties : le raisonnement clinique, la réflexivité puis le professionnalisme. Je détaillerai dans chaque partie le cheminement intellectuel qui fut le mien et présenterai le cadre conceptuel des différents travaux réalisés. Une dernière partie de synthèse et perspective viendra conclure le document.

## PREMIÈRE PARTIE : LE RAISONNEMENT CLINIQUE

### DE LA DÉCOUVERTE DE L'INCERTITUDE DANS LES SOINS AU RAISONNEMENT CLINIQUE

L'incertitude dans la pratique de la médecine ne m'a réellement interpellé que lors de mon premier stage en médecine générale. Ce stage a eu lieu en 8<sup>e</sup> année (deuxième année d'internat). Il m'est apparu que l'incertitude n'était pas uniquement liée à un défaut de connaissance de ma part mais était constitutive de l'exercice de la médecine générale. Pour soigner nos patients, il fallait donc tolérer cette incertitude.

J'ai immédiatement trouvé le sujet passionnant et j'ai décidé d'y consacrer ma thèse d'exercice.

Mes recherches documentaires m'ont initialement amené à considérer l'incertitude en médecine générale sous l'angle des symptômes médicalement inexpliqués (SMI). Ceux-ci sont définis comme des symptômes vagues, qui, malgré un interrogatoire et un examen médical approfondi, ne permettent pas de retenir une hypothèse diagnostique (22). Une étude réalisée en Hollande au début des années 2000 concluait que 12,2 % des consultations de médecine générales concluent à un SMI (23). De plus, 8,5% des consultations seraient également considérées comme « incertaines » au niveau du diagnostic sans rentrer strictement dans les critères des SMI (23). En médecine générale, une consultation sur cinq se déroulerait donc en situation d'incertitude diagnostique (23). Tolérer cette incertitude diagnostique est souhaitable pour plusieurs raisons pour le Collège des Médecins Généralistes des Pays-Bas (22) :

- la majorité des SMI vus en soins primaires disparaîtront rapidement spontanément sans traitement ;
- la probabilité pré-test des examens complémentaires est faible ;
- le risque de faux-positif est important et la conduite à tenir en soins primaire est peu étayée par la littérature.

Face à cette incertitude prégnante, divers auteurs anglo-saxons appellent dès la fin des années 1970 à développer une forme nécessaire de tolérance à l'incertitude en médecine générale (24,25).

En France, le travail de Géraldine Bloy (26) représente une contribution remarquable à l'étude de l'incertitude en médecine générale. Cette sociologue a réalisé un travail de synthèse autour de cette

question dans une démarche de théorisation ancrée. Elle a notamment réalisé des dizaines d'observations de consultations et conduit 200 entretiens avec des médecins généralistes (26).

Elle affirme dans son travail que l'incertitude est constitutive de l'exercice de la médecine générale en lien avec deux caractéristiques principales du métier : le rapport du généraliste au savoir médical et sa place dans le système de soins (26).

Concernant le rapport au savoir médical, Géraldine Bloy reprend les travaux de Renée Fox (27) qui considèrent que l'incertitude médicale dépend de trois niveaux : un premier niveau lié à une maîtrise incomplète ou imparfaite du savoir médical disponible, un deuxième niveau lié aux limites propres à la connaissance médicale et un dernier niveau qui tient à la difficulté pour le médecin de faire la part du premier niveau et du second.

L'incertitude de niveau 1 est ainsi particulièrement répandue en médecine générale. Le médecin généraliste est confronté à un champ de connaissances à maîtriser potentiellement le plus étendu (26). Le médecin généraliste peut être considéré comme celui qui sait "*un peu de tout*" à l'inverse d'un spécialiste d'organe qui saurait "*tout de un peu*".

Concernant la place de la médecine générale dans le système de santé en France, il s'agit pour la majorité des patients du premier recours dans l'accès aux soins (26). Cette position de première ligne confronte le médecin généraliste à la plus grande variété de motifs de consultations qu'il soit. Cette variété renforce l'incertitude de niveau 1 puisqu'il est impossible de maîtriser tout le savoir médical disponible pour l'ensemble des motifs de consultations rencontrés en médecine générale. Enfin, cette position de premier recours conduit les médecins généralistes à rencontrer les pathologies à des stades précoces et souvent indifférenciés. Une fièvre isolée pourra deux jours plus tard correspondre à une banale grippe ou se révéler une rare maladie auto-immune. L'incertitude diagnostique est souvent moindre pour les médecins exerçant en deuxième et troisième ligne du fait de l'apparition secondaire fréquente d'autres signes et symptômes ou de la disponibilité de résultats d'examens complémentaires. L'incertitude en médecine générale est donc aussi liée à sa position spécifique dans le système de santé en France.

À partir de ces lectures sur l'incertitude en médecine générale et la nécessité d'y faire face, je me suis intéressé à la notion de tolérance à l'incertitude.

L'un des abords possibles est la notion inverse : l'intolérance à l'incertitude. En psychologie, l'intolérance à l'incertitude est définie comme « un état résultant de croyances négatives sur l'incertitude et ses implications » (28). Elle se caractérise par un désir de prédictibilité, une recherche de certitude et une impossibilité de réfléchir et de réagir face à une situation d'incertitude (28,29). L'intolérance à l'incertitude semble impliquée dans le développement de troubles anxieux et de dépressions (30–32). Le souhait d'éviter de se confronter à l'incertitude conduirait en outre certains étudiants pré-gradués à éviter de choisir comme spécialité la médecine générale (33). Chez les médecins généralistes, l'intolérance à l'incertitude est par ailleurs associée à un risque plus élevé de burnout (34). L'intolérance à l'incertitude conduit les jeunes médecins à "cacher" l'incertitude aux patients (35).

À l'opposé, dans ces situations d'incertitude diagnostique, de nombreux médecins vont prendre des décisions sans ressentir le besoin de prescrire des examens complémentaires ou de demander d'avis spécialisé (23,36). Il s'agit d'un mécanisme de gestion de l'incertitude diagnostique, parfois appelé « tolérance à l'incertitude » (25,37,38). La tolérance à l'incertitude serait ainsi l'acceptation d'un niveau d'incertitude dans la résolution d'un problème clinique. Thomson proposait dès 1978 que des facteurs propres au médecin déterminait « un seuil personnel de tolérance à l'incertitude » (25). Pourtant, il n'existait pas à ma connaissance d'études ayant exploré les déterminants de la tolérance à l'incertitude chez les médecins généralistes. J'ai donc réalisé un travail d'identification de ces déterminants en réalisant une analyse thématique d'entretiens avec des médecins généralistes (39). Les déterminants de la tolérance à l'incertitude étaient la gravité de la situation, le contexte, les caractéristiques du patient et celles du médecin (39). En comparant ces données avec celles de la littérature, je me suis intéressé à la notion de prise de décision (*decision-making* en anglais). Celle-ci est un sous-processus d'un ensemble plus vaste et crucial pour tout professionnel de santé : le raisonnement clinique.

## LE RAISONNEMENT CLINIQUE

Le raisonnement clinique est un concept dont on retrouve de nombreuses définitions dans la littérature. A l'échelle même des experts du domaine, la définition du processus de raisonnement clinique, sa nature et ses déterminants ne font pas consensus (40).

Une définition historique du raisonnement clinique est celle de Barrows et Tamblyn de 1980 : "*l'ensemble des processus cognitifs qui permettent au professionnel de santé d'évaluer et de prendre en*

*charge le ou les problèmes de santé d'un patient*" (41). Les modèles actuels du raisonnement clinique vont désormais au-delà du seul aspect cognitif et postulent qu'il s'agit d'un processus complexe, contextuel, multifactoriel et non-linéaire (42-45).

Le raisonnement clinique est fréquemment décrit comme la compétence centrale de l'ensemble des professionnels de santé (40,46,47). La plupart des référentiels de formation des professionnels de santé lui accordent désormais une place importante. Dans les référentiels en médecine, le raisonnement clinique reçoit une attention particulière. Il est une compétence dans le référentiel du deuxième cycle français présenté précédemment (9). Il est présenté dans le référentiel CANMEDS au Canada en tant que concept fondamental du rôle "expert médical" (48). Il apparaît dans les référentiels du *Accreditation Council for Graduate Medical Education* aux États-Unis sous la forme de compétences dans les domaines "patient care" (49) ou "medical knowledge" (50). A l'échelle européenne, le projet *"Tuning Learning Outcomes/Competences for Undergraduate Medical Education in Europe"* décrit dans les habiletés "Carry out a consultation with a patient" et "Assess clinical presentations, order investigations, make differential diagnoses, and negotiate a management plan" des aspects centraux du raisonnement clinique comme la génération de diagnostics différentiels ou la négociation d'un plan de prise en charge d'un patient (51).

Dans d'autres milieux de formation, le raisonnement clinique est également central. Chez les infirmiers en France, il s'agit de la première compétence du référentiel de formation : "Evaluer une situation clinique et établir un diagnostic dans le domaine infirmier" (52). Au Canada, le raisonnement clinique fait partie des compétences de base des infirmières et infirmiers à développer lors de la formation (53).

Chez les orthophonistes en France, le raisonnement clinique est la compétence n°1 du référentiel de formation : "Analyser, évaluer une situation et élaborer un diagnostic orthophonique" (54). Au Canada, le raisonnement clinique transparait dans le rôle d'expert clinique (notamment dans la compétence essentielle "Planifier, mener et adapter une évaluation") (55).

Il en est de même pour de nombreuses autres professions de santé : pharmacie, kinésithérapie, maïeutique, ostéopathie, ergothérapie, dentaire, etc... L'ouvrage de référence de Joy Higgs *et al.* "Clinical reasoning in health professions" (47) consacre ainsi un chapitre dédié aux spécificités du raisonnement clinique dans ces différentes professions.

De plus en plus des projets nationaux et internationaux se développent, visant à créer des curriculums interprofessionnels pour développer le raisonnement clinique des professionnels de santé. Les institutions prennent conscience de la nécessité d'accompagner de tels programmes, un bon exemple étant le financement européen du programme "*Development of a longitudinal interprofessional clinical reasoning curriculum*" porté par Hege *et al.* (56).

Pourquoi le raisonnement clinique est-il donc si important chez les professionnels de santé ? Une étude australienne menée dans les années 90 a conclu que plus de la moitié des évènements cliniques indésirables survenus en une année étaient liés à un défaut de raisonnement clinique (57). Les erreurs de raisonnement clinique semblent fréquentes notamment en médecine. Une revue de la littérature par Berner et Graber estimait en 2008 entre 10 et 15% le taux d'erreurs diagnostiques dans les spécialités médicales cliniques (58).

Dans la formation des professionnels de santé, les difficultés de raisonnement clinique sont également une préoccupation importante. On estime qu'entre 5 et 15% des étudiants rencontrent des difficultés dans leur formation – essentiellement à cause de difficultés de raisonnement clinique (59,60). Ces difficultés de raisonnement clinique sont souvent identifiées tardivement (59,61). Plusieurs travaux tendent à indiquer que l'identification tardive et l'absence de remédiation des difficultés de raisonnement clinique chez les futurs professionnels de santé pourraient conduire à un risque pour la qualité des soins aux patients (59,62,63).

Le raisonnement clinique est fréquemment enseigné et analysé sous l'angle de la théorie du double processus. Ainsi, en situation de résolution de problème, le professionnel de santé se forge une représentation mentale du problème présenté par le patient (45,64). Cette représentation mentale est un modèle cognitif qui met en lien les informations de la situation clinique avec les connaissances antérieures du professionnel de santé (64). Elle est construite dans une démarche de génération et de vérifications d'hypothèses (42).

Deux types de processus cognitifs sont à l'œuvre dans cette démarche : des processus intuitifs et des processus analytiques. Ces processus sont fréquemment désignés respectivement comme « système 1 » et « système 2 » (42,65). Les processus intuitifs consistent à reconnaître une similarité avec des cas rencontrés précédemment ou une configuration caractéristique de données (65,66). Ces processus sont

rapides, sans efforts conscients et irrépessibles (67–69). Ils utilisent les informations immédiatement disponibles telles que les données contextuelles et sont très dépendants de l'état émotionnel du clinicien (65,69). En opposition aux processus intuitifs, les processus analytiques sont présentés comme des processus rationnels et délibérés de collecte active d'information puis d'application de règles logiques apprises au cours de la formation professionnelle (69,70). Ces processus sont plus lents et consomment davantage de ressources intellectuelles (65).

Il est intéressant de noter qu'une partie de la communauté des chercheurs préfère le terme de « non-analytique » pour désigner les processus intuitifs (42,70,71). Le terme « intuition » peut revêtir une dimension « mystique » dans sa dimension théologique voir induire une confusion avec « l'instinct ». Ce choix d'une dichotomie processus analytiques versus processus non-analytiques serait alors plus didactique et plus « scientifique ». Les partisans du terme « intuition » mettent en avant le lien évident avec les émotions induit par ce choix sémantique. Une des définitions possibles de l'intuition est « *un mode de connaissance, de pensée ou jugement, perçu comme immédiat (au sens de direct) - une manière rapide d'évaluer une situation en la mettant en rapport avec des situations similaires déjà connues* » (72). Je préfère à titre personnel l'emploi du terme « intuition » pour qualifier ce mode de génération d'hypothèses diagnostiques et j'utiliserai donc le terme de « processus intuitifs » pour désigner les mécanismes cognitifs s'y rapportant dans la suite du manuscrit.

Comment ces deux types de processus s'articulent-ils entre eux ? La vision historique du raisonnement clinique postulait une succession chronologique des processus intuitifs servant à générer des hypothèses puis des processus analytiques pour vérifier celles-ci (73). Les modèles récents tendent à considérer que ces deux types de processus fonctionnent en parallèle avec une prédominance de l'un ou l'autre type selon des facteurs tels que le niveau d'incertitude, la complexité de la situation clinique, le temps disponible et l'importance des enjeux liés à la résolution du problème (74–76). La complexité de la situation est une appréciation individuelle basée notamment sur l'expérience (77). Ce modèle d'interaction entre les processus intuitifs et analytiques est né dans les années 90 dans le champ de la psychologie cognitive (78,79). Il est appelé « *dual-process* » ou « double processus ». Les individus en général priorisent les solutions intuitives plutôt qu'analytiques (80). Les processus analytiques exercent alors une fonction de « monitoring » en veillant à ce que le système intuitif produise une réponse correcte. La figure 1 illustre (en anglais) ce double processus.

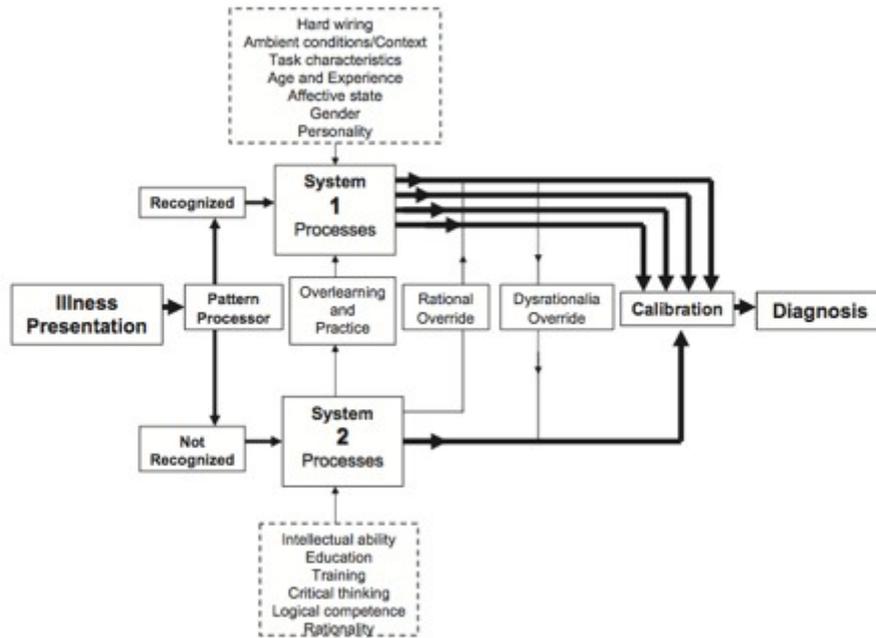


FIGURE 2 : REPRÉSENTATION DU DOUBLE PROCESSUS SELON CROSKERRY (EN ANGLAIS)

Cette représentation illustre bien les différents déterminants du processus de raisonnement clinique. D'une part le contexte, l'âge, l'expérience, les émotions, la personnalité, la nature du problème à résoudre, le genre et le « câblage » pour le système intuitif. D'autre part les capacités intellectuelles, le niveau d'éducation, la formation, la réflexivité, l'esprit de logique et pragmatisme pour le système analytique. Elle représente également que ces deux systèmes agissent simultanément avec habituellement une prédominance des processus intuitifs. Retenir un diagnostic équivaut à trouver un « point d'équilibre » entre ces deux processus : le système analytique ne détecte pas de « conflit » au niveau des hypothèses générées (69,75).

Ce double processus mobilise des connaissances cliniques spécifiquement organisées pour la pratique. On distingue au sein des modèles d'organisation des connaissances, d'une part des connaissances théoriques, et d'autre part des connaissances cliniques (16,45,73,77). Les connaissances théoriques sont de nature causale et visent à expliquer les différentes maladies et leurs symptômes en les mettant en rapport avec des données physiopathologiques ou psychosociales (16). Ces connaissances sont typiquement acquises en début de cursus et vont au contact des patients s'encapsuler aux connaissances cliniques des pathologies (16,77,81). Les connaissances cliniques sont de nature associative et visent à lier les signes et les symptômes d'une maladie, les examens, les traitements, etc... (16,81). Ces liens associatifs entre connaissances vont permettre de développer des prototypes de maladie (association caractéristique d'un ensemble de signes ou de symptômes) ainsi que des scripts

cliniques (des réseaux élaborés de connaissances permettant de déterminer les informations à collecter et de porter des micro-jugements sur les hypothèses diagnostiques) (16,16,42,46,77). Les prototypes sont issus de la mémorisation à long terme des signes ou symptômes les plus fréquemment rencontrés lors des rencontres cliniques du professionnel de santé (16). Il s'agit d'une étape d'abstraction des données les plus utiles à la génération d'une hypothèse diagnostique (16,64,82). La génération d'une hypothèse conduit à l'activation du script clinique qui lui est lié et donc à un ensemble de signes et de symptômes à rechercher, de facteurs déclenchants, de moyens d'investigation, de traitements associés (77). Cette organisation des connaissances permet de porter en permanence des micro-jugements sur la cohérence des informations cliniques recueillies avec le script clinique (77).

Ainsi, le script « pneumonie bactérienne » évoque au médecin généraliste entre autres la présence de fièvre, de difficultés respiratoires, la visualisation d'un foyer pulmonaire à la radiographie thoracique et la nécessité d'un traitement antibiotique. Constaté la présence de fièvre dans le contexte est jugé comme renforçant la probabilité de l'hypothèse « pneumonie bactérienne ». A l'inverse, une radiographie thoracique normale est jugée comme diminuant la probabilité de cette hypothèse. Chaque donnée supplémentaire est analysée à la lumière des scripts cliniques activés pour tester les hypothèses générées.

Notre compréhension actuelle du développement du raisonnement clinique est basée sur l'acquisition de prototypes et de scripts cliniques performants (73). En début de cursus, confronté à la résolution d'un problème clinique, l'étudiant novice utilise majoritairement un raisonnement causal pour déterminer ce qui explique un signe ou un symptôme clinique (73). Il adosse ainsi de manière analytique ce signe ou symptôme à ses connaissances théoriques antérieures. Ces rencontres permettent aux étudiants d'acquérir des connaissances cliniques. Progressivement, le lien entre connaissances théoriques et cliniques se fait de plus en plus rapidement et les connaissances théoriques s'encapsulent. Les prototypes et les scripts se créent puis s'affinent en gagnant en efficacité.

Au-delà du professionnel de santé, les théories et modèles du raisonnement clinique tendent à évoluer vers une perspective sociocognitive de la question. Le raisonnement clinique est ainsi "situé" : l'accent est mis sur les interactions complexes entre le professionnel de santé, le patient, l'entourage, l'environnement dans un contexte singulier (43,45,83,84). Des modèles plus complexes, sortant le processus de raisonnement clinique de la "tête du professionnel de santé", permettent de mieux aborder

son enseignement et la recherche sur le sujet. Je développerai cette perspective sur le raisonnement clinique plus loin dans le tapuscrit en prenant pour exemple le modèle de Montréal (45).

## ÉVALUER LE RAISONNEMENT CLINIQUE

Plusieurs capacités fondamentales peuvent être mises en avant pour décrire la qualité du raisonnement clinique : la capacité à bien se représenter la situation clinique en tous moments de la rencontre clinique (notamment précocement), la capacité à réaliser une transformation sémantique et la capacité à confronter les hypothèses (16).

La capacité à se forger précocement une représentation mentale est essentielle dans la performance diagnostique (85). Un indicateur est la possibilité pour l'étudiant de résumer en une phrase la problématique du patient. Cette demande de synthèse peut avoir lieu pendant une rencontre clinique en interrogeant l'étudiant ou a posteriori dans le cadre d'un entretien d'explicitation (16). Une telle démarche permet également de vérifier à quel moment les hypothèses de l'étudiant sont apparues. Cette capacité dépend d'une association pertinente d'un nombre restreint d'informations et donc de la capacité à se forger précocement une représentation mentale de la situation clinique.

La capacité à réaliser une transformation sémantique renvoie à l'action de traduire en termes médicaux les données recueillies en les agençant selon des axes sémantiques (le sens donné aux signes ou aux symptômes) (16,86). Cette transformation sémantique permet de donner du sens aux informations recueillies pour ainsi activer les prototypes et scripts liés (86). Un indicateur de cette capacité consiste à analyser le vocabulaire utilisé par un étudiant devant rapporter un cas clinique ou expliciter sa représentation mentale de la situation. Georges Bordage distingue ainsi plusieurs types de discours selon la richesse sémantique et syntaxique (l'étendue du discours) (86). L'expert aura tendance à utiliser un discours riche sémantiquement mais réduit syntaxiquement. C'est ce que l'on qualifie de discours « compilé » ou « encapsulé » (86). L'évaluation du processus de raisonnement clinique peut ainsi s'appuyer sur l'analyse d'un discours rapportant la représentation mentale d'une situation clinique.

Enfin, la capacité à comparer les différentes hypothèses générées lors d'une rencontre clinique témoigne d'un raisonnement clinique efficace (87). Un indicateur possible de cette capacité est de demander à l'étudiant de hiérarchiser ses différentes hypothèses en argumentant ce classement. Par la

suite, analyser l'impact de nouvelles données sur ce classement permet d'étayer cette capacité au travers des micro-jugements permanents que le clinicien porte en se basant sur ses scripts cliniques.

Ces trois capacités sont étroitement liées. Une représentation mentale précoce performante utilise une sémantique adaptée. Cette sémantique permet l'activation efficace des scripts et prototypes. Le résultat du processus de comparaison des hypothèses est une représentation mentale de qualité.

## DES OUTILS ET DES MÉTHODES POUR ÉVALUER LE PROCESSUS DE RAISONNEMENT CLINIQUE

De nombreux outils et méthodes sont décrits dans la littérature pour évaluer le processus de raisonnement clinique (44). Je distingue ici les méthodes qui seraient selon le Larousse une "*manière de mener une activité*", des outils qui seraient un "*élément d'une activité*".

Daniel *et al.* ont publié en 2019 dans la revue *Academic Medicine* une revue de la littérature des différents outils et méthodes d'évaluation du raisonnement clinique (44). Décrire en détail les différents outils présentés dans ce travail est au-delà de l'objectif de cette thèse. Je vais simplement en faire l'inventaire afin de mieux situer les travaux présentés par la suite dans le panorama scientifique de l'évaluation du processus de raisonnement clinique.

Ils ont distingué les outils et méthodes évaluant le raisonnement clinique en milieu simulé, en stage (à partir du milieu professionnel réel) et "en salle de classe" (44).

Concernant les outils "en salle de classe", les auteurs listaient les questions à choix, les *extended matching questions*, les questions rédactionnelles, les questions rédactionnelles séquentielles, les *patients management problems*, les problèmes à éléments clés, le test de concordance de script, les *clinical integrative puzzles*, les cartes conceptuelles et les examens oraux (44).

Concernant le milieu simulé, les auteurs mentionnaient les ECOS et la méthode de la simulation "améliorée par la technologie" (*technology-enhanced simulation*) (44).

Enfin, les auteurs présentaient en milieu de stage la méthode de l'observation directe avec l'emploi d'une variété d'outils (ex : le mini-CEX), les évaluations globales - le jugement d'un superviseur sur la qualité du raisonnement clinique d'un étudiant à l'échelle d'une période de stage, les présentations orales de cas, les rapports écrits (d'une situation clinique vécue par l'étudiant), les discussions aidées par le dossier (*chart stimulated recall*), les protocoles de pensée à voix haute et les microanalyses des apprentissages auto-régulés (*self-regulated learning microanalysis*) (44).

Les auteurs mettent en avant un message clair : aucun outil, aucune méthode ne peut à elle seule correctement évaluer le processus de raisonnement clinique (44). Il faut combiner une variété d'outils et de méthodes pour évaluer correctement le raisonnement clinique (44).

#### UN MODÈLE UNIVERSEL DU RAISONNEMENT CLINIQUE : LE MODÈLE MOT

Afin de favoriser le développement du raisonnement clinique, il semble utile de baser les formations sur un modèle théorique. Parmi les modèles décrits dans la littérature, l'un a particulièrement retenu mon attention lors de mes recherches : le modèle de Montréal ou modèle MOT (45).

À partir d'un processus de recherche action participative, l'équipe de Bernard Charlin a proposé en 2012 une représentation graphique de la complexité du processus de raisonnement clinique (45). Ils ont pour cela utilisé un langage de représentation des connaissances : la modélisation par objets typés (MOT). La MOT est un langage de représentation graphique et semi-formel de la connaissance, développé par le laboratoire en informatique cognitive et environnement de formation (LICEF) à Montréal au Canada. La MOT assure une typologie des objets du savoir (les connaissances) et des liens entre eux (88). Comme tout langage, la MOT il comprend un alphabet, une grammaire et une sémantique.

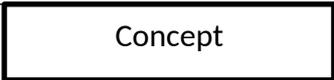
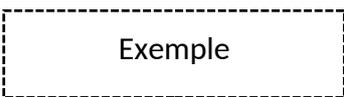
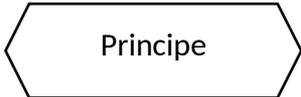
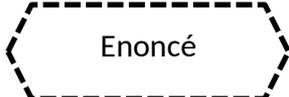
Dans la MOT, la connaissance est divisée en trois types : déclarative, action et stratégie. Les connaissances déclaratives sont ainsi le "quoi" des choses, les connaissances d'action le "comment" et les connaissances stratégiques le "pourquoi". Un parallèle peut être fait ici avec la taxonomie des connaissances de Tardif (89). Cette taxonomie identifie trois registres de la connaissance : les connaissances déclaratives (le quoi), les connaissances procédurales (le comment) et les connaissances conditionnelles (le pourquoi mais aussi le "quand") (89).

Les connaissances déclaratives sont un ensemble de faits et de règles qu'un individu est capable d'énoncer. Les connaissances procédurales sont des connaissances d'action qui permettent à un individu d'effectuer une tâche ou un geste en respectant les différentes étapes. Enfin, les connaissances conditionnelles renvoient à la capacité d'un individu d'identifier les situations où il doit effectuer une tâche et à justifier ses décisions (89,90). J'utiliserai plus loin dans cette synthèse cette taxonomie des connaissances de la MOT pour discuter les résultats de certains des articles présentés.

Appliqué à l'alphabet de la MOT, les connaissances déclaratives sont représentées par des rectangles, les connaissances d'action par des ovales et les connaissances stratégiques par des hexagones. A l'intérieur de ces différents types de connaissances, la MOT distingue les connaissances factuelles (représentées en pointillés) des connaissances abstraites (représentées en traits pleins) :

- Une connaissance déclarative abstraite est appelée « concept » alors qu'une connaissance factuelle est appelée « exemple » ;
- Une connaissance d'action abstraite est appelée « procédure » alors qu'une connaissance factuelle est appelée « trace » ;
- Une connaissance stratégique abstraite est appelée « principe » alors qu'une connaissance factuelle est appelée « énoncé ».

Cette catégorisation de la MOT est nécessaire à ce langage pour exister. Comme toute catégorisation, elle peut cependant être critiquée au vu du côté parfois artificiel des dichotomies qu'elle impose. Le tableau n°1 récapitule les différentes représentations des connaissances selon la technique de modélisation des objets typés.

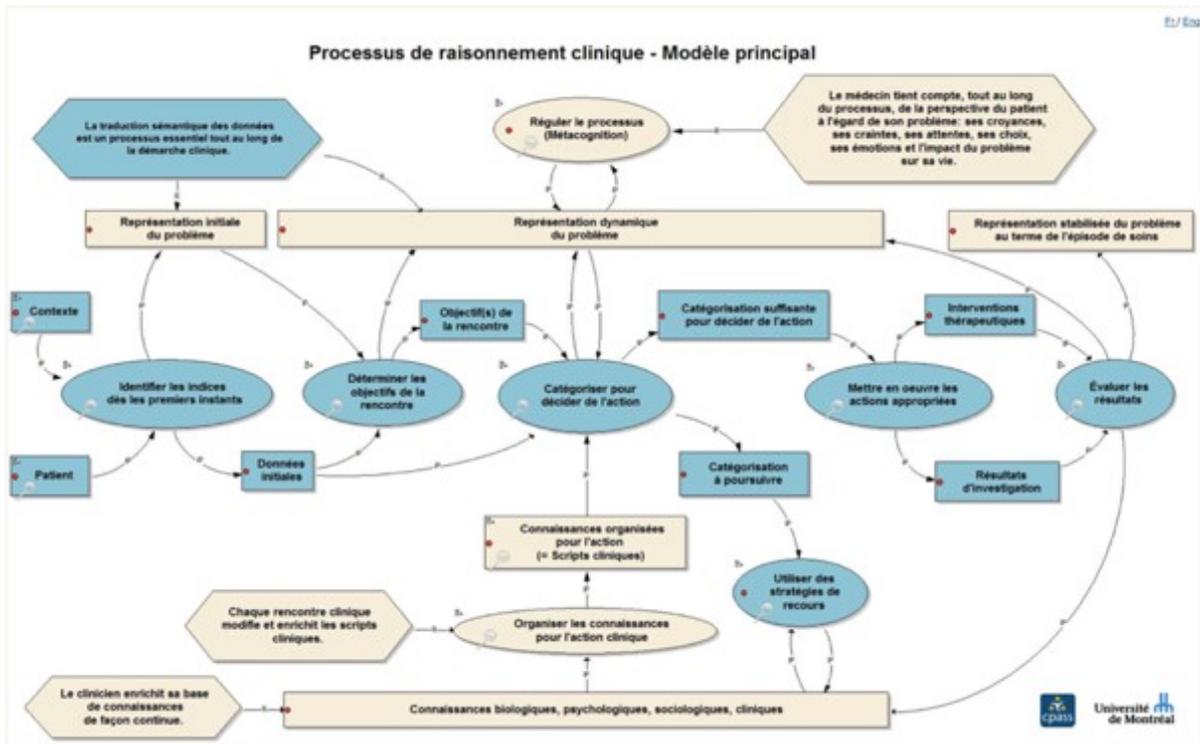
	Connaissances abstraites	Connaissances factuelles
Connaissance déclarative	 Concept	 Exemple
Connaissance d'action	 Procédure	 Trace
Connaissance stratégique	 Principe	 Enoncé

**TABLEAU 1 : REPRÉSENTATIONS DES CONNAISSANCES SELON LA MODÉLISATION PAR OBJETS TYPÉS**

Il existe 7 types de liens entre les connaissances :

- Le lien de composition (C). Il sert à représenter les composants d'une connaissance. Par exemple : une voiture est composée d'une carrosserie et de roues ;
- Le lien de spécialisation (S). Il sert à représenter la spécialisation d'une connaissance de même type l'une par rapport à l'autre. Par exemple : un labrador est une sorte de chien ;
- Le lien de régulation (R). Il sert à représenter la relation de régulation entre un principe et une connaissance. Par exemple : la crise économique commande une réduction des dépenses ;
- Le lien d'instanciation (I). Il sert à représenter le lien entre une connaissance factuelle et une connaissance abstraite du même type. Par exemple : si la température de votre enfant dépasse les 38,5°C, vous pouvez lui donner du paracétamol. La température de mon enfant est de 39,1°C, je peux lui donner du paracétamol ;
- Le lien intrant/produit (IP). Il sert à représenter le lien entre une procédure et un concept pour désigner les composants nécessaires à la réalisation de la procédure ainsi que les connaissances produites. Par exemple : il faut de l'essence pour faire rouler une automobile (intrant), faire rouler une automobile relâche des gaz à effet de serre (produit) ;
- Le lien de précédence (P). Il sert à représenter la séquence d'exécution des procédures ou l'ordonnancement de l'application de principes. Par exemple : je mets ma carte de crédit dans un distributeur de billet, je compose mon code, je retire de l'argent ;
- Le lien d'attribut (A). Il sert à représenter le fait qu'un concept est un attribut d'un autre, une caractéristique. Exemple : la couleur est un attribut d'une voiture.

La figure 2 présente le modèle principal du processus de raisonnement clinique élaboré selon les règles de la modélisation par objets typés (source : <https://edugenie.com/prc2015/>).



**FIGURE 3: MODÈLE PRINCIPAL DU PROCESSUS DE RAISONNEMENT CLINIQUE SELON LA MOT**

Ce modèle principal rend compte de la complexité du raisonnement clinique dans sa globalité avec ses multiples boucles de rétrocontrôle / rétroaction. Nous pouvons constater que cette représentation du raisonnement clinique en tant que processus abstrait ne représente que des connaissances abstraites (représentées par les formes aux contours pleins). Au niveau du code couleur, les items bleus représentent la partie clinique du processus de raisonnement clinique tandis que les items beiges représentent la partie gestion des connaissances. Nous allons détailler les principales procédures à l'œuvre dans le modèle MOT ; c'est-à-dire les catégories d'opérations ou d'actions pouvant être accomplies durant le processus du raisonnement clinique. Ces procédures sont au nombre de cinq :

1. Identifier les indices dès les premiers instants
2. Déterminer les objectifs de la rencontre
3. Catégoriser pour décider de l'action
4. Mettre en œuvre les actions appropriées
5. Évaluer les résultats

Lors d'une rencontre clinique, la première procédure du processus de raisonnement clinique est nommée « identifier les indices dès les premiers instants ». Cette procédure utilise les concepts de « patient » et de « contexte » pour produire les concepts de « représentation initiale du problème » et « données initiales ». La figure 3 illustre cette procédure (source : <https://edugenie.com/prc2015/>).

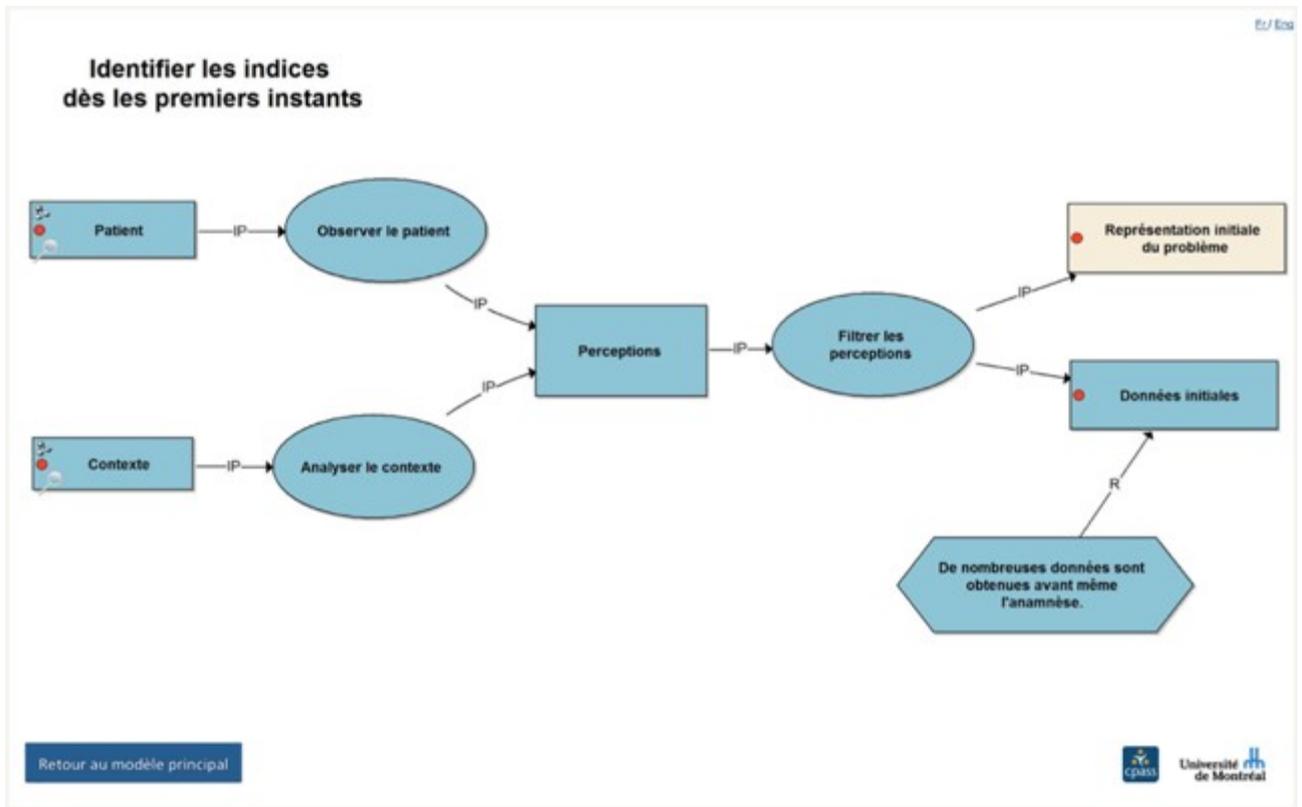


FIGURE 4: DÉTAIL DE LA PROCÉDURE "IDENTIFIER LES INDICES DÈS LES PREMIERS INSTANTS"

Dans cette procédure, l'élément central est le concept de « perceptions ». Ce concept se réfère aux connaissances abstraites déclaratives produites par les procédures d'observation du patient et d'analyse du contexte de la rencontre clinique. Ces perceptions sont ensuite filtrées pour aboutir aux données initiales et à la représentation initiale du problème.

Le principe « de nombreuses données sont obtenues avant l'anamnèse » illustre le fait que certaines données permettant au clinicien de débiter son processus de raisonnement clinique lui sont accessibles avant même d'avoir commencé à interroger son patient. C'est là un point important de ce modèle : les professionnels de santé commencent à raisonner dès qu'ils obtiennent des informations sur un patient.

Cette procédure souligne que les données visuelles et contextuelles sont à la base des données initiales qui permettent au professionnel de santé de se forger une première représentation du problème

de santé présenté par un patient. Ce constat implique d'accorder une place prépondérante au contexte et à l'observation du patient dans tout cas clinique.

Dans le modèle MOT, le concept de contexte est composé de l'histoire médicale du patient, du contexte épidémiologique, des compétences et limites du clinicien mais aussi de l'environnement clinique. Cet environnement est lui-même composé du lieu, des caractéristiques de la rencontre médicale et des conditions de cette rencontre. La figure 4 illustre les composants du contexte et leurs éventuelles spécialisations (source : <https://edugenie.com/prc2015/>).

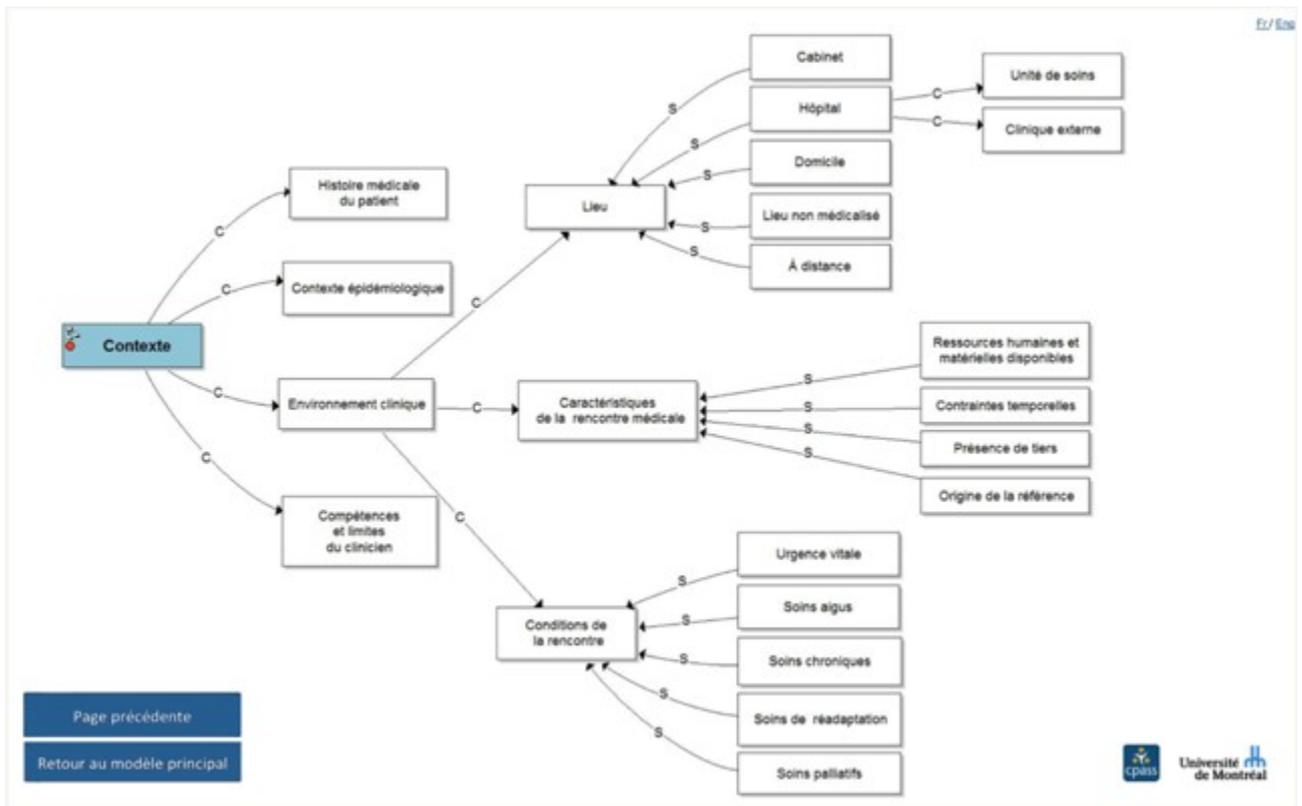


FIGURE 5 : COMPOSANTS ET SPÉCIALISATIONS DU CONCEPT DE "CONTEXTE"

Le concept de patient a quant à lui les attributs d'aspect physique, d'aspect psychologique et d'aspect social. L'aspect physique est composé notamment de l'état de santé apparent du patient et de sa morphologie. L'aspect psychologique se recoupe avec l'aspect physique pour composer l'expression faciale et les signes de détresse. L'aspect social se combine avec l'aspect psychologique pour composer l'attitude générale. Enfin, ces trois aspects composent le concept d'habillement et niveau d'hygiène. Aussi on constate aisément que ces données utilisées pour se forger une représentation du problème de santé du patient sont bien difficiles à retranscrire dans un texte. La figure 5 illustre tous les attributs et composants du concept « patient » (source : <https://edugenie.com/prc2015/>).

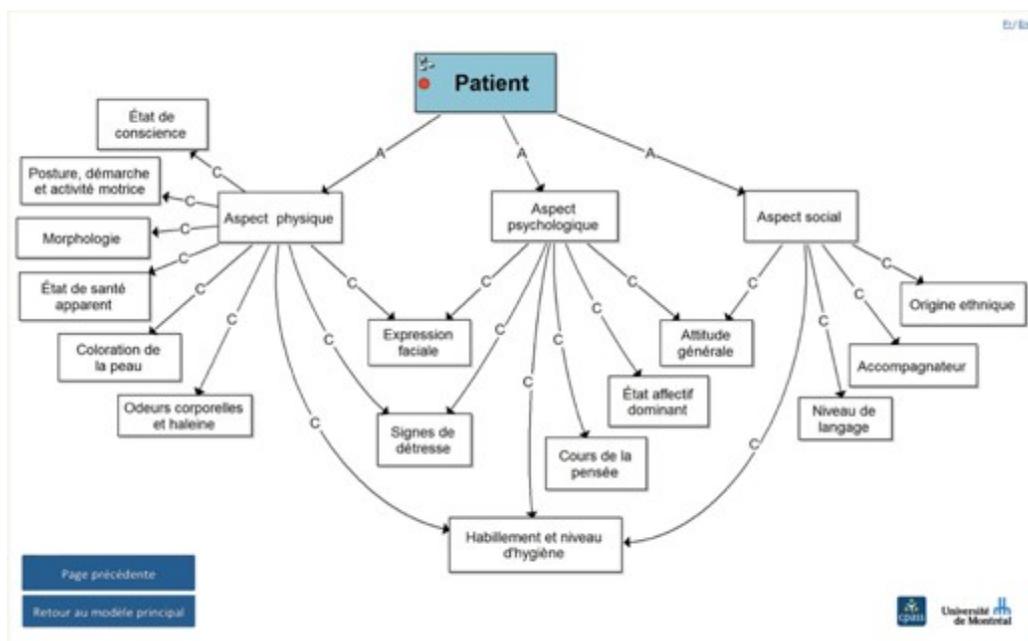


FIGURE 6 : ATTRIBUTS ET COMPOSANTS DU CONCEPT DE "PATIENT"

À partir des données initiales et de la représentation initiale du problème, la procédure « déterminer les objectifs de la rencontre » est une sorte de « deuxième étape » du processus de raisonnement clinique. Cette procédure utilise les concepts de « données initiales » et de « représentation initiale du problème » pour aboutir aux concepts de « représentation dynamique du problème » et des « objectifs de la rencontre ».

À regarder le détail de cette procédure, ces concepts sont produits via la procédure « établir les priorités ». Cette dernière procédure est définie par le principe suivant : « pour établir les priorités, on doit tenir compte de l'urgence, de la pertinence et de l'efficacité ». Plusieurs procédures sont des intrants possibles du concept « établir les priorités ». La demande du patient s'associe à la représentation initiale du problème au sein de la procédure « clarifier la demande du patient » pour produire le concept « besoins exprimés par le patient ». S'il y a une demande d'un médecin référent ou d'un professionnel, la procédure « prendre connaissance de la demande » aboutit aux « besoins exprimés par le médecin référent ou un professionnel ». S'il y a une « information provenant d'un tiers », la procédure « prendre connaissance de l'information » produit les « renseignements fournis par des tiers ». Enfin, les « données initiales » passent au travers de la procédure « identifier les besoins du patient » pour produire les « besoins du patient tels que perçus par le médecin ».

Tous ces concepts vont ainsi participer à construire la « représentation dynamique du problème ». Cette représentation est donc le fruit d'un raisonnement complexe et multifactoriel dont les différentes procédures constitutives sont difficiles à explorer par écrit. La figure 4 illustre ces connaissances et les liens entre elles (source : <https://edugenie.com/prc2015/>).

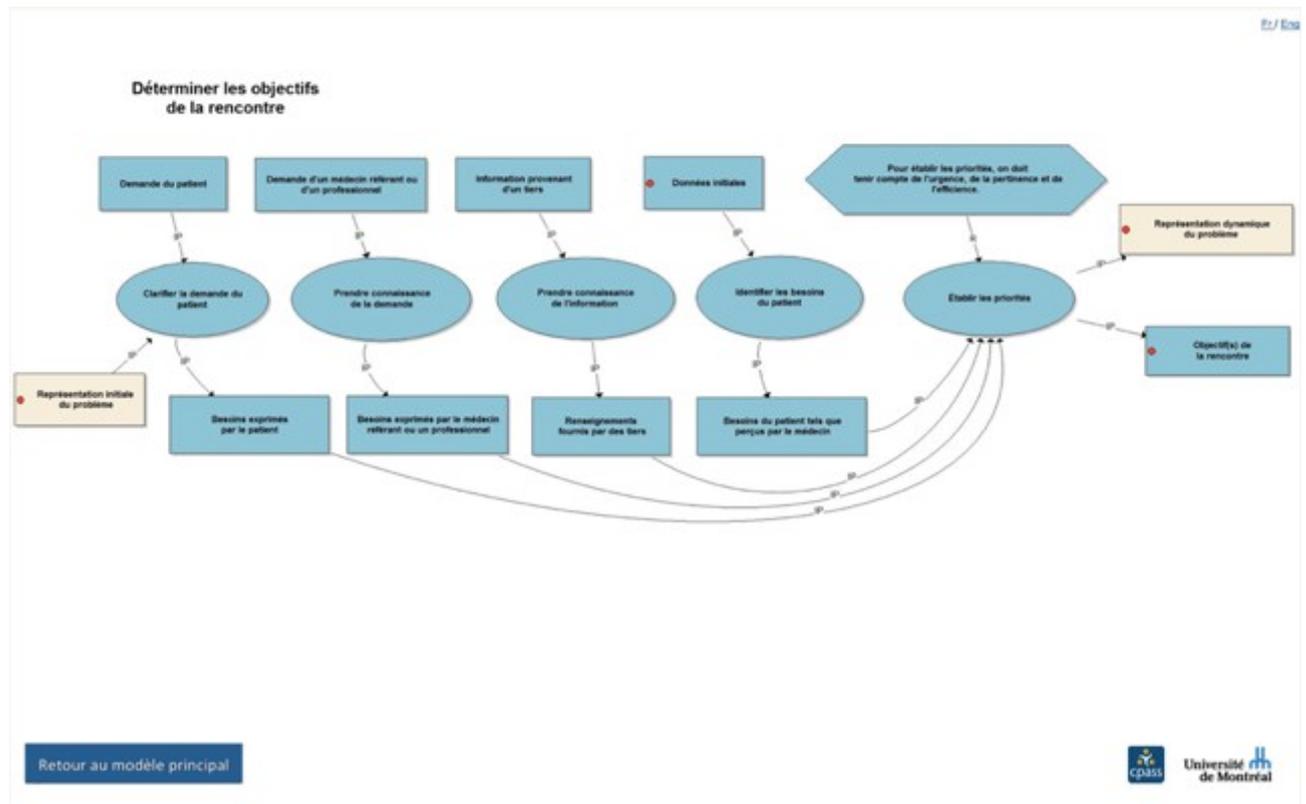


FIGURE 7 : SOUS-PROCÉDURES ET INTRANTS DU CONCEPT DE "DÉTERMINER LES OBJECTIFS DE LA RENCONTRE"

La procédure « catégoriser pour décider de l'action » est l'élément central du processus de raisonnement clinique. Catégoriser renvoie du point de vue du raisonnement clinique au fait d'avoir une représentation mentale du problème suffisamment fine pour proposer au patient un ensemble d'actions appropriées. La catégorisation diffère d'un diagnostic puisqu'elle n'est pas forcément une entité précise. Cette catégorisation pourrait être par exemple « insuffisance respiratoire aiguë » qui appellerait des actions comme « débuter une oxygénothérapie » ou « prise en charge urgente ». Le corollaire à cette catégorisation est que l'action est déclenchée parfois sans attendre une génération d'hypothèses précises. Les hypothèses diagnostiques doivent à ce stade encore être précisées pour affiner la prise en charge.

De nombreux intrants vont déterminer cette catégorisation : les données initiales, les objectifs de la rencontre, la représentation dynamique et les scripts cliniques. La catégorisation produira en retour soit le concept « catégorisation suffisante pour décider de l'action », soit le concept catégorisation à poursuivre. Une catégorisation suffisante débouchera sur la procédure « mettre en œuvre les actions appropriées ». Une catégorisation à poursuivre conduira à la procédure « utiliser des stratégies de recours ».

Il est à noter que la procédure de catégorisation donne lieu à une sorte de rétrocontrôle sur la représentation dynamique du problème puisque ces connaissances sont à la fois des intrants et des produits l'une de l'autre. Je vais expliciter les sous-procédures qui la composent.

La première « étape » est la procédure « identifier les données pertinentes » qui se base sur les données initiales, les objectifs de la rencontre, la représentation dynamique du problème et les scripts cliniques. Cette procédure va produire le concept « données cliniques retenues initialement ».

Ces données cliniques retenues initialement sont l'intrant de la procédure « reconnaître un pattern ou un cas analogue ». Cette procédure va produire les concepts soit « hypothèses initialement reconnues » si cette reconnaissance est positive, soit « aucun pattern ou cas analogue identifié » dans le cas contraire.

Dans le cas où les hypothèses sont retenues, la procédure « s'assurer qu'il n'existe pas d'autres hypothèses explicatives » est activée. Son produit est le concept « ensemble d'hypothèses avec leurs scripts activés ». L'étape suivante est la procédure « rechercher de façon orientée des données pour infirmer ou confirmer les hypothèses », qui va produire le concept « données cliniques significatives ». La dernière procédure est « analyser l'adéquation entre les données et les scripts activés en vue de la catégorisation ». Cette procédure va produire les concepts résultats finaux de la catégorisation pour décider de l'action : catégorisation suffisante pour décider de l'action ou catégorisation à poursuivre. Le parallèle avec le double processus et la théorie des scripts décrits en introduction est ici évident.

Cette procédure d'analyse de l'adéquation entre les données et les scripts exerce un rétrocontrôle sur la représentation dynamique du problème du clinicien. Une adéquation forte entre le script d'une pathologie et les données recueillies va renforcer la probabilité diagnostique de cette hypothèse au sein de la représentation mentale du professionnel de santé.

Dans le cas où la reconnaissance de pattern ou de cas analogue n'aboutirait pas à reconnaître un de ces éléments, la procédure « rechercher de façon analytique des hypothèses explicatives » est activée.

Elle conduit, soit à reconnaître des hypothèses et à réintégrer le déroulement normal de la catégorisation, soit à poursuivre la catégorisation si des hypothèses ne sont pas identifiées. C'est là encore un aspect clairement décrit dans la littérature du double-processus.

Pour conclure, la procédure de catégorisation repose sur les scripts cliniques en sus des concepts de données initiales, des objectifs de la rencontre et de la représentation dynamique du problème. Explorer la procédure de catégorisation consiste indirectement à explorer les scripts cliniques du clinicien. Aussi vraisemblablement, des outils comme les tests de concordance de script explorent une sous-partie de cette procédure.

La figure 7 illustre les sous-procédures et intrants de la procédure « catégoriser pour décider de l'action » (source : <https://edugenie.com/prc2015/>).

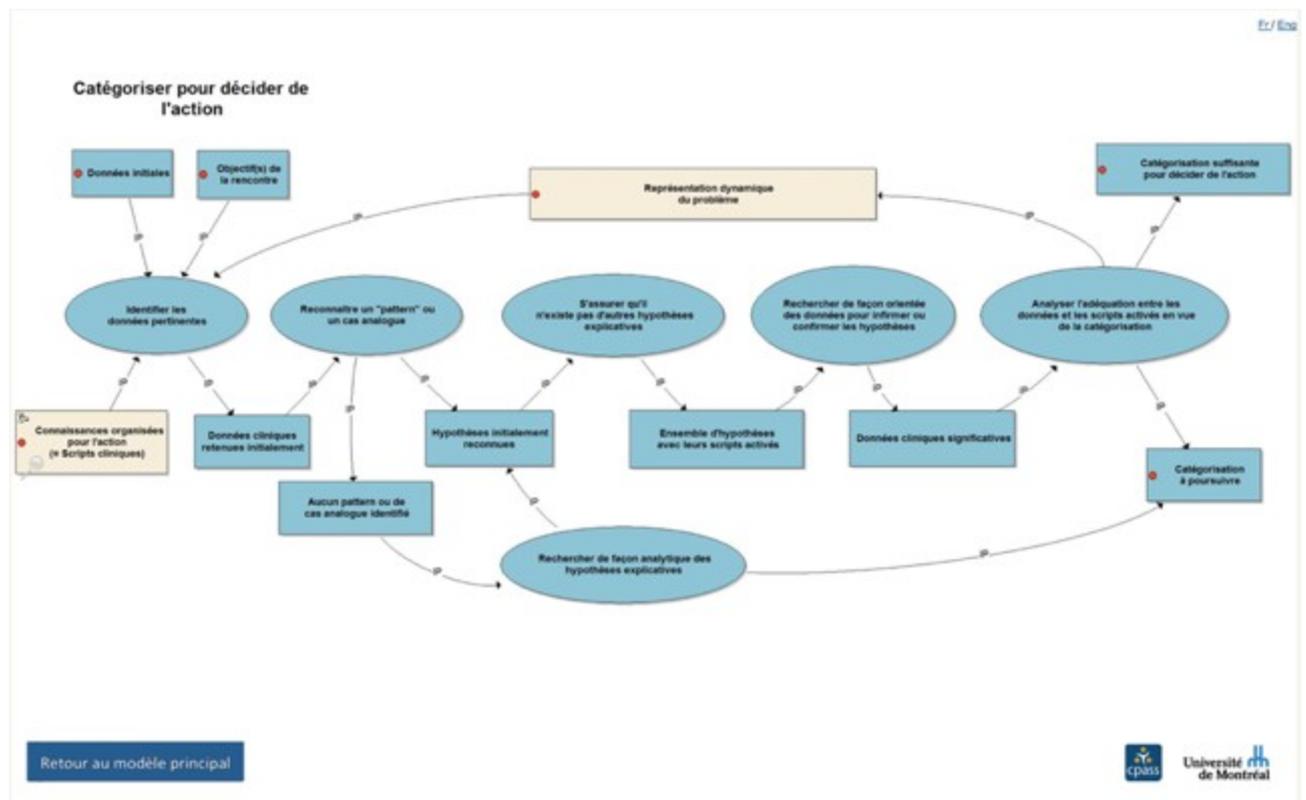


FIGURE 8 : SOUS-PROCÉDURES ET INTRANTS DE LA PROCÉDURE « CATÉGORISER POUR DÉCIDER DE L'ACTION »

En cas de « succès » de la catégorisation (la catégorisation est suffisante pour décider de l'action), le professionnel de santé va déclencher la procédure « Mettre en œuvre les actions appropriées ». Cette procédure produit deux concepts : « interventions thérapeutiques » et « résultats d'investigation ». Je vais détailler les procédures sous-jacentes à la mise en place des actions appropriées.

Nous avons vu au paragraphe précédent qu'un ou plusieurs scripts avaient été activés. La première étape ici est de « réactiver les volets investigation et intervention des scripts activés ». Cette procédure va produire le concept « options d'investigation et d'intervention activées ».

La procédure suivante est plus complexe. Il va s'agir de « rechercher de façon orientée des données qui pourraient influencer le choix des options ». Pour ce faire, le clinicien va s'appuyer également sur des concepts de « considération d'efficacité et d'efficience », de « données probantes », « singularité du patient » et « expérience du clinicien ». Le produit de cette procédure va être de sélectionner des « options pertinentes ». La procédure « vérifier l'adéquation entre les attributs des options pertinentes et l'ensemble des données » va permettre de sélectionner les « options retenues ».

Une phase qui consiste à « valider les options avec le patient » va débiter. Les investigations sont divisées en « plan d'investigation » visant à produire des résultats d'investigation au travers de la procédure « appliquer le plan d'investigation » et en plan d'intervention thérapeutique qui produira les « interventions thérapeutiques » au travers de la procédure « appliquer le plan d'intervention thérapeutique ».

Au total, cette procédure de mise en œuvre des actions appropriées va permettre au clinicien d'obtenir, si nécessaire, des résultats d'investigation et une intervention thérapeutique à partir de sa catégorisation en tenant compte d'éléments extérieurs (singularité du patient, avis du patient, données probantes, etc...) dans sa décision. Là encore, ces éléments sont bien difficiles à explorer en utilisant les traditionnels cas cliniques textuels largement répandus dans les formations des professionnels de santé. La figure 8 illustre les différentes sous-procédures et intrants de la procédure « mettre en œuvre les actions appropriées » (source : <https://edugenie.com/prc2015/>).

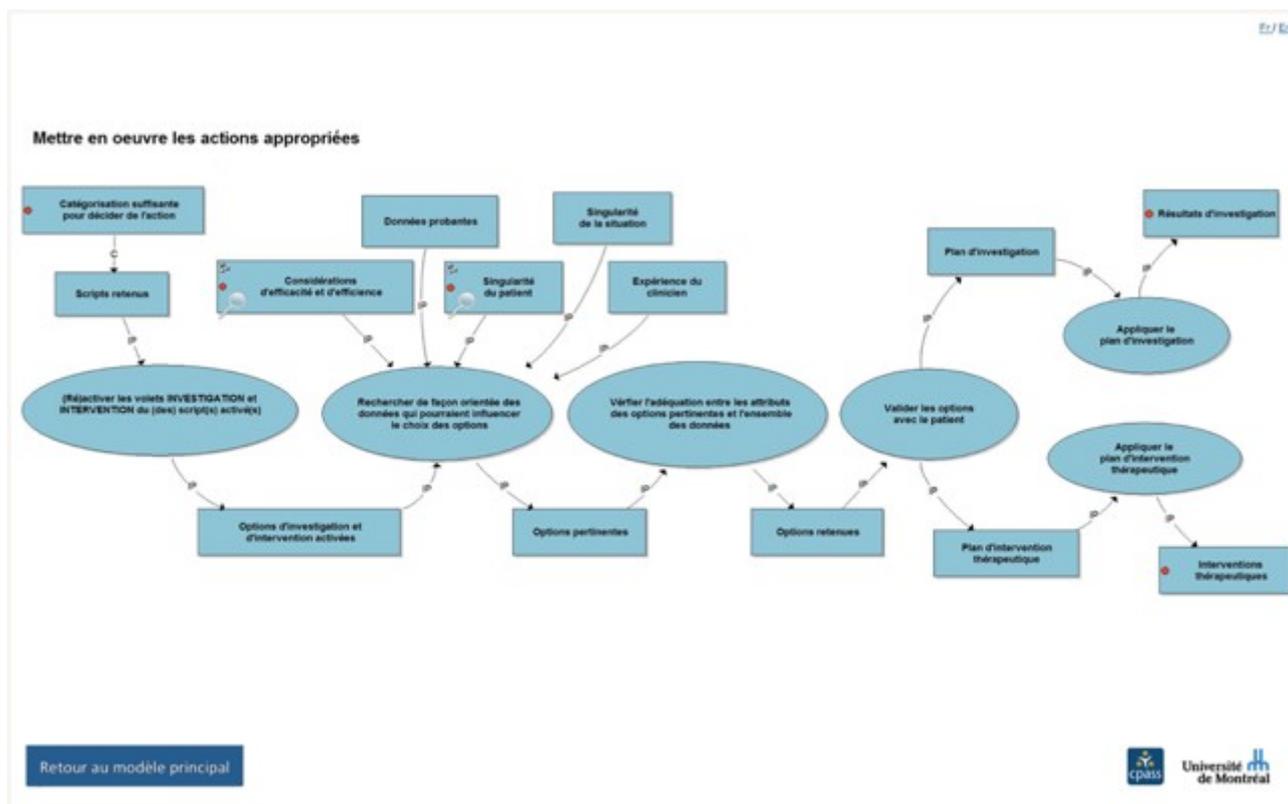


FIGURE 9 : SOUS-PROCÉDURES ET INTRANTS DE LA PROCÉDURE « METTRE EN ŒUVRE LES ACTIONS APPROPRIÉES »

L'étape finale du processus de raisonnement clinique est nommée « évaluer les résultats ». Cette procédure consiste, au travers des procédures « analyser les résultats d'investigation » et « analyser les résultats d'intervention thérapeutique », à l'obtention du concept « nouvelles données cliniques ». Ces nouvelles données cliniques alimentent la procédure « réévaluer la catégorisation » qui produit alors une nouvelle représentation dynamique du problème, de nouvelles « connaissances biologiques, psychologiques, sociologiques, cliniques » et la « représentation stabilisée du problème au terme de l'épisode de soin ». Il s'agit à la fois de « clore » la rencontre clinique et d'utiliser les données de cet épisode de soin pour réorganiser ses connaissances cliniques (scripts, prototypes) dans un processus d'autorégulation.

Si le produit de cette procédure est une nouvelle représentation dynamique du problème car les résultats des investigations ou interventions ne permettent pas de « clore » l'épisode de soin, la catégorisation doit alors reprendre. Il s'agit dans ce cas d'un « retour à l'étape précédente » pour proposer une nouvelle prise en charge. La figure 9 illustre les sous-procédures et intrants de la procédure « évaluer les résultats » (source : <https://edugenie.com/prc2015/>).

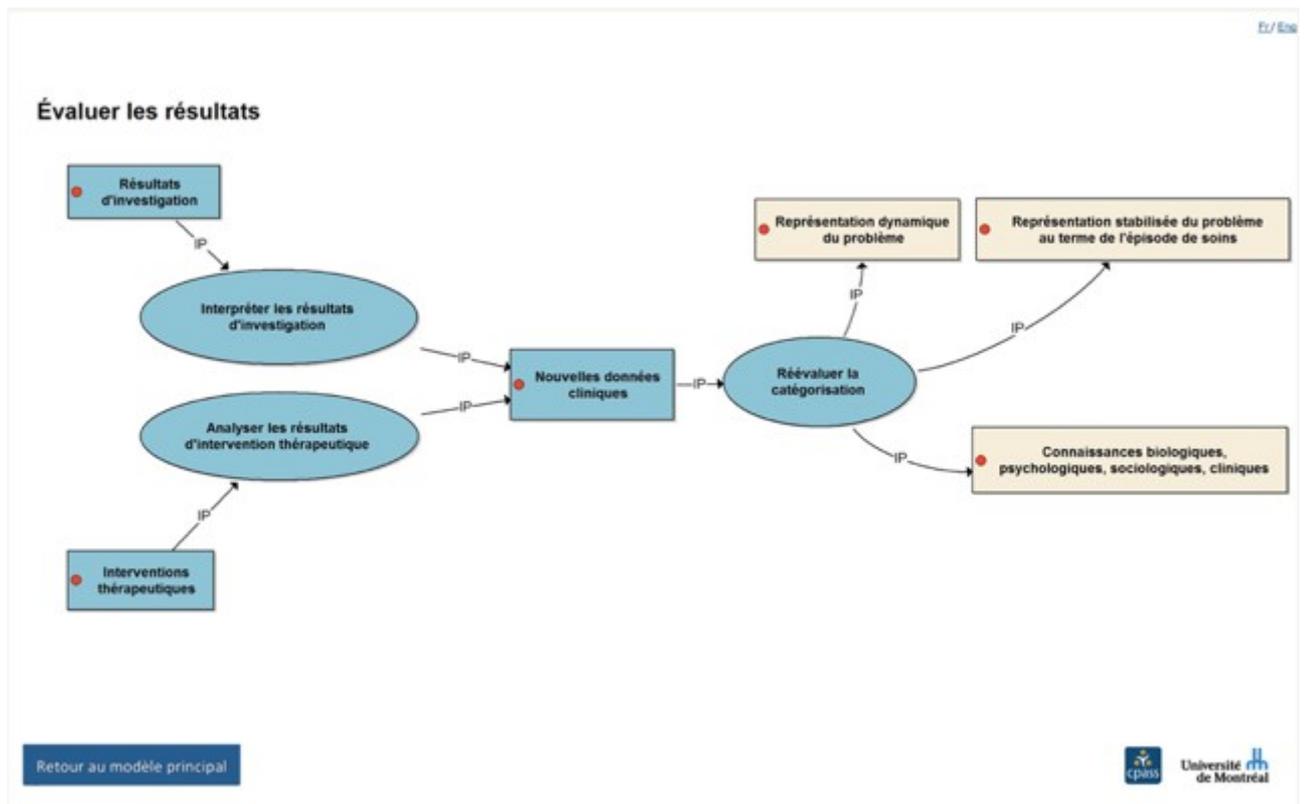


FIGURE 10: SOUS-PROCÉDURES ET INTRANTS DE LA PROCÉDURE « ÉVALUER LES RÉSULTATS »

Dans le cas où la catégorisation serait à poursuivre, le clinicien peut « utiliser des stratégies de recours ». Cette procédure mobilise les connaissances biologiques, psychologiques, sociologiques, cliniques dans une boucle d'interactions mutuelles puisque l'un est à la fois l'intrant et le produit de l'autre. En d'autres termes, le clinicien va chercher via une variété de ressources internes et externes à proposer une catégorisation suffisante. Dans le processus, il va réorganiser ses connaissances spécifiques.

Si on détaille les spécialisations de l'utilisation des stratégies de recours, plusieurs procédures sont possibles : « mobiliser des ressources externes », « observer l'évolution clinique sans faire d'intervention thérapeutique », « faire un essai thérapeutique » ou « mobiliser ses ressources internes ». Cette dernière procédure comporte plusieurs spécialisations : « analyser selon une perspective étiologique », « analyser selon une perspective physiopathologique », « analyser selon une perspective anatomique » ou « analyser selon une perspective psychosociale ». La figure 8 illustre les sous-procédures, intrants et spécialisations de la procédure « utiliser des stratégies de recours » (source : <https://edugenie.com/prc2015/>).

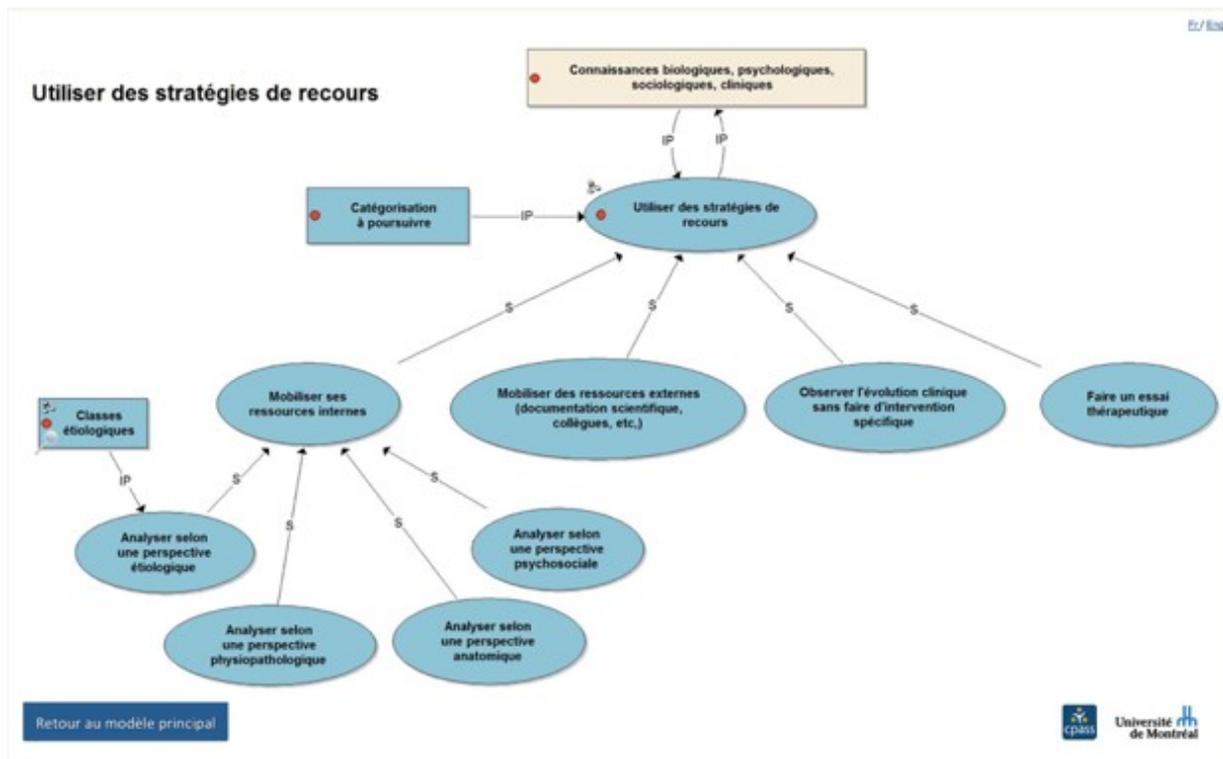


FIGURE 11 : SOUS-PROCÉDURES, INTRANTS ET SPÉCIALISATIONS DE LA PROCÉDURE « UTILISER DES STRATÉGIES DE RECOURS »

Après avoir détaillé les différentes étapes du modèle MOT, nous pouvons dégager une sorte de « chronologie globale » ainsi que les éléments principaux utilisés par les professionnels de santé pour raisonner.

Les informations contextuelles et l’observation du patient vont permettre au professionnel de santé de se faire une idée précoce du problème présenté par le patient et d’extraire les données initiales. Une représentation dynamique du problème plus complexe va suivre selon l’analyse de ces données initiales et la détermination des objectifs de la rencontre. Le professionnel de santé va ensuite évaluer s’il peut débiter l’action si sa catégorisation est suffisante. Cela reposera sur l’adéquation des données avec les scripts cliniques du professionnel de santé. Les actions appropriées selon les scripts retenus seront effectuées après négociation avec le patient et leurs résultats évalués afin d’obtenir une représentation stabilisée du problème. Ce processus réglera en continu les connaissances spécifiques du professionnel de santé dans une démarche réflexive.

Le modèle MOT reste cependant une approximation de la réalité du processus de raisonnement clinique. Il est probable que la relative linéarité présentée dans le modèle MOT soit soumise à plus d’aléas dans la réalité. Ce modèle semble néanmoins poser des bases intéressantes pour proposer des cas

cliniques qui, en explorant ces différentes "étapes" du processus de raisonnement clinique, pourraient optimiser son développement.

## DES OUTILS PÉDAGOGIQUES VISANT À DÉVELOPPER LE RAISONNEMENT CLINIQUE ET À ILLUSTRER L'INCERTITUDE DANS LES SOINS : LES FORMATIONS PAR CONCORDANCE

Mes lectures sur le raisonnement clinique m'ont mené à m'intéresser aux outils et méthodes de son évaluation. J'emploierai pour la suite de ma thèse le terme de méthode pour me référer à une manière de mener l'évaluation en tant qu'activité, tandis que le terme d'outil fera référence à un élément de l'activité d'évaluation. Ainsi, je considérerai la simulation comme une méthode et les ECOS comme un outil d'évaluation.

En 2019, Daniel *et al.* (44) ont identifié dans la revue de la littérature pour chaque méthode et outil d'évaluation du raisonnement clinique quelles en étaient les "composantes" et détaillé lesquelles étaient correctement évaluées. L'idée principale de ce travail est qu'aucune méthode ou outil seul ne permet d'évaluer correctement l'ensemble du processus de raisonnement clinique (44). La seule manière de "bien" évaluer le raisonnement clinique est d'utiliser différentes méthodes et outils (44). Dans cette perspective, la réforme du deuxième cycle des études médicales propose d'utiliser une combinaison de divers outils d'évaluation des apprentissages et notamment des problèmes à éléments clés. Cet outil étant relativement méconnu en France, j'ai écrit un article didactique sur le sujet (91) pour faciliter son appropriation par la communauté universitaire (voir annexe n°2). L'article a été publié dans la revue *exercer* octobre 2021 (91).

Mes travaux et lectures de ces dernières années ont principalement été consacrés à la question de l'évaluation des apprentissages et particulièrement l'évaluation du raisonnement clinique. Le test de concordance de script (TCS) fut ma porte d'entrée dans ce domaine.

### LE TEST DE CONCORDANCE DE SCRIPT (TCS)

Le TCS est un outil d'évaluation des apprentissages né il y a une vingtaine d'années au Canada (92). Une particularité du TCS est qu'il existe souvent plusieurs "bonnes réponses" à une même question. Un panel de référence répondant aux mêmes questions que les étudiants étant utilisé pour créer le barème de notation, les points sont attribués selon le nombre de membre du panel ayant choisi cette option de

réponse. Cette méthode permet une chose rare en évaluation des apprentissages en sciences de la santé : refléter l'incertitude.

Comme je l'exposerai plus bas, si le TCS est la "tête de gondole" de l'utilisation de la concordance par son ancienneté et son acceptation scientifique, c'est réellement la méthode de la concordance dans l'évaluation qui sera le cœur de mon travail.

Pour présenter le TCS à ce stade du tapuscrit, je vais largement me référer à l'article didactique que j'ai rédigé dans la revue *exercer* pour permettre à la communauté universitaire de ma discipline de s'approprier cet outil dans le cadre de la réforme du deuxième cycle des études médicales (93) (voir annexe n°3).

Le TCS est basé sur le modèle d'organisation des connaissances des scripts cliniques (46). Dans la théorie des scripts cliniques, l'évocation d'une hypothèse diagnostique par le professionnel de santé active un réseau de connaissances. Le professionnel de santé va ensuite interpréter l'ensemble des données cliniques du patient en regard de ces scripts : les données "collent-elles" avec le script activé ? Si oui, l'hypothèse en sort renforcée. Dans le cas contraire, il est probablement nécessaire de considérer une autre hypothèse.

Voici un exemple issu de la pratique de la médecine générale : devant une jeune patiente de 22 ans présentant une douleur abdominale prédominant en fosse iliaque droite, le médecin évoque notamment l'hypothèse de l'appendicite. L'évocation de cette hypothèse active le script "appendicite". Celui-ci contient les éléments attendus à l'interrogatoire devant une appendicite, les signes cliniques à rechercher à l'examen, les résultats attendus des examens complémentaires, la prise en charge à proposer, etc... Le médecin demande à la patiente s'il elle a eu de la fièvre. La patiente répond qu'elle a eu 38,7°C cette nuit avec des frissons. Cette donnée "colle" avec le script "appendicite" : il est fréquent qu'une fièvre accompagne une appendicite aiguë. A l'examen clinique, le médecin s'attarde particulièrement à palper la fosse iliaque droite. L'absence de défense ou de contracture qu'il constate "ne colle pas" avec son script "appendicite". Il s'attendait à trouver de tels éléments.

Le principe du test de concordance de script est celui-ci : dans une vignette clinique, on demande à l'étudiant de juger de l'impact d'une donnée sur le script d'une hypothèse ou d'une option (une "option d'investigation" ou une "option thérapeutique"). L'étudiant répond via une échelle de Likert à cinq

modalités : ne change rien, impact positif, impact fortement positif, impact négatif, impact fortement négatif. Le tableau n°2 présente un exemple de questions d'un TCS.

Mme Hillinger est une étudiante de 22 ans dans le domaine du textile. Elle se présente à votre cabinet pour des douleurs abdominales prédominant en fosse iliaque droite apparues cette nuit. Elle n'a pas d'antécédents particuliers et ne prend pas de contraception.			
N°	Si vous pensiez à ...	... et qu'alors vous trouvez ...	... l'impact sur votre hypothèse ou option est ...
Q1	Une appendicite	Une température à 38,7°C	-2   -1   0   +1   +2 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Q2	Une grossesse extra-utérine	Que la patiente ne se souvient plus de la date de ses dernières règles	-2   -1   0   +1   +2 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Q3	Proposer de réaliser une échographie abdominale	Que la patiente est enceinte	-2   -1   0   +1   +2 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

-2 = Fortement négatif, - 1 = Négatif, 0 = Ne change rien, +1 = Positif, + 2 = Fortement positif

**TABLEAU 2 : EXEMPLE DE QUESTIONS DE TCS**

Cet exemple permet d'illustrer plusieurs des difficultés fréquentes rencontrées par les étudiants avec les TCS. En premier lieu, les hypothèses ou options ("si vous pensiez à...") sont imposées par le format. Il n'est donc pas demandé à l'étudiant de discuter le caractère pertinent ou non de celles-ci au regard de la vignette clinique. Si l'hypothèse ou option ne semble pas pertinente pour l'étudiant dans cette situation clinique, il lui sera très difficile de répondre à la question.

Deuxièmement, les questions sont indépendantes. Ainsi, vous ne connaissez pas la température de la patiente dans la question n°2. Ceci soulève la problématique de la charge cognitive de l'étudiant confronté à une succession de questions indépendantes sur une même vignette (94). Il peut être pour les étudiants difficile "d'oublier" les informations précédentes.

Enfin, la nuance entre les modalités "impact fortement négatif" et "impact négatif" ou entre "impact fortement positif" et "impact positif" relève d'une certaine finesse dans le raisonnement clinique. Même chez des étudiants en fin de cursus, les déterminants de ce choix sont inconstants (95). Certains

auteurs proposent alors de n'utiliser que des TCS à 3 modalités notamment en début de formation (impact négatif – ne change rien – impact positif) (96,97).

La plus grande originalité des TCS vient de leur mode de notation. Les points sont attribués selon la proximité des réponses des étudiants à celles d'un panel de référence. Si nous prenons la question n°1 de notre exemple, un panel de dix médecins généralistes a répondu. Si huit membres ont répondu "impact positif" (+1), cette réponse vaudra pour les étudiants  $8/8=1$  point. Si les deux membres restant du panel ont répondu "impact fortement positif" (+2), cette réponse vaudra  $2/8=0,25$  point. Les autres réponses vaudront  $0/8=0$  point. De nombreuses autres méthodes de notation ont été proposées dans la littérature mais celle-ci fait consensus actuellement (96–99).

Ce mode de notation a un effet extrêmement intéressant : c'est la proximité entre le raisonnement clinique de l'étudiant et celui d'un panel de référence qui détermine son score. Un étudiant qui obtient un bon score aux TCS est considéré comme ayant un raisonnement clinique proche d'un groupe expert. Il s'agit d'une manière normative de déterminer ce qui constitue la réussite en matière de raisonnement clinique. Ce qui est fait par la majorité du groupe de référence est considéré comme la "bonne réponse". C'est aussi un moyen puissant d'illustrer l'incertitude dans la pratique de la médecine : les membres d'un groupe expert ne sont pas toujours d'accord entre eux.

Le seuil de réussite d'un TCS est habituellement fixé à 60/100 (98). Ce seuil a été déterminé à partir de données statistiques et correspond à une performance considérée comme "trop éloignée" de celle du groupe de référence pour être acceptable (98). Il s'agit là aussi d'une différence notable avec le classique seuil de réussite de 10/20 encore dominant dans les modalités habituelles d'évaluation des apprentissages en médecine.

De ces considérations docimologiques découle des problèmes d'acceptabilité par les étudiants du TCS. Il leur est souvent difficile d'admettre en début de formation qu'il existe parfois plusieurs réponses "valables" pour un même problème en médecine. Le mode de sélection des membres du panel de référence est également une source d'inquiétude : un TCS n'est valable que si les membres sont réellement experts – s'ils possèdent les "bons" scripts cliniques. Un problème de légitimité fréquemment évoqué par les étudiants concerne le recrutement d'universitaires en médecine qui n'ont qu'une activité clinique marginale peu représentative. Il semble important pour les enseignants de bien communiquer

sur ces considérations et les moyens utilisés pour lutter contre ces menaces afin de favoriser l'acceptabilité du TCS.

Le TCS explore uniquement une sous-partie du raisonnement clinique. Selon la théorie du double processus (42), le TCS explore la partie analytique du raisonnement clinique : la phase de raisonnement logique, consciente, qui monitore les processus intuitifs. Au travers du TCS, c'est l'efficacité de l'organisation des connaissances cliniques de l'étudiant qui est évaluée (ses scripts cliniques).

Il est donc nécessaire de garder à l'esprit que le TCS n'est pas l'outil miracle en évaluation du raisonnement clinique. Son utilisation doit être combinée à d'autres méthodes et outils pour correctement évaluer le raisonnement clinique des étudiants (44).

#### UN EXEMPLE DE RECHERCHE SUR LE RAISONNEMENT CLINIQUE UTILISANT LE TCS : LE BURNOUT AFFECTE-T-IL LE RAISONNEMENT CLINIQUE ? UNE ÉTUDE OBSERVATIONNELLE CHEZ DES INTERNES DE MÉDECINE GÉNÉRALE

Comme précisé en introduction, les TCS étaient utilisés dans le cadre des évaluations certificatives de fin de DES de médecine générale. Alors que dans littérature est débattu le lien entre le syndrome d'épuisement professionnel et le raisonnement clinique, nous avons utilisé les résultats de ces évaluations pour apporter quelques données supplémentaires sur cette question.

Ce travail a été réalisé dans le cadre du mémoire du master Ingénierie de la Formation et des Compétences (IFC) de l'Université de Strasbourg par Philippe Guillou. Philippe Guillou était un de mes collègues du Département de Médecine Générale de la Faculté de Médecine et il m'a sollicité pour encadrer son travail. Thierry Pelaccia fut sollicité dans un deuxième temps pour apporter son expertise méthodologique et son expérience de valorisation des travaux scientifiques en sa qualité de responsable du parcours "pédagogie en sciences de la santé" du master IFC. Enfin, Marie-Frédérique Bacqué est psychologue et travaille notamment sur la question du burnout au sein de son laboratoire SuLiSoM à l'Université de Strasbourg. J'ai participé pour ce travail à toutes les étapes, de la conception de l'étude, au recueil et l'analyse des données, jusqu'à la publication.

## Contexte de l'étude

Herbert J. Freudenberger a défini le burnout comme "le fait de s'épuiser en sollicitant de manière excessive [...] ses ressources au travail" (100). Il comporte trois dimensions : l'épuisement émotionnel (EE), la dépersonnalisation (DP) et la réduction de l'accomplissement professionnel (AP) (101,102). La prévalence du burnout chez les médecins a été signalée comme pouvant atteindre 85 %, et jusqu'à 75 % chez les internes, en fonction de leur spécialité médicale (103–106). Des tests psychométriques ont montré que le burnout est associé à des détériorations de l'attention, de la mémoire et des fonctions exécutives dans la population générale (107). Les effets de l'épuisement professionnel sur la capacité de raisonnement des médecins n'ont pas été étudiés de manière approfondie et les résultats sont contradictoires.

Si certaines études ont montré un impact négatif du burnout sur les performances du raisonnement clinique des médecins (108), d'autres n'ont pas réussi à démontrer un tel impact (109–111). Afin de mieux comprendre le lien entre le raisonnement clinique et le burnout, nous avons mené une étude visant à rechercher une association entre le statut de burnout et une moindre performance en matière de raisonnement clinique dans une population d'internes en médecine générale.

## Matériel et méthodes

Nous avons conçu une étude observationnelle transversale à la recherche d'une association statistique entre les scores de la version française du *Maslach Burnout Inventory-Human Services Survey* (MBI-HSS) (112) et un TCS (92). Nous avons supposé qu'un score élevé d'épuisement professionnel pouvait être associé à un score plus faible au TCS.

Nous avons recueilli des données sur les examens de certification des internes en 2017 et 2019 à l'Université de Strasbourg, France. Les participants ont reçu une présentation de l'étude, un formulaire de consentement écrit et le questionnaire MBI-HSS. Nous avons informé oralement les internes de l'objectif de la recherche et nous nous sommes assurés du caractère volontaire de leur participation avant le début de l'examen. L'examen TCS a duré 90 minutes, et les internes devaient rester jusqu'à la fin. La participation ou la non-participation n'a pas influencé les programmes de formation ou les évaluations des internes.

Le MBI-HSS est l'outil le plus couramment utilisé pour évaluer le burnout dans la population médicale (106,113). Il s'agit d'un questionnaire d'auto-évaluation validé mesurant les trois dimensions de l'épuisement professionnel : épuisement émotionnel, dépersonnalisation et réduction de l'accomplissement personnel (101). Le MBI-HSS donne un score dans chacune des trois dimensions du burnout. Il fut demandé aux internes d'indiquer, sur une échelle de Likert, leur degré d'accord avec les énoncés d'une échelle d'épuisement émotionnel à 9 items (par exemple, "J'ai l'impression de travailler trop dur dans mon travail"), d'une échelle de dépersonnalisation à 5 items (par exemple, "Je ne me soucie pas vraiment de ce qui arrive à certains patients") et d'une échelle d'accomplissement personnel à 8 items (par exemple, "J'ai accompli beaucoup de choses valables dans ce travail"). Chaque item était noté de 0 à 6 sur une échelle de fréquence de Likert : 0 = jamais, 1 = au moins quelques fois par an, 2 = au moins une fois par mois, 3 = quelques fois par mois, 4 = une fois par semaine, 5 = quelques fois par semaine et 6 = tous les jours. Nous avons utilisé la version française du MBI-HSS de *Mind Garden, Inc.* avec une licence de reproduction.

Nous avons choisi d'utiliser deux des principaux seuils de burnout décrits dans la littérature, car il existe un débat intense sur la meilleure façon de les définir (106,113,114). Aucune méthode ne semble faire l'objet d'un consensus.

Tout d'abord, les seuils de sévérité du score de burnout dans cette étude sont basés sur les normes originales du manuel d'inventaire du burnout de Maslach, révisées dans cette population française (101,104,112,115) (voir tableau 3). Nous appellerons par la suite cette méthode "l'approche conventionnelle". La définition des seuils de chaque dimension reste controversée, car ils peuvent différer d'une population à l'autre (106,113,116,117). Nous avons choisi ces seuils, car ils ont été utilisés il y a 10 ans dans une population similaire pour évaluer le burnout (104).

Le nombre de dimensions fortement affectées définissait le burnout : faible avec un score élevé dans une seule dimension, modéré avec des scores élevés dans deux dimensions, et élevé avec des scores élevés dans les trois dimensions.

	Faible	Modéré	Élevé
Épuisement émotionnel	≤17	18-29	≥30
Dépersonnalisation	≤5	06-11	≥12

Accomplissement personnel	≥40	34-39	≤33

**TABEAU 3: SEUILS DES DIMENSIONS DU BURNOUT DU MBI-HSS DANS L'APPROCHE CONVENTIONNELLE**

Deuxièmement, nous avons analysé le niveau d'épuisement global à l'aide de la règle "épuisement +1" appliquée aux scores du MBI-HSS : les individus peuvent être considérés comme épuisés lorsque, par rapport à un groupe de référence, ils rapportent un épuisement émotionnel élevé combiné à une dépersonnalisation élevée ou un faible accomplissement personnel (114). Dans cette approche, "élevé" signifie que le score se situe dans le 75e percentile ou plus, tandis que "faible" fait référence au score dans le 25e percentile ou moins (114). Nous avons utilisé cette méthode pour déterminer les seuils de burnout, car elle considère le burnout davantage comme un continuum que comme un seuil prédéfini (113).

#### *Le raisonnement clinique*

L'évaluation du processus de raisonnement clinique est une tâche compliquée, sans outil d'évaluation unique idéal (44). Le TCS est actuellement l'un des outils les plus performants disponibles pour évaluer le raisonnement clinique (118). Nous avons créé un TCS de 90 items sur la médecine générale (119). Nous avons utilisé le guide de Lubarsky et al. pour développer ce SCT (99) et les recommandations de Dory et al. pour recruter les membres du panel et analyser les scores (120). Les scores sont exprimés en pourcentages pour l'analyse. Un seuil de réussite à 60% est généralement recommandé. Ce seuil de réussite est déterminé par le score moyen des membres du panel (généralement environ 80 %). Un score <60% est alors considéré comme trop éloigné de la "bonne" réponse pour être accepté. Une question de ce SCT est présentée à titre d'exemple dans le tableau n°4.

Une femme de 85 ans a chuté ce matin et souffre de la hanche droite.			
N°	Si vous pensiez à...	Et que vous trouvez...	L'impact sur votre hypothèse ou option est :
Q1	Une fracture fémorale	Aucune impotence fonctionnelle	-2 -1 0 +1 +2

	droite	de la hanche droite	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Q2	Prescrire une radiographie de la hanche droite	Que la patiente marche parfaitement bien.	-2 -1 0 +1 +2 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

-2 = Fortement négatif, - 1 = Négatif, 0 = Ne change rien, +1 = Positif, + 2 = Fortement positif

**TABLEAU 4 : EXEMPLE DE QUESTIONS DE TCS EMPLOYÉES DANS L'ÉTUDE**

## Résultats

### Population

En 2017, 111 (86,7 %) internes ont accepté de participer. En 2019, 88 (66,7 %) internes ont accepté, soit un total de 199 participants.

Le score au TCS moyen de nos participants était de 76,44% (IC 95% : 75,77-77,10). Le score minimal était de 61,30 % et le score maximal de 88,96 % (écart-type = 4,78). Le score global du TCS de tous les internes pour ces deux années (n=259) était de 76,41% (IC 95% : 75,77-77,04). Il n'y avait pas de différence significative entre les scores moyens globaux de tous les internes et des participants (p=0,4738).

Dans l'approche conventionnelle des seuils de burnout, 126 participants (63,31%) n'avaient pas de burnout, 37 (18,59%) avaient un burnout léger, 23 (11,56%) un burnout modéré et 13 (6,53%) un burnout sévère. Les scores moyens étaient de 20,88 pour l'épuisement émotionnel, 9,57 pour la dépersonnalisation et 38,52 pour l'accomplissement personnel. Le tableau n°5 illustre la distribution des scores des dimensions du burnout selon l'approche conventionnelle des seuils.

	Score faible	Score modéré	Score élevé
Épuisement émotionnel	82 (41.21%)	73 (36.68%)	44 (22.11%)
Dépersonnalisation	57 (28.64%)	70 (35.18%)	82 (41.21%)
Accomplissement personnel	94 (47.24%)	69 (69.70%)	36 (18.09%)

**TABLEAU 5 : RÉPARTITION DES SCORES DES DIMENSIONS DU BURNOUT DANS L'APPROCHE CONVENTIONNELLE**

Dans l'approche "épuisement +1", 38 résidents avaient un statut de burnout (19,10%). Les seuils élevés et bas des scores des dimensions du burnout sont présentés dans le tableau n°6.

	Score faible	Score modéré	Score élevé
Épuisement émotionnel	≤12	13-27	≥28
Dépersonnalisation	≤5	5-12	≥13
Accomplissement personnel	≥43	34-42	≤35

**TABLEAU 6 : SEUILS DES DIMENSIONS DU BURNOUT DU MBI-HSS DANS L'APPROCHE " ÉPUISEMENT +1 "**

Il n'y avait pas de corrélation statistiquement significative entre les scores TCS et le statut de burnout dans l'approche conventionnelle des seuils ( $p=0.6509$ ). Les détails des scores SCT concernant la sévérité du burnout sont présentés dans le tableau n°7.

	Score TCS moyen (IC95%)	p
Absence de burnout	73.29% (72.32-74.29)	p=0.6509
Burnout léger	71.45% (69.08-73.83)	
Burnout modéré	74.96% (72.43-77.49)	
Burnout sévère	71.83 (70.31-73.36)	

**TABLEAU 7 : COMPARAISON DES SCORES MOYENS DU SCT SUR LA SÉVÉRITÉ DU BURNOUT DANS L'APPROCHE CONVENTIONNELLE**

Il n'y avait pas non plus de différence statistiquement significative entre les scores TCS et le statut de burnout dans l'approche "épuisement +1" : le score TCS moyen avec burnout était de 73,48% [IC 95% : 71,76-75,21] contre 72,95% [IC 95% : 72,01-73,88] sans burnout ;  $p=0,6136$ .

## Discussion

Nos résultats sont en accord avec plusieurs études qui n'ont montré aucun effet négatif du burnout sur le raisonnement clinique dans divers contextes (109-111). Plusieurs facteurs pourraient expliquer ces résultats. Dans l'approche "épuisement émotionnel +1", le burnout est un syndrome émotionnel. Alors que les émotions sont connues pour avoir un impact sur le raisonnement clinique (121,122), certains auteurs soutiennent que les émotions ont un impact plus important sur le système 1 (123). Une tâche de raisonnement spécifique au TCS consistant à analyser l'impact des informations sur une hypothèse ou une option d'investigation est un processus hypothético-déductif qui favorise et explore le système 2 (42,47). Par conséquent, une spéculation concernant nos résultats est que le burnout pourrait ne pas affecter le système 2.

Une autre hypothèse serait que le burnout n'affecte pas du tout le raisonnement clinique. Une telle hypothèse remettrait en question la valence affective du raisonnement intuitif. De nombreux auteurs, tel que Croskerry, soutiennent depuis plus d'une décennie que les émotions ont un impact sur le raisonnement clinique (124,125). En tant que tel, d'autres études pourraient viser à explorer spécifiquement l'impact de l'épuisement professionnel sur le raisonnement clinique système 1 pour tester l'impact de l'épuisement professionnel sur le raisonnement clinique plus spécifiquement.

La version publiée en janvier 2021 dans la revue *BMC Medical Education* (126) se trouve ci-dessous.

RESEARCH ARTICLE

Open Access

# Does burnout affect clinical reasoning? An observational study among residents in general practice



Philippe Guillou<sup>1</sup>, Thierry Pelaccia<sup>2,3</sup>, Marie-Frédérique Bacqué<sup>4</sup> and Mathieu Lorenzo<sup>1,2\*</sup>

## Abstract

**Background:** Burnout results from excessive demands at work. Caregivers suffering from burnout show a state of emotional exhaustion, leading them to distance themselves from their patients and to become less efficient in their work. While some studies have shown a negative impact of burnout on physicians' clinical reasoning, others have failed to demonstrate any such impacts. To better understand the link between clinical reasoning and burnout, we carried out a study looking for an association between burnout and clinical reasoning in a population of general practice residents.

**Methods:** We conducted a cross-sectional observational study among residents in general practice in 2017 and 2019. Clinical reasoning performance was assessed using a script concordance test (SCT). The Maslach Burnout Inventory for Human Services Survey (MBI-HSS) was used to determine burnout status in both original standards of Maslach's burnout inventory manual (conventional approach) and when individuals reported high emotional exhaustion in combination with high depersonalization or low personal accomplishment compared to a norm group ("emotional exhaustion +1" approach).

**Results:** One hundred ninety-nine residents were included. The participants' mean SCT score was 76.44% (95% CI: 75.77–77.10). In the conventional approach, 126 residents (63.31%) had no burnout, 37 (18.59%) had mild burnout, 23 (11.56%) had moderate burnout, and 13 (6.53%) had severe burnout. In the "exhaustion + 1" approach, 38 residents had a burnout status (19.10%). We found no significant correlation between burnout status and SCT scores either for conventional or "exhaustion + 1" approaches.

**Conclusions:** Our data seem to indicate that burnout status has no significant impact on clinical reasoning. However, one speculation is that SCT mostly examines the clinical reasoning process's analytical dimension, whereas emotions are conventionally associated with the intuitive dimension. We think future research might aim to explore the impact of burnout on intuitive clinical reasoning processes.

**Keywords:** Burnout, Clinical reasoning, General practice, Medical education, Script concordance test

\* Correspondence: [mlorenzo@unistra.fr](mailto:mlorenzo@unistra.fr)

<sup>1</sup>Departement of General Practice, Medicine Campus, University of Strasbourg, 4, rue Kirschleger, 67085 Strasbourg Cedex, France

<sup>2</sup>Center for Training and Research in Health Sciences Education, Medicine Campus, University of Strasbourg, 4, rue Kirschleger, 67085 Strasbourg Cedex, France

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Herbert J. Freudenberger defined burnout as “becoming exhausted by making excessive demands on [ ...] resources at work” [1]. The prevalence of burnout among physicians has been reported to be up to 85%, and up to 75% in residents, depending on their medical specialty [2–5]. Burnout has been found, through psychometric tests, to be associated with deteriorations in attention, memory, and executive functions in the general population [6]. The effects of burnout on a physician’s ability to reason have not been extensively studied, and the results are contradictory.

Clinical reasoning encompasses the range of cognitive processes necessary to evaluate and treat patients [7]; it lies “at the core of health care practice and education” [8]. Multiple clinical reasoning components can be identified: information gathering, hypothesis generation, forming a problem representation, generating a differential diagnosis, selecting a leading or working diagnosis, providing diagnostic justification, and developing a management or treatment plan [9]. One of the current main theoretical models assumes the existence of two cognitive processes commonly used by physicians to perform these steps [10, 11]: intuitive processes (system 1) and analytical processes (system 2). Both systems are jointly involved in most physicians’ decisions: reasoning always starts intuitively (system 1), generating one or more possible solutions, and then the analytical system (system 2) will allow confirmation or invalidation of the relevance of these (hypotheses selection) [12].

While some studies have shown a negative impact of burnout on physicians’ clinical reasoning performance [13], others have failed to demonstrate any such impact [14–16]. Residents seemed to be more susceptible to burnout effects than faculty in a study by Durning et al. [13]. Residents had different blood oxygen level-dependent (BOLD) signals detected by functional magnetic resonance when answering and reflecting upon clinical problems [13]. Higher depersonalization scores were associated with a lower BOLD signal in some areas of the brain, and higher emotional exhaustion scores were associated with stronger BOLD signals in others [13].

However, another study on pediatric residents showed no statistically significant association between burnout and harmful, nonharmful, or total errors [14]. A study on internist residents showed no difference in diagnostic and therapeutic accuracy compared with certified internists, despite significantly higher burnout scores [15]. Even more confusing, residents with high burnout scores were reported to have a small decrease in medical errors compared with burnout-free residents in a study on internal medicine residents [16].

To better understand the link between clinical reasoning and burnout, we carried out a study looking for an

association between burnout status and lower clinical reasoning performance in a population of residents in general practice.

## Methods

We designed a cross-sectional observational study looking for a statistical association between the scores on the French version of the Maslach Burnout Inventory-Human Services Survey (MBI-HSS) [17] and a script concordance test (SCT) [18]. We assumed that high burnout scores could be associated with a lower SCT rating. We followed the Strengthening The Reporting of Observational studies in Epidemiology (STROBE) checklist on what should be included in an accurate and complete report of an observational study [19].

## Setting

We collected data on residents’ certification examinations in 2017 and 2019 at the University of Strasbourg, France. Participants were given a presentation of the study, a written consent form, and the MBI-HSS questionnaire. We informed the residents of the research goal orally and ensured the voluntary nature of their participation before the start of the examination. The SCT examination lasted 90 min, and the residents had to stay until the end. Two of the researchers (PG and ML) were faculty for the residents.

## Population

The study population consisted of general practice residents in their final year (third postgraduate year) at the University of Strasbourg. To participate, residents had to give written informed consent. Participation or nonparticipation did not influence residents’ training programs or assessments.

As participants could realize that they suffered from burnout, they received the contact details of adequate support resources with the consent form before the study.

## Variables under consideration

We collected information on sex, age, marital and parental status, and current residency workplace, as these variables might influence burnout and clinical reasoning, as shown in previous studies [13, 20].

## Burnout

The MBI-HSS is the most commonly used tool to assess burnout in the medical population [5, 21]. It is a validated self-report questionnaire measuring the three dimensions of burnout: emotional exhaustion, depersonalization, and reduced personal accomplishment [22]. The MBI-HSS gives a score in each of burnout’s three dimensions. Residents were asked to indicate, on a Likert scale, their

degree of agreement with statements on a 9-item emotional exhaustion scale (e.g., “I feel I’m working too hard at my job”), a 5-item depersonalization scale (e.g., “I do not truly care what happens to some patients”), and an 8-item personal accomplishment scale (e.g., “I have accomplished many worthwhile things in this job”). Each item scores from 0 to 6 on a Likert frequency scale: 0 = never, 1 = at least a few times a year, 2 = at least once a month, 3 = a few times a month, 4 = once a week, 5 = a few times a week, and 6 = every day. We used the French version of the MBI-HSS from Mind Garden, Inc., with a license to reproduce.

We chose to use two of the main burnout cutoffs described in the literature, as there is an intense debate on the best way to define them [5, 21, 23]. No single method seems to enjoy a consensus.

First, burnout score severity cutoffs in this study were based on the original standards of Maslach’s burnout inventory manual, revised in this French population [3, 17, 22, 24] (see Table 1). We will refer later to this method as the “conventional approach.” Defining each dimension’s cutoff remains controversial, as they may differ from one population to another [5, 21, 25, 26]. We chose these cutoffs, as they were used 10 years ago in a similar population to evaluate burnout [3].

The number of highly affected dimensions defined burnout: low with a high score in only one dimension, moderate with high scores in two dimensions, and high with high scores in all three dimensions.

Second, we analyzed the overall burnout level using the “exhaustion + 1” rule applied to MBI-HSS scores: individuals can be considered burned out when, compared to a norm group, they report high emotional exhaustion in combination with high depersonalization or low personal accomplishment [23]. In this approach, “high” means scoring in the 75th percentile or higher, while “low” refers to scoring in the 25th percentile or lower [23]. We used this method to determine burnout cutoffs, as it considers burnout more as a continuum than a predefined cutoff [21].

### Clinical reasoning

Evaluation of the clinical reasoning process is a complicated task with no ideal single assessment tool [27]. Some authors consider that testing clinical reasoning in the context of uncertainty and respecting the possibility

of more than one good option are two core principles [28]. The script concordance test (SCT) is currently one of the most powerful tools available to assess clinical reasoning under these principles [28]. SCTs are meant to measure the degree of concordance between examinees and a reference panel of experts concerning clinical decisions and actions under uncertainty. For each item, a clinical case is presented (a vignette), containing either insufficient information to solve the clinical problem (diagnostic, treatment) or ambiguous data. A series of questions is related to the case. Each contains an option relevant to the clinical problem, followed by the presentation of new information. The examinees’ task is to assess the effect this new information has on the option’s status. It mostly examines the hypothesis selection stage in the clinical reasoning process [27, 29]. It is thought to explore system 2 rather than system 1 [18]. SCT is used in many curricula worldwide to evaluate clinical reasoning among pre- and postgraduate medical students.

We used SCT scores to evaluate clinical reasoning in our population. We created a 90-item SCT on general practice [30]. We used Lubarsky et al.’s guide to develop this SCT [31] and recommendations from Dory et al. to recruit panel members and analyze scores [32]. Scores are expressed as percentages for the analysis. A pass cutoff at 60% is usually recommended. This passing cutoff is determined by the mean score of panel members (usually approximately 80%). A score < 60% is then considered too far from what is the “good” answer to be accepted.

We chose to use universal anchors to be able to mix every type of SCT question in each vignette when wanted. One question from this SCT is presented as an example in Table 2.

### Statistical analysis

For 80% power and an alpha risk of 5%, 20 or more participants from each group (with or without burnout) were required to show a difference of 1 point or more in SCT scores.

The sample was subjected to descriptive statistical analysis. Means are presented in the results with 95% confidence intervals and the standard deviation or the minimum and maximum values.

The mean SCT scores were compared for qualitative variables of sex, marital status, and parental status using the Mann-Whitney test. We studied the statistical association between the SCT results and the quantitative age variables and the different MBI scores for each burnout dimension by calculating the Pearson correlation coefficients (Rho). Finally, we compared the means of the SCT score in the subgroups classified according to burnout severity using the Kruskal-Wallis test for the conventional

**Table 1** MBI-HSS burnout dimension cut-offs in the conventional approach

	Low	Moderate	High
Emotional exhaustion	≤17	18–29	≥30
Depersonalization	≤5	06–11	≥12
Personal accomplishment	≥40	34–39	≤33

**Table 2** Example of SCT questions

85 years old woman had felt this morning and suffers from right hip pain.			
N°	If you were thinking of:	And then you were to find:	The effect of the new information on the hypothesis is:
Q1	A right hip fracture	No functional disability on the right leg	-2 -1 0 +1 +2 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Q2	Prescribe a right hip X-ray	This patient walks perfectly fine	-2 -1 0 +1 +2 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

-2: strongly negative; -1: negative; 0: no effect; +1: positive; +2: strongly positive

burnout method and Student’s t-test for the burnout “exhaustion + 1” method. Statistical analysis was performed using R through the GMRC Shiny Stats application and RStudio version 1.2.1335.

**Results**

**Population**

In 2017, 111 (86.7%) residents agreed to participate. In 2019, 88 (66.7%) residents agreed, for a total of 199 participants. Informed consent was obtained for all participants.

Almost two-thirds of the participants were women (n = 123). Most of the participants were in a relationship (n = 139; 83.2%; 32 missing data) and had no children (n = 113; 86.3%; 68 missing data). The participants’ average age was 28 [min-max: 26–42; two missing data].

**SCT scores**

The mean SCT score of our participants was 76.44% (95% CI: 75.77–77.10). The minimum score was 61.30%, and the maximum score was 88.96% (SD = 4.78).

The global SCT score of all residents for these 2 years (n = 259) was 76.41% (95% CI: 75.77–77.04). There was no significant difference between the global mean scores of all residents and participants (p = 0.4738).

**Burnout assessment**

In the conventional burnout cutoff approach, 126 residents (63.31%) had no burnout, 37 (18.59%) had mild burnout, 23 (11.56%) had moderate burnout, and 13 (6.53%) had severe burnout. The mean scores were 20.88 for emotional exhaustion, 9.57 for depersonalization, and 38.52 for personal accomplishment.

Table 3 illustrates the distribution of burnout dimension scores in the conventional cutoff approach.

In the “exhaustion + 1” approach, 38 residents had a burnout status (19.10%). High and low cutoffs for burnout dimension scores are presented in Table 4.

**SCT and MBI score association**

There was no statistically significant correlation between SCT scores and burnout status in the conventional cutoff approach (p = 0.6509). Details of SCT scores concerning burnout severity are presented in Table 5.

There was also no statistically significant difference between SCT scores and burnout status in the “exhaustion + 1” approach: mean SCT with burnout was 73.48% [95% CI: 71.76–75.21] versus 72.95% [95% CI: 72.01–73.88] without burnout; p = 0.6136.

**Discussion**

**Comparison with the literature**

In the conventional approach, the mean burnout scores found in our study in each dimension were slightly lower than those found in a recent literature review conducted by Erschens et al. among medical residents for emotional exhaustion (20.58 versus 22.9, respectively) and personal accomplishment (38.53 versus 35.1, respectively). However, it was higher for depersonalization (9.57 versus 8.9, respectively) [5]. Similarly, our mean burnout scores were comparable with those of a national study in France among general practice residents from 2011: 20.0 for emotional exhaustion, 9.7 for depersonalization, and 34.8 for personal accomplishment [3]. The prevalence of burnout was lower in our population than in the authors’ study: 36.68% versus 48.1% [3]. We found fewer residents with burnout in the conventional approach than in a study from 2009 at the Strasbourg medical school (46%) [24]. Our resident population seemed to suffer less from burnout than other studies.

**Table 3** Repartition of burnout dimension scores in the conventional approach

	Low score	Moderate score	High score
Emotional exhaustion	82 (41.21%)	73 (36.68%)	44 (22.11%)
Depersonalization	57 (28.64%)	70 (35.18%)	82 (41.21%)
Personal accomplishment	94 (47.24%)	69 (69.70%)	36 (18.09%)

**Table 4** MBI-HSS burnout dimensions cut-offs in the "exhaustion +1" approach

	Low score	Moderate score	High score
Emotional exhaustion	≤12	13–27	≥28
Depersonalization	≤5	5–12	≥13
Personal accomplishment	≥43	34–42	≤35

SCT scores were comparable with the values expected for such a postgraduate examination [32]. The mean SCT score for residents at the end of their formation is 75% [32].

Our results contradict our initial hypothesis, assuming that high burnout scores could be associated with a lower SCT rating. There seems to be no significant association between clinical reasoning measured by SCT and burnout within our experimental conditions.

Therefore, our results are in line with several studies that showed no negative effect of burnout on clinical reasoning in various settings [14–16]. Several factors could explain these results. In the "emotional exhaustion +1" approach, burnout is an emotional syndrome. As emotions are known to impact clinical reasoning [33, 34], some authors argue that emotions have a greater impact on the system 1 reasoning process [35]. An SCT-specific reasoning task consisting of analyzing the impact of information on a hypothesis or an investigation option is a hypothetical-deductive process that fosters and explores system 2 [8, 11]. Consequently, one speculation regarding our results is that burnout might not affect system 2.

However, no specific data on clinical reasoning behaviors with SCT are available. One study by Surry et al. examined clinical reasoning behaviors in a 210-item clinical-vignette MCQ test based on dual-process theory [36]. The results showed that both systems 1 and 2 processes were elicited for nearly all test questions (100 and 97.1%, respectively) in a small sample of subjects [36]. Further studies are needed to explore system 1 and system 2 reasoning use during an SCT to support the assumption that SCT mostly explores system 2. Finally, our findings illustrate some of the difficulties in studying the links between clinical reasoning and burnout.

Another hypothesis would be that burnout does not affect clinical reasoning at all. Considering this would question the affective valence of intuitive reasoning.

**Table 5** Comparison of SCT mean scores on burnout severity in the conventional approach

	Means SCT score (95%CI)	p-value
Burnout absence	73.29% (72.32–74.29)	$p = 0.6509$
Mild burnout	71.45% (69.08–73.83)	
Moderate burnout	74.96% (72.43–77.49)	
Severe burnout	71.83 (70.31–73.36)	

Many authors, such as Croskerry, have argued for over a decade that emotions do impact clinical reasoning [10, 37]. As such, other studies might aim to specifically explore the impact of burnout on clinical reasoning system 1 to test the impact of burnout on clinical reasoning more specifically.

#### Strengths and weaknesses

Our study's main strength is its originality, as few data on the links between burnout and clinical reasoning performance are available. To the best of our knowledge, no other studies used SCT to study the links between burnout and clinical reasoning.

The use of a standardized and validated burnout questionnaire given to the whole study population has reduced subjectivity bias. The oral and written study presentation included no hypothesis but rather a broad research question to avoid influencing the participants' answers to the MBI-HSS questionnaire.

Burnout is impacted by the environment and context of the individual. We assume that being out of the real professional environment in a quiet classroom may have reduced tension for some residents. Clinical reasoning in an SCT question probably does not imply as much emotion as with a real patient. This could be another limitation of our findings. Having the subject perform a clinical reasoning task under more realistic workplace conditions could improve our research's validity.

We noticed a slight decrease in SCT scores with increased burnout levels. There might be an association that is not reflected in our study because it lacks power. A larger sample size could probably show a statistically significant difference between the groups. However, even if such a difference existed, the estimated effect size would be much lower than a standard deviation.

Measuring burnout remains challenging today. No assessment tool seems ideal, and there is much debate on the best definition of burnout [21]. The heterogeneity of research on burnout is called into question, as no less than 142 unique definitions of burnout were found in a recent review [21]. Thus, the clinical validity of burnout definitions are not certain [21]. The fourth edition of the MBI manual removed the cutoffs [38] of the classical approach and preconized the calculation of individuals' latent profiles for burnout [39]. These latent burnout profiles could be used for future research.

Likewise, assessing clinical reasoning with a single tool such as SCT may not be sufficiently valid for such a complex process [27]. Future studies on clinical reasoning and burnout may use a wide range of assessment methods, such as SCT + OSCE + direct observations + global assessments + think-aloud techniques [27].

Last, it is unclear whether nonparticipants and participants in the studied population are equivalent concerning burnout.

## Conclusions

Our data with SCT scores seem to indicate that burnout status has no significant impact on clinical reasoning among GP residents. However, burnout may either truly have no impact on clinical reasoning performance or impact system 1 processes. Overall, researching the links between burnout and clinical reasoning is complex. Further studies could explore such an impact of burnout on clinical reasoning system 1, but this will not be an easy task. To our knowledge, there is no valid easy method capable of separating these two closely related processes. Mixed methods approaches, where qualitative studies are integrated with quantitative studies, might be a fruitful avenue for future research.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12909-020-02457-y>.

**Additional file 1.** All data of participants including MBI-HSS responses, SCT scores, and descriptive statistical analysis.

## Acknowledgments

The authors thank Chloé Delacour and Léa Charton who helped us collect the data, François Severac and Boris Gass who gave support for statistical analysis and presentation of results. This article received support for English translation from the Maison Interuniversitaire des Sciences de l'Homme d'Alsace (MISHA) and the Excellence Initiative of the University of Strasbourg.

## Authors' contributions

PG, MFB, and ML designed the study, collected data, analyzed the data, drafted the manuscript, and critically revised subsequent versions. TP helped to design the study, writing the article, and revising subsequent versions. All authors approved the final manuscript and agreed both to be personally accountable for their own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which they were not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature.

## Authors' information

Philippe Guillou is MD, Med. He works as a general practitioner and faculty. He works on clinical reasoning, burnout, and learning environment. Thierry Pelaccia is MD, Ph.D. He is the director of the Centre for Training and Research in Health Sciences Education (CFRPS), Faculty of Medicine, University of Strasbourg (France), whose main field of research deals with clinical reasoning and decision-making. Marie-Frédérique Bacqué is PsyD, Ph.D. She is director of the UR3071 laboratory Psychology Faculty, University of Strasbourg. She works in the field of death studies, psycho-oncology, burnout, and mental health. Mathieu Lorenzo is MD, Med, Ph.D. candidate. He works as a general practitioner and faculty. He works in the field of clinical reasoning assessment and virtual patients.

## Funding

None.

## Availability of data and materials

The dataset supporting the conclusions of this article is included in the article in Additional file 1.

## Ethics approval and consent to participate

The protocol was approved by the ethical committee of the Faculty of Medicine and Dentistry, the Schools of Nursing, Physiotherapy, and Maieutics, and the Strasbourg University Hospital. We obtained informed consent to participate from all participants with a written consent form.

All methods were carried out in accordance with the STROBE guidelines.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>Departement of General Practice, Medicine Campus, University of Strasbourg, 4, rue Kirschleger, 67085 Strasbourg Cedex, France. <sup>2</sup>Center for Training and Research in Health Sciences Education, Medicine Campus, University of Strasbourg, 4, rue Kirschleger, 67085 Strasbourg Cedex, France. <sup>3</sup>Prehospital Emergency Care Service, Strasbourg University Hospital, University of Strasbourg, 1, place de l'hôpital, BP 426, 67091 Strasbourg Cedex, France. <sup>4</sup>EA3071, Psychology Faculty, University of Strasbourg, 12, rue Goethe, 67000 Strasbourg, France.

Received: 11 September 2020 Accepted: 11 December 2020

Published online: 07 January 2021

## References

- Freudenberger HJ. Staff Burn-Out. *J Soc Issues*. 1974;30(1):159–65.
- Ishak WW, Lederer S, Mandili C, Nikravesh R, Seligman L, Vasa M, et al. Burnout during residency training: a literature review. *J Grad Med Educ*. 2009;1(2):236–42.
- Galam E, Komly V, Tourneur AL, Jund J. Burnout among French GPs in training: a cross-sectional study. *Br J Gen Pract*. 2013;63(608):e217.
- Dyrbye L, Shanafelt T. A narrative review on burnout experienced by medical students and residents. *Med Educ*. 2016;50(1):132–49.
- Erschens R, Keifenheim KE, Herrmann-Werner A, Loda T, Schwillke-Kiuntke J, Bugaj TJ, et al. Professional burnout among medical students: systematic literature review and meta-analysis. *Med Teach*. 2019;41(2):172–83.
- Deligkaris P, Panagopoulou E, Montgomery AJ, Masoura E. Job burnout and cognitive functioning: a systematic review. *Work Stress*. 2014;28(2):107–23.
- Barrows HS, Tamblyn RM. Problem-based learning: an approach to medical education. New York: Springer Publishing Company; 1980. p. 225.
- Higgs J, Jensen GM, Loftus S, Christensen N. Clinical reasoning in the health professions. Edinburgh: Elsevier-Health Sciences; 2018. p. 532.
- Gruppen LD, Frohna AZ. Clinical Reasoning. In: Norman GR, van der Vleuten CPM, Newble DI, DHJM D, Mann KV, Rothman A, et al., editors. *International Handbook of Research in Medical Education*. Dordrecht: Springer Netherlands; 2002. p. 205–30. Available from: [http://link.springer.com/10.1007/978-94-010-0462-6\\_8](http://link.springer.com/10.1007/978-94-010-0462-6_8). [cited 2019 Jul 14].
- Croskery P. A universal model of diagnostic reasoning. *Acad Med J Assoc Am Med Coll*. 2009;84(8):1022–8.
- Pelaccia T, Tardif J, Tribby E, Charlin B. An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory. *Med Educ Online*. 2011;16:5890. <https://doi.org/10.3402/meo.v16i0.5890>.
- Hogarth RM. Deciding analytically or trusting your intuition? The advantages and disadvantages of analytic and intuitive thought. In: *The routines of decision making*. 1st ed. New York: Psychology Press; 2004. p. 18.
- Durning SJ, Costanzo M, Artino AR, Dyrbye LN, Beckman TJ, Schuwirth L, et al. Functional neuroimaging correlates of burnout among internal medicine residents and faculty members. *Front Psychiatry*. 2013;4:131. <https://doi.org/10.3389/fpsy.2013.00131>.
- Brunsborg KA, Landrigan CP, Garcia BM, Petty CR, Sectish TC, Simpkin AL, et al. Association of pediatric resident physician depression and burnout with harmful medical errors on inpatient services. *Acad Med*. 2019;94(8):1150–6.
- McBee E, Ratcliffe T, Picho K, Schuwirth L, Artino AR, Yepes-Rios AM, et al. Contextual factors and clinical reasoning: differences in diagnostic and therapeutic reasoning in board certified versus resident physicians. *BMC Med Educ*. 2017;17(1):211.
- Kwah J, Weintraub J, Fallar R, Ripp J. The effect of burnout on medical errors and professionalism in first-year internal medicine residents. *J Grad Med Educ*. 2016;8(4):597–600.
- Maslach C, Jackson SE, Leiter MP. *Maslach burnout inventory manual*. 3rd. Palo Alto: Consulting Psychologists Press; 1996. p. 191–218.

18. Charlin B, Roy L, Brailovsky C, Goulet F, van der Meuten C. The script concordance test: a tool to assess the reflective clinician. *Teach Learn Med*. 2000;12(4):189–95.
19. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The strengthening of reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med*. 2007;147(8):573–7.
20. Richter A, Kostova P, Harth V, Wegner R. Children, care, career - a cross-sectional study on the risk of burnout among German hospital physicians at different career stages. *J Occup Med Toxicol Lond Engl*. 2014;9(1):41.
21. Rotenstein LS, Torre M, Ramos MA, Rosales RC, Guille C, Sen S, et al. Prevalence of burnout among physicians: a systematic review. *JAMA*. 2018; 320(11):1131–50.
22. Maslach C, Jackson SE. The measurement of experienced burnout. *J Organ Behav*. 1981;2(2):99–113.
23. Brenninkmeijer V, VanYperen N. How to conduct research on burnout: advantages and disadvantages of a unidimensional approach in burnout research. *Occup Environ Med*. 2003;60(suppl 1):i16–20.
24. Ernst M. Le syndrome de burnout des internes en médecine générale à la Faculté de Médecine de Strasbourg: prévalence et analyse d'entretiens. France: Université de Strasbourg. Faculté de médecine; 2009.
25. Schaufeli WB, Van Dierendonck D. A cautionary note about the cross-national and clinical validity of cut-off points for the Maslach burnout inventory. *Psychol Rep*. 1995;76(3 Pt 2):1083–90.
26. Kovess-Masfety V, Saunderson L. Le burnout : historique, mesures et controverses. *Arch Mal Prof Environ*. 2017;78(1):16–23.
27. Daniel M, Rencic J, Durning SJ, Holmboe E, Santen SA, Lang V, et al. Clinical reasoning assessment methods: a scoping review and practical guidance. *Acad Med*. 2019;94(6):902.
28. Cooke S, Lemay J-F. Transforming medical assessment: integrating uncertainty into the evaluation of clinical reasoning in medical education. *Acad Med*. 2017;92(6):746–51.
29. Charlin B, Gagnon R, Pelletier J, Coletti M, Abi-Rizk G, Nasr C, et al. Assessment of clinical reasoning in the context of uncertainty: the effect of variability within the reference panel. *Med Educ*. 2006;40(9):848–54.
30. Attali C, Huez J-F, Valette T, Lehr-Drylewicz A-M. Les grandes familles de situations cliniques. *Exercer*. 2013;24(108):165–9.
31. Lubarsky S, Dory V, Duggan P, Gagnon R, Charlin B. Script concordance testing: from theory to practice: AMEE guide no. 75. *Med Teach*. 2013;35(3): 184–93.
32. Dory V, Gagnon R, Vanpee D, Charlin B. How to construct and implement script concordance tests: insights from a systematic review. *Med Educ*. 2012; 46(6):552–63.
33. McConnell MM, Eva KW. The role of emotion in the learning and transfer of clinical skills and knowledge. *Acad Med J Assoc Am Med Coll*. 2012;87(10): 1316–22.
34. Kozłowski D, Hutchinson M, Hurley J, Rowley J, Sutherland J. The role of emotion in clinical decision making: an integrative literature review. *BMC Med Educ*. 2017;17(1):255.
35. Croskerry P, Norman G. Overconfidence in clinical decision making. *Am J Med*. 2008;121(5 Suppl):S24–9.
36. Surry LT, Torre D, Durning SJ. Exploring examinee behaviours as validity evidence for multiple-choice question examinations. *Med Educ*. 2017;51(10): 1075–85.
37. Croskerry P, Abbass AA, Wu AW. How doctors feel: affective issues in patients' safety. *Lancet Lond Engl*. 2008;372(9645):1205–6.
38. Mind Garden. The Problem with Cut-Offs for the Maslach Burnout Inventory. Available from: <https://www.mindgarden.com/documents/MBI-Cutoff-Caveat.pdf>. Accessed 11 Nov 2020.
39. Leiter MP, Maslach C. Latent burnout profiles: a new approach to understanding the burnout experience. *Burn Res*. 2016;3(4):89–100.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)



## Synthèse de l'étude

Ce travail illustre une utilisation potentielle d'un outil tel que le TCS dans une perspective de recherche sur le processus de raisonnement clinique. Ces résultats ont également une application concrète sur le versant de la formation : si un lien avait été mis en évidence entre la présence d'un burnout et des difficultés de raisonnement clinique, les étudiants avec de faibles scores aux TCS auraient pu bénéficier d'une recherche ciblée d'un burnout et de l'accompagnement dédié le cas échéant.

Une telle association aurait eu également un impact important sur les soins puisque la prévention et la prise en charge du burnout n'aurait été plus seulement liée au bien-être de l'étudiant mais aussi à la sécurité du patient et à l'efficacité des soins proposés.

Les résultats négatifs de cette étude n'en sont pas moins importants. Il est tout d'abord indispensable dans une démarche de respect des bonnes pratiques en recherche de diffuser tout autant les résultats négatifs que positifs. Ensuite, ils soulignent la difficulté à mener des travaux sur le raisonnement clinique en l'absence de "*gold-standard*" pour l'évaluer. Enfin, si la théorie du double processus postule l'existence de deux systèmes de raisonnement concurrents/complémentaires, prétendre évaluer séparément un système de l'autre relève à ce jour de l'impossible.

Appliquer la taxonomie des connaissances de la MOT et le modèle de Montréal à ce travail est également intéressant. La représentation dynamique du problème de santé au sein du processus de raisonnement clinique a comme intrant la procédure "réguler le processus". Cette dernière est sous l'influence de divers concepts comme les influences d'origine cognitive, celles d'origine relationnelle ou celles d'origine émotionnelle. La question de l'impact du burnout sur la performance du raisonnement clinique pourrait ainsi se voir comme celle des influences d'origine émotionnelle sur la procédure "catégoriser pour décider de l'action" via la représentation dynamique du problème de santé. Ainsi, ces influences d'origine émotionnelle pourraient ne pas impacter significativement la connaissance d'action que constitue la sous-procédure "analyser les données et les scripts activés en vue de la catégorisation". Pourquoi la tâche cognitive d'un TCS ne serait-elle pas impactée par le burnout ? Le modèle MOT postule qu'une connaissance stratégique qui "régule la régulation" du processus de raisonnement clinique est que le clinicien tient compte, tout au long du processus, de la perspective du patient sur son problème (notamment ses émotions, ses craintes). Alors qu'une des dimensions du burnout est la dépersonnalisation, il pourrait impacter le raisonnement clinique dans la "vraie vie" mais pas dans un TCS. L'outil TCS ne mobilise pas cette connaissance stratégique. Ceci soulève alors un questionnement sur la

validité écologique de l'utilisation du TCS dans une telle recherche sur le processus de raisonnement clinique.

#### AU-DELÀ DU TCS : LES FORMATIONS PAR CONCORDANCE

Le TCS étant basé sur le calcul d'un score, ses propriétés psychométriques ont pu être étudiées sous tous les angles (95-98,127,128). Il a donné naissance à une catégorie à part entière de formations : les formations par concordance (129). Selon le Centre de Pédagogie Appliquée aux Sciences de la Santé (CPASS) de l'Université de Montréal, trois types de formations par concordance existent (129,130) :

- La concordance de raisonnement (FpC-raisonnement) qui permet de former au raisonnement clinique ;
- La concordance de jugement (FpC-jugement) qui permet de former dans les domaines de l'éthique ou du professionnalisme ;
- La concordance de perception (FpC-perception) qui permet de former à la perception et l'interprétation dans les domaines avec imagerie.

En lien avec les travaux présentés dans le cadre de ma thèse, je vais m'attarder sur les deux premiers types de FpC : les FpC-raisonnement et les FpC-jugement.

Les FpC-raisonnement ont classiquement le format le plus proche du TCS. Il est demandé à l'étudiant après avoir sélectionné une option de réponse concernant l'impact de l'hypothèse ou option de justifier son choix (129,130). Le feedback est en général triple selon le CPASS : "*(i) les réponses données par les membres du panel (dans quelle mesure sa réponse concorde avec celle des membres du panel, (ii) les justifications données par ces derniers, (iii) un message synthèse (ce dernier venant selon les cas, soit à la fin de chaque question, soit à la fin du cas)*". La constitution du panel de référence se fait sur le critère de l'expertise professionnelle dans le domaine évalué (129). De ce point découle une question délicate : comment définir l'expertise professionnelle ? Des auteurs comme K.A Ericsson ont consacré des carrières entières à définir le concept d'expertise professionnelle, son développement et son maintien dans divers milieux professionnels dont la médecine (131,132). Rentrer dans un débat sur l'expertise n'est pas l'objet de ce travail. Une solution proposée concernant les panels "experts" des TCS et FpC-raisonnement consiste à reconnaître comme expert celui qui est considéré comme tel par ses pairs. Ainsi, un vote pour être organisé au sein d'un échantillon de professionnels en demandant de désigner une ou plusieurs

personnes qui pourraient être considérées comme "expertes" du domaine. Ce mode de désignation du panel n'est pas sans poser des problèmes de faisabilité notamment en termes d'acceptabilité dans les communautés professionnelles peu familières avec ce genre d'outils pédagogiques. Bien souvent dans le contexte des sciences de la santé en France, ce critère d'expertise est remplacé par celui "d'années d'expérience" dans le recrutement des panels de référence. Les membres d'un panel de référence avec un nombre d'années d'expérience élevé produisent des barèmes de notation sensiblement plus "durs" (dans le sens de l'éloignement avec le profil de réponse des étudiants) (133).

Un exemple de vignette de FpC-raisonnement est proposé ci-dessous dans la figure n°11.

Alors que vous marchez dans la rue, vous observez un homme d'âge mûr tomber brutalement au sol. Vous vous approchez rapidement de lui pour vous assurer qu'il va bien. Il ne répond pas à vos questions et semble avoir perdu connaissance			
N°	Si vous pensiez à ...	... et qu'alors vous trouvez ...	... l'impact sur votre hypothèse ou option est ...
Q 1	Un arrêt cardio-respiratoire	Qu'il ne respire pas	-2   -1   0   +1   +2 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

-2 = Fortement négatif, - 1 = Négatif, 0 = Ne change rien, +1 = Positif, + 2 = Fortement positif

Comment justifiez-vous votre réponse ?

FIGURE 12 : EXEMPLE DE VIGNETTE DE FpC-RAISONNEMENT

Sous une forme proche, les FpC-jugement proposent de *"se prononcer sur le caractère adéquat ou inadéquat de comportements touchant le professionnalisme ou l'éthique"* (130). Deux formats différents de questionnement existent après la description de la situation.

Le premier pose la question "dans quelle mesure le comportement proposé est-il acceptable ?" puis propose un comportement. L'étudiant doit sélectionner une option de réponse sur une échelle de Likert à quatre modalités : *Totalement inacceptable/Difficilement acceptable/Plutôt acceptable/Totalement acceptable*. Ce choix est accompagné d'une justification écrite. La figure n°12 ci-dessous présente un exemple de FpC-jugement utilisant ce premier format (format "A") issu de ma banque personnelle.

Format A		
<p>Vous recevez en fin de journée à la demande de sa mère la petite Cécile, 10 ans qui vomit et a mal au ventre. Vous découvrez rapidement qu'elle avait réalisé un examen d'urines il y a une semaine dans le cadre d'un bilan de fièvre qui montrait des signes évidents d'infection urinaire.</p> <p>Ce jour, votre examen est en faveur d'une pyélonéphrite (complication potentiellement grave de son infection urinaire).</p> <p>La mère de Cécile est surprise du diagnostic car elle dit avoir reçu un appel de votre collègue la rassurant sur la normalité des examens complémentaires. Vous vous demandez si votre collègue s'est trompé dans l'interprétation de l'analyse d'urine ou s'il disposait d'une version tronquée des résultats.</p>		
Vous pensez...		Cette attitude, ce comportement serait...
Q1	Dire à Cécile et sa mère qu'elle aurait dû recevoir un traitement antibiotique la semaine passée qui aurait évité cette complication	<input type="checkbox"/> Totalement inacceptable <input type="checkbox"/> Difficilement acceptable <input type="checkbox"/> Plutôt acceptable <input type="checkbox"/> Totalement acceptable
Q2	Dire à Cécile et à sa mère que les examens de la semaine passée étaient normaux pour couvrir votre collègue	<input type="checkbox"/> Totalement inacceptable <input type="checkbox"/> Difficilement acceptable <input type="checkbox"/> Plutôt acceptable <input type="checkbox"/> Totalement acceptable
Q3	Proposer un traitement adapté à Cécile sans mentionner l'épisode de la semaine précédente	<input type="checkbox"/> Totalement inacceptable <input type="checkbox"/> Difficilement acceptable <input type="checkbox"/> Plutôt acceptable <input type="checkbox"/> Totalement acceptable

Vos justifications :

Q1 :

Q2 :

Q3 :

Le deuxième format pose la question "à quel point est-il légitime de (action), sachant que...". Chaque question propose une information qui vient interroger la légitimité de l'action précédemment mentionnée. L'étudiant répond sur la même échelle de Likert et y ajoute également une justification écrite.

La figure n°13 ci-dessous présente un exemple de questions utilisant ce deuxième format (format "B").

Format B		
<p>Vous recevez en fin de journée à la demande de sa mère la petite Cécile, 10 ans qui vomit et a mal au ventre. Vous découvrez rapidement qu'elle avait réalisé un examen d'urines il y a une semaine dans le cadre d'un bilan de fièvre qui montrait des signes évidents d'infection urinaire.</p> <p>Ce jour, votre examen est en faveur d'une pyélonéphrite (complication potentiellement grave de son infection urinaire).</p> <p>La mère de Cécile est surprise du diagnostic car elle dit avoir reçu un appel de votre collègue la rassurant sur la normalité des examens complémentaires. Vous vous demandez si votre collègue s'est trompé dans l'interprétation de l'analyse d'urine ou s'il disposait d'une version tronquée des résultats.</p>		
A quel point est-il légitime de proposer un traitement adapté à Cécile sans mentionner l'épisode de la semaine précédente ?		
Q1	Considérant que Cécile va bien	<input type="checkbox"/> Totalement inacceptable <input type="checkbox"/> Difficilement acceptable <input type="checkbox"/> Plutôt acceptable <input type="checkbox"/> Totalement acceptable
Q2	Considérant que sa mère vous montre une copie des résultats d'urine qu'elle a reçue au domicile et qu'il y est noté « infection urinaire »	<input type="checkbox"/> Totalement inacceptable <input type="checkbox"/> Difficilement acceptable <input type="checkbox"/> Plutôt acceptable <input type="checkbox"/> Totalement acceptable
Q3	Considérant que l'état de Cécile nécessite l'appel du SAMU pour un transfert médicalisé aux urgences devant des signes de défaillance viscérale	<input type="checkbox"/> Totalement inacceptable <input type="checkbox"/> Difficilement acceptable <input type="checkbox"/> Plutôt acceptable <input type="checkbox"/> Totalement acceptable

Vos justifications :

Q1 :

Q2 :

Q3 :

FIGURE 14 : EXEMPLE DE QUESTIONS D'UNE VIGNETTE DE FPC-JUGEMENT UTILISANT LE FORMAT "B"

Le feedback consiste à présenter les réponses des membres du panel et les justifications de ces choix. Le participant analyse la concordance entre son choix de réponse et ceux du panel ainsi que la concordance des justifications. Les exemples ci-dessous illustrent une présentation possible de ce feedback pour les exemples de questions précédemment présentées (figures n°14 et n°15). On notera au passage que si le format d'une FpC-jugement peut adopter deux types différents, le feedback est présenté de la même manière.

Question 1	Totalement inacceptable 1/3	#1 : Il n'est pas acceptable de dire du mal d'un collègue, ce n'est pas utile aux soins et le patient peut se sentir mal d'avoir été mal soigné ce qui peut nuire à son état de santé.
	Difficilement acceptable 1/3	#1 : Cette formule va culpabiliser la mère et mettre en difficulté le premier médecin
	Plutôt acceptable 0/3	
	Totalement acceptable 1/3	#1 : L'enfant et sa mère doivent pouvoir comprendre la pathologie et son évolution, c'est de l'éducation thérapeutique du patient. C'est important pour la qualité de la relation médecin/patient.
Question 2	Totalement inacceptable 2/3	#1 : Le mensonge n'est pas acceptable #2 : Il n'est pas acceptable de mentir à une patiente
	Difficilement acceptable 1/3	#1 : C'est délétère de mentir pour la qualité de la relation médecin-patient. Je ne vois pas l'intérêt de rentrer dans la collusion du mensonge.
	Plutôt acceptable 0/3	
	Totalement acceptable 0/3	
Question 3	Totalement inacceptable 0/3	
	Difficilement acceptable 2/3	#1 : Mentionner l'épisode se fera en fonction de la demande de la patiente. #2 : L'omission masque une information importante à l'éducation pour les patients. Point positif, le traitement est proposé pour avancer !
	Plutôt acceptable 1/3	#1 : Cela pourrait se concevoir pour ne pas mettre en difficulté le collègue et si la mère de la patiente ne pose pas de questions.
	Totalement acceptable 0/3	

FIGURE 15 : EXEMPLE DE FEEDBACK DE FpC-JUGEMENT CONCERNANT LES QUESTIONS PRÉSENTÉES DANS LE "FORMAT A"

Question 1	Totalement inacceptable 0/3	
	Difficilement acceptable 1/3	#1 : Acceptable pour protéger le collègue mais non acceptable dans l'idée de concerner les patients à leur démarche de santé. La prochaine fois, la mère ne se sentira pas « concernée ».
	Plutôt acceptable 2/3	#1 : L'objectif est que pour la patiente l'omission n'ai pas conséquence directe. #2 : Il n'est pas nécessaire de parler d'une erreur d'un collègue. Cela n'apporte rien aux soins. Vous n'avez pas les éléments pour savoir ce qui s'est passé.
	Totalement acceptable 0/3	
Question 2	Totalement inacceptable 1/3	#1 : La qualité de la relation de soins doit rester primordiale. Le mensonge n'en fait pas partie.
	Difficilement acceptable 1/3	#1 : Si le document est présenté, un commentaire du résultat est souhaité.
	Plutôt acceptable 1/3	#1 : La mention de l'épisode se fera à la demande de la patiente. Il est impératif de répondre aux questions des patients.
	Totalement acceptable 0/3	
Question 3	Totalement inacceptable 1/3	#1 : La qualité de la relation de soins doit rester primordiale. Le mensonge n'en fait pas partie.
	Difficilement acceptable 2/3	#1 : Là aussi la protection du collègue peut se concevoir mais l'hypothèse la plus probable étant que le processus évolutif de la pathologie sera reformulé, la non-mention de l'épisode à la mère passera également pour la collusion du mensonge. #2 : Là il serait sûrement utile d'appeler l'autre collègue pour l'encourager à parler de son « erreur ». Ça reste anti-déontologique de dévaloriser un collègue.
	Plutôt acceptable 0/3	
	Totalement acceptable 0/3	

FIGURE 16 : DEUXIÈME EXEMPLE DE FEEDBACK DE FpC-JUGEMENT CONCERNANT LES QUESTIONS PRÉSENTÉES DANS LE "FORMAT B"

A la suite des feedbacks des panels sur les questions, la synthèse vient constituer un troisième feedback pour les participants (voir exemple dans la figure n°16 ci-dessous).

### Synthèse

Ce cas traite de la divulgation d'informations liée à un aléa thérapeutique. Il renvoie à l'article 35 du code de déontologie médicale sur l'information des patients : « *Le médecin doit à la personne qu'il examine, qu'il soigne ou qu'il conseille une information loyale, claire et appropriée sur son état, les investigations et les soins qu'il lui propose. Tout au long de la maladie, il tient compte de la personnalité du patient dans ses explications et veille à leur compréhension. Toutefois, lorsqu'une personne demande à être tenue dans l'ignorance d'un diagnostic ou d'un pronostic, sa volonté doit être respectée, sauf si des tiers sont exposés à un risque de contamination* » (1).

C'est aussi la notion du mensonge qui est abordée. Certains médecins considèrent que "*la vraie information, c'est de répondre à la demande des malades*" (2). Il s'agit pourtant d'un mensonge par omission qui est contraire à l'article 35 du code de santé publique (1). Selon les travaux de Sylvie Fainzang, environ 43% des médecins français seraient prêts à ne pas révéler une erreur médicale si elle ne nuit pas au patient. Par contre seulement 16% ne révéleraient pas une erreur si celle-ci affecte le patient (2). Les médecins justifient le recours au mensonge par le fait que les patients ne sont pas capables d'entendre la "vérité" (2). Ici, c'est davantage le principe de confraternité "*les médecins se doivent assistance dans l'adversité*" qui est évoqué (3).

Il faut dans un cas comme celui-là privilégier l'information « claire et loyale » du patient et éviter de mentir. Les informations manquent pour expliquer totalement ce dysfonctionnement et il sera utile d'approfondir cet épisode de soins avec le médecin concerné.

### Pour aller plus loin

1. Code de la santé publique - Article R4127-35. Code de la santé publique.
2. Fainzang S. La relation médecins-malades : information et mensonge. Paris (6, avenue Reille 75685): Presses Universitaires de France - PUF; 2006. 159 p.
3. Décret n°2004-802 du 29 juillet 2004 relatif aux parties IV et V (dispositions réglementaires) du code de la santé publique et modifiant certaines dispositions de ce code. - Article R. 4127-56. 2004-802 juill 29, 2004.

FIGURE 17 : EXEMPLE DE SYNTHÈSE D'UN CAS DE FPC-JUGEMENT

Mon expérience d'utilisation des FpC-jugement est qu'il s'agit d'un outil extrêmement efficace pour engager un groupe d'étudiants dans un débat et une réflexion collective autour des questions de professionnalisme, d'éthique et de déontologie en sciences de la santé. La discussion s'engage en proposant à quelques étudiants de partager en plénière leur choix de réponse pour une question. En demandant "qui a répondu autre chose ?", d'autres étudiants s'engagent dans la discussion, comparent leurs justifications et en débattent. Une deuxième phase de réflexion se produit en donnant le feedback du panel de référence que les étudiants critiquent, approuvent, questionnent. Une troisième phase de réflexion se produit en donnant la synthèse du cas.

L'utilisation décrite dans la littérature de ces deux type de FpC est davantage asynchrone en ligne (129,134,135). Si le calcul d'un score est possible sur le mode du TCS dans ces FpC (129), son intérêt me semble marginal et notamment dans une perspective d'utilisation dans une évaluation formative. Ce score ne peut rendre compte de la richesse des apprentissages réalisées. L'étudiant compare certes son choix de modalité avec les choix des membres du panel de référence, mais c'est surtout la comparaison des justifications, l'appropriation de la synthèse et des éventuelles références qui permettent les apprentissages les plus profonds. Dans les travaux qui utiliseront des formations par concordance, je ferai le choix de délaissé le calcul d'un quelconque score pour éviter de réduire la richesse des apprentissages à un simple chiffre.

#### LE DÉVELOPPEMENT DE L'OUTIL HELLO-CLINICAL : UN OUTIL PERMETTANT D'UTILISER DES PATIENTS VIRTUELS POUR ÉVALUER LE PROCESSUS DE RAISONNEMENT CLINIQUE

A partir de mes lectures sur le modèle MOT et des formations par concordance, j'ai souhaité développer un logiciel informatique qui permettent de proposer une évaluation formative du processus de raisonnement clinique en utilisant un système de concordance. J'ai donc établi un cahier des charges de ce que devrait être un tel outil.

J'ai proposé de découper une rencontre clinique simulée selon les « étapes » (procédures) du MMOT de recueillir les représentations dynamique et stabilisée du problème sous forme d'hypothèses diagnostiques accompagnées d'une justification. J'ai choisi de recueillir la représentation initiale sous forme d'un texte libre. J'ai également choisi de recueillir dès les premiers instants sous forme de texte libre les données issues de l'observation du patient.

Le feedback supportant l'évaluation formative du raisonnement clinique est apporté en invitant l'étudiant à comparer sa réponse à celles de panels de référence. Il s'agit d'une auto-évaluation basée sur le principe des formations par concordance (98,135,136). Les étudiants sont invités à comparer leur réponse avec un échantillon des réponses de leurs pairs ainsi que les réponses d'un panel de référence. Ce panel de référence est à constituer selon le critère d'expertise clinique. Aussi, les professionnels invités à constituer ce panel de référence devraient être idéalement désignés par un vote de leurs pairs qui les reconnaîtraient comme « experts ».

Ce mode de retro-information par concordance permet d'illustrer l'incertitude fréquemment présente en situation de soins (tous les membres du panel ne sont pas systématiquement d'accord) et

qu'il existe bien souvent plusieurs réponses « acceptables » face à un même problème de santé. Ce "reflet de l'incertitude" est considéré comme une qualité dans la littérature sur l'évaluation du raisonnement clinique (16,42,137).

J'ai opté pour l'usage d'enregistrements vidéo pour simuler le patient dans cet outil informatique. Ce choix se justifie par le fait que la vidéo est ce qui se rapproche le plus d'un patient réel en termes d'informations disponibles à l'observation selon le modèle MOT : état de santé apparent, aspect corporel, niveau d'élocution, etc... (45). Aussi cette proximité entre la vidéo et la réalité permet une meilleure authenticité des consultations simulées présentées. J'estime ainsi que la transférabilité des apprentissages réalisés via cet outil en est meilleure qu'avec du texte, des photographies ou des modèles 3D. Certains aspects du modèle MOT ne peuvent pas être présentés pour des limitations techniques comme l'odorat et le toucher.

Dans la même logique d'authenticité, j'ai souhaité que les apprenants puissent interroger les patients simulés d'une manière la plus authentique possible. L'idéal est alors d'avoir un système de reconnaissance vocale des questions posées oralement par les apprenants et un algorithme permettant une reconnaissance de la question posée pour apporter la réponse pertinente sous forme de vidéo. Ce type de procédé reste complexe technologiquement parlant de nos jours. J'ai donc opté pour un système de questionnement textuel : les apprenant tapent une question dans une barre de recherche vierge. Un algorithme tente de reconnaître la question et y apporte la réponse qu'il estime appropriée. Il me semblait important d'éviter les méthodes type « choix dans une liste de questions possibles » pour éviter tout biais d'indicage. De la même manière, les apprenants ont la possibilité d'examiner le patient simulé en entrant l'examen qu'ils souhaitent réaliser dans une barre de recherche vierge. Ce mode de questionnement et d'examen suppose de prévoir en amont des scénarios les questions et examens possibles. Dans un souci de simplification, j'ai donc établi une liste des examens cliniques possibles à réaliser dans un contexte de médecine générale ainsi qu'une liste des questions habituelles posées dans les prises en charges courantes en médecine générale. Cette liste est évolutive, les questions et examens inconnus demandés par les apprenants permettent de la compléter au fur et à mesure des scénarios enregistrés. Il est demandé dans la mesure du possible d'interpréter les résultats des examens. Plutôt que d'annoncer le résultat d'une otoscopie, une image de tympan est présentée à l'apprenant. Ce dernier doit alors estimer le caractère normal ou pathologique de cette image. Cette tâche se rapproche des

conditions réelles d'un examen clinique. Au moment de l'écriture de ce document, plus de 800 questions et demandes d'examens différentes sont reconnues et traitées par deux modules de traitement du langage naturel (*chatbots*).

Pour tourner ces vidéos, j'ai opté pour l'embauche de comédiens professionnels ou semi-professionnels afin de renforcer l'authenticité des consultations simulées. J'ai également opté pour l'embauche d'un réalisateur professionnel pour assurer la meilleure qualité possible à la vidéo et à la prise de son. Ce même réalisateur est chargé de découper les vidéos selon la liste des questions et examens. Les informations contextuelles sont apportées de diverses manières au sein du logiciel. Lors du choix du scénario, le métier du professionnel de santé est défini : médecin, infirmier, kinésithérapeute, etc... Pour les médecins, la spécialisation est également à définir : médecine générale, cardiologie, rhumatologie, etc... La première page d'un scénario apporte quelques informations contextuelles supplémentaires sur la rencontre clinique : lieu de la rencontre, moment, contexte professionnel. Les apprenants reçoivent aussi des informations sur la nature de l'autoévaluation formative qu'ils auront à accomplir. La figure n°17 ci-dessous montre une capture d'écran du message d'accueil affiché aux participants à l'ouverture d'un scénario.

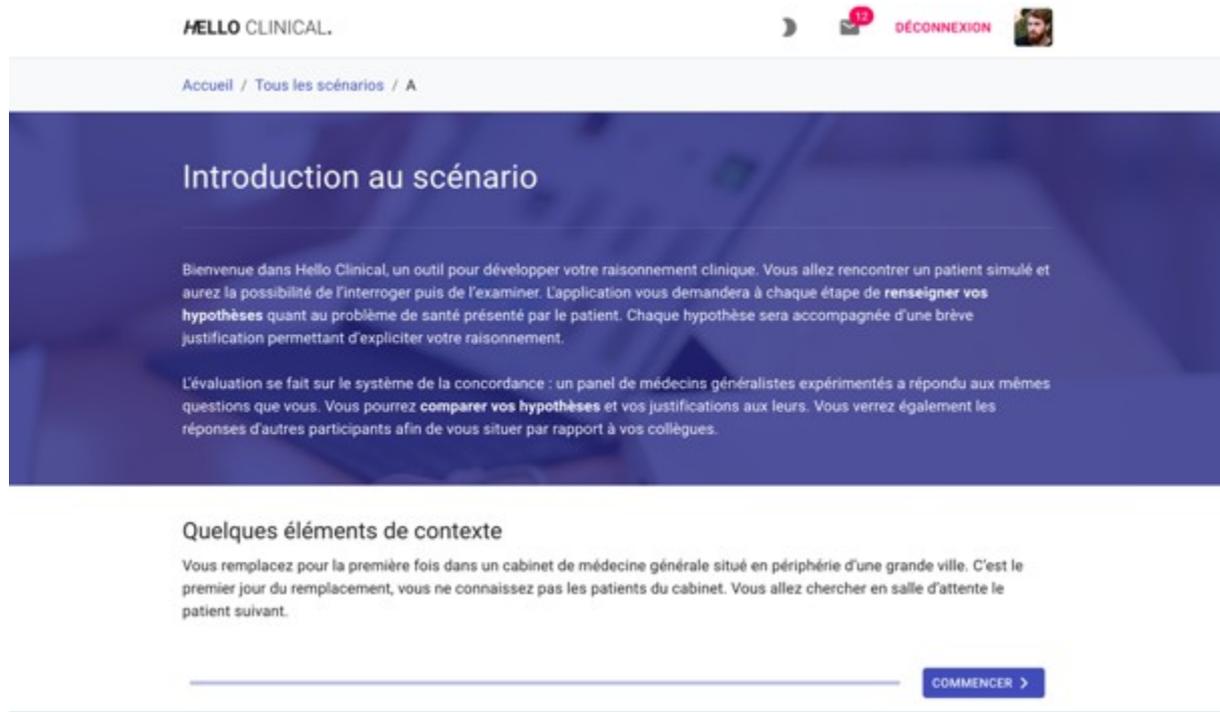


FIGURE 18 : CAPTURE D'ÉCRAN DE LA PAGE D'ACCUEIL D'UN SCÉNARIO DU LOGICIEL HELLO-CLINICAL

Les apprenants accèdent après les premiers recueils et retro-informations à un dossier médical simulé. Ils y trouvent les informations habituelles : données administratives, antécédents, traitements habituels, biométries, historique des consultations. Ce dossier apparaît en même temps que les premières explications du patient sur les raisons de sa consultation dans le cadre d'une rencontre en médecine générale. La figure n°18 ci-dessous illustre ce dossier simulé.

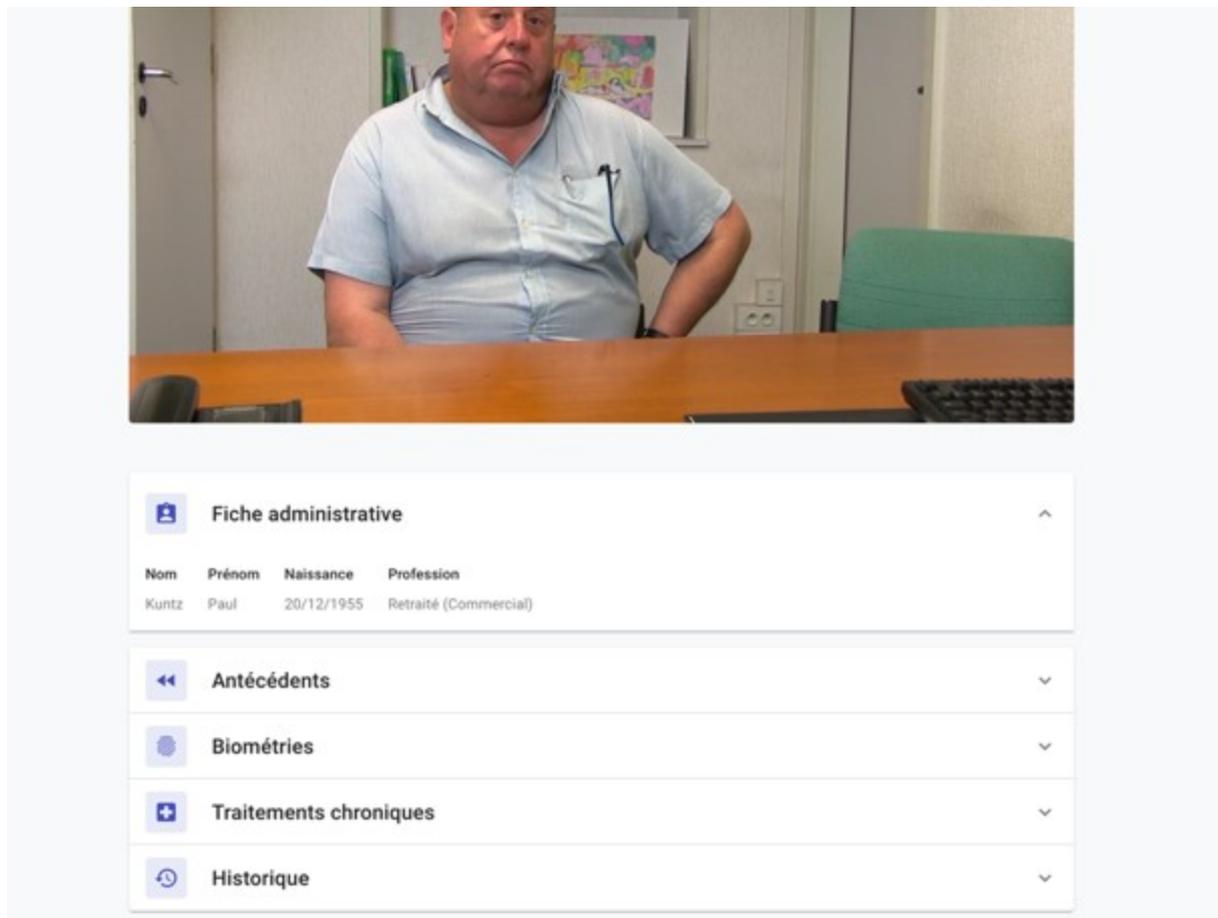


FIGURE 19 : CAPTURE D'ÉCRAN DE L'AFFICHE DU DOSSIER MÉDICAL SIMULÉ

L'écriture des scénarios est basée sur des situations réelles pour qu'ils soient les plus authentiques possibles. Chaque scénario nécessite de définir les éléments du contexte, les caractéristiques du patient, les consignes aux acteurs, un historique de consultation, les différentes réponses à apporter aux différentes questions, les résultats des différents examens possibles. Chaque scénario nécessite ainsi au minimum une à deux heures d'écriture. Les tournages sont en conséquence relativement longs. Chaque scénario a nécessité entre deux et trois heures de tournage.

Ce projet de création d'un outil informatique permettant d'administrer des consultations simulées aux professionnels de santé sur la base du modèle MOT dans une optique d'évaluation formative du processus de raisonnement clinique a reçu un double financement « Initiative d'Excellence » (IdEx) par l'université de Strasbourg : 20 000€ pour le développement de l'outil et la réalisation de scénarios en 2015 puis 3 000€ en 2017 pour compléter le développement.

Ce projet a également obtenu un financement de 36 000€ en juin 2018 sur le programme Samuel de Champlain visant à explorer les utilisations potentielles de l'outil et à initier un partenariat d'expertises entre les Universités de Montréal et de Strasbourg. Le projet fut porté par le Pr Emmanuel Triby et le Pr Thierry Pelaccia pour le volet français et par les Prs Nicolas Hernandez et Joseph-Omer Dyer au Canada. Ce montant a servi à financer des études sur l'outil et à organiser des séjours de travail dans les pays respectifs des partenaires.

Enfin, ayant identifié un certain nombre de limites de l'outil après la première version, un nouveau développement a été entrepris à l'été 2020 avec l'inclusion de *chatbots* pour parfaire le traitement du langage naturel dans les phases de questionnement et d'examen du patient virtuel. Ce projet est en cours de financement à hauteur de 30 000€ par la société d'accélération du transfert de technologie (SATT) Conectus au jour de rédaction de ce document. Le projet consiste à rechercher des partenaires commerciaux afin de financer un développement intégral de l'outil et la création de nouveaux scénarios à des fins de formation et de recherche.

Un certain nombre de limites peuvent être relevées concernant cet outil. Plusieurs données utilisées pour raisonner par les professionnels de santé ne sont pas administrables à l'heure actuelle : odorat, toucher. Le système de questionnement est artificiel au travers de questions tapées. Le système d'examen est lui aussi simulé. Les résultats des examens cliniques liés au toucher ou à l'odorat sont ainsi retranscrits en texte : examen normal, pas de masse à la palpation, etc... Les résultats liés à l'audition sont le plus possible apportés sous format audio + vidéo ou vidéo+ insert de texte (exemple : auscultation cardiaque). Des limitations techniques dans la capture de sons d'auscultations réelles limitent les possibilités de résultats au format audio. Par défaut, un texte accompagne la vidéo de l'examen clinique. Les limitations matérielles concernant possibilités de capture vidéo des examens type inspection pharyngée ou otoscopie nous conduisent à fréquemment présenter des images au lieu de vidéos.

De manière générale, le système de questionnement et d'examen reste artificiel malgré les efforts pour optimiser son authenticité. Plusieurs erreurs de reconnaissances des questions posées dans le *chatbot* actuel grèvent l'authenticité de l'expérience des utilisateurs à l'heure actuelle. Le développement complémentaire et l'entraînement progressif des *chatbots* sur de larges bases de données devraient permettre de résoudre ce problème à l'avenir.

L'écriture et la réalisation de scénarios est chronophage et complexe. Elle nécessite une formation approfondie au fonctionnement de l'outil. Ces points limitent la diffusion de l'outil et son appropriation par les professionnels de santé formateurs.

Le besoin en financement est important, de l'ordre de 500€ par scénario pour une qualité optimale.

Il existe une tension entre une perspective sur raisonnement clinique comme compétence et la standardisation du contexte clinique qu'offre cet outil. Si la standardisation est parfois souhaitable en Sciences de la Santé car porteuse de valeurs comme la sécurité du patient, l'équité et la fidélité des évaluations, les données de la littérature montrent que les scripts cliniques sont hautement idiosyncratiques et que l'évaluation de la compétence raisonnement clinique ne peut s'exonérer de la prise en compte du contexte (10,42,45,46,138). Aussi, cet outil devrait n'être utilisé qu'à visée formative dans les cursus par compétences.

Dans la même lignée, la méthode de questionnement proposée par texte dans cet outil est inauthentique : le *chatbot* reconnaît plus aisément les questions fermées. Les enchaînements de questions ne sont pas encore pris en compte (ex : "vous avez mal ?" – "oui, souvent" – "où ?"). L'apprentissage de l'interrogatoire en particulier et de la communication avec les patients en général ne doit donc pas reposer sur de telles méthodes. Il ne s'agit donc pas d'un outil visant à développer la communication des professionnels de santé mais uniquement le raisonnement clinique. Les apprenants devront être mis en garde sur ces limites.

Plusieurs points méritent cependant d'être mis en avant concernant cet outil et son utilisation potentielle. Il s'agit d'un logiciel en ligne permettant des formations asynchrones. Il permet un accès flexible sur des supports variés : smartphones, ordinateurs et tablettes. Il permet aux apprenants de

gérer leurs temps de formation selon leurs disponibilités. Il s'agit d'un outil novateur parmi les simulateurs de patients virtuels puisqu'il combine ces différents points :

- Il utilise des vidéos d'acteurs ;
- Il propose un système de questionnement et d'examen basé sur un traitement du langage naturel (par opposition au choix d'une question ou d'un examen dans une liste) ;
- Il est le premier à utiliser un système de concordance pour les retro-informations ;
- Il est le seul à être bâti sur un modèle théorique du processus de raisonnement clinique ;
- Il possède une architecture visant à rendre son utilisation universelle. Tout professionnel de santé pourrait en théorie l'utiliser à l'avenir.

Son utilisation en recherche sur le processus de raisonnement clinique permettra de standardiser la situation clinique afin d'en étudier les déterminants dans la prise de décision. C'est d'ailleurs dans cette optique d'aborder la recherche sur le processus de raisonnement clinique qu'a été réalisée l'étude suivante.

#### UN EXEMPLE DE RECHERCHE SUR LE RAISONNEMENT CLINIQUE UTILISANT HELLO-CLINICAL : LIEN ENTRE EXPÉRIENCE ET PRESCRIPTIONS COMPLÉMENTAIRES EN MÉDECINE GÉNÉRALE. UNE ÉTUDE OBSERVATIONNELLE UTILISANT DES PATIENTS VIRTUELS

À la suite du développement de la première version de l'outil Hello-Clinical, j'ai souhaité étudier l'impact de l'expérience sur le raisonnement clinique en situation d'incertitude. Je me suis intéressé à un point débattu dans la littérature : le lien entre expérience et prescriptions d'examens complémentaires en médecine générale.

J'ai donc conçu une étude qui permettrait d'apporter de nouvelles données sur cette question en adoptant un mode de recueil original : un patient virtuel standardisé. Pour ce travail, j'ai collaboré avec Hubert Maisonneuve qui est un confrère médecin généraliste universitaire travaillant à Lyon et à Genève pour concevoir l'étude. Je me suis également entouré de mes deux directeurs de thèse Emmanuel Tribby et Nicolas Fernandez, de mon ancien directeur Bernard Charlin et de Thierry Pelaccia afin de bénéficier de leurs expertises méthodologiques et de leurs conseils sur la rédaction de l'article. Enfin, deux internes m'ont aidé sur le recueil de données et la revue de la littérature (Romain Krider et Yunus Can) et François Lefevbre a réalisé les analyses statistiques.

## Contexte de l'étude

Développer la tolérance à l'incertitude chez les futurs médecins est actuellement l'un des principaux défis dans la formation des Médecins (36). En situation d'incertitude, les médecins généralistes proposent parfois des prescriptions complémentaires (PC), telles que des tests biologiques ou des examens radiologiques, ou orientent les patients vers un autre spécialiste afin d'aider à la prise de décision (45). La gestion optimale des PC est un défi notamment en termes d'efficacité des soins et d'utilisation des ressources du système de santé (139). Cela est particulièrement important en médecine générale, où au moins 20% des consultations concernent des situations d'incertitude (140,141). L'expérience semble être un déterminant important du recours aux PC (139,142-144).

Dans les études traitant des liens entre l'expérience et les PC en situation d'incertitude, les données ont été obtenues à l'aide de questionnaires (145), d'analyses rétrospective de bases de données (146) ou d'entretiens avec des médecins généralistes (142-144). À notre connaissance, aucune étude n'a été réalisée dans des conditions standardisées avec une grande authenticité. Ainsi, afin de mieux comprendre le lien entre l'expérience et les PC dans la prise de décision en situation d'incertitude, nous pensons que l'utilisation d'une situation standardisée, présentée par des patients virtuels, est une approche intéressante.

L'objectif de cette étude était d'explorer la relation entre l'expérience et les PC en situation d'incertitude, en utilisant des consultations virtuelles de médecine générale.

## Matériel et méthodes

Nous avons mené une étude observationnelle sur l'association entre l'expérience et les PC dans un groupe de médecins généralistes. Nous avons choisi un seuil de dix ans d'expérience basé sur la littérature sur le développement de l'expertise pour dichotomiser notre échantillon (19,20).

Nous avons recruté des médecins généralistes travaillant dans la région du Grand Est de la France. Nous avons stratifié le recrutement des participants pour inclure autant d'hommes que de femmes afin de refléter les proportions observées parmi les médecins généralistes en France. Nous avons recruté autant de médecins ayant plus et moins de 10 ans d'expérience. Nous avons utilisé la méthode de la boule de neige pour le recrutement : chaque participant était invité, après la collecte des données, à donner le nom, le numéro de téléphone ou l'e-mail d'un à trois autres médecins généralistes qui pourraient accepter de participer à l'étude.

Nous avons utilisé via Hello-Clinical le scénario d'un homme de 69 ans souffrant d'une dyspnée depuis plusieurs jours. Ce scénario correspondait à une forme atypique d'embolie pulmonaire.

Nous avons collecté le sexe des participants, leur âge, le nombre d'années d'expérience, le lieu d'exercice (rural/semi-rural/urbain), le type de cabinet (seul/association), le nombre de patients "médecin traitant" et la durée moyenne des consultations. Nous avons mesuré le niveau d'anxiété des participants à l'aide de la dimension anxiété d'état de l'échelle STAI-Y (*State-Trait Anxiety Inventory*) dans sa version française (147). Nous avons recueilli les réponses à la version française du questionnaire *Physician Reaction to Uncertainty* (PRU) (148,149). Les hypothèses diagnostiques et leurs justifications ont été recueillies à l'étape finale du cas clinique. La sévérité de ces hypothèses a été évaluée sur une échelle de 1 à 5 sur la base de la classification clinique en médecine d'urgence (CCMU). Enfin, les PC à l'issue de la consultation ont été enregistrées, regroupées et chiffrées selon la tarification française en vigueur (CNAMTS).

En posant l'hypothèse empirique selon laquelle les généralistes ayant plus de 10 ans d'expérience auraient recours à un nombre moyen de PC de 5 (écart-type : 2,5) et les généralistes ayant moins de 10 ans d'expérience un nombre moyen de PC de 10 (écart-type : 6,0), nous avons dû inclure 20 généralistes dans chaque groupe pour obtenir une puissance de 90 % avec une erreur de type I de 5 %.

## Résultats

Quarante médecins généralistes ont participé à l'étude entre juillet et septembre 2017. La moitié d'entre eux étaient des femmes. L'âge moyen des participants était de 43 ans [min 28-max 68, déviation standard (DS) : 13,8], l'âge moyen était de 37 ans. Le nombre moyen d'années d'expérience était de 15,1 (min 1-max 40 ; DS : 13,6). La moitié des participants avaient plus de dix ans d'expérience. Les participants avaient en moyenne 451 patients médecin traitant (min 0 - max 2000 ; DS : 473,5). Quatorze participants étaient des remplaçants et n'avaient donc pas de patients médecin traitant. Le nombre moyen de patients médecin traitant des participants installés était de 695 (écart-type : 417,1).

Le score moyen au STAI-Y était de 37 (IC 95% : 34,2 ; 39,8 et DS : 8,6), indiquant un faible niveau d'anxiété. Dix-sept participants présentaient une anxiété minimale (42,5 %), 17 une anxiété faible (42,5 %) et 6 une anxiété légère (15 %). Le score PRU moyen était de 17,6 (écart-type : 6,9 et IC 95% : 15,4 ; 19,8). Les participants ont évalué l'incertitude du scénario à 3,1/5 (DS 0,9) et son authenticité à 4,4/5 (DS 0,8).

Les participants ont proposé en moyenne 3 hypothèses de diagnostic (écart-type : 0,920, min 2 - max 5), avec un score CCMU moyen de 2,5 (écart-type : 0,413, min 1,5 - max 3,7). Ils ont écrit une moyenne de

10,3 CP (min 1 - max 16 ; DS 3,6), avec 34 propositions différentes. Le prix moyen des CP par participant était de 141,80 € (min 17,50 - max 440,50 ; DS 81,6). Les détails de notre population sont présentés dans le tableau n°8.

Le groupe ayant moins de 10 ans d'expérience avait en moyenne moins de patients médecin traitant (158 contre 745 ;  $p < 0,001$ ) et des scores STAI-Y et PRU plus élevés (respectivement 32,7 contre 41,3 ;  $p < 0,001$  et 14,4 contre 20,7 ;  $p=0,002$ ).

Aucune différence statistiquement significative n'a été observée concernant le sexe, le lieu et le type de pratique, les notes d'incertitude et d'authenticité et la durée moyenne des consultations. Les détails des comparaisons entre les groupes sont présentés dans le tableau n°8.

	Total n (%)	< 10 ans d'expérience	> 10 ans d'expérience	Valeur du p
Age moyen (années)	43	31.3	54.9	p <0.001
Femme	20 (50)	7	13	p = 0.060
Expérience moyenne (années)	15.1	3.6	26.5	p <0.001
Type de cabinet	40 (100)			p =0.520
Seul	26 (65)	14	12	
Association	14 (35)	6	8	
Lieu d'exercice	40 (100)			p =0.340
Urbain	25 (62.5)	11	14	
Semi-urbain	15 (37.5)	9	6	
Nombre de patients médecin traitant	451	158	745	p <0.001
Durée moyenne des consultations	40 (100)			p = 1
< 10 min	2 (5)	1	1	
10-20 min	32 (80)	16	16	
20-30 min	6 (15)	3	3	
Score d'incertitude	3.1	3.3	2.9	p = 0.232
Score d'authenticité	4.4	4.50	4.30	p = 0.442
Score STAI-Y moyen	37	41.3	32.7	p = < .001
Score PRU moyen	17.6	20.7	14.4	p = 0.002
Nombre de prescriptions complémentaires	10.3	11.5	9.2	p = 0.0285
Prix moyen des prescriptions complémentaires (euros)	141.8	153.8	129.6	p = 0.356
Nombre d'hypothèses diagnostiques	3.0	3.1	2.9	p = 0.612
Score CCMU	2.5	2.5	2.5	p = 0.798

**TABLEAU 8 : ANALYSE UNIVARIÉE DES CARACTÉRISTIQUES DES GROUPES**

Les participants ayant moins de 10 ans d'expérience professionnelle ont eu recours en moyenne à 10,2 PC [ET 3,4 - IC95% : 8,9 ; 11,7] contre 8,1 [ET 3,7 - IC95% : 6,5 ; 9,9] pour les participants ayant plus de 10 ans d'expérience (p = 0,0285) en analyse univariée. La taille de l'effet était de 0,603 (-0,052 ;

1,257). Aucune différence statistiquement significative sur le prix des PC n'a été trouvée dans l'analyse univariée (153,84 - SD 91,4 contre 129,67 - DS 70,4 ;  $p=0,2766$  avec une taille d'effet de 0,296 [-0,348 ; 0,939]).

En analyse multivariée, il y avait une association significative sur le nombre de PC avec l'âge (risque relatif : 0,9770828, 95%CI : 0,9606 à 0,994 ;  $p$  ajusté=0,00745) et le score moyen STAI-Y (RR : 0,984, 95%CI : 0,968 à 0,999 ;  $p$  ajusté=0,0351) en considérant l'expérience comme catégorielle (plus ou moins de 10 ans).

Nous avons donc effectué une autre analyse multivariée sur le nombre de PC en considérant l'expérience comme continue avec ces résultats sur l'âge. L'expérience était alors significativement associée au nombre de PC (RR : 0,977, 95%CI : 0,965 à 0,988 ;  $p<0,0001$  ajusté). Le nombre de patients réguliers en cabinet était également associé de manière significative au nombre de PC (RR 1,000, IC95 % : 1,000 à 1,001 ;  $p$  ajusté = 0,0486).

## Discussion

Comme le suggèrent certaines données de la littérature (139,142-144), une plus grande expérience semble être associée à une diminution du nombre de PC en médecine générale. C'est la seule association qui a persisté dans l'analyse multivariée. Cette association pourrait être liée à une tolérance croissante de l'incertitude au cours de la pratique de la médecine générale : plus les médecins généralistes sont confrontés à l'incertitude, plus ils reçoivent de rétro-information sur le manque d'utilité des PC dans certaines situations d'incertitude (36,142).

Une autre explication possible de cette association est que les jeunes générations de professionnels de la santé s'appuient davantage sur les PC que sur la démarche clinique pour poser un diagnostic (150). Les jeunes générations sont peut-être mieux formées aux recommandations de bonne pratique, tandis que les médecins généralistes expérimentés sont peut-être mieux formés pour identifier les limites des recommandations.

La pertinence de l'association entre l'expérience et les PC dans notre étude mérite d'être nuancée. La diminution du nombre de PC ne s'est pas accompagnée d'une diminution significative de leur coût entre nos deux groupes. Cela peut être dû à une puissance insuffisante de notre étude ou signifier que la réduction du nombre de PC ne se traduit pas nécessairement par des économies pour le système de santé.

L'enseignement du raisonnement clinique devrait inclure des sessions sur des cas cliniques impliquant la décision de faire des PC ou non en situation d'incertitude (151). Un effort devrait également être fait pour améliorer les pratiques des superviseurs en troisième cycle visant à offrir un retour d'expérience spécifique pour renforcer la tolérance à l'incertitude des étudiants (36,142). Ce type d'interventions pourrait contribuer à renforcer la tolérance à l'incertitude, ce qui permettrait un meilleur recours aux PC en médecine générale.

L'article intégral dans la version soumise à la revue *Medical Education* se trouve ci-dessous.

### **Experience and complementary prescriptions under uncertainty in general practice - An observational study using virtual patients**

Mathieu Lorenzo<sup>a,b,c</sup>, Thierry Pelaccia<sup>b</sup>, Emmanuel Triby<sup>c</sup>, François Lefebvre<sup>d</sup>, Yunus Can<sup>a</sup>, Romain Krider<sup>a</sup>, Bernard Charlin<sup>e</sup>, Nicolas Fernandez<sup>e</sup>, Hubert Maisonneuve<sup>f</sup>

<sup>a</sup> Department of General practice, Faculty of Medicine, Strasbourg, France

<sup>b</sup> Centre for Training and Research in Health Sciences Education (CFRPS), Faculty of Medicine, Strasbourg, France

<sup>c</sup> Faculty of education and lifelong learning (ESPE), Strasbourg, France

<sup>d</sup> Public Health Department, Faculty of Medicine, Strasbourg, France

<sup>e</sup> Centre for Applied Pedagogy in Health Sciences (CPASS), University of Montreal, Montreal, Canada

<sup>f</sup> Primary care unit, Faculty of medicine, University of Geneva, Geneva, Switzerland

Correspondence: Dr M. Lorenzo – Department of General practice, 4 rue Kirschleger, 67000 Strasbourg, France - [mlorenzo@unistra.fr](mailto:mlorenzo@unistra.fr)

Medical Education



### **Experience and complementary prescriptions under uncertainty in general practice**

Journal:	<i>Medical Education</i>
Manuscript ID	MED-2021-0984
Manuscript Type:	Original Research
Keywords:	Computers, Simulation, new technology, Primary Care Education, Evaluation/assessment of Clinical Performance, Personal Characteristics/Attitudes
Research Method:	Quantitative Research Methods

SCHOLARONE™  
Manuscripts

1  
2  
3 **Experience and complementary prescriptions under uncertainty in general practice - An observational study**  
4  
5 **using virtual patients**  
6

7  
8 **Abstract**  
9

10  
11 **Context:** In situations of uncertainty, general practitioners (GPs) sometimes order complementary prescriptions  
12 (CPs), such as biological testing, radiological analyses, or offer to refer patients to another specialist, to help  
13 decision-making.  
14

15  
16  
17 **Objective:** To explore the relation between experience and CPs in uncertain situations in general practice.  
18

19  
20 **Methods:** We conducted an observational study using virtual patients to compare CPs in a group of GPs with  
21 over 10 years of experience vs. a group with under 10 years of experience. We collected the number and costs  
22 of CPs in the virtual patient case, years of experience, STAI-Y anxiety score and Physician Reaction to Uncertainty  
23 (PRU) scores.  
24  
25  
26  
27

28  
29 **Results:** Participants with under 10 years of professional experience ordered more CPs on average (10.2 [IC95%:  
30 8.9; 11.7] vs. 8.1 [IC95%: 6.5; 9.9];  $p = 0.0285$ ). There was no with no statistically significant difference on costs  
31 (€129.67 vs €153.84;  $p=0.2766$ ). The group with under 10 years of experience had significantly higher STAI-Y and  
32 PRU scores than the group with over ten years of experience (respectively 32.7 vs. 41.3;  $p < 0.001$  and 14.4 vs  
33 20.7;  $p=0.002$ ). Experience was meaningfully associated with the number of CPs in multivariate analysis (RR:  
34 0.977, 95%CI: 0.965 to 0.988; adjusted  $p < 0.001$ ).  
35  
36  
37  
38  
39  
40  
41

42 **Conclusions:** Greater experience appears to be associated with a decrease in the number of CPs in general  
43 practice. This association could be linked to growing tolerance of uncertainty in the course of practicing medicine:  
44 the more GPs face uncertainty, the more they get feedback on the lack of usefulness of CPs in some situations  
45 of uncertainty. An effort could be made in offering specific feedback on experience aimed at building up  
46 tolerance of uncertainty during medical training. This type of interventions might help building up tolerance of  
47 uncertainty resulting in best-use of CPs in general practice.  
48  
49  
50  
51  
52  
53

54  
55 **Word count:** 2732  
56  
57  
58  
59  
60

## Introduction

Building up tolerance of uncertainty in future practitioners is currently one of the main challenges in medical education.<sup>1</sup> Uncertainty can be defined as the “conscious, metacognitive awareness of ignorance”.<sup>2</sup> In situations of uncertainty, general practitioners (GPs) sometimes order complementary prescriptions (CPs), such as biological testing, radiological analyses, or offer to refer patients to another specialist, to help decision-making.<sup>3</sup> Optimal management of CPs is an objective for cost-effectiveness.<sup>4</sup> Some authors argue that the younger generations of practitioners use too many CPs.<sup>5,6</sup> This is a particularly important challenge in general practice, where at least 20% of consultations are in situations of uncertainty.<sup>7,8</sup>

Experience appears to be a key determinant in CPs ordering in several studies.<sup>4,9-11</sup> In addition to experience, several factors have been singled out in the literature for their possible influence on CPs ordering in situations of uncertainty. Some relate to conditions of practice, such as the time available for each patient or the availability of tests<sup>9-12</sup> or the existence and provision of forms for requesting tests.<sup>13</sup> Others include societal factors, such as the fear of possible prosecution or the low cost of tests in some health care systems,<sup>11</sup> practitioners' clinical reasoning,<sup>9,11,13-17</sup> practitioners' emotional reaction to uncertainty, fear of regrets and anxiety,<sup>9,11,13-15,18</sup> influence of the patient,<sup>7,9,11</sup> practitioners' expectations of performance<sup>11</sup> or practitioners' gender<sup>12</sup>. Anxiety in particular, had been showed to be correlated with CPs ordering in uncertainty.<sup>9,15</sup>

In studies addressing the links between experience and CPs in situations of uncertainty, data were obtained using either questionnaires,<sup>15</sup> database analyses<sup>12</sup> or interviews with general practitioners.<sup>9-11</sup> To our knowledge, no study has been carried out under standardized conditions with high authenticity. Thus, in order to gain a better understanding of the connection between experience and CPs in decision making in situations of uncertainty, we believe that using a standardized situation, presented by virtual patients, is a sound approach.

The objective of this study was to explore the relation between experience and CPs in uncertain situations, using virtual general practice consultations.

## Methods

We conducted an observational study to compare the association between experience and CPs in a group of GPs. We selected a ten years threshold based on literature on development of expertise to dichotomize our sample.<sup>19,20</sup>

1  
2  
3 *Measurement tool*  
4  
5  
6 At the University of X, we have developed a computer tool (Hello-Clinical) that makes it possible to simulate  
7  
8 virtual consultations to assess the decision-making process of health care professionals for training or research  
9  
10 purposes under standardized conditions.

11  
12  
13 We divided our virtual clinical consultations into stages according to the modeling using typing objects (MOT)  
14  
15 model of clinical reasoning,<sup>3</sup> to collect the student's diagnostic hypotheses in an open text field without  
16  
17 suggestions for responses. MOT model is a hierarchical model depicting the multifaceted processes of clinical  
18  
19 reasoning.<sup>3</sup> The virtual patients were presented in video recordings. After the virtual consultation, students could  
20  
21 write in CPs they deem appropriate in an open text field.

22  
23  
24 *The population under study*  
25  
26  
27 We recruited GPs working in X without randomization. We stratified participant recruitment to include as many  
28  
29 men as women in order to mirror the proportions observed among GPs in X. We recruited as many GPs with over  
30  
31 and under 10 years of experience. We used a snowball method for recruitment: each participant was asked, after  
32  
33 data collection, to give the name, the phone number or e-mail of one to three other GPs who might agree to  
34  
35 participate in the study. We started by reaching by mail two GPs X knew. Potential future participants were then  
36  
37 reached by phone or e-mail. Oral consent was obtained before the interview for each participant. No exclusion  
38  
39 criteria were defined. At first, the only inclusion criteria were to be an active GP. When we included a few GPs,  
40  
41 we had to stratify our sample on sex and experience. We then specifically looked for future potential participants  
42  
43 based on sex or experience needs for stratification.

44  
45  
46 *Virtual clinical case*  
47  
48  
49 We used the script of a 69-year-old man consulting his GP for dyspnea that had progressed for several days: an  
50  
51 atypical case of pulmonary embolism. The clinical picture was deliberately vague, and it was not possible to reach  
52  
53 this diagnosis with certainty in light of relevant diagnostic criteria. The goal was to contextualize the participants'  
54  
55 clinical reasoning by attempting to reproduce an authentic context as closely as possible. The patient was a  
56  
57 retired physics and chemistry teacher. Details of main characteristics are given in Table 1. His last consultation  
58  
59 logged in the record had taken place five months previously, to reassess his antihypertensive treatment.  
60

1  
2  
3 Participants were able to familiarize themselves with the software beforehand by tackling a tutorial script on a  
4  
5 60-year-old man with recent, unexplained asthenia.  
6

#### 7 8 *Variables collected* 9

10  
11 A review of the literature enabled us to identify collectable variables influencing ordering of CP in situations of  
12  
13 uncertainty: participants' gender, age, number of years of experience, place of practice, type of practice, average  
14  
15 number of in-office patients and average duration of consultations.<sup>7,9-15</sup> Experience was defined as the number  
16  
17 of years of practice since graduation. Place of practice (urban, rural or semi-urban), type of practice (alone or  
18  
19 association), number of regular in-office patients and average duration of consultation were related to practice  
20  
21 conditions.<sup>9-12</sup> These data were collected in a questionnaire submitted after the simulated consultations. No  
22  
23 confounding patient-related factors were collected, given the use of a standardized patient. Test availability was  
24  
25 set to maximum in this virtual clinical case. We did not collect data on the practitioners' expectation of  
26  
27 performance or societal factors. We measured the participants' anxiety levels using the State-Trait Anxiety  
28  
29 Inventory (STAI-Y) scale<sup>21</sup> in its French version, observing only the state-anxiety dimension to save time. We  
30  
31 chose this instrument as it is a valid anxiety measurement tool used worldwide with a validated French  
32  
33 version.<sup>22,23</sup> It has been employed in GPs trainees.<sup>24</sup> We gathered responses to the French version of the  
34  
35 Physician Reaction to Uncertainty (PRU) questionnaire.<sup>25,26</sup> This questionnaire has been used previously in GPs  
36  
37 populations.<sup>26,27</sup> We then asked participants to rate the uncertainty and authenticity of the virtual clinical case  
38  
39 from 1 to 5.  
40

41  
42 Diagnostic hypotheses and their justifications were gathered at the final stage in the virtual clinical case. The  
43  
44 severity of these hypotheses was assessed on a 1-to-5 scale on the basis of the xx.  
45

46  
47 Lastly, CPs at the end of the consultation were recorded, grouped and cost-related under the xx.  
48

#### 49 50 *Statistical analysis* 51

52  
53 Given our empirical hypothesis that GPs with over 10 years of experience would order a mean number of CPs of  
54  
55 5 (SD: 2.5) and the GPs with under 10 years of experience a mean number of CP of 10 (SD: 6.0), we had to include  
56  
57 20 GPs in each group to obtain a 90 % power with a 5% type I error with a ratio 1:1. We used a two-sample t-test  
58  
59 for the power analysis.  
60

1  
2  
3 To conduct the descriptive analysis of quantitative variables, for each variable we compiled position parameters  
4 (average, mean, minimum, maximum, first and third quartiles) and dispersion parameters (variance, standard  
5 deviation, range, midspread). To describe qualitative variables such as gender or type of practice, we compiled  
6  
7 the headcounts and proportions for each modality in the sample.  
8  
9

10  
11 Comparisons of quantitative variables between groups were performed using either a Poisson regression model  
12 for counting variables, a linear model for gaussian variables or a nonparametric Mann-Whitney-Wilcoxon test  
13 for non-gaussian variables. For comparisons between more than two groups, we relied either on analysis of  
14 variance (for the Gaussian data) or the Kruskal-Wallis test (non-Gaussian data). The correlation between the  
15 quantitative variables was measured using the Pearson correlation coefficient or the Spearman correlation  
16 coefficient.  
17  
18

19  
20 A multivariate analysis was performed with all variables. We used a stepwise selection method based on the  
21 minimization of Akaike's information criterion. Analyses were performed using R software, release 3.2.2 (The R  
22 project). Full data are available on demand to the corresponding author.  
23  
24

## 25 **Results**

### 26 *Population*

27  
28 Forty general practitioners participated in the study between July and September 2017. Half of the participants  
29 were women. The average age of participants was 43 [min 28-max 68, SD: 13,8], the mean age was 37. The  
30 average number of years of experience was 15.1 (min 1-max 40; SD: 13.6). Half of the participants had over ten  
31 years of experience. Participants had on average 451 regular patients (min 0 – max 2,000; SD: 473.5). Fourteen  
32 participants were substitute general practitioners and as such had no regular patients. The average number of  
33 regular in-office patients for established participants was 695 (SD 417.1).  
34  
35

36  
37 The average STAI-Y score was 37 (IC 95%: 34.2; 39.8 and SD: 8.6), indicating a low level of anxiety. Seventeen  
38 participants presented minimal anxiety (42.5%), 17 low anxiety (42.5%), and 6 mild anxiety (15%). The average  
39 PRU score was 17.6 (SD: 6,9 and IC 95%: 15.4; 19.8). Participants rated the script's uncertainty at 3.1/5 (SD 0.9)  
40 and its authenticity at 4.4/5 (SD 0.8).  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Participants proposed an average of 3 diagnostic hypotheses (SD: 0,920, min 2 – max 5), with an average CCMU  
4 score of 2.5 (SD: 0.413, min 1.5 – max 3.7). They wrote an average of 10.3 CPs (min 1 – max 16; SD 3,6), with 34  
5 different proposals. The average price of CPs per participant was €141.80 (min 17.50-max 440.50; SD 81,6).  
6  
7  
8  
9 Details of our population are given in Table 2.

10  
11  
12 The group with under 10 years of experience had on average fewer regular patients (158 vs. 745;  $p < 0.001$ ) and  
13 higher STAI-Y and PRU scores (respectively 32.7 vs. 41.3;  $p < .001$  and 14.4 vs 20.7;  $p = 0.002$ ).

14  
15  
16  
17 No statistically significant difference was observed regarding gender, place and type of practice, uncertainty and  
18 authenticity ratings and average duration of consultations. Details of comparisons between groups are given in  
19  
20  
21 Table 2.

#### 22 23 24 *Primary endpoint*

25  
26  
27 Participants with under 10 years of professional experience ordered an average of 10.2 tests [SD 3.4 - IC95%: 8.9;  
28 11.7] vs. 8.1 [SD 3.7 - IC95%: 6.5; 9.9] for participants with over 10 years of professional experience ( $p = 0.0285$ )  
29 in univariate analysis. The effect size was 0.603 (-0.052; 1.257). No statistically significant difference on price of  
30  
31  
32 CPs was found in univariate analysis (153.84 - SD 91.4 versus €129.67 – SD 70.4;  $p = 0.2766$  with effect size 0.296  
33 [-0.348; 0.939]).

34  
35  
36  
37  
38 In multivariate analysis, there was a meaningful association on the number of CPs with age (relative risk:  
39 0.9770828, 95%CI: 0.9606 to 0.994; adjusted  $p = 0.00745$ ) and mean STAI-Y score (RR: 0.984, 95%CI: 0.968 to  
40  
41  
42 0.999; adjusted  $p = 0.0351$ ) considering the experience as categorical (more or less than 10 years).

43  
44  
45 We thus performed another multivariate analysis on the number of CPs considering the experience as continuous  
46  
47  
48 with these results on age. Experience was then meaningfully associated with the number of CPs (RR: 0.977,  
49 95%CI: 0.965 to 0.988; adjusted  $p < 0.0001$ ). Number of regular in-office patients was also meaningfully associated  
50  
51  
52 with the number of CPs (RR 1.000, 95%CI: 1.000 to 1.001; adjusted  $p = 0.0486$ ). Details of initials models are in  
53  
54  
55 supplementary materials and finals models in table 3.

#### 56 57 58 59 60 *Secondary endpoints*

1  
2  
3 Consultations lasting 20 to 30-minutes increased the cost of tests compared to 10 to 20-minute-long  
4 consultations (+92€, 95%CI: 19 to 164; adjusted p=0.00799). The number of hypothesis was also meaningfully  
5 associated with a higher cost of CPs (+31€, 95%CI: 3 to 59; adjusted p= 0.0318).  
6  
7

8  
9  
10 No meaningful association was found between the uncertainty rating and the number of diagnostic hypotheses,  
11 the duration of consultations or the STAI-Y, PRU or CCMU scores.  
12  
13

#### 14 **Discussion**

15  
16  
17 The group with under 10 years of experience had significantly higher STAI-Y and PRU scores than the group with  
18 over ten years of experience. The number of years of experience was associated with a lower number of CPs,  
19 without significant effects on costs. This association persisted in the multivariate analysis.  
20  
21  
22

23  
24 To our knowledge, this study is the first to test the links between CPs and experience on virtual patients in  
25 situations of uncertainty. With this tool, we were able to offer a standardized clinical situation to participants.  
26  
27

28  
29 Several limitations should be noted however: the authenticity of the scripted case is limited by the extent of the  
30 software's possibilities and current technical constraints. The STAI-Y score only included the measurement of  
31 state-anxiety (not trait-anxiety). CPs were declarative and may not correctly predict prescriptions in real life.  
32 Writing down hypotheses may be a bias in clinical reasoning. Experience was only defined by years of practice  
33 and correlation in medicine had been showed to be weak between gross measures of expertise as years of  
34 experience and performance.<sup>19</sup> Qualitative insights on experience could have enriched our comprehension.  
35 Snowball sampling method used may be a bias for our results: it increased the chance of finding associations in  
36 this population that do not exist in general population.<sup>28</sup> Further studies may consider use a random sampling  
37 method. We chose to focus on only a portion of the potential determinants of CPs in our data collection for  
38 feasibility reasons.  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

50  
51 As some literature data suggest,<sup>4,9-11</sup> greater experience appears to be associated with a decrease in the number  
52 of CPs in general practice. This was the only association that persisted in the multivariate analysis. This  
53 association could be linked to growing tolerance of uncertainty in the course of practicing medicine: the more  
54 GPs face uncertainty, the more they get feedback on the lack of usefulness of CPs in some situations of  
55 uncertainty.<sup>1,9</sup> Feedback on CPs is provided with practice, seeing that few of these prescriptions were useful, or  
56  
57  
58  
59  
60

1  
2  
3 by trained educators highlighting uncertainty in general practice to students.<sup>1</sup> This is in line with  
4  
5 recommendations on the management of medically unexplained symptoms in general practice,<sup>29</sup> which call for  
6  
7 avoiding multiple investigations. These data are also consistent with script theory in clinical reasoning.<sup>3,30</sup> As  
8  
9 experience is conducive to the optimization of the knowledge organization among GPs, patient care proposals  
10  
11 are less dispersed, and fewer additional prescriptions are ordered. A study designed to measure the evolution of  
12  
13 the duration of consultations with experience, exploring the participants' clinical scripts, could be helpful to  
14  
15 support this hypothesis.

16  
17  
18 Another possible explanation for this association is that young generations of health care professionals rely more  
19  
20 on tests than on clinical examination to perform diagnoses.<sup>5</sup> Young generations may be better trained in  
21  
22 guidelines whereas experienced GPs may be better trained to identify the limits of the guidelines.

23  
24  
25 The relevance of the association between experience and CPs in our study deserves nuancing. The lower number  
26  
27 of CPs ordered did not come with a significant decrease in their cost between our two groups. This may be caused  
28  
29 by an insufficient power of our study or may mean that reducing the number of CPs ordered may not necessarily  
30  
31 result in savings for the healthcare system.

32  
33  
34 The decrease in the cost of additional prescriptions observed with shorter consultations should be discussed in  
35  
36 light of the study conducted by Bugter-Maessen *et al.* in the Netherlands in 1996.<sup>10</sup> The authors posited that  
37  
38 experienced GPs worked faster.<sup>10</sup> In our study, no such association with shortened duration of consultations and  
39  
40 experience was found.

41  
42  
43 An hypothesis for the association between short consultations and the lower cost of CPs is that GPs who do short  
44  
45 consultations minimize uncertainty in their practice.<sup>31</sup> Bloy's work has shown that some GPs give little  
46  
47 consideration to the patients' complaints and symptoms to reduce uncertainty in their consultations.<sup>31</sup> The  
48  
49 absence of correlation between the duration of consultations and uncertainty ratings in our study does not  
50  
51 support this hypothesis. A qualitative study designed to investigate the clinical reasoning process of GPs in a  
52  
53 similar protocol could be useful to explore this hypothesis.

54  
55  
56 Unlike the study by Allison *et al.* in a hospital setting,<sup>15</sup> we found no significant association between the STAI-Y  
57  
58 score and the cost of test orders. Our results are, however, consistent with those of Zaat *et al.*,<sup>13</sup> who found that  
59  
60 the anxiety level had no effect on test ordering in a panel of Dutch GPs.<sup>13</sup>

1  
2  
3 Further studies should test the link between experience and CPs from the angle of the development of  
4 participants' clinical knowledge organization. The link between the costs of CPs and experience should also be  
5 investigated in studies with greater statistical power.  
6  
7

8  
9  
10 The teaching of clinical reasoning should include practical work on clinical cases involving the decision to make  
11 CPs in uncertain situations.<sup>32</sup> An effort should also be made to improve the practices of post-graduated  
12 supervisors by offering specific feedback on experience aimed at building up tolerance of uncertainty.<sup>1,9</sup> This type  
13 of interventions might help building up tolerance of uncertainty resulting in best-use of CPs in general practice.  
14  
15  
16  
17  
18  
19  
20

21  
22 **Additional information**

23  
24 **Authors contributions:** X

25  
26  
27 All authors read and approved the final manuscript.  
28

29  
30 **Data sharing statement:** data sets are available on demand to the corresponding author.  
31

32  
33 **Ethical approval:** We declared this study to the xx and its protocol was approved by the Ethics Committee of X  
34 on 20 June 2017.  
35

36  
37  
38 **Acknowledgements:** X  
39

40  
41  
42  
43 **Funding statement:** blinded for review.  
44

45  
46 **Keywords:** Clinical Decision-Making; Uncertainty; Patient Simulation; General Practice Anxiety; Diagnostic Tests;  
47 Referral and Consultation  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## References

1. Simpkin AL, Schwartzstein RM. Tolerating Uncertainty — The Next Medical Revolution? *N Engl J Med*. 2016;375(18):1713-1715. doi:10.1056/NEJMp1606402
2. Smithson null. Conflict Aversion: Preference for Ambiguity vs Conflict in Sources and Evidence. *Organ Behav Hum Decis Process*. 1999;79(3):179-198. doi:10.1006/obhd.1999.2844
3. Charlin B, Lubarsky S, Millette B, et al. Clinical reasoning processes: unravelling complexity through graphical representation. *Med Educ*. 2012;46(5):454-463. doi:10.1111/j.1365-2923.2012.04242.x
4. Logan RL, Scott PJ. Uncertainty in clinical practice: implications for quality and costs of health care. *Lancet Lond Engl*. 1996;347(9001):595-598.
5. Feddock CA. The lost art of clinical skills. *Am J Med*. 2007;120(4):374-378. doi:10.1016/j.amjmed.2007.01.023
6. Bordage G. Where are the history and the physical? *CMAJ Can Med Assoc J*. 1995;152(10):1595-1598.
7. van der Weijden T, van Velsen M, Dinant G-J, van Hasselt CM, Grol R. Unexplained Complaints in General Practice: Prevalence, Patients' Expectations, and Professionals' Test-Ordering Behavior. *Med Decis Making*. 2003;23(3):226-231. doi:10.1177/0272989X03023003004
8. Kassirer JP. Our stubborn quest for diagnostic certainty. A cause of excessive testing. *N Engl J Med*. 1989;320(22):1489-1491. doi:10.1056/NEJM198906013202211
9. Lorenzo M. Tolérance à l'incertitude en médecine générale: une démarche de théorisation ancrée à partir de 14 entretiens. Published online 2013.
10. Bugter-Maessen AM, Winkens RA, Grol RP, et al. Factors predicting differences among general practitioners in test ordering behaviour and in the response to feedback on test requests. *Fam Pract*. 1996;13(3):254-258.
11. van der Weijden T, van Bokhoven MA, Dinant G-J, van Hasselt CM, Grol RPTM. Understanding laboratory testing in diagnostic uncertainty: a qualitative study in general practice. *Br J Gen Pract*. 2002;52(485):974-980.
12. Kristiansen IS, Hjortdahl P. The general practitioner and laboratory utilization: why does it vary? *Fam Pract*. 1992;9(1):22-27.
13. Zaat JO, van Eijk JT. General practitioners' uncertainty, risk preference, and use of laboratory tests. *Med Care*. 1992;30(9):846-854.
14. Koch H, Bokhoven MA van, Riet G ter, et al. What makes general practitioners order blood tests for patients with unexplained complaints? A cross-sectional study. *Eur J Gen Pract*. 2009;15(1):22-28. doi:10.1080/13814780902855762
15. Allison JJ, Kiefe CI, Cook EF, Gerrity MS, Orav EJ, Centor R. The association of physician attitudes about uncertainty and risk taking with resource use in a Medicare HMO. *Med Decis Mak Int J Soc Med Decis Mak*. 1998;18(3):320-329. doi:10.1177/0272989X9801800310
16. Miyakis S, Karamanof G, Liontos M, Mountokalakis TD. Factors contributing to inappropriate ordering of tests in an academic medical department and the effect of an educational feedback strategy. *Postgrad Med J*. 2006;82(974):823-829. doi:10.1136/pgmj.2006.049551
17. Amar E, Pereira C. *Les prescriptions des médecins généralistes et leurs déterminants*. Dress n°440; 2005.

- 1
- 2
- 3 18. Tsalatsanis A, Hozo I, Vickers A, Djulbegovic B. A regret theory approach to decision curve analysis: A novel
- 4 method for eliciting decision makers' preferences and decision-making. *BMC Med Inform Decis Mak.*
- 5 2010;10(1):51. doi:10.1186/1472-6947-10-51
- 6
- 7 19. Norman GR, Grierson LEM, Sherbino J, Hamstra SJ, Schmidt HG, Mamede S. Expertise in medicine and
- 8 surgery. In: *The Cambridge Handbook of Expertise and Expert Performance.* Cambridge University Press;
- 9 2018:331-355.
- 10
- 11 20. Ericsson KA. Deliberate Practice and the Acquisition and Maintenance of Expert Performance in Medicine
- 12 and Related Domains. *Acad Med.* 2004;79(10):S70.
- 13
- 14 21. Spielberger CD, Gorsuch RL, Lushene RE. Manual for the State-Trait Anxiety Inventory. Published online
- 15 1970. Accessed June 12, 2020. <http://ubir.buffalo.edu/xmlui/handle/10477/2895>
- 16
- 17 22. Spielberger CD, Bruchon-Schweitzer M, Paulhan I. *Inventaire d'anxiété état-trait forme Y (STAI-Y).* Éditions
- 18 du centre de psychologie appliquée, DL 1993; 1993.
- 19
- 20 23. Gauthier J, Bouchard S. Adaptation canadienne-française de la forme révisée du State-Trait Anxiety
- 21 Inventory de Spielberger. *Can J Behav Sci.* 1993;25(4):559-578. doi:10.1037/h0078881
- 22
- 23 24. Kerrien M, Pougnet R, Garlantézec R, et al. Prévalences des troubles anxieux, de l'humeur et de la
- 24 souffrance au travail des internes de médecine. *Presse Médicale.* 2015;44(4, Part 1):e84-e91.
- 25 doi:10.1016/j.lpm.2014.06.042
- 26
- 27 25. Gerrity MS, White KP, DeVellis RF, Dittus RS. Physicians' Reactions to Uncertainty: Refining the constructs
- 28 and scales. *Motiv Emot.* 1995;19(3):175-191. doi:10.1007/BF02250510
- 29
- 30 26. Le Roux V. Validation culturelle et psychométrique du PRU, Physicians' Reactions to Uncertainty scales,
- 31 questionnaire américain de mesure des réactions affectives des médecins face à l'incertitude. Published
- 32 online September 17, 2015.
- 33
- 34 27. Cooke GPE, Doust JA, Steele MC. A survey of resilience, burnout, and tolerance of uncertainty in Australian
- 35 general practice registrars. *BMC Med Educ.* 2013;13:2. doi:10.1186/1472-6920-13-2
- 36
- 37 28. Patton MQ. *Qualitative Research & Evaluation Methods: Integrating Theory and Practice.* SAGE
- 38 Publications; 2014.
- 39
- 40 29. Dutch College of General Practitioners. *Medically unexplained symptoms.* Dutch College of General
- 41 Practitioners; 2013.
- 42
- 43 30. Charlin B, Boshuizen HPA, Custers EJ, Feltovich PJ. Scripts and clinical reasoning. *Med Educ.*
- 44 2007;41(12):1178-1184. doi:10.1111/j.1365-2923.2007.02924.x
- 45
- 46 31. Bloy G. L'incertitude en médecine générale : sources, formes et accommodements possibles. *Sci Soc Santé.*
- 47 2008;26(1):67-91. doi:10.3406/sosan.2008.1881
- 48
- 49 32. Chamberland M. Comment exploiter les résultats de la recherche pour améliorer nos activités
- 50 d'enseignement et d'apprentissage du raisonnement clinique ? *Pédagogie Médicale.* 2005;6(4):197-199.
- 51 doi:10.1051/pmed:2005023
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

1  
2  
3 **Tables**  
45  
6 **Table 1. Main characteristics of the virtual patient.**  
7

8	
9	<b><i>Medical history</i></b>
10	<hr/>
11	Smoking ceased in 2012. 20 years-pack
12	High blood pressure since 2010
13	Right ulnar head fracture in 1989
14	Peritonitis in 1974
15	
16	
17	
18	<b><i>Biometrics in the virtual case</i></b>
19	<hr/>
20	85kgs
21	
22	1. 86m
23	
24	BMI 24,5
25	
26	Blood pressure 132/76mmHg
27	
28	Pulse 110
29	
30	<b><i>Treatment</i></b>
31	<hr/>
32	Hydrochlorothiazide 25mg 1/day
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	

**Table 2. Univariate analysis of group of GPs with under 10 years of experience versus more than 10 years**

	Total n (%)	< 10 years of experience	> 10 years of experience	p value
Mean age (years)	43	31.3	54.9	p<0.001
Women	20 (50)	7	13	p = 0.060
Mean experience (years)	15.1	3.6	26.5	p<0.001
Practice setting	40 (100)			p=0.520
Alone	26 (65)	14	12	
Association	14 (35)	6	8	
Practice location	40 (100)			p=0.340
Urban	25 (62.5)	11	14	
Semi-urban	15 (37.5)	9	6	
Number of regular in-office patients	451	158	745	p<0.001
Average duration of consultations	40 (100)			p=1
< 10 min	2 (5)	1	1	
10-20 min	32 (80)	16	16	
20-30 min	6 (15)	3	3	
Average uncertainty	3.1	3.3	2.9	p = 0.232
Average authenticity	4.4	4.50	4.30	p = 0.442
Average STAI-Y score	37	41.3	32.7	p = < .001
Average PRU score	17.6	20.7	14.4	p = 0.002
Number of complementary prescriptions	10.3	11.5	9.2	p = 0.0285
Average Price of complementary prescriptions (euros)	141.8	153.8	129.6	p = 0.356
Number of diagnostic hypotheses	3.0	3.1	2.9	p = 0.612
CCMU score	2.5	2.5	2.5	p = 0.798

**Table 3. Final models of the number of CPs after stepwise backward selection**

	Relative Risk	95% CI	Adjusted p-value
<b>Experience as categorial (R-squared: 0.308)</b>			
Experience (>10)	1.042	[0.673-1.611]	0.8549
Age	0.980	[0.9634-0.997]	0.0223
Number of regular in-office patients	1.000	[0.999-1.000]	0.1505
STAI-Y score	0.984	[0.968-0.999]	0.0351
Sex (Men)	0.842	[0.667-1.064]	0.1492
Number of hypothesis	1.093	[0.980-1.218]	0.1091
<b>Experience as continuous (R-squared: 0.272)</b>			
Experience	0.9766374	[0.965-0.988]	<0.0001
Number of regular in-office patients	1.0002539	[1.000-1.001]	0.0486
STAI-Y score	0.9868940	[0.973-1.001]	0.0784

Pre-review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Synthèse

Ce travail est intéressant sous deux aspects principaux. Premièrement, il démontre le potentiel de l'utilisation en recherche sur le raisonnement clinique d'un tel outil proposant des consultations simulées via des patients virtuels. Un tel outil permet d'appréhender la recherche dans le champ du raisonnement clinique à la fois dans des conditions standardisées mais également fortement authentiques. Ainsi, des déterminants précis du processus de raisonnement clinique peuvent être étudiés comme les caractéristiques personnelles des professionnels de santé. Ce genre d'approche a le potentiel pour améliorer la compréhension du processus de raisonnement clinique.

Deuxièmement, ce travail apporte des éléments de réflexion sur le développement de la tolérance à l'incertitude en médecine générale. Cette tolérance est-elle "innée" et nos groupes différaient-ils sur ce point ? Ou bien la tolérance à l'incertitude est-elle "acquise", au contact des problématiques de santé qui placent quotidiennement le médecin généraliste face à l'incertitude ? Dans cette dernière option, des stratégies spécifiques visant à développer la tolérance à l'incertitude doivent être testées afin de mieux former les futurs médecins généralistes. Sur ce point, un lien se fait avec l'article précédent sur le lien entre burnout et raisonnement clinique. Le TCS est un outil puissant pour illustrer l'incertitude dans les soins. Utiliser des outils comme les TCS précocement dans la formation des futurs médecins pourrait-il avoir un impact favorable sur le développement de la tolérance à l'incertitude ? Cette hypothèse est plausible puisque les étudiants pourraient réaliser plus tôt dans leur formation que l'incertitude est prégnante et qu'il faut composer avec elle dans sa pratique. Ceci devra faire l'objet de travaux prospectifs ultérieurs pour tenter de tester cette hypothèse.

Comme pour l'article sur le burnout et le raisonnement clinique, il est possible d'analyser le présent article sous l'angle de la MOT et du modèle MOT. Les prescriptions complémentaires sont la connaissance déclarative (concept) "plan d'investigation". Alors que dans le modèle MOT ce plan d'investigation est le produit d'une validation des options retenues avec le patient, cette validation n'était pas possible dans le logiciel utilisé. Ainsi, les prescriptions complémentaires discutées ici correspondraient probablement davantage au concept "options retenues" au sein de la procédure globale "mettre en œuvre les actions appropriées".

Comme présenté dans l'article, l'expérience du clinicien est un déterminant théorique de ces options d'investigation dans le modèle MOT. Sous cet angle, l'analyse du lien entre expérience et prescriptions complémentaires via l'outil Hello-Clinical possède une bonne validité de construit.

L'utilisation d'une situation pleinement standardisée permet de neutraliser tout ou partie des intrants "singularité de la situation" et "singularité du patient" alors que l'on cherche à analyser un déterminant isolément. La même réserve en matière de validité écologique que sur dans le travail sur le lien entre raisonnement clinique et burnout s'applique cependant. Cette force de l'outil qui standardise une consultation est une limite sur le plan de l'authenticité.

## DEUXIÈME PARTIE : LA RÉFLEXIVITÉ COMME MOTEUR DU DÉVELOPPEMENT DU RAISONNEMENT CLINIQUE

### DE LA MÉTACOGNITION À LA RÉFLEXIVITÉ

Le modèle MOT du processus de raisonnement clinique contient une procédure centrale influençant la représentation dynamique du problème que se forge le médecin face à un patient : "réguler le processus (métacognition)". Le principe accolé à cette procédure stipule que *"le clinicien tient compte, tout au long du processus, de la perspective du patient à l'égard de son problème : ses croyances, ses craintes, ses attentes, ses choix, ses émotions et l'impact du problème sur sa vie"* (45). Cette procédure de métacognition se compose de trois sous-procédures : analyser ses propres activités cognitives, évaluer son propre processus cognitif et intervenir sur son processus cognitif.

Dans la littérature, la métacognition est un concept polysémique dont les définitions ont beaucoup varié dans le temps (152,153). La définition de Flavell de 1976 (154) est considérée comme un jalon important des travaux dans le domaine : *"[...] la métacognition désigne la connaissance que l'on a de ses propres processus et produits cognitifs ou de tout ce qui s'y rapporte [...]. Par exemple, je fais de la métacognition [...] si je remarque que j'ai plus de difficulté à apprendre A que B ; si je suis frappé par le fait que je devrais révérifier C avant de l'accepter comme un fait [...] La métacognition fait référence, entre autres, à la surveillance active et à la régulation et l'orchestration consécutive à ces processus [...] habituellement au service d'un but ou d'un objectif concret."*

Cette définition a fait l'objet de débats critiques dans la littérature principalement sur son côté "flou" (152,153). Une définition plus récente et qui semble mieux coller avec la procédure du modèle MOT est celle de Noël (153,155) : *"La métacognition est un processus mental dont l'objet est soit une activité cognitive, soit un ensemble d'activités cognitives que le sujet vient d'effectuer ou est en train d'effectuer, soit un produit mental de ces activités cognitives. La métacognition peut aboutir à un jugement (habituellement non exprimé) sur la qualité des activités mentales en question ou de leur produit et éventuellement à une décision de modifier l'activité cognitive, son produit ou même la situation qui l'a suscité"*.

Dans le champ de l'éducation médicale et dans les référentiels de compétences régissant la formation des médecins en France, c'est plus souvent le concept de réflexivité qui est mis en avant pour désigner une forme particulière de métacognition. Le concept de réflexivité a été abordé par de nombreux auteurs, issus de disciplines différentes (17). Historiquement, les premiers écrits traitant du lien entre réflexivité et apprentissage remontent à Aristote dans "*Éthique à Nicomaque*". Par la suite, le premier auteur contemporain majeur à traiter la question semble avoir été John Dewey. Ce psychologue et philosophe américain développe dès 1910 dans son ouvrage "*How we think*" le concept de pensée réflexive (156). Dewey nomme pensée réflexive une pensée pour laquelle "*le motif ou le fondement d'une croyance est activement recherché et sa pertinence à soutenir la croyance est examinée*" (156).

À partir des années 1970, un autre psychologue et pédagogue américain du nom de David Kolb va apporter une contribution importante au concept de réflexivité via ses travaux sur l'apprentissage expérientiel (157). Dans sa boucle de l'apprentissage expérientiel, l'observation réfléchie (*reflective observation*) permet à partir d'une expérience concrète de prendre du recul sur celle-ci, d'en expliciter l'implicite, de rendre consciente l'expérience (158). La figure n°19 ci-dessous illustre la boucle de l'apprentissage expérientiel selon Kolb.

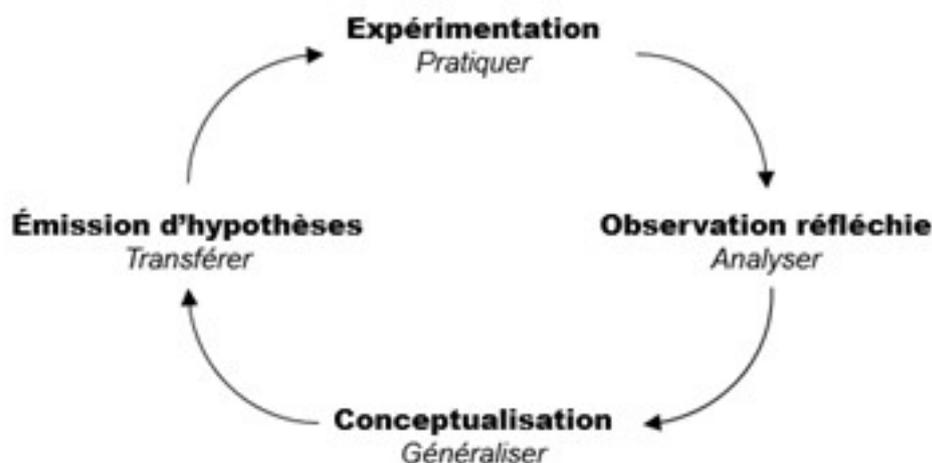


FIGURE 20: BOUCLE DE L'APPRENTISSAGE EXPÉRIENTIEL SELON KOLB

Dans les années 1980, ce sont David Boud, Rosemary Keogh et David Walker qui vont argumenter le rôle de la réflexivité dans l'apprentissage dans leur ouvrage "*Reflection: Turning Experience into Learning*" (159). À partir notamment des travaux de Dewey et Kolb, Boud *et al.* vont utiliser le terme de

"réflexivité" et développer celui-ci comme un processus d'aller-retours avec l'expérience visant des résultats concrets (159). C'est notamment la place des émotions dans le processus réflexif qui est une particularité du modèle d'apprentissage réflexif de Boud *et al.* La figure n°20 ci-dessous illustre le processus d'apprentissage réflexif de Boud, Keogh et Walker.

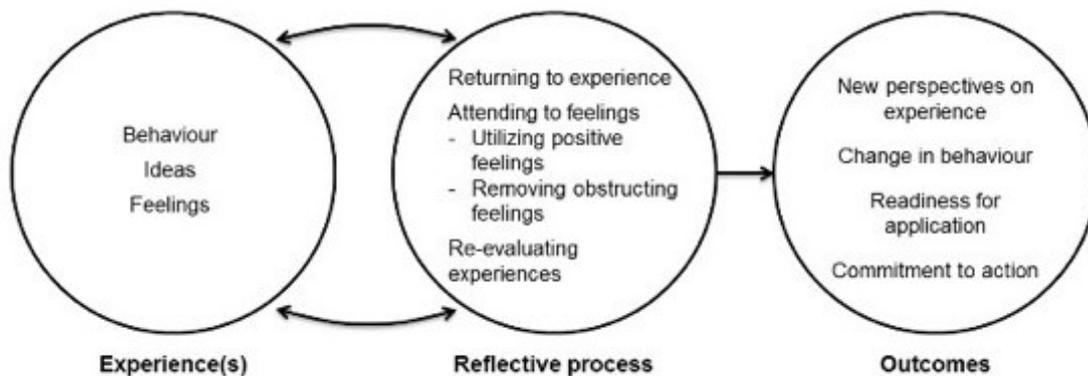


FIGURE 21 : L'APPRENTISSAGE RÉFLEXIF SELON BOUD, KEOGH ET WALKER

Enfin, le dernier auteur majeur sur la réflexivité est Donald Schön. Cet américain professeur de planification urbaine et philosophe, développa le concept de pratique réflexive à partir du début des années 1980 (160,161). Pour Schön, les professionnels efficaces sont ceux qui réfléchissent "dans l'action" (161). Cette réflexion est selon Schön à la fois "sur" l'action et "pour" l'action. Les professionnels s'améliorent en prenant un certain recul sur leur pratique (161).

À partir des travaux de ces auteurs et en réalisant une revue systématique de la littérature, Nguyen *et al.* ont proposé une analyse conceptuelle de la réflexivité en 2014 (162). Pour Nguyen *et al.*, la réflexivité désigne "le processus [qui consiste à] s'engager dans des interactions attentives, critiques, exploratoires et itératives avec ses pensées et ses actions, et leurs cadres conceptuels sous-jacents, en visant à changer ces différentes composantes et en examinant le changement lui-même" (162). Le modèle de Nguyen *et al.* distingue ainsi cinq composantes de la réflexivité :

1. Les pensées et actions
2. La composante attentive, critique, exploratoire et itérative (ACEI)
3. Les cadres conceptuels sous-jacents
4. La vue sur le changement (VC)

## 5. Le soi

Nguyen et Raymond-Carrier proposent une définition simple de chaque composante dans le chapitre *"Un professionnel qui se pose des questions : la réflexivité"* de l'ouvrage *"Comment (mieux) former et évaluer les étudiants en médecine et en sciences de la santé"* (17). Nous proposons de les définir ici pour mieux les utiliser dans la suite de notre tapuscrit.

La composante "Pensées et actions" désigne *"le contenu de départ la réflexivité, c'est-à-dire ce à quoi on réfléchit. Lorsqu'on s'engage dans une activité réflexive, on questionne, on analyse, on interagit avec nos pensées et nos actions. Il s'agit donc du matériel de base à partir duquel la réflexivité se développe. La composante PA se veut très englobante. Elle inclut, par exemple, des idées, des opinions, des émotions, des habitudes, des problèmes, des erreurs ou des rencontres"*.

La composante "ACEI" décrit *"le processus de la réflexivité, c'est-à-dire la manière de réfléchir. Pour qu'un processus cognitif devienne réflexif, il faut qu'il soit qualitativement différent des pensées habituelles. Il s'agit ainsi de réfléchir de façon attentive, critique, exploratoire et itérative. Un processus « attentif » permet d'analyser le contenu de la réflexivité de manière consciente et délibérée. Les aspects « critiques » et « exploratoires » assurent que cette réflexivité remettra en question la vision actuelle de la situation, afin d'envisager ses multiples perspectives, ramifications, alternatives et conséquences potentielles. Le processus doit aussi être « itératif », dans la mesure où il se poursuit en continu et qu'il intègre à chaque fois les résultats du processus réflexif précédent"*.

Les "cadres conceptuels sous-jacents" sont *"les schèmes et les scripts, qui sous-tendent notre façon de penser et d'agir. Ils peuvent tout autant être conscients qu'inconscients, et explicites ou implicites. Le fait de les inclure dans le processus réflexif permet de comprendre et d'expliquer nos PA. Une activité réflexive devient en effet beaucoup plus riche lorsqu'on s'intéresse aux raisons de nos comportements, de nos habitudes, de nos croyances et de nos erreurs"*.

La composante "VC" donne *"un objectif au processus réflexif. Celui-ci resterait en effet stérile s'il n'induisait pas, chez l'individu qui s'y engage, une possibilité de changement. Ainsi, dans un processus réflexif, la préoccupation consistant à trouver de nouvelles perspectives et à identifier des pistes de solution doit être manifeste. D'ailleurs, en visant et en produisant un changement (qu'il concerne les pensées, les actions ou les cadres conceptuels), celui-ci pourra lui-même être questionné. Il deviendra alors le contenu de la réflexivité et pourra être réflexif"*.

La composante du "soi" désigne le fait de "ramener et de rattacher à soi l'activité réflexive. Un processus cognitif ne deviendra ainsi pleinement réflexif que lorsque le soi interagit avec les quatre autres composantes de la réflexivité. Être réflexif implique en effet de s'intéresser à soi-même (ses pensées, ses actions, ses cadres conceptuels et son changement) et de mettre le soi au service de sa réflexivité, en puisant dans la connaissance de soi et dans ses ressources, afin d'aller au bout du processus réflexif".

La figure n°21 ci-dessous illustre l'articulation entre les composantes de la réflexivité selon Nguyen *et al.*

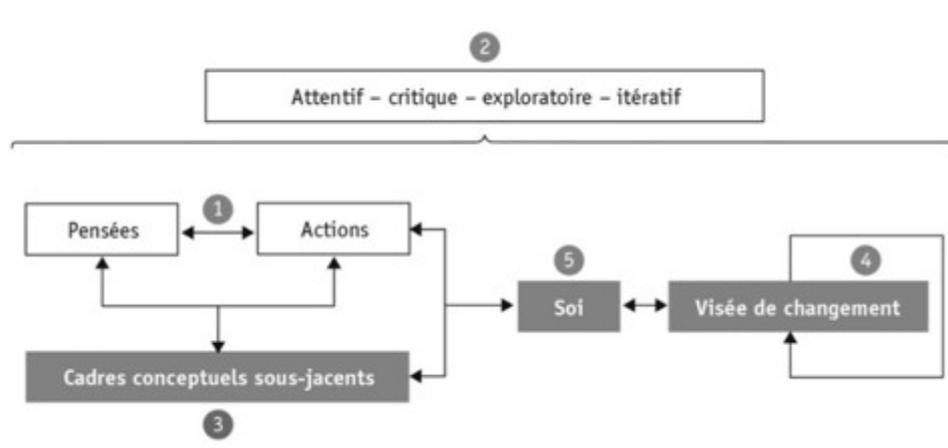


FIGURE 22 : LES COMPOSANTES DE LA RÉFLEXIVITÉ SELON NGUYEN ET AL.

#### ACCOMPAGNER LE DÉVELOPPEMENT DE LA RÉFLEXIVITÉ VIA DES ÉCRITS RÉFLEXIFS

Appliqué à la formation des médecins généralistes, ces travaux m'ont permis d'aborder différemment la question de l'accompagnement du développement de la réflexivité. Les étudiants de troisième cycle inscrits en DES de médecine générale en France doivent déposer des traces écrites d'apprentissages dans leurs portfolios. Ces traces se veulent réflexives afin à la fois de développer la réflexivité lors de la tâche (visée formative) mais également visent à certifier la réflexivité des futurs médecins généralistes (visée certificative). La forme la plus codifiée et la plus aboutie de ces écrits réflexifs se nomme "récit de situation complexe authentique" (RSCA). Je vais m'inspirer dans la section ci-dessous d'un court texte que j'ai rédigé à destination des internes de médecine générale de Strasbourg.

Le RSCA est un travail de réflexion de l'apprenant sur une situation qu'il a personnellement vécue au sein de sa pratique professionnelle. Ce RSCA est utilisé comme outil de formation du médecin réflexif. Il s'agit d'une trace d'apprentissage observable. Un RSCA comporte 4 parties : le récit, l'analyse, la description des tâches d'apprentissages induites et la synthèse.

Dans le récit, la situation est décrite sans interprétation à posteriori. Le récit s'attache autant aux faits qu'au ressenti de l'apprenant. L'analyse vient ensuite à distance de la situation interpréter ce qui s'est passé dans le récit. Pourquoi l'apprenant s'est-il comporté ainsi ? Pourquoi cela a-t-il été fait ? etc... Il s'agit d'une auto-évaluation de l'apprenant. Il est demandé aux apprenants de dégager quelques problématiques qui vont être traitées dans la partie suivante. Ces problématiques sont organisées selon les thèmes suivants : biomédical, psycho-social, environnemental, administratif et éthique, médico-légal. Le thème « environnemental » renvoie à l'environnement de soin (l'entourage, les interactions avec les professionnels de santé, etc...). Chacune de ces problématiques est contextualisée : la question se rapporte au cas exemplaire et non à un "cas général".

La partie "trace d'apprentissage" ou "auto-formation" explicite les résultats des recherches réalisées par l'étudiant pour chacune des problématiques. Une re-contextualisation des résultats des recherches au cas exemplaire est demandée.

L'apprenant reprend finalement les points essentiels de l'analyse et de l'autoformation réalisée pour écrire une synthèse qui décrit l'impact de ce travail en termes de développement des compétences, d'acquisition de connaissances et d'impact sur les pratiques ultérieures.

L'évaluation proposée au niveau national de ces RSCA utilise une grille s'intéressant plus à la forme qu'au fond (163,164). La réflexivité n'y est pas explicitée comme objectif d'apprentissage. Le détail des items de la grille est détaillé dans le tableau n°9 ci-dessous.

## GRILLE D'ÉVALUATION du RSCA

### Grille d'évaluation d'une trace écrite d'apprentissage

#### Grille d'évaluation formative unique des traces d'apprentissage

**Préambule :** Cette grille est un outil d'aide pour évaluer le travail d'écriture : elle permet une auto-évaluation et une évaluation par le tuteur. Les notes aident à mesurer le décalage entre ces deux évaluations (source d'échanges), ainsi que la progression de l'interne au cours des trois années du DES.

	Pertinent	Attendu	Satisfaisant	Améliorable	Limite	Insuffisant
<b>Narration de la situation</b>	Le récit comprend : - présentation du patient (physiques, biographiques) - présentation du contexte - les différents temps de la consultation = tempi - une description des émotions manifestées par le patient et ressenties par l'interne, émotions propres à l'interne. - une description des comportements, interactions de l'interne et du patient.			Il manque un ou plusieurs critères de la colonne de gauche		Le récit est similaire à une prise d'observation clinique : description impersonnelle et inadéquate d'un cas clinique, sans fait marquant, sans éléments permettant de suivre et comprendre la complexité de la démarche
<b>/10</b>	<b>10</b>	<b>9</b>	<b>8</b>	<b>7</b>	<b>6</b>	<b>5</b>
				<b>4</b>		<b>3</b>
						<b>2</b>
						<b>1</b>

<b>Problématisation</b>	-identifie les problèmes, - hiérarchise les problèmes, - et précise les interactions. -Fait un diagnostic de situation dans le cadre d'une approche globale, -les questions posées sont précises et adaptées à la situation, en adéquation avec les compétences de MG.	-L'approche globale est ébauchée mais insuffisante. - Soit l'identification est incomplète. - Soit la hiérarchisation est incomplète.	-se limite au diagnostic de maladie, centré maladie, uniquement dans un seul registre (ex : biomédical) alors que la situation appelle à plusieurs axes
<b>/10</b>	<b>10</b>	<b>9</b>	<b>8</b>
			<b>7</b>
			<b>6</b>
			<b>5</b>
			<b>4</b>
			<b>3</b>
			<b>2</b>
			<b>1</b>
<b>Qualité de la recherche et synthèse documentaire</b>	- La recherche est adaptée aux questions posées - Chaque axe fait l'objet d'une synthèse, dont les références sont valides et actualisées. - Une discussion nuancée selon les niveaux de preuve est présente.	- un ou deux des trois items de la colonne de gauche est/sont absent(s)	- il n'y a pas d'adéquation avec les axes de recherche, - il n'existe pas de synthèse permettant de répondre à la question posée, - les références sont de niveau de preuve insuffisant, - les références sont non actualisées.
<b>/10</b>	<b>10</b>	<b>9</b>	<b>8</b>
			<b>7</b>
			<b>6</b>
			<b>5</b>
			<b>4</b>
			<b>3</b>
			<b>2</b>
			<b>1</b>

<b>Identification des apprentissages</b>	- le travail sur la situation et son exposé montre que l'étudiant a bien identifié les apprentissages prioritaires (en fonction de la situation et/ou en fonction des besoins) - l'étudiant renseigne sa progression. - il propose des solutions partielles ou globales aux problèmes initialement posés, - il argumente en quoi elles sont plus adaptées.	- L'étudiant a identifié quelques apprentissages, - il a du mal à proposer de nouvelles solutions - il a du mal à renseigner sa progression	- L'étudiant n'arrive pas à identifier ses apprentissages, - Il ne propose pas de solution nouvelle et plus adaptée, - Il ne renseigne pas sa progression
<b>/10</b>	<b>10 9 8</b>	<b>7 6 5 4</b>	<b>3 2 1</b>
<b>Acquisition des compétences soulevées par la situation</b>	- les compétences décrites par l'étudiant sont bien celles mises en œuvre pour résoudre la situation, avec une prise de conscience et description de leurs composantes.  - Les compétences décrites sont bien celles de la « Marguerite »	- les compétences ne sont que listées, - les compétences ne sont pas détaillées dans ces composantes.	- les compétences ne sont pas citées, - les compétences ne sont pas à propos.
<b>/10</b>	<b>10 9 8</b>	<b>7 6 5 4</b>	<b>3 2 1</b>

**TABLEAU 9 : GRILLE NATIONALE D'ÉVALUATION DES RSCA**

A partir de cette discordance entre l'outil d'évaluation proposé et la finalité des RSCA, j'ai travaillé la question de l'évaluation de la réflexivité des traces écrites d'apprentissage.

#### ÉVALUER LA RÉFLEXIVITÉ DES TRACES ÉCRITES D'APPRENTISSAGES : LA TRADUCTION DE LA GRILLE REFLECT

Dans le cadre du travail de mémoire du master Pédagogie des Sciences de la Santé de Racha Onaisi, nous avons cherché à proposer un outil validé d'évaluation de la réflexivité des traces écrites d'apprentissage. Plutôt que de développer un nouvel outil, nous avons identifié via une recension des écrits la grille REFLECT (165). Cette grille est validée en langue anglaise. N'ayant pas identifié de tel outil en langue française, nous avons décidé d'entamer un travail de traduction/adaptation culturelle de la grille REFLECT.

#### Contexte de l'étude

La réflexivité n'est pas une capacité intuitive ou innée, notamment en début de cursus médical (162,165,166). Afin de guider son développement et son évaluation, plusieurs auteurs ont tenté d'élaborer des outils qui diffèrent tant dans leur forme que dans le cadre théorique sur lequel ils s'appuient.

Wald *et al.* (165) ont élaboré un outil d'hétéro-évaluation en anglais afin de favoriser une évaluation fiable en vue de la rétroaction proposée aux étudiants durant les groupes d'écriture réflexive : la grille REFLECT, pour *Reflection Évaluation for Learners' Enhanced Competencies Tool*, que l'on pourrait traduire

par "Outil d'évaluation de la réflexivité pour améliorer les compétences des apprenants". Le cadre conceptuel de développement de cette grille s'appuie sur plusieurs auteurs de référence des théories de la réflexivité : Mann *et al.*, Schön, Boud *et al.*, Moon et Mezirow (165). Ces derniers décrivent la réflexivité comme une analyse critique du savoir et de l'expérience, intégrant les émotions et le ressenti, afin d'atteindre un niveau approfondi de compréhension et de donner du sens, selon un processus complexe, qui nécessite une capacité de conceptualisation des bases de la réflexion (« *premise reflection* ») en vue de transformer et/ou renforcer les apprentissages(162,165,167–169). Le processus intègre une dimension itérative (concept de boucle réflexive) et une dimension verticale hiérarchisant divers niveaux de réflexion (162,165,167–169). Le tableau n°10 ci-dessous présente la grille REFLECT originale.

Criterion	Level				Axis II for critical reflection	
	Habitual action (Nonreflective)	Thoughtful action or Introspection	Reflection	Critical reflection	Transformative reflection and learning	Confirmatory learning
<b>Writing spectrum</b>	Superficial descriptive writing approach (fact reporting, vague impressions) without reflection or introspection	Elaborated descriptive writing approach and impressions without reflection	Movement beyond reporting or descriptive writing to reflecting (i.e., attempting to understand, question, or analyze the event)	Exploration and critique of assumptions, values, beliefs, and/or biases, and the consequences of action (present and future)	Frames of reference or meaning structures are transformed. Requires critical reflection	Frames of reference or meaning structures are confirmed. Requires critical reflection
<b>Presence</b>	Sense of writer being partially present	Sense of writer being partially present	Sense of writer being largely or fully present	Sense of writer being fully present	Integration of new learning into one's identity, informing future perceptions, emotions, attitudes, insights, meanings, and actions. Conveys a clear sense of a breakthrough	
<b>Description of conflict or disorienting dilemma</b>	No description of the disorienting dilemma, conflict, challenge, or issue of concern	Absent or weak description of the disorienting dilemma, conflict, challenge, or issue of concern	Description of the disorienting dilemma, conflict, challenge, or issue of concern	Full description of the disorienting dilemma, conflict, challenge, or issue of concern that includes multiple perspectives, exploring alternative explanations, and challenging assumptions		
<b>Attending to emotions</b>	Little or no recognition or attention to emotions	Recognition but no exploration or attention to emotions	Recognition, exploration, and attention to emotions	Recognition, exploration, attention to emotions, and gain of emotional insight		
<b>Analysis and meaning making</b>	No analysis or meaning making	Little or unclear analysis or meaning making	Some analysis and meaning making	Comprehensive analysis and meaning making		
<b>Optional minor criterion: Attention to assignment (when relevant)</b>	Poorly addresses the assignment question and does not provide a compelling rationale for choosing an alternative	Partial or unclear addressing of assignment question; does not provide a compelling rationale for choosing an alternative	Clearly answers the assignment question or, if relevant, provides a compelling rationale for choosing an alternative	Clearly answers the assignment question or, if relevant provides a compelling rationale for choosing an alternative		

TABLEAU 10 : GRILLE REFLECT DANS SA VERSION ORIGINALE

D'autres outils ont été développés dans le cadre de l'éducation médicale. Ainsi, l'outil *Reflection-on-Action*, développé à des fins de recherche pour l'évaluation des traces écrites d'apprentissage également, unidimensionnel (une seule entité définie comme étant la capacité réflexive et évaluée en six niveaux) et qui est comparé à la grille REFLECT dans l'étude de Miller-Kuhlmann et al. (170). Au-delà du cadre de l'écriture réflexive, d'autres outils ont également été développés, comme la *Groningen Reflection Ability Scale* (GRAS) qui est une échelle unidimensionnelle couvrant trois aspects d'intérêt selon le cadre conceptuel d'élaboration (171,172). Boenink et al. (173) ont pour leur part développé un outil semi-structuré d'hétéro-évaluation de la démarche réflexive des étudiants portant sur des vignettes

cliniques. Autre exemple, la grille StARS® d'évaluation de la démarche réflexive des étudiants guidée par un questionnaire de 6 questions suite au visionnage de vidéos de rencontres cliniques, qui considère trois éléments constitutifs de la réflexivité (174).

Il existe également des outils développés en dehors du contexte médical. On peut par exemple citer le travail de thèse de Christian Michaud, portant sur l'évaluation du niveau de réflexivité du portfolio de futurs enseignants, à l'aide d'une grille déterminant un indice réflexif (175,176).

Il ne semble pas exister d'outil validé d'évaluation de la réflexivité des traces écrites d'apprentissage appliqué au milieu des sciences de la santé en français. Dans ce contexte, nous avons fait le choix de travailler à la traduction de la grille REFLECT pour l'utiliser dans une perspective d'évaluation formative des RSCA. Nous avons réalisé une étude visant à adapter la grille REFLECT pour un usage en langue française.

#### Matériel et méthodes

Nous avons pour cela mené un processus de double traduction puis double rétro-traduction afin de garantir une adaptation interculturelle respectant le sens de la version originale. Nous avons ensuite effectué des tests pilotes sur la version française préliminaire avec une évaluation de la validité de contenu. La méthode a été élaborée à partir des recommandations de bonne pratique proposées par Sousa et Rojjanasrirat (177). Six étapes se sont succédé :

1. Une double traduction vers le français a été effectuée par deux traducteurs bilingues indépendants ;
2. Une comparaison des deux traductions par une troisième personne bilingue ayant une expérience en traduction ;
3. Une rétro-traduction, par deux autres traducteurs bilingues indépendants, de langue maternelle anglaise, et ayant également une double culture anglo-saxonne et française ;
4. La résolution des discordances par discussion entre les auteurs (version pré-finale)
5. Des tests pilotes sur un panel de futurs utilisateurs composé d'étudiants et d'enseignants de troisième cycle de médecine générale puis une évaluation de la clarté des items par un panel d'experts
6. Une présentation de l'outil lors d'un atelier de formateurs de médecine générale

Enfin, la grille finale a été soumise à un panel d'utilisateurs pour évaluer un échantillon de traces écrites d'apprentissage.

### Résultats

Lors de la traduction de l'anglais vers le français (étape 1), certains termes spécifiques ont posé des difficultés aux traducteurs. Il s'agissait principalement de termes relativement spécifiques du champ des sciences de l'éducation ou du construit (« *meaning structures* », « *transformative reflection and learning* » ou encore « *confirmatory learning* » par exemple). Certaines expressions spécifiquement anglo-saxonnes sans équivalent exact en français étaient également concernées (par exemple « *insight* » ou « *meaning making* »).

La version pré-finale a été évaluée par un panel de quatorze volontaires : sept enseignants et sept étudiants. Trente et un items sur les 39 évalués ont été considérés comme insuffisamment ou pas du tout clairs par plus de 20% des répondants et ont été modifiés, à l'aide des suggestions proposées par les répondants. À l'issue de cette étape, le critère optionnel de la grille a été supprimé. Seuls 21% des répondants ont trouvé que ce critère était insuffisamment ou pas du tout clair, mais ils étaient 29% à trouver la description du niveau 3 insuffisamment ou pas du tout claire et 36% pour le niveau 4.

Les six experts ont évalué les 30 items, en lien avec le cadre conceptuel choisi et les cinq composantes essentielles de la réflexivité : 80% des items ont un indice de validité de contenu à 1 et 20% ont un indice de validité de contenu à 0.83. L'indice de validité de contenu global (à l'échelle de la grille) est de 0.97.

Les items et les intitulés ont été soumis aux experts : 16 (47%) ont été évalués « insuffisamment ou pas du tout clairs » mais seulement 5 (15%) l'ont été par 50% des experts ou plus. Pour ces derniers, les propositions pour clarifier l'intitulé étaient consensuelles pour trois items (60%). Pour les 11 items évalués « insuffisamment ou pas du tout clairs » par 33% des experts, les remarques étaient consensuelles pour seulement trois d'entre eux (27%). Seuls les items pour lesquels les commentaires concordent n'ont pas nécessité d'arbitrage particulier. Nous avons mené une dizaine d'arbitrages à cette étape.

Un échantillon de 10 futurs utilisateurs a accepté de tester la grille sur trois traces d'apprentissage longues. Six réponses n'étaient pas exploitables et ont été exclues. L'analyse de la concordance retrouvait un coefficient kappa = 0.521 [-0.133 ; 0.714].

## Discussion

Ce travail a permis d'aboutir à une version française de la grille REFLECT. Suite aux retours effectués par l'échantillon d'enseignants et d'étudiants, la majorité des descripteurs de niveaux sont formulés de façon plus synthétique et ont été simplifiés par rapport à la traduction initiale. À l'inverse, certains descripteurs de niveaux ont été étoffés, notamment lorsqu'ils paraissaient trop abstraits ou insuffisamment explicites pour les évaluateurs. Nous avons également été amenés à modifier certains descripteurs pour marquer la différence entre deux niveaux. Il est important de noter que la différenciation parfois insuffisante entre les niveaux fait partie des critiques qui ont pu être formulées à l'encontre de la grille REFLECT originale (170). Toutefois ces modifications n'ont pas dénaturé l'esprit de la version anglophone de la grille REFLECT. Pour mémoire, pour chaque niveau de chacun des critères était proposée une définition descriptive et c'était donc à l'étudiant et/ou à la personne évaluatrice de rechercher et documenter les extraits narratifs permettant d'argumenter le niveau atteint. Ces extraits constituent des indicateurs. La version française garde ce format et ne prescrit aucun indicateur précis.

Le critère optionnel a été totalement supprimé et la mise en page de la grille modifiée afin de faire apparaître plus clairement le processus d'évaluation en deux étapes, bien qu'il soit explicité dans les instructions d'utilisation de la grille. Par ailleurs, nous avons choisi de joindre un glossaire pour expliciter certains termes ou expressions qui nous paraissaient complexes ou bien utilisés dans un sens très spécifique. Tous ces éléments constituent des différences importantes avec la grille d'origine. Toutefois, cela est nécessaire pour rendre la grille accessible à l'ensemble des utilisateurs, notamment les étudiants, sans nécessiter une formation spécifique.

Une autre différence est le choix du cadre conceptuel sous-jacent, toutefois la définition de Nguyen *et al.* (162) englobe le cadre théorique de développement de la version originale. En plus d'être englobante, elle est plus récente que la grille REFLECT et fait référence dans le milieu de l'éducation médicale francophone (17).

Par ailleurs, l'indice de validité de contenu tant au niveau de chaque item qu'au niveau global démontre que la version française présente une validité de contenu élevée.

De plus, la formulation de la question posée aux experts permet également d'apprécier en partie la validité de construit, puisque nous interrogeons la pertinence par rapport à la définition de la réflexivité proposée par Nguyen *et al.* (162). Toutefois, l'absence d'outil de référence ne permet pas d'effectuer

d'analyse de la convergence. Il pourrait être intéressant d'évaluer la validité divergente par rapport à d'autres outils n'analysant pas le même construit. La version intégrale de l'article publié dans la revue *Pédagogie Médicale* en mars 2021 (178) se trouve ci-dessous.

## Évaluer la réflexivité à travers les traces écrites d'apprentissage des étudiants en santé : traduction et adaptation interculturelle de la grille REFLECT

*Evaluating health sciences students' reflexivity through written works: translation and cultural adaptation of the REFLECT rubric*

Racha ONAISI<sup>1,2,\*</sup>, François SEVERAC<sup>3</sup>, et Mathieu LORENZO<sup>1,4</sup>

<sup>1</sup> Centre de formation et de recherche en pédagogie des sciences de la santé (CFRPS), Université de Strasbourg, Strasbourg, France

<sup>2</sup> Département de médecine générale, Université de Bordeaux, Bordeaux, France

<sup>3</sup> Laboratoire de biostatistique et informatique médicale, Université de Strasbourg, Strasbourg, France

<sup>4</sup> Département de médecine générale, Université de Strasbourg, Strasbourg, France

Manuscrit reçu le 22 février 2020 ; commentaires éditoriaux formulés aux auteurs le 24 octobre et le 7 janvier 2020 ; accepté pour publication le 11 janvier 2021

**Résumé-Contexte :** La réflexivité est considérée par de nombreux auteurs comme une composante essentielle du professionnalisme, tout en étant au cœur du développement de ce dernier et de l'identité professionnelle. Le portfolio, outil de soutien de l'apprentissage et d'évaluation en médecine générale, comporte des traces écrites d'apprentissage à visée réflexive. Il n'existe pas d'outil validé en langue française permettant d'évaluer la réflexivité à travers des traces écrites d'apprentissage. **But :** L'objectif de cette étude est donc d'adapter la grille REFLECT, validée en anglais, pour un usage en langue française. **Méthode :** Un processus de double traduction puis double rétro-traduction a été mené afin de garantir une adaptation interculturelle respectant le sens de la version originale. Des tests pilotes ont été menés sur la version française préliminaire avec évaluation de la validité de contenu. **Résultats :** L'évaluation de la clarté de la version française a conduit à une version finale comportant 5 critères ayant chacun 4 niveaux d'évaluation possibles. La version française de la grille REFLECT présente une validité de contenu élevée. **Conclusion :** L'utilisation de la grille REFLECT en français permettra vraisemblablement d'optimiser le développement de la réflexivité chez les internes en médecine générale et de faciliter la rétroaction par les superviseurs.

**Mots clés :** réflexivité, évaluation, portfolio, écriture réflexive, validation

**Abstract. Introduction:** Reflection is considered by many authors as an essential component of professionalism, at the heart of its development and the construction of professional identity. Portfolios, when used as learning and assessment support, usually includes reflective written works. There is no tool validated in French to assess students' reflection through written works. **Objective:** To translate and adapt the REFLECT rubric, validated in English, for it to be used in French. **Methods:** We conducted a double forward translation, then a double reverse translation to ensure cultural adaptation consistent with the original version. We then ran pilot tests on the preliminary French version, to assess content-validity. Finally, we ran validation tests. **Results:** The final French version is a four-levels and five criteria rubric. The French version has a high content validity index. **Conclusion:** The French REFLECT Rubric will help improve the development of students' reflection and will help supervisors provide a more precise feedback.

**Keywords:** Reflection, assessment, portfolio, reflective writing, validation

\*Correspondance et offprints : Racha ONAISI, Département de médecine générale, Université de Bordeaux, 146 Rue Léo Saignat, Collège santé Case 148, 33076 Bordeaux Cedex, France.  
Mailto: [racha.onaisi@u-bordeaux.fr](mailto:racha.onaisi@u-bordeaux.fr).

## Introduction

### Problématique et cadre conceptuel

La réflexivité est considérée par de nombreux auteurs comme étant une composante essentielle du professionnalisme et de l'identité professionnelle [1–9] : « *la dimension réflexive est au cœur de toutes les compétences professionnelles, elle est constitutive de leur fonctionnement et de leur développement* » [10].

Nguyen *et al.* [7] proposent une définition synthétique basée sur le référencement, à l'issue d'une revue méthodique de la littérature, des quinze auteurs les plus cités dans les travaux de recherche relatifs à la réflexivité en sciences de l'éducation et en éducation médicale.

Ils distinguent cinq composantes au cœur du processus réflexif. La réflexivité peut alors être définie comme étant « *le processus qui consiste à engager (i) le soi dans des (ii) interactions attentives, critiques, exploratoires et itératives avec ses (iii) pensées et ses actions, et leurs (iv) cadres conceptuels sous-jacents, en visant à (v) changer ces différents composantes et en examinant le changement lui-même* » [11].

L'utilisation d'outils comme les portfolios a notamment pour objectif de développer la posture réflexive des étudiants en sciences de la santé [12,13]. Ainsi, selon Naccache *et al.*, le portfolio « *encourage [l'apprenant] à planifier ses activités d'apprentissage à partir des tâches professionnelles authentiques auxquelles il est exposé et à s'autoévaluer dans une démarche de pratique critique et réflexive* » [12]. Le recours à la production de traces écrites d'apprentissage permettrait, en lui-même, de promouvoir l'activité réflexive du fait du recul qu'elle permet à l'apprenant de prendre par rapport à la situation clinique vécue [6,12]. Il est par ailleurs indispensable que la trace écrite d'apprentissage soit rédigée selon la propre perspective de l'étudiant, telle que la situation a été vécue. En effet, lorsque les étudiants ont la consigne de rédiger un écrit réflexif selon une perspective de médecin senior et non la leur, cela ne stimule pas la démarche réflexive [1].

L'écriture réflexive fait ainsi partie des outils pédagogiques promus pour accompagner les étudiants dans le développement de leurs compétences professionnelles [5,14] et évaluer leurs apprentissages [12].

La réflexivité n'est pas une capacité intuitive ou innée, notamment en début de cursus médical [5–7]. Afin de guider son développement et son évaluation, plusieurs auteurs ont tenté d'élaborer des outils qui diffèrent tant dans leur forme qu'au regard du cadre théorique sur lequel ils s'appuient.

Wald *et al.* [6] ont élaboré un outil d'hétéro-évaluation en anglais afin de favoriser une évaluation fiable en vue de la rétroaction proposée aux étudiants durant les groupes d'écriture réflexive : la grille REFLECT, pour *Reflection Évaluation for Learners' Enhanced Competencies Tool*, que l'on pourrait traduire par « *Outil d'évaluation de la réflexivité pour améliorer les compétences des apprenants* ». Le cadre conceptuel de développement de cette grille s'appuie sur plusieurs auteurs de référence des

théories de la réflexivité : Mann *et al.* [15], Schön [16], Boud *et al.* [17], Moon [18] et Mezirow [19,20]. Ces derniers décrivent la réflexivité comme une analyse critique du savoir et de l'expérience, intégrant les émotions et le ressenti, afin d'atteindre un niveau approfondi de compréhension et de donner du sens, selon un processus complexe, qui nécessite une capacité de conceptualisation des bases de la réflexion (« *premise reflection* ») en vue de transformer et/ou renforcer les apprentissages [6,7,15,16,19]. Le processus intègre une dimension itérative (concept de boucle réflexive) et une dimension verticale hiérarchisant divers niveaux de réflexion [6,7,15,16,19].

D'autres outils ont été développés dans le cadre de l'éducation médicale.

Ainsi, l'outil *Reflection-on-Action*, a été développé à des fins de recherche pour l'évaluation, également, des traces écrites d'apprentissage ; il est unidimensionnel (une seule entité définie comme étant la capacité réflexive et évaluée en six niveaux) et a été comparé à la grille REFLECT dans l'étude de Miller-Kuhlmann *et al.* [21]. Au-delà du cadre de l'écriture réflexive, d'autres outils ont également été développés, comme la *Groningen Reflection Ability Scale* (GRAS) qui est une échelle unidimensionnelle couvrant trois aspects d'intérêt selon le cadre conceptuel d'élaboration [4,22]. Boenink *et al.* [1] ont pour leur part développé un outil semi-structuré d'hétéro-évaluation de la démarche réflexive des étudiants portant sur des vignettes cliniques. Autre exemple, la grille StARS<sup>®</sup> d'évaluation de la démarche réflexive des étudiants guidée par un questionnaire de six questions, suite au visionnage de vidéos de rencontres cliniques, considère trois éléments constitutifs de la réflexivité [23].

Il existe également des outils développés en dehors du contexte médical. On peut par exemple citer le travail de thèse de Michaud, portant sur l'évaluation du niveau de réflexivité du portfolio de futurs enseignants, à l'aide d'une grille déterminant un indice réflexif [24,25].

Divers arguments ont appuyé notre choix de travailler à partir de la grille REFLECT. Avant tout, elle a été développée dans un contexte de formation médicale, notamment de futurs médecins généralistes [6], au sein duquel les étudiants doivent rédiger des écrits réflexifs portant sur une situation clinique vécue, authentique. Or, dans notre contexte, à savoir la formation des internes en médecine générale, les traces écrites d'apprentissage attendues dans le portfolio visent également le développement de la réflexivité [26,27]. La rédaction des « *réécits de situation complexe authentique* » (RSCA), selon un modèle proposé par le Collège national des généralistes enseignants et appliqué dans la majorité des départements de médecine générale en France, a pour objectif de favoriser le développement des compétences des étudiants et leur autonomie d'apprentissage par le développement des capacités réflexives [28]. Ces RSCA sont des traces écrites d'apprentissage réflexives généralement longues (10–20 pages). Ils comprennent un récit à la première personne d'une situation vécue en stage par l'étudiant.

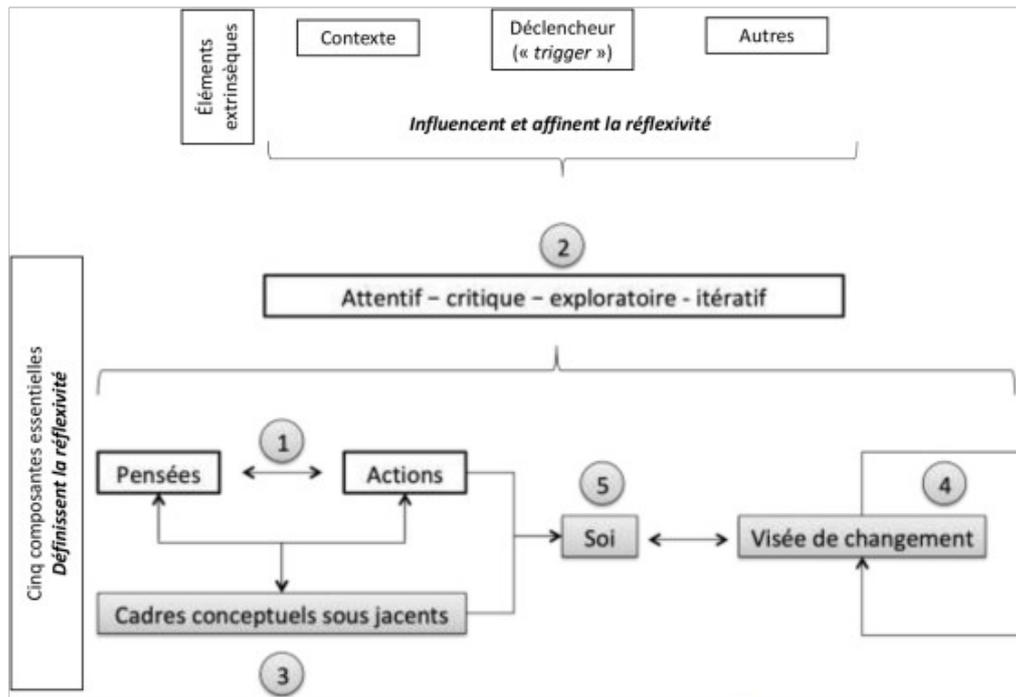


Fig. 1. Modèle conceptuel de la réflexivité d'après Nguyen *et al.* [7] (traduction libre).

Une section « analyse » vient ensuite, à distance de la situation, interpréter ce qui est rapporté dans le récit. Deux à trois problématiques contextualisées sont dégagées par l'étudiant. Une section « autoformation » présente ensuite les résultats des recherches de l'étudiant en regard de ces problématiques. Les résultats sont recontextualisés à la situation exemplaire. Une synthèse décrit finalement l'impact de ce travail sur les compétences de l'étudiant.

Les critères d'évaluation utilisés dans la grille REFLECT nous ont semblé particulièrement adaptés aux consignes de rédaction des RSCA et la grille d'appropriation plus aisée par des formateurs en santé que d'autres outils d'évaluation développés en dehors du contexte de l'éducation médicale.

Enfin, la grille REFLECT a été soumise à une démarche de validation et elle présente des caractéristiques intéressantes, notamment en termes de fiabilité [6,14]. Il n'y a, à notre connaissance, aucun outil d'évaluation de la réflexivité au travers des traces écrites d'apprentissage qui soit validé en langue française dans le contexte de l'éducation médicale.

Nous avons par ailleurs choisi d'ancrer notre travail dans la définition et le modèle théorique proposés par Nguyen *et al.* [7]. En effet, ils intègrent, du fait de la méthode mise en œuvre, les auteurs sur lesquels se sont appuyés Wald *et al.* [6,7] et, en cela, notre choix ne nous semble pas trahir les choix effectués lors de l'élaboration de la grille REFLECT en anglais. Le caractère synthétique de la définition et du modèle théorique nous semblait plus adapté pour la démarche d'évaluation de la validité conceptuelle de la traduction en français (voir Fig. 1).

## Objectif

En lien avec la problématique développée ci-dessus, l'objectif de cette étude était donc de réaliser une traduction et une adaptation culturelle de la grille REFLECT en français et d'en initier la validation.

## Méthodes

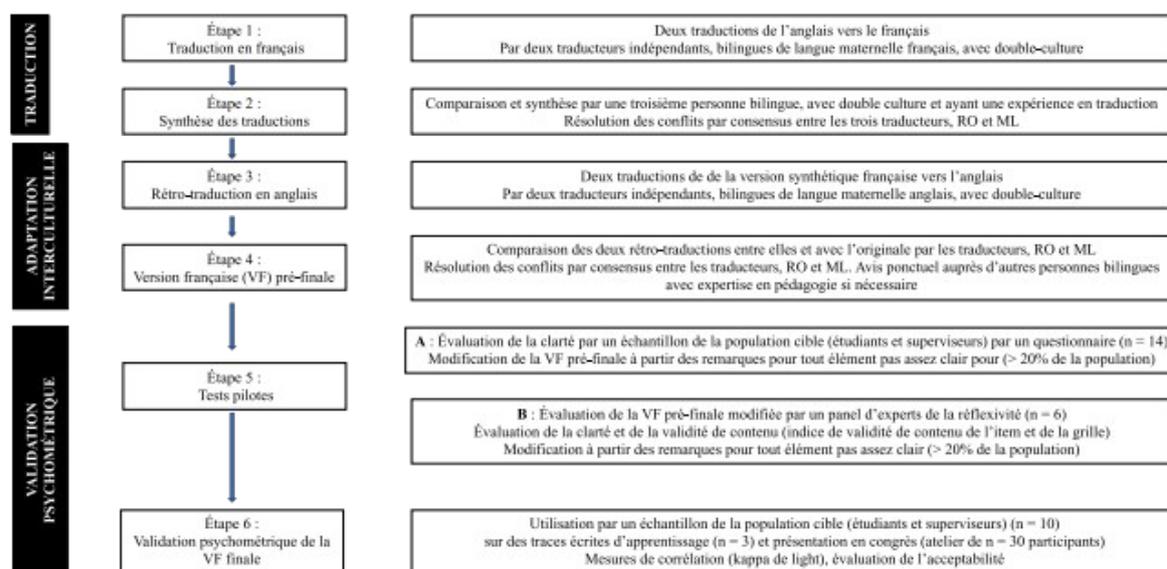
### L'outil : la version anglaise princeps de la grille REFLECT

La grille REFLECT [6] est un outil qualitatif évaluant six critères (propriété ayant la fonction de norme, qui exprime une qualité et qui permet de porter un jugement de valeur et de fonder une décision). L'un des critères est optionnel (respect des consignes). Les critères sont organisés en quatre niveaux selon un gradient de réflexivité allant d'« absence de réflexivité » à « réflexivité » (voir Tab. 1). Pour chaque niveau de chaque critère, une définition descriptive est donnée pour aider l'évaluateur à déterminer le niveau dont rend compte la trace écrite d'apprentissage.

Si la trace écrite d'apprentissage relève d'un niveau global « *critical reflection* », soit le niveau le plus élevé dans la grille en anglais, alors un deuxième axe d'évaluation est proposé. Ce dernier porte sur les conséquences de la réflexivité. Il permet de déterminer si le processus réflexif a conduit à des apprentissages transformateurs (AT) et/ou une confirmation des apprentissages antérieurs (CAA).

**Tableau I.** Schéma de la grille REFLECT originale.

	Axe I				Axe II	
	Niveau 1	Niveau 2	Niveau 3	Niveau 4	AT	CAA
Critère 1						
Critère 2						
Critère 3						
Critère 4						

**Fig. 2.** Diagramme de flux méthodologique.

Comme cela a déjà été indiqué, l'évaluation du dernier critère (critère 6) est optionnelle : il concerne le respect d'éventuelles consignes thématiques concernant le travail d'écriture (ce qui est le cas dans l'université de développement de la grille REFLECT). S'il n'y a pas de telles consignes, ce critère optionnel n'est pas pris en compte.

L'objectif de la grille REFLECT est de promouvoir le développement de la posture réflexive des étudiants à l'aide d'une évaluation formative et de l'enrichissement du *feedback* fourni par les superviseurs [6,21], ainsi que par l'auto-évaluation permettant une planification des apprentissages.

### Processus de traduction et d'adaptation culturelle

La méthode a été élaborée à partir des recommandations de bonne pratique proposées par Sousa et Rojjanasrirat [29]. Le processus suivi, comportant six étapes, est résumé à l'aide du diagramme de flux ci-dessus (Fig. 2).

Une double traduction vers le français a été effectuée par deux traducteurs bilingues indépendants, dont la

langue maternelle est le français mais ayant une double culture française et anglo-saxonne (étape 1). Leurs deux traductions ont ensuite été comparées par une troisième personne bilingue, ayant une expérience en traduction. Les discordances posant problème ont été résolues par consensus entre les traducteurs et les auteurs (RO et ML). Un avis ponctuel a été demandé pour certains termes à des experts en éducation médicale, bilingues (étape 2).

La version française (VF) préliminaire issue de ces deux premières étapes a été soumise à une étape de rétro-traduction, par deux autres traducteurs bilingues indépendants, de langue maternelle anglaise, et ayant également une double culture anglo-saxonne et française (étape 3). Cela permettait de s'assurer de l'absence de perte de sens lors de l'étape de traduction et mettait en lumière les termes ou expressions nécessitant une éventuelle clarification. Les discordances ont été résolues par consensus entre les traducteurs et les auteurs (RO et ML).

Ce cycle de traduction et rétro-traduction a permis d'élaborer une VF pré-finale.

Les tests pilotes (étape 5) ont été effectués en deux sous-étapes (5A et 5B).

Un panel de futurs utilisateurs constitué d'enseignants et d'étudiants de troisième cycle a été recruté pour évaluer la clarté de la VF pré-finale (étape 5A). Tous étaient spécialistes ou en cours de spécialisation en médecine générale, en raison de facilités de recrutement. Par ailleurs, la grille REFLECT avait aussi été développée et validée auprès des résidents en médecine de famille de l'Université de Brown [6].

Il leur était demandé, pour chacun des critères, intitulés des niveaux et descripteurs des niveaux, ainsi que pour la présentation des axes, si la formulation leur paraissait claire, insuffisamment ou pas du tout claire. Lorsqu'un participant répondait que la formulation était « insuffisamment ou pas du tout claire », il était demandé de proposer une modification qui rendrait la formulation plus intelligible. Lorsqu'un intitulé était considéré « insuffisamment ou pas du tout clair » par au moins 20 % des répondants, il était réévalué par RO et ML à l'aide des suggestions faites par l'échantillon.

Cette version pré-finale modifiée a alors été soumise à un panel de six experts concernant la réflexivité (étape 5B), pour évaluation de la clarté, à l'aide du même questionnaire que précédemment. Il leur a également été demandé d'évaluer la validité de contenu en lien avec le cadre théorique de Nguyen *et al.* [7,11], à l'aide d'une échelle ordinale. Il s'agissait d'une échelle de Lickert en quatre niveaux selon la classification suivante : 1 = *tout à fait pertinent*, 2 = *pertinent mais nécessite des ajustements*, 3 = *impossible de déterminer la pertinence*, 4 = *non pertinent*. Nous avons choisi d'adosser la version française de la grille REFLECT à la définition proposée par Nguyen *et al.* [7].

Les participants ont répondu à l'aide de questionnaires réalisés sur le logiciel LimeSurvey®.

La version finale (étape 6) a été présentée en atelier lors du 19<sup>e</sup> congrès national du Collège national des généralistes enseignants en novembre 2019 [30] et quelques modifications mineures ont été apportées suite aux retours des participants à l'atelier.

Nous avons évalué l'acceptabilité de cet outil en proposant une zone de commentaires libres dans le questionnaire LimeSurvey® ainsi qu'auprès des participants à l'atelier lors du congrès.

Le **tableau II** présente les caractéristiques des personnes ayant participé aux différentes étapes du travail de traduction et d'adaptation culturelle.

### Analyses statistiques

La validité de contenu a été évaluée par la mesure de l'index de validité de contenu (CVI) à l'échelle de l'item (I-CVI) et à l'échelle de la grille (S-CVI). L'I-CVI correspondait au nombre d'experts ayant évalué l'item comme pertinent (sans ou avec modifications mineures) divisé par le nombre total d'experts. Un seuil à 0,78 était visé. Le S-CVI correspondait à la moyenne des I-CVI, le seuil étant ici fixé à au moins 0,90.

## Résultats

### Traduction et adaptation interculturelle

Lors de la traduction de l'anglais vers le français (étape 1), certains termes spécifiques ont posé des difficultés aux traducteurs. Il s'agissait principalement de termes relativement spécifiques du champ des sciences de l'éducation ou du construit (« *meaning structures* », « *transformative reflection and learning* » ou encore « *confirmatory learning* » par exemple). Certaines expressions spécifiquement anglo-saxonnes sans équivalent exact en français étaient également concernées (par exemple « *insight* » ou « *meaning making* »). Les traducteurs ont proposé une traduction littérale de la grille, n'étant pas experts du champ de l'éducation médicale.

Lors de l'étape 2 (comparaison des deux traductions en vue d'une synthèse), RO et ML ont apporté l'expertise pédagogique, avec l'aide ponctuelle de correspondants bilingues, afin de réduire le risque de déformation de sens tout en recherchant une formulation adaptée au terrain d'utilisation future de la grille. Une dizaine d'arbitrages ont été réalisés.

La rétro-traduction a permis de s'assurer que les arbitrages et adaptations réalisés n'entraînaient pas de perte de sens voire de contresens par rapport à la version originale. Il n'y avait que peu de discordances entre les deux rétro-traductions ; elles correspondaient principalement à l'utilisation de synonymes ou expressions de signification similaire. La comparaison avec la version originale n'a pas décelé de contresens ou de différences majeures en dehors des termes spécifiquement anglo-saxons qui n'ont pas été rétro-traduits à l'identique par rapport à la version originale.

### Tests pilotes

#### Évaluation de la clarté de la grille

La VF pré-finale (matériel éditorial complémentaire ; document 1) a été évaluée par un panel de quatorze volontaires : des enseignants ( $n=7$ ) et des étudiants de troisième cycle ( $n=7$ ) en médecine générale.

Trente et un items sur les 39 évalués ont été considérés comme insuffisamment ou pas du tout clairs par plus de 20 % des répondants et ont donc été modifiés, à l'aide des suggestions proposées par les répondants. Ils apparaissent surlignés sur le document 1.

Il est intéressant de noter que, là encore, parmi les éléments ayant posé des difficultés aux volontaires, qu'ils soient enseignants ou étudiants, on retrouvait notamment les traductions proposées pour « *meaning structures* », « *transformative reflection and learning* », « *confirmatory learning* », ainsi que « *insight* » ou « *meaning making* », qui avaient été sources de difficultés pour l'étape de traduction, malgré le glossaire que nous avons proposé pour clarifier ces éléments.

À l'issue de cette étape, le critère optionnel a été supprimé. Seuls 21 % des répondants ont trouvé que ce critère était insuffisamment ou pas du tout clair, mais ils

**Tableau II.** Tableau synoptique des personnes impliquées à chaque étape.

Étape	Statut	
Étape 1 : traduction en Français	Traducteur 1	Médecin généraliste, ancienne CCU-MG, bilingue, langue maternelle Français et double culture franco-britannique
	Traducteur 2	Avocate, bilingue, langue maternelle Français et double culture franco-britannique
Étape 2 : synthèse	Traducteur 3	Ancienne traductrice professionnelle Français-Anglais, double culture franco-américaine
	Traducteur 4	Journaliste, bilingue, langue maternelle Anglais, double culture franco-britannique
Étape 3 : rétro-traduction	Traducteur 5	Enseignant d'Anglais à l'Université, traducteur, bilingue, langue maternelle Anglais, double culture franco-britannique
Étape 5A : tests pilotes		Sept enseignants de médecine générale issus de différents DMG Sept internes de médecine générale issus de différentes subdivisions
	Expert 1	Enseignante de médecine générale, recherches sur l'évaluation de la réflexivité
	Expert 2	Cadre supérieur de santé formatrice, titulaire M2 ingénierie de la formation et des compétences, responsable d'un centre d'enseignement des soins d'urgences, enseignante en pédagogie des sciences de la santé
	Expert 3	Professeur Émérite, faculté des sciences de l'éducation. Axes et thématiques recherche : pédagogie de l'enseignement supérieur, des sciences de la santé, auto-formation en langues
	Expert 4	Docteur en sciences de l'éducation, Formatrice à l'entretien d'explicitation, auteure d'un ouvrage portant sur la pratique réflexive
	Expert 5	MD, MA, MPH, PhD candidate, chercheur dans le champ de l'éducation médicale, thématique de recherche : réflexion et pratique réflexive en pédagogie médicale
Étape 5B : test pilotes (experts)	Expert 6	PhD. en sciences de l'éducation, Professeur à la faculté des sciences de l'éducation,
		Sept enseignants de médecine générale issus de différents DMG Trois internes de médecine générale issus de différentes subdivisions

CCU-MG = Chef de clinique des universités en médecine générale ; DMG = Département de médecine générale ; M2 = Master 2 ; MD. = *Medicinæ Doctor* ; MA = *Maitrise ès Arts* ; MPH = *Master of Public Health* (Master de santé publique) ; PhD candidate = *Doctorant* (candidat au titre de *Philosophiæ Doctor*) ; PhD = *Philosophiæ Doctor*.

étaient 29 % à trouver la description du niveau 3 insuffisamment ou pas du tout claire et 36 % pour le niveau 4. Leurs commentaires et suggestions montraient un risque d'incompréhension et donc de réponse erronée à ce critère (« *je ne comprends pas vraiment ce que veut dire « choix d'une question alternative »* », proposition de remplacer par « *suis les recommandations [de bonne pratique des soins] applicables* »). Ce critère n'était pas directement en lien avec la réflexivité, nous avons fait le choix de le supprimer de la version française.

Nous avons également modifié la mise en page de la grille, en raison de plusieurs commentaires suggérant de mieux marquer la différence entre l'évaluation du niveau de réflexivité en elle-même et l'axe II d'évaluation des impacts de la réflexivité sur les apprentissages. C'est pour cela que nous avons marqué la séparation entre les deux axes d'évaluation par un espace vide, et ajouté une flèche précisant les conditions d'évaluation de l'axe II.

#### Évaluation de la validité de contenu et de la clarté par les experts

La grille soumise aux experts (matériel éditorial complémentaire ; document 3) était issue des modifications réalisées suite aux tests pilotes auprès de l'échantillon de futurs utilisateurs.

Après discussion entre les auteurs, il n'a pas semblé pertinent d'évaluer les intitulés des quatre niveaux de réflexivité (« *absence de réflexion sur l'action* », « *ébauche de réflexion sur l'action* », « *réflexion sur l'action* » et « *réflexion approfondie sur l'action (réflexivité)* ») par rapport au cadre conceptuel.

Les six experts ont donc évalué 30 items, en lien avec le cadre conceptuel choisi et les cinq composantes essentielles de la réflexivité [7] : 80 % des items ont un indice de validité de contenu à 1 et 20 % ont un indice de validité de contenu à 0,83. L'indice de validité de contenu global (à l'échelle de la grille) est de 0,97.

Ils ont également évalué la clarté de 34 items car, pour cette évaluation, les intitulés des quatre niveaux de réflexivité ont été soumis aux experts : 16 (47 %) ont été évalués « *insuffisamment ou pas du tout clairs* » mais seulement 5 (15 %) l'ont été par 50 % des experts ou plus. Pour ces derniers, les propositions pour clarifier l'intitulé étaient consensuelles pour trois items (60 %). Pour les 11 items évalués « *insuffisamment ou pas du tout clairs* » par 33 % des experts, les remarques étaient consensuelles pour seulement trois d'entre eux (27 %). Seuls les items pour lesquels les commentaires concordaient n'ont pas nécessité d'arbitrage particulier.

Nous avons mené une dizaine d'arbitrages. Quatre items n'ont pas été modifiés après réévaluation, soit car la modification aurait été en incohérence avec le cadre conceptuel, soit parce que les propositions s'opposaient aux commentaires faits dans l'étape 5A. Lorsque les remarques des experts étaient trop divergentes pour un même item, nous n'avons retenu que les propositions de modification en cohérence avec le cadre conceptuel et l'esprit de la version originale.

### Éléments de validation psychométrique de la version francophone de la grille REFLECT

Utilisation par un échantillon de la population cible

Un échantillon de 10 futurs utilisateurs a accepté de tester la grille sur trois traces d'apprentissage. Six réponses n'étaient pas exploitables et ont été exclues. L'analyse de la concordance retrouvait un coefficient  $kappa = 0,521 [-0,133 ; 0,714]$ .

Présentation de la grille en congrès

Nous avons précisé la formulation du critère « *Profondeur du récit* » en raison d'incompréhensions liées à la spécificité des traces réflexives demandées en médecine générale (le terme « *récit* » fait référence à un chapitre de la trace écrite d'apprentissage) par rapport au contexte de développement de la grille (où le mot « *récit* » renvoie à l'ensemble de la trace d'apprentissage qui, comme le montre l'exemple disponible en matériel éditorial complémentaire (document 2), n'est pas divisée en chapitres).

La plupart des enseignants et internes en médecine générale ayant participé à ce travail ainsi que des participants à l'atelier durant le Congrès du Collège national des généralistes enseignants ont exprimé des retours positifs quant à l'utilité perçue et souhaitent intégrer cet outil à leur pratique. Une partie d'entre eux, des étudiants mais aussi enseignants, ont également souligné un impact positif sur leur compréhension de la posture réflexive.

## Discussion

### Une version francophone de la grille REFLECT pour évaluer la réflexivité

Suite aux retours effectués par l'échantillon d'enseignants et d'étudiants, la majorité des descripteurs de niveaux sont formulés de façon plus synthétique et ont été

simplifiés par rapport à la traduction initiale. À l'inverse, certains descripteurs de niveaux ont été étoffés, notamment lorsqu'ils paraissaient trop abstraits ou insuffisamment explicites pour les évaluateurs. Nous avons également été amenés à modifier certains descripteurs pour marquer la différence entre deux niveaux. Il est important de noter que la différenciation parfois insuffisante entre les niveaux fait partie des critiques qui ont pu être formulées à l'encontre de la grille REFLECT originale [21]. Toutefois ces modifications n'ont pas dénaturé l'esprit de la version anglophone de la grille REFLECT. Pour mémoire, pour chaque niveau de chacun des critères était proposée une définition descriptive et c'était donc à l'étudiant et/ou à la personne évaluatrice de rechercher et documenter les extraits narratifs permettant d'argumenter le niveau atteint. Ces extraits constituent des indicateurs. La version française garde ce format et ne prescrit aucun indicateur précis.

Le critère optionnel a été totalement supprimé et la mise en page de la grille modifiée afin de faire apparaître plus clairement le processus d'évaluation en deux étapes, bien qu'il soit explicité dans les instructions d'utilisation de la grille (matériel éditorial complémentaire ; document 3). Par ailleurs, nous avons choisi de joindre un glossaire pour expliciter certains termes ou expressions qui nous paraissaient complexes ou bien utilisés dans un sens très spécifique. Tous ces éléments constituent des différences importantes avec la grille d'origine. Toutefois, cela est nécessaire pour rendre la grille accessible à l'ensemble des utilisateurs, notamment les étudiants, sans nécessiter une formation spécifique.

Une autre différence est le choix du cadre conceptuel sous-jacent ; toutefois la définition de Nguyen *et al.* [7] englobe le cadre théorique de développement de la version originale. En plus d'être englobante, elle est plus récente que la grille REFLECT et fait référence dans le milieu de l'éducation médicale francophone [11].

Par ailleurs, l'indice de validité de contenu, tant au niveau de chaque item qu'au niveau global, démontre que la version française présente une validité de contenu élevée.

De plus, la formulation de la question posée aux experts permet également d'apprécier en partie la validité de construit, puisque nous interrogeons la pertinence par rapport à la définition de la réflexivité proposée par Nguyen *et al.* [7]. Toutefois, l'absence d'outil de référence ne permet pas d'effectuer d'analyse de la convergence. Il pourrait être intéressant d'évaluer la validité divergente par rapport à d'autres outils n'analysant pas le même construit.

### Forces et limites de l'étude

La mise en œuvre d'une méthode rigoureuse pour les étapes de traduction et d'adaptation interculturelle constitue la force majeure de cette étude. Cela permet d'établir une équivalence conceptuelle, sémantique et de contenu dès le départ entre l'outil d'origine et la traduction [29].

Ensuite, nous avons inclus des étudiants dans l'échantillon de la population cible, ce qui est novateur. En effet, s'agissant d'un outil d'évaluation développé dans une visée formative, il doit être accessible non seulement pour les enseignants mais également pour des étudiants n'ayant pas forcément de maîtrise du domaine pédagogique.

Toutefois, l'étude n'est pas exempte de limites.

En effet, aucun des traducteurs n'avait d'expertise dans le domaine des sciences de l'éducation. Ainsi, cela a pu conduire à une traduction trop littérale de certaines expressions, avec un risque de perte de nuances ou de sens. Toutefois, ce risque a été limité par la vérification point par point par les auteurs RO et ML, et par le recours à des experts en pédagogie médicale bilingues pour les termes posant des difficultés particulières, que ce soit d'ordre conceptuel ou purement sémantique.

Lors de la rétro-traduction, nous obtenions des versions qui différaient de la version originale notamment sur les termes problématiques, qui ont nécessité des ajustements. Certains termes étaient trop spécifiques de l'anglais et la traduction a consisté en une paraphrase à défaut de pouvoir proposer un terme ou une expression correspondant à une traduction *stricto sensu*.

Par ailleurs, aucun des traducteurs participants n'était certifié et seul le traducteur de la deuxième étape avait une expérience professionnelle dans le domaine de la traduction. Cela a pu majorer le risque d'inexactitudes dans la traduction.

La taille de l'échantillon de la population cible correspond aux recommandations de bonne pratique [29]. Concernant les experts, la taille de l'échantillon est aussi en adéquation avec les recommandations de Sousa et Rojjanasrirat [29] mais n'atteint pas la taille idéale de dix personnes. Cela nous a mis en difficulté pour l'évaluation de la clarté de la grille par les experts car le seuil de 20 % était rapidement atteint. De plus, lorsque seulement deux experts sur six trouvaient la formulation « insuffisamment ou pas du tout claire » (11 items/16), dans plus de 70 % des cas le motif invoqué et la proposition de modification ne concordait pas. Nous avons donc été contraints de réaliser un nombre important d'arbitrages lors de la réévaluation de ces items. Par ailleurs, deux experts étaient Québécois, ce qui a pu influencer l'évaluation de la clarté puisque le français québécois présente des singularités par rapport au français dit international (français de Paris) [31]. À l'inverse, cela peut aussi renforcer la possibilité de diffuser cette grille dans d'autres pays francophones, même si nous semble que pour l'affirmer avec certitude, un travail supplémentaire d'adaptation culturelle serait nécessaire.

Notre travail de traduction, rigoureux, constitue donc un préalable indispensable, pour pouvoir utiliser la grille REFLECT en France, et éventuellement dans d'autres pays francophones.

Il ne s'agit toutefois que d'un travail préliminaire, qui ne permet pas de conclure au regard de l'ensemble des indicateurs de crédibilité, fiabilité et transférabilité. Pour ce faire, un travail de validation approfondi reste nécessaire afin de s'assurer que notre outil présente des garanties suffisantes. Le choix du modèle de validation sera important

à considérer, afin qu'il soit adapté aux caractéristiques de l'outil. Il pourrait par exemple être intéressant de s'appuyer sur le modèle de Kane [32], ou un autre modèle adapté à la démarche de validation d'outils d'évaluation qualitatifs.

### Considérations concernant l'évaluation de la réflexivité des traces écrites d'apprentissage

L'évaluation est une composante essentielle du parcours de professionnalisation en sciences de la santé : non seulement elle doit répondre à l'objectif ultime d'attester que les étudiants sont devenus, à l'issue de leur cursus, des professionnels compétents, mais ses modalités déterminent la façon dont l'apprenant va étudier [5,14,33].

Ainsi, si l'on considère que la réflexivité est une compétence indispensable à acquérir et développer en vue d'un exercice professionnel équilibré, il semble utile de développer des outils permettant de la documenter et de l'évaluer [4], d'autant qu'une rétroaction peut être nécessaire pour faciliter ou permettre l'acquisition d'une véritable posture réflexive [5].

Toutefois, pour certains auteurs [2,34], l'impératif académique peut décourager les étudiants de s'engager dans une démarche honnête et ouverte d'analyse réflexive, voire les pousser à ne choisir que des situations d'analyses dont ils estiment qu'elles sont acceptables selon le cadre professionnel, ou à inventer des situations. Ils peuvent même encore simplement simuler la démarche réflexive par l'usage de termes et expressions permettant de faire illusion : il ne faut pas oublier que, s'agissant d'un processus métacognitif, les traces écrites d'apprentissage ne sont qu'un reflet indirect de la démarche engagée par l'apprenant [35]. Il s'agirait alors d'une démarche contre-productive sur le plan réflexif.

La non-prescription d'indicateurs précis et forcément limités en nombre par la grille REFLECT permet de limiter ce risque. En effet, c'est à l'étudiant de réfléchir et de se questionner sur les indicateurs présents dans ses traces écrites d'apprentissage, qui permettent de documenter la démarche réflexive menée sur la situation professionnelle relatée.

Par ailleurs, il ne faut pas négliger le risque de biais d'évaluation, les évaluateurs risquant notamment de prendre en compte le contenu de la prise en charge effectuée plus que le processus réflexif en lui-même [2,7]. Certains auteurs en concluent même qu'évaluer la réflexivité revient à tenter de marier deux processus incompatibles [2].

Sans avoir un avis aussi tranché, il semble important de garder en tête que l'influence d'une évaluation sommative sur la nature et la qualité des apprentissages recueillis dans le portfolio est encore insuffisamment documentée, y compris pour l'écriture réflexive [12,14], bien que le portfolio soit un outil d'évaluation recommandé notamment pour les construits de haut niveau taxonomique [36]. Ces questions se posent également en cas d'usage à visée purement formative, les conséquences des outils d'évaluation en termes de développement de la réflexivité et d'engagement des étudiants dans la démarche réflexive étant encore peu étudié [21].

Autre élément non négligeable, l'écriture réflexive requiert des compétences rédactionnelles et narratives qui ne relèvent pas des capacités réflexives de l'apprenant [35], frein qui a été soulevé par l'un des experts ayant évalué la version française de la grille REFLECT.

D'autres modalités d'évaluation de la démarche réflexive pourraient donc être envisagées, en complément de celle s'appuyant sur les traces écrites d'apprentissage. Cela nécessiterait de s'appuyer sur d'autres outils, adaptés, par exemple, à l'utilisation au cours d'entretiens de supervision directe ou indirecte en stage, ou encore lors d'échanges avec les tuteurs.

Par ailleurs, il importe de garder en tête que le processus réflexif est complexe et multidimensionnel : son évaluation ne doit pas se contenter d'une dichotomie entre réflexif et non réflexif. Chaque composante du modèle théorique et le processus réflexif de façon générale doivent être développés et envisagés selon un continuum de progression jusqu'à la démarche réflexive aboutie intégrant les cinq composantes [7], sans réduire la démarche réflexive au fait de cocher des cases [34].

Enfin, l'un des enjeux majeurs de l'évaluation de la réflexivité est lié à la diversité des modèles théoriques et définitions [35]. Notre choix de cadre théorique [7] constitue de ce point de vue un atout. En effet, sa méthodologie d'élaboration en fait un modèle synthétique, regroupant et combinant les principales perspectives et théories des quinze auteurs les plus influents sur le sujet. Bien que nous n'ayons pas élaboré un outil *de novo* à partir de cette théorie, les composantes essentielles de la réflexivité [7] trouvent une traduction concrète dans la version française de la grille REFLECT. Cela est conforté par l'indice de validité de contenu élevé que nous mettons en évidence. Toutefois, la validité de construit de la version française de la grille REFLECT nécessiterait une évaluation plus approfondie.

Il est intéressant de noter que les étudiants comme les enseignants ayant évalué la clarté de la version française nous ont fait part de leur intérêt pour l'outil REFLECT. Les étudiants y voient un outil de guidage de la démarche réflexive et de clarification des attentes face au terme « *écriture réflexive* », tandis que les enseignants y voient une réponse au moins partielle aux difficultés qu'ils rencontrent à évaluer correctement la réflexivité et surtout à fournir ensuite une rétroaction argumentée s'appuyant sur un « *langage commun avec les étudiants* ».

En lien avec la démarche de validation de la version française de la grille REFLECT, l'évaluation de son impact sur le développement effectif de la réflexivité des étudiants sera à documenter par des travaux ultérieurs.

## Conclusion

La version française de la grille REFLECT est un outil d'évaluation de la réflexivité au travers des traces écrites d'apprentissage en contexte d'éducation médicale. Les résultats préliminaires concernant la validité de contenu sont tout à fait satisfaisants et c'est un outil qui permettra vraisemblablement d'optimiser le développement de la réflexivité chez les internes en médecine générale, de faciliter

la rétroaction pour les superviseurs et d'harmoniser les pratiques d'évaluation de la réflexivité au travers du portfolio.

Des études dans d'autres disciplines médicales et paramédicales seront nécessaires pour évaluer la généralisabilité de nos résultats.

## Contributions

Racha Onaisi a participé à la conception du protocole de recherche, au recueil des données, à l'interprétation des résultats et à l'écriture du manuscrit. François Severac a participé à la conception du protocole, à l'analyse statistique, à l'écriture et à la relecture du manuscrit. Mathieu Lorenzo a participé à la conception du protocole de recherche, au recueil des données, à l'interprétation des résultats, à l'écriture et à la relecture du manuscrit.

## Liens d'intérêts

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

## Approbation éthique

L'étude a obtenu un avis favorable de la part du comité d'éthique de la Société internationale francophone d'éducation médicale (SIFEM) en date du 20/02/2019.

Elle a fait l'objet d'une déclaration sur le site de l'Université de Strasbourg concernant l'utilisation des données des participants.

## Valorisation académique

Ce travail a été élaboré dans le cadre d'un travail académique (mémoire de master de pédagogie des sciences de la santé, Université de Strasbourg) et la version française de la grille REFLECT a fait l'objet d'une présentation lors d'un atelier pendant le 19<sup>e</sup> Congrès du Collège national des généralistes enseignants le 29 Novembre 2019 à Nantes.

## Remerciements

Un immense merci à Clara, Emily, Caroline, Andrew et Darragh d'avoir accepté d'effectuer ce travail de traduction. Merci au Pr Loye et ses étudiants, ainsi qu'au Pr Acker pour leur aide face à certains casse-têtes.

Merci également à l'ensemble des internes et enseignants de médecine générale ayant répondu à l'appel pour la validation de cet outil en français, et aux experts, Pr Nguyen, Pr Fernandez, Pr Poteaux, Dr Cabannes, Mme Balas-Chanel et Mme Brunstein, de nous avoir accordé de leur temps pour répondre aux questionnaires.

Merci au Pr Hedy Wald pour son accord concernant la traduction de la grille REFLECT et ses conseils, ainsi qu'au Pr Shmuel Reis.

Merci au Pr Dory et au Pr Dionne pour leurs conseils concernant la démarche de validation d'un outil d'évaluation.

## Annexe 1: Version française finale de la grille REFLECT.

Axe I : Niveaux de réflexivité de l'étudiant au travers de la trace écrite d'apprentissage					Axe II : Impacts de la réflexivité sur les apprentissages	
Critère					Apprentissages transformateurs	Confirmation des apprentissages antérieurs
	Absence de réflexion sur l'action	Ébauche de réflexion sur l'action	Réflexion sur l'action	Réflexion approfondie sur l'action (réflexivité)		
<b>Profondeur du récit (et de l'analyse)</b>	Description superficielle (description de faits, impressions vagues) sans réflexion ou introspection	Description détaillée de la situation et impressions sans réflexion	Écriture au-delà du rapport de faits ou de la description approchant de la réflexion (tentative de compréhension, de remise en question, ou d'analyse de l'événement sans analyse des conséquences de l'action)	Exploration et critique des préjugés, valeurs, croyances, biais, et des conséquences de l'action (immédiates et futures)	Les cadres de référence* ou structures de raisonnement* sont transformés, le champ de compétences est élargi.	Les cadres de référence et structures de raisonnement sont validés dans la situation en question.
<b>Implication de l'auteur</b>	Impression que le récit est impersonnel, l'auteur étant peu ou pas impliqué	Impression que l'auteur n'est que partiellement impliqué dans le récit	Impression que l'auteur est largement impliqué, permettant au moins partiellement d'accéder à ses pensées et d'identifier le rôle qu'il/elle a eu	Impression que l'auteur est pleinement impliqué tout au long du récit, permettant d'accéder à ses pensées et d'identifier le rôle qu'il/elle a eu	Intégration perceptible de nouveaux apprentissages, façonnant ses futures perceptions, ses émotions, ses attitudes, ses capacités de compréhension intuitive*, ses actions et le sens donné aux situations.	
<b>Description des problématiques</b>	Aucune description des problèmes, questionnements et/ou difficultés soulevés par la situation en question	Description insuffisante des problèmes, questionnements et/ou difficultés soulevés par la situation en question	Description détaillée des problèmes, questionnements et/ou difficultés soulevés par la situation en question	Description complète des problèmes, questionnements et/ou difficultés soulevés par la situation en question avec présentation de plusieurs perspectives, recherche de différentes explications et mise à l'épreuve d'hypothèses	Dégage une impression d'avancée majeure.	
<b>Attention portée à ses propres émotions</b>	Peu ou pas de prise en compte ou d'attention portée à ses émotions	Reconnaissance de ses émotions sans prise en compte ou exploration	Reconnaissance, exploration, et prise en compte de ses émotions	Reconnaissance, exploration, prise en compte et amélioration des capacités de gestion de ses émotions		
<b>Analyse et recherche de sens*</b>	Aucune analyse ou recherche de sens	Ébauche d'analyse ou de recherche de sens	Analyse et recherche de sens incomplètes	Analyse et recherche de sens complètes et approfondies		



## Glossaire

Certains des termes de la grille REFLECT en Français renvoient à des concepts qu'il est utile d'expliquer. Les définitions ci-dessous permettent aux formateurs et aux apprenants d'acquérir un vocabulaire commun pour évaluer la réflexivité et favoriser son développement.

**Cadres de référence :**

La notion de cadres de référence renvoie à un ensemble d'éléments objectifs (exemples : faits, connaissances, bases théoriques...) ou subjectifs (préjugés, croyances...) acceptés par une personne et qui influencent sa perception d'une situation, ses décisions, opinions, ou comportements.

**Capacités de compréhension intuitive :**

Là encore, cette expression renvoie dans la version anglaise au terme d'*insight*, cette fois utilisé seul. Face à une situation donnée, c'est la capacité d'appréhender la nature profonde de cette situation et/ou d'en avoir une perception intuitive pour la comprendre et agir en conséquence. Toutefois, à la différence d'une simple intuition, la notion d'*insight* repose aussi sur les connaissances du sujet, qui alimentent cette compréhension intuitive. La « compréhension intuitive » renvoie aussi à ce caractère très singulier, chaque sujet appréhendant un problème de manière personnelle afin de le résoudre.

**Recherche de sens :**

L'analyse de la situation doit permettre de donner une signification personnelle aux événements, selon ses cadres de référence. Cette recherche/construction de sens contribue ainsi à une meilleure compréhension de la situation.

**Structures de raisonnement :**

C'est la façon dont le sujet construit son raisonnement et sa réflexion face à une situation ou un problème pour lui donner un sens, une signification. Ces structures ou schémas de raisonnement se transforment par la réflexion et l'expérience d'apprentissage.

## Guide d'utilisation

La grille REFLECT est une traduction validée d'un outil développé en langue anglaise pour évaluer le niveau de réflexivité des étudiants au travers de leurs traces écrites d'apprentissage. Elle permet d'offrir une rétroaction de qualité qui optimise l'apprentissage de la réflexivité.

L'application de la grille REFLECT à une trace d'apprentissage réflexive s'opère en quatre étapes :

1. Lecture de l'ensemble du récit
2. Fragmentation : « zoomez » sur les détails du récit (expressions, phrases) pour évaluer la présence et la qualité de tous les critères. Déterminez quel niveau correspond à chaque critère.
3. Réévaluation du récit dans son ensemble : Prenez du recul pour considérer le récit dans son ensemble (tout en prenant en compte l'analyse détaillée que vous venez de réaliser à l'étape 2). Déterminez le niveau atteint du récit en entier. Déterminez si l'un ou l'autre des impacts de la démarche réflexive en termes d'apprentissage (apprentissages transformateurs et/ou confirmation des apprentissages antérieurs) ont été également atteints.
4. Argumentation du choix des niveaux atteints et impacts sur les apprentissages en citant des extraits du texte. Ne « lisez pas entre les lignes »

## Matériel éditorial complémentaire

*Document 1:* Version française pré-finale (à l'issue de l'étape 4).

*Document 2:* Traduction de l'exemple d'utilisation de la grille REFLECT en médecine générale proposé par Wald et al. adapté avec la terminologie de la version française.

*Document 3:* Version française pré-finale modifiée à l'issue de l'étape 5A (version soumise aux experts).

Le matériel éditorial complémentaire est disponible sur <http://www.pedagogie-medicale.org/10.1051/pmed/2021001/olm>.

## Références

- Boenink AD, Oderwald AK, De Jonge P, Van Tilburg W, Smal JA. Assessing student reflection in medical practice. The development of an observer-rated instrument: reliability, validity and initial experiences. *Med Educ* 2004;38:368-77.
- Hargreaves J. So how do you feel about that? Assessing reflective practice. *Nurse Educ Today* 2004;24:196-201.
- Sobral DT. Medical students' mindset for reflective learning: a revalidation study of the reflection-in-learning scale. *Adv Health Sci Educ Theory Pract* 2005;10:303-14.
- Aukes LC, Geertsma J, Cohen-Schotanus J, Zwierstra RP, Slaets JPJ. The development of a scale to measure personal reflection in medical practice and education. *Med Teach* 2007;29:177-82.
- Sandars J. The use of reflection in medical education: AMEE Guide No. 44. *Med Teach* 2009;31:685-95.
- Wald HS, Borkan JM, Taylor JS, Anthony D, Reis SP. Fostering and evaluating reflective capacity in medical education: developing the REFLECT rubric for assessing reflective writing. *Acad Med* 2012;87:41-50.
- Nguyen QD, Fernandez N, Karsenti T, Charlin B. What is reflection? A conceptual analysis of major definitions and a proposal of a five-component model. *Med Educ* 2014;1176-89.
- Chouinard I, Caron J. Le recours aux approches réflexives dans les métiers relationnels : modélisation des conceptions de la réflexivité. *Phronesis* 2015;4:11-21.
- Vacher Y. Construire une pratique réflexive : comprendre et agir. Bruxelles : De Boeck supérieur, 2015.
- Perrenoud P. Développer la pratique réflexive dans le métier d'enseignant : professionnalisation et raison pédagogique. Issy-les-Moulineaux : ESF, 2012.
- Nguyen QD, Raymond-Carrier S. Un professionnel de santé qui se pose des questions : la réflexivité. In : Pelaccia T (Dir.). Comment (mieux) former et évaluer les étudiants en médecine et en sciences de la santé ? Louvain-La-Neuve : De Boeck supérieur, 2016;45-62.
- Naccache N, Samson L, Jouquan J. Le portfolio en éducation des sciences de la santé : un outil d'apprentissage, de développement professionnel et d'évaluation. *Pédagogie Médicale* 2006;7:110-27.
- Challis M. AMEE Medical Education Guide No.11 (revised): Portfolio-based learning and assessment in medical education. *Med Teach* 1999;21:370-86.
- Moniz T, Arntfield S, Miller K, Lingard L, Watling C, Regehr G. Considerations in the use of reflective writing for student assessment: issues of reliability and validity. *Med Educ* 2015;49:901-8.
- Mann K, Gordon J, MacLeod A. Reflection and reflective practice in health professions education: a systematic review. *Adv Health Sci Educ Theory Pract* 2009;14:595-621.
- Schön DA. Le praticien réflexif : à la recherche du savoir caché dans l'agir professionnel. Montréal : Les Éditions Logiques, 1994.
- Boud D, Keogh R, Walker D. (Eds.). Reflection: Turning Experience Into Learning. London: Routledge, 1985.
- Moon JA. A Handbook of Reflective and Experiential Learning: Theory and Practice. London: Routledge, 2013.
- Lundgren H, Poell RF. On Critical Reflection: A Review of Mezirow's Theory and Its Operationalization. *Hum Resour Dev Rev* 2016;15:3-28.
- Mezirow J. Transformative Dimensions of Adult Learning. San Francisco: Jossey-Bass, 1991.
- Miller-Kuhlmann R, O'Sullivan PS, Aronson L. Essential steps in developing best practices to assess reflective skill: A comparison of two rubrics. *Med Teach* 2016;38:75-81.
- Andersen NB, O'Neill L, Gormsen LK, Hvidberg L, Moreke AM. A validation study of the psychometric properties of the Groningen Reflection Ability Scale. *BMC Med Educ* 2014;14:214.
- Koole S, Dornan T, Aper L, De Wever B, Scherpbier A, Valcke M, et al. Using video-cases to assess student reflection: development and validation of an instrument. *BMC Med Educ* 2012;12:22.
- Michaud C. Le portfolio : un en-(je)u de formation et de développement professionnel [thèse]. Lyon : Université Claude Bernard - Lyon I, 2010.
- Michaud C. Le portfolio, quel lien entre les écrits réflexifs et les compétences ? *Mesure et évaluation en éducation* 2012;35:9-38.
- CNGE - Collège National des Généralistes Enseignants. Concepts et principes pédagogiques. 2013 [On-line]. Disponible sur : [https://www.cnge.fr/la\\_pedagogie/concept\\_s\\_et\\_principes\\_pedagogiques/](https://www.cnge.fr/la_pedagogie/concept_s_et_principes_pedagogiques/)
- CNGE - Collège National des Généralistes Enseignants. Introduction à l'évaluation. 2012 [On-line]. Disponible sur : [https://www.cnge.fr/la\\_pedagogie/troisieme\\_cycle/introduction\\_levaulation/](https://www.cnge.fr/la_pedagogie/troisieme_cycle/introduction_levaulation/)
- Taha A. RSCA : récit de situation complexe authentique : de l'idée à la réalisation. Paris : Global Media Santé/CNGE productions, 2018.
- Sousa VD, Rojjanasirath W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline: Validation of instruments or scales. *J Eval Clin Pract* 2011;17:268-74.
- Onaisi-Chantrel R, Lorenzo M. Mesurer la réflexivité : la grille REFLECT. Atelier organisé lors du 19<sup>e</sup> Congrès national du Collège national des généralistes enseignants, 27-29 novembre 2019, Nantes. 2019. [On-line] Résumé disponible sur : <http://abstracts.congrescngc.fr/>
- Bigot D, Papen RA. Sur la «norme» du français oral au Québec (et au Canada en général). *Langage et Société* 2013;146:115-32.
- Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med Educ* 2015;49:560-75.
- Pelaccia T (Dir.). Comment (mieux) former et évaluer les étudiants en médecine et en sciences de la santé ? Louvain-la-Neuve : De Boeck Supérieur, 2016.
- de la Croix A, Veen M. The reflective zombie: Problematising the conceptual framework of reflection in medical education. *Perspect Med Educ* 2018;7:394-400.

35. Koole S, Dornan T, Aper L, Scherpbier A, Valcke M, Cohen-Schotanus J, *et al.* Factors confounding the assessment of reflection: a critical review. *BMC Med Educ* 2011;11:104.
36. Shumway JM, Harden RM. AMEE Guide No. 25: The assessment of learning outcomes for the competent and reflective physician. *Med Teach* 2003;25:569-84.

**Citation de l'article** : Onaisi R, Severac F, Lorenzo M. Évaluer la réflexivité à travers les traces écrites d'apprentissage des étudiants en santé : traduction et adaptation interculturelle de la grille REFLECT. *Pédagogie Médicale* 2021;22:15-26

## Synthèse

Ce travail a produit un outil concret utile en pédagogie des sciences de la santé. Si l'on s'accorde à dire que la réflexivité fait partie des qualités (du profil) souhaité chez les futurs professionnels de santé, les outils validés pour accompagner son développement sont rares.

La grille REFLECT a ainsi le potentiel pour favoriser ce développement. En permettant à l'étudiant d'identifier clairement dès le début de sa formation ce qui est attendu d'un écrit réflexif. En offrant au lecteur d'un tel écrit réflexif une base concrète pour évaluer la réflexivité d'un étudiant puis en guidant la rétro-action. Dans ces deux cas, l'évaluation dans une perspective formative pourrait être plus performante (dans le sens permettre de mieux développer) et pourrait permettre de former des professionnels de santé davantage réflexif.

Plusieurs perspectives de recherche découlent de ces hypothèses. Il pourrait s'agir tout d'abord de vérifier si l'utilisation de la grille REFLECT par les étudiants leur permet de produire des écrits réflexifs plus conformes aux attentes des formateurs. Il pourrait ensuite s'agir de vérifier si l'utilisation de la grille REFLECT permet de former des professionnels de santé davantage réflexifs par comparaison à un groupe qui ne l'utiliserait pas. Il pourrait enfin s'agir de regarder sur des critères "durs" comme la qualité des soins et l'impact sur la santé des populations d'une telle intervention. Ces trois pistes de recherche sont de difficulté croissante.

En appliquant le modèle MOT à ce travail, plusieurs réflexions supplémentaires émergent. La métacognition apparaît dans la procédure "réguler le processus" qui influence la représentation dynamique du problème que se forge le clinicien tout au long d'une rencontre avec un patient. Cette procédure métacognitive a comme composantes trois sous-procédures : "analyser ses propres activités cognitives", "évaluer son propre processus cognitif" et "intervenir sur son processus cognitif". Bien que le modèle MOT n'explique pas suffisamment les types de connaissances sous-tendant ce processus métacognitif, il est possible d'établir un parallèle avec le concept de réflexivité et son évaluation via la grille REFLECT. Le point commun est que la réflexivité et la procédure "réguler le processus" demandent au "soi" d'examiner ses "pensées et actions" et dans une certaine mesure les "cadres conceptuels sous-jacents". Si ce n'est explicite dans le modèle MOT, l'intervention sur le processus cognitif semble "viser le changement" – son optique est d'améliorer le processus de raisonnement clinique. Ainsi, l'utilisation de la grille REFLECT a le potentiel en favorisant le développement de la réflexivité des professionnels de

santé d'améliorer leur raisonnement clinique. Ces hypothèses théoriques devront bien entendu être testées dans de futures études.

#### LE LIEN ENTRE LES FORMATIONS PAR CONCORDANCE ET LA RÉFLEXIVITÉ

Parmi les outils cités dans différents ouvrages et recommandations visant à accompagner le développement de la réflexivité en pédagogie des sciences de la santé, aucun ne mentionne les formations par concordance (17,166,168,179–182). Pourtant, un article de 2016 de Fernandez *et al.* affirme que les formations par concordance développent la réflexivité des étudiants (129). Cette affirmation se base sur l'analyse effectuée par les auteurs des commentaires libres d'un questionnaire d'évaluation proposé aux étudiants ayant réalisé une formation par concordance à l'Université de Montréal (129). D'un point de vue méthodologique, cette affirmation semble fragile : l'analyse qualitative de questions ouvertes d'un tel questionnaire est en général peu robuste (183).

Dans ce contexte, j'ai souhaité réaliser une étude qui viserait à documenter le développement potentiel de la réflexivité au travers des formations par concordance. Pour cela, j'ai proposé à une étudiante du master en pédagogie des sciences de la santé (Elodie Hernandez) de collaborer sur cette étude. Ce travail a été possible grâce à l'aide de Léa Charton et Fabien Rougerie qui m'ont aidé à monter la formation par concordance pour laquelle l'impact sur la réflexivité a été étudiée.

#### Contexte de l'étude

Nous avons développé un programme de formation par concordance de raisonnement (FpC-raisonnement) en ligne sur l'interprétation de l'électrocardiogramme (ECG) en médecine générale à l'Université de Strasbourg. Une vignette clinique textuelle est présentée aux apprenants, relatant une situation réelle impliquant un ECG. Il leur est demandé d'interpréter l'ECG et de formuler des hypothèses diagnostiques sur la situation clinique. Une fois cette étape franchie, ils accèdent à une deuxième page où ils doivent comparer leur interprétation et leurs hypothèses avec celles proposées par un panel de cardiologues experts, et remplir une proposition de prise en charge du patient en tenant compte des avis des cardiologues experts. Sur la troisième page, ils doivent comparer leur proposition de prise en charge avec celles d'un panel de médecins généralistes experts. Enfin, ils accèdent à une synthèse sur le cas, rédigée par un médecin généraliste expert dans le domaine de l'interprétation des ECG, complétée par des liens vers des références sur le même thème.

Nous avons mené une étude exploratoire afin de documenter l'impact de cette FpC-raisonnement à l'interprétation de l'ECG sur la réflexivité des participants.

### Matériel et méthodes

La formation a été proposée en ligne à tous les étudiants post-gradués de médecine générale de la Faculté de Médecine de Strasbourg ainsi que sur la base du volontariat à ceux de la Faculté de Médecine de Besançon. Les étudiants étaient libres d'effectuer la formation quand ils le souhaitaient. Il était proposé aux étudiants de participer à des focus groupes pour explorer la nature de leurs apprentissages à la fin de la formation.

Deux focus groupes ont été menés avec les participants volontaires. Les participants ne connaissaient que le thème général de la recherche. Le seul critère d'inclusion était le fait d'avoir complété la formation. Il n'y avait pas de critère d'exclusion. L'enquêtrice présentait son travail de recherche en expliquant la méthode qualitative, la préservation de l'anonymat des participants, recueillait le consentement oral pour l'enregistrement audio du focus group et sa retranscription.

Le guide d'entretien a été conçu au préalable et soumis à un groupe d'expert en recherche qualitative avant le recueil des données mais n'avait pas été testé au préalable. Le guide a été adapté en cours d'étude selon l'analyse des données du premier focus group. L'enquêtrice a utilisé un enregistrement audio afin de pouvoir réaliser une retranscription intégrale des verbatim. Des notes de terrain ont été prises avant les entretiens afin de décrire les participants. Les participants pouvaient demander à ce que les entretiens leur soient transmis pour modification ou annotation mais aucun ne l'a demandé.

Une analyse thématique déductive et inductive a été réalisée sur les retranscriptions de ces focus groupes par ML et EH. Le codage déductif consistait à rechercher la présence des cinq composantes de la réflexivité selon Nguyen *et al.*

### Résultats

Trente-huit internes ont été volontaires pour participer aux focus groupes. Deux focus groupes ont été réalisés en mai et juin 2018, regroupant 10 participants au total. Le premier focus groupe a duré 1h25 et le second 1h50.

L'analyse thématique comportait 203 codes regroupés selon quatre composantes de la réflexivité : la composante pensées et actions, la composante attentive, critique, exploratoire et itérative (ACEI), la composante visant le changement et la composante du soi.

#### Discussion

Sur la base du modèle de Nguyen *et al.* (162), l'analyse de ces données démontre la présence de quatre des cinq composantes de la réflexion. Pendant et après la FpC-raisonnement, les participants semblent s'être engagés dans l'interprétation ECG dans des interactions attentives, critiques, exploratoires et itératives entre leurs pensées et leurs actions. Ces interactions étaient orientées vers le changement (s'améliorer). Ces résultats ne confirment que partiellement les données de la littérature sur l'impact des formations par concordance sur le développement de la réflexivité (129). La composante des " cadres conceptuels sous-jacents " n'a pas été retrouvée dans nos données. Il est donc impossible de conclure que les interactions observées dans notre étude sont réellement réflexives, compte tenu de cette composante manquante (162). Il est probable que la méthodologie utilisée dans notre étude n'était pas propice à l'exploration des cadres conceptuels sous-jacents. Des entretiens d'explicitation auraient pu nous permettre de mieux rechercher la présence de cette composante (184).

Les formations par concordance pourraient donc soit augmenter réellement la réflexivité chez les apprenants, soit simplement faciliter des processus de pensée métacognitifs génériques (179). Comme la réflexivité doit être considérée comme un continuum et non comme un processus dichotomique (170,185-188), ces données suggèrent malgré tout que les spécificités des formations par concordance pourraient favoriser le développement de la réflexivité (129).

L'article intégral dans sa version actuelle (en anglais) se trouve ci-dessous.

1 **Reflection and Learning-by-Concordance**

2

3 **A qualitative exploratory study on the impact of an ECG learning by concordance**  
4 **training program on reflection**

5

6 Elodie Hernandez<sup>1,2,+</sup>, Léa Charton<sup>3,+</sup>, Fabien Rougerie<sup>3</sup>, Mathieu Lorenzo<sup>2,3</sup>

7

8 <sup>1</sup> Département de Médecine Générale de Besançon, Faculté de Médecine de Besançon, France

9 <sup>2</sup> Centre de Formation et de Recherche en Pédagogie des Sciences de la Santé (CFRPS), Faculté  
10 de Médecine de Strasbourg, France

11 <sup>3</sup> Département de Médecine Générale de Strasbourg, Faculté de Médecine de Strasbourg,  
12 France

13

14 <sup>+</sup> Léa Charton and Elodie Hernandez contributed equally to this work.

15

16 Corresponding author: Mathieu Lorenzo, Centre de Formation et de Recherche en Pédagogie  
17 des Sciences de la Santé (CFRPS), Faculté de Médecine de Strasbourg, 4 rue Kirschleger,  
18 67000 Strasbourg, France

19 Telephone: +33 (0)3 88 11 69 66

20 Fax number: +33 (0)3 88 11 69 67

21 E-mail: [mlorenzo@unistra.fr](mailto:mlorenzo@unistra.fr)

22

23

24

25 **Abstract**

26

27 *Introduction*

28 Learning-by-Concordance (LbC) is a fairly recent approach that introduces learners to the  
29 complexity and uncertainty of clinical practice. Some data on LbC suggest that it stimulates  
30 reflection in future clinicians. As we developed an online LbC training program on  
31 electrocardiogram (ECG) interpretation in general practice at the University of Strasbourg,  
32 France, we conducted an exploratory qualitative study to document the impact of this ECG  
33 learning-by-concordance training program on reflection in participants.

34 *Materials and methods*

35 We created a 18 clinical vignettes LbC training program on ECG interpretation based on a  
36 review of the literature on frequent and serious cardiovascular diseases that can be identified  
37 with an ECG in general practice. The training program was delivered online to postgraduate  
38 general practice students in two faculties of medicine. We conducted a qualitative study based  
39 on the thematic analysis of two focus groups. Inductive and deductive coding were performed.  
40 The five major components of reflection of Nguyen model were used in the deductive coding:  
41 (i) thoughts and actions, (ii) attentive, critical, exploratory and iterative processes (ACEI), (iii)  
42 underlying conceptual frame, (iv) change and (v) self.

43 *Results*

44 Two focus groups were convened with a total of ten participants. The qualitative analysis  
45 comprised 203 codes divided into four groups based on components of reflection in the Nguyen  
46 model: (i) the self, (ii) attentive, critical, exploratory and iterative interactions with (iii) one's  
47 thoughts and actions, with (iv) a view on the change itself. Inductive coding revealed interesting  
48 thoughts about the impact of the identity of the panel members and the absence of scoring  
49 system.

50 *Discussion*

51 This study only partially supports the assumption that the use of LbC applied to ECG  
52 interpretation could foster reflection in future general practitioners. Further studies designed to  
53 explore the underlying conceptual frames of the reflection process should be undertaken.

54

55 **Keywords**

56 Reflection, Learning by concordance, Clinical reasoning, Assessment, ECG interpretation

57 **Introduction**

58

59 Learning-by-Concordance (LbC) is a fairly recent approach that introduces learners to the  
60 complexity and uncertainty of clinical practice.<sup>1</sup> Best known format is the script concordance  
61 test which is broadly recognize as an useful tool for assessing clinical reasoning.<sup>1-3</sup> In the past  
62 5 years, other LbC format have been described with more emphases on justification of  
63 reasoning than on reasoning result.<sup>1</sup> Some data on these type of LbC suggest that it stimulates  
64 reflection in future clinicians.<sup>1</sup> As described by Fernandez et al., all LbC tools shares common  
65 features.<sup>1</sup> A real-life clinical situation is described briefly, in a short statement with a picture or  
66 a video. A question a professional would entertain in that situation is presented. Participants'  
67 answers are registered and targeted feedback by experts is generated. The feedback includes  
68 justifications given by the experts about their responses and key-messages provided by a  
69 recognized scholar in the field, including hyper-links to complementary resources (e.g.  
70 scientific paper, website, etc.). Expert responses may reflect various relevant solutions to the  
71 same situation.<sup>1</sup>

72

73 Reflection is therefore defined according to Nguyen *et al.* as “the process of engaging the self  
74 in attentive, critical, exploratory and iterative interactions with one’s thoughts and actions, and  
75 their underlying conceptual frame, with a view to changing them and a view on the change  
76 itself”.<sup>4</sup> Reflection is a desired skill for caregivers, as it could contribute to improve their  
77 performance,<sup>5-8</sup> any other definitions of reflection are used in medical education since almost  
78 one century. Definitions of reflection from preeminent authors like Dewey<sup>9</sup> or Schön<sup>10</sup> are  
79 widely employed. However, as reflection is a complex construct, some authors argued that  
80 using a common explicit understanding of reflection could help teaching and conducting

81 research project on it. The Nguyen *et al.* definition and model of reflection is now commonly  
82 used in our medical education community.<sup>11</sup>

83

84 Electrocardiogram (ECG) interpretation is a simple paraclinical test for checking for heart  
85 diseases. ECG interpretation performance is poor among both medical students and doctors  
86 with possible impact on patient outcomes.<sup>12,13</sup> Interventions as checklists to try help learners did  
87 not have an overall effect in reducing diagnostic errors in ECG interpretation.<sup>14</sup> No single  
88 approach to or format of ECG teaching seem to be superior when compared each to the other.<sup>15</sup>  
89 However, self-directed learning seems to be associated with poorer interpretation performance  
90 and computer-based learning may be advantageous.<sup>15</sup>

91

92 Considering this, we developed an online LbC training program on electrocardiogram (ECG)  
93 interpretation in general practice at the University of Strasbourg, France. A textual clinical  
94 vignette was presented to learners, relating a real-life situation involving an ECG. They are  
95 asked to interpret the ECG and to formulate diagnostic hypotheses on the clinical situation.  
96 Once this step was successfully completed, they accessed a second page where they had to  
97 compare their interpretation and hypotheses with those offered by a panel of expert  
98 cardiologists, and fill out a patient care proposal, taking the opinions of the expert cardiologists  
99 into consideration. On the third page, they compared their care proposal with those made by a  
100 panel of expert family doctors. Lastly, they had to read a synthesis on the case, drafted by a  
101 family physician with expertise in the field of ECG interpretation, complete with links to  
102 bibliographical references.

103

104 We conducted an exploratory study to document the impact of this ECG learning-by-  
105 concordance training program on reflection in participants.

106

107 **Materials and methods**

108

109 *ECG LbC design*

110

111 LC and ML conducted a review of the literature on frequent and serious cardiovascular diseases  
112 that can be identified with an ECG in general practice. Twenty-four clinical vignettes pertaining  
113 to these diseases as well as two vignettes featuring a normal ECG were drafted by LC, FR and  
114 another experienced family physician. These vignettes were first submitted to a seven-member  
115 panel at the general practice department (GPD) of the University of Strasbourg's Faculty of  
116 Medicine. The panel was asked to assess the authenticity of the vignettes on a Likert scale from  
117 0 (no authenticity) to 4 (maximal authenticity) and to suggest possible changes. Two vignettes  
118 with an authenticity score below 2.5/4 were removed, and two others were rewritten according  
119 to suggestions made by members of the panel.

120 The remaining vignettes were submitted to a panel of eight volunteer cardiologists, who filled  
121 out their interpretation of the ECG on the basis of the clinical vignette. Two vignettes were  
122 removed at that stage due to significant disagreements in the interpretation of ECG strips.

123 Four members of the Strasbourg GPD who had access to the cardiologists' ECG interpretations  
124 then drafted care proposals. Two vignettes were removed at that stage due to significant  
125 disagreements in the proposals.

126 Lastly, FR who had an expertise in the use of ECGs in general practice drafted a synthesis and  
127 listed bibliographic references on the theme of the vignette. The training program in its final  
128 form comprised eighteen clinical vignettes.

129

130 One of the eighteen vignettes was used as a tutorial at the beginning of the training, with  
131 suggestions of responses and advice on how to use expert panel feedback. The goal of this  
132 tutorial was to address difficulties in the appropriation of LbC as described in the literature.<sup>16</sup>  
133 Unlike in the LbC tools previously described in the literature,<sup>1,16,17</sup> no score is calculated. This  
134 program is purely training-oriented and aimed at improving the knowledge of students on  
135 situations involving an ECG and encourage reflection. We therefore choose to not calculate any  
136 score. We did not display the identities of panel members to the learners as we did not collect  
137 their permissions to do so.

138

#### 139 *Selection of expert panels*

140

141 The panel of cardiologists was assembled by randomly contacting thirty independent and  
142 hospital cardiologists in the Alsace and Rhône-Alpes regions of France. Eight agreed to  
143 participate in the study.

144 The panel of family physicians was assembled on a voluntary basis among the twelve members  
145 of the GPD with an expertise in ECG interpretation. Four agreed to participate in the study.

146

#### 147 *Participants*

148

149 The training program was offered online to all postgraduate general practice students at the  
150 University of Strasbourg's Faculty of Medicine, and University of Besançon's Faculty of  
151 Medicine (both in France). Students were free to do the training when they wished for a period  
152 of three months. They were asked to participate in focus groups to explore the nature of what  
153 they had learned by the end of the program.

154

155 *Qualitative study of reflection*

156

157 A qualitative study was conducted, based on the thematic analysis of two focus groups  
158 convened in the Faculties of Medicine of the Universities of Strasbourg and Besançon. We  
159 choose focus groups to encourage participants to confront their points of view about the impact  
160 of the LbC training program on reflection and identifying convergent opinions.<sup>18</sup>

161

162 EH, was the junior lecturer in general practice at the Faculty of Medicine of Besançon. She  
163 coordinated the focus groups. She was trained in qualitative research and had been involved in  
164 focus groups both as participant and investigator. She maintained a logbook throughout the  
165 research.

166

167 EH knew some of the participants in the Faculty of Medicine of Besançon. She specified her  
168 role within the framework of the focus group and the distinction between her roles as a  
169 researcher and a teacher. The age difference between the participants and the investigator was  
170 small. A scribe who the participants did not know participated in the second focus group only.

171

172 Focus group participants were only aware of the general theme of the research. The only  
173 criterion for inclusion was having completed the training program and volunteering. There were  
174 no exclusion criteria. The investigator presented her research by explaining the qualitative  
175 method, the preservation of anonymity for participants, secured their oral consent for the audio  
176 recording and transcription of the focus group.

177

178 The students who were available to participate at the proposed dates were contacted via email.  
179 The data were collected in a restaurant in April 2018 (first focus group) and in the University  
180 of Besançon's Faculty of Medicine in May 2018 (second focus group).

181

182 The interview guide was conceived beforehand and submitted to a group of qualitative research  
183 experts before the data collection phase, but not tested. The guide was adapted during the  
184 analysis of data from the first focus group (see appendices). The investigator used an audio  
185 recording to transcribe the discussions extensively. Field notes were taken before the interviews  
186 to describe the participants. The participants were notified that they could ask for the transcripts  
187 for changes or comments, but none requested them. All the verbatim was anonymized.

188

189 EH coded the entirety of the data using the NVivo ® software in a thematic approach. Co-  
190 coding was performed by ML on a random sample of 15 percent of transcripts. A 90 percent  
191 consensus on coding was pursued through discussion between the authors.<sup>19</sup> Inductive and  
192 deductive coding were performed. The five major components of reflection of Nguyen model  
193 was used in the deductive coding: (i) thoughts and actions, (ii) attentive, critical, exploratory  
194 and iterative processes (ACEI), (iii) underlying conceptual frame, (iv) change and (v) self.<sup>4</sup>  
195 Data saturation was not sought as we were in an exploratory perspective and for pragmatic  
196 considerations, such as participant availability. Inductive coding was inspired by grounded-  
197 theory approach. It was performed to allow unexpected aspect of the impact of LbC on  
198 reflection emerge from our data.

199 The data processing begins in April 2018 with the first focus group and finish in December  
200 2018 with the consensus on coding.

201

202 *Ethics approval and consent to participate*

203 This study was approved by the ethics committee of the Groupe Hospitalier Mulhouse Sud-  
204 Alsace on 28 May 2018. A declaration was filed to the French National Commission on  
205 Information Technology and Liberties (CNIL) on 27 April 2018. Informed consents to  
206 participate in the study were obtained from all participants. All identifying details were  
207 anonymized.

208

## 209 **Results**

210

211 Thirty-eight students volunteered to participate in the focus groups. The two focus groups were  
212 convened in May and June 2018, bringing together a total of ten participants. The first focus  
213 group lasted 1hr 25 min and the second 1hr 50 min. The single motive for non-participation  
214 was scheduling issues. Table 1 presents the characteristics of the participants.

215

### 216 *Deductive coding*

217 The qualitative analysis comprised 203 codes divided into four groups based on components of  
218 reflection in Nguyen model: thoughts and actions ; attentive, critical, exploratory and iterative  
219 process (ACEI) ; change and self. We had 79 codes in the ACEI component, 63 in self  
220 component, 25 for the change component and 36 for the thoughts and actions component. We  
221 had no code for the underlying conceptual frames component.

222

### 223 *The “thoughts and actions” component*

224 As expected, the thoughts and actions of participants were subjects of reflection: “*you realize*  
225 *that there are several possible answers and that it’s not because yours is a bit different that*  
226 *you’ve got it wrong, necessarily*” (participant no. 8).

227 *“About the ECG, since it’s normal, I don’t do anything, but obviously you reassure [the*  
228 *patient]. Of course, right? Now [during the internship] obviously you always reassure people,*  
229 *you don’t even realize, actually”* (participant no. 3).

230

231

232 *The ACEI component*

233 Many codes derived from the analysis of interview transcripts correspond to an attentive,  
234 critical, exploratory and iterative thought process: *“you start reflecting on things, really*  
235 *reflecting”* (attentive thought - participant no.4).

236 *“I would put myself in the situation, I’d tell myself, I’m the doctor, this patient is my patient,*  
237 *and yeah, I was really totally in the case”* (attentive thought – participant no. 8)

238 *“So, the thing is, it made me ask myself by the end of the 18 cases: have you been dangerous*  
239 *or not?”* (critical attitude - participant n°10).

240 *“You’re going to be judged, they’ll say: “wait a minute, X wrote this, she missed the infarcts,*  
241 *she didn’t send him [to the emergency ward] he’d be dead [...],” but the fact that this is*  
242 *completely anonymous [for the participants], I think that for a training program like this it*  
243 *really is very important”* (exploratory thought – participant no. 3).

244 *“I’d reviewed that same case during my internship, I’d actually reread [the case presented*  
245 *during the training], because I’d downloaded it, I’d clicked on the link again, sent it to myself*  
246 *over email”* (iterative thought - participant no. 1).

247 *“Well, I adjusted to the answers, anyway – for instance, about reassurance there was a first*  
248 *case where I hadn’t written it down, and in the answer, he put reassurance, and so in the last*  
249 *case where it was anxiety and all that, I put reassurance”* (iterative thought – participant no. 9).

250

251 The anonymity of panel members was a particular concern for participants, unsure of what  
252 value to attribute to the answers offered: “*they tell us cardiology specialists but we don’t know*  
253 *who they are at all*” (participant no. 2) – “*The thing is, we know cardiologists, who... you*  
254 *know... [do a bad job] [...] so, people we don’t trust completely*” (critical thought - participant  
255 no. 1).

256

### 257 *The change component*

258 A number of the utterances made by participants illustrated a desire to change their thought  
259 processes: “*for instance, you see [atrial fibrillation] [...] I told myself, ‘OK, we have the right*  
260 *to handle that ourselves’*” (participant no. 1).

261 “*I made sure to do things a little bit more seriously, because, you know, there are actually*  
262 *important things to monitor, so yeah, that and being thorough when I read my ECGs*”  
263 (participant no. 4) – “*Same here, yeah, it’s true that often we tend to just glance at the thing,*  
264 *maybe we don’t necessarily, you know [laughs] do it as well as we did when we were students*”  
265 (participant no. 7) – “*do a better job of read the ECG*” (participant no. 5).

266

### 267 *The self component*

268 The participants easily connected their thoughts to their ‘selves’ during and after the training  
269 program: “*My way of reasoning is fairly logical, compared to the others, I’m a little bit*  
270 *reassured, I’ve gained a little bit of trust in myself*” (participant no. 3).

271 The absence of grades and the purely learning-oriented goal of the training program was  
272 important for this component: “*I think it isn’t bad, doing it like this, because in the end when I*  
273 *did it I wasn’t under any pressure, and so you know, you’re not stressed out. When you’re being*  
274 *graded or evaluated, you get stressed out, whereas this is really just for us*” (participant no. 5).

275

276 *Inductive coding*

277 Inductive coding revealed interesting thoughts about the impact of the identity of the panel  
278 members and the absence of scoring system.

279 Participants were unsure of what value to attribute to the answers offered. This seems to foster  
280 a critical distance with the role model that recognized panel members could have: “*they tell us*  
281 *cardiology specialists but we don’t know who they are at all [in the ECG LbC training program*  
282 *panel]”* (participant no. 2) – “*The thing is, we know cardiologists, who... you know... [do a bad*  
283 *job] [...] so, people we don’t trust completely [the cardiologists panel]”*. These parts were also  
284 coded as critical thoughts in the ACEI component.

285

286 The absence of scoring system and the assessment for learning goal of the training program  
287 was important for the participants: “*I think it isn’t bad [the ECG LbC training program], doing*  
288 *it like this, because in the end when I did it, I wasn’t under any pressure, and so you know,*  
289 *you’re not stressed out. When you’re being graded or evaluated, you get stressed out, whereas*  
290 *this is really just for us”* (participant no. 5). This part was also coded as an attentive thought of  
291 the ACEI component.

292

### 293 **Discussion**

294

295 Based on the model in Nguyen et al.,<sup>4</sup> analysis of the data from this study demonstrates the  
296 presence of four out of five components of reflection. During and after the LbC program,  
297 participants appear to have engaged in ECG interpretations characterized by attentive, critical,  
298 exploratory and iterative interactions between their thoughts and actions. These interactions  
299 were change-oriented. These findings only partially support the literature data on the impact of  
300 LbC in stimulating reflection.<sup>1</sup> The “underlying conceptual frame” component was not found



301 in our data. It is therefore impossible to conclude that the interactions observed in our study are  
302 actually reflective, considering this missing component.<sup>4</sup> It is likely that the methodology used  
303 in our study was not conducive to exploring underlying conceptual frames. Elicitation  
304 interviews could have allowed us to probe the presence of this component.<sup>20</sup>

305

306 Learning-by-Concordance could thus either actually increase reflection among learners,<sup>1</sup> or  
307 simply facilitate generic metacognitive thought processes.<sup>21</sup> As reflection must be considered  
308 as a continuum, not a dichotomic process,<sup>22-26</sup> these data suggest that the specificities of LbC  
309 could favor the development of reflection.<sup>1</sup>

310

311 Two methodological choices appear to deserve particular consideration when the objective of  
312 a LbC program is to boost reflection among learners: the anonymity of panel members and the  
313 calculation of a score. While the identity of panel members is usually announced at the  
314 beginning of the proceedings,<sup>1,27</sup> not disclosing it appears to have reinforced the ACEI  
315 component of reflection in our study. Other studies will have to investigate the importance of  
316 this parameter, but we advise trainers to reflect on this aspect when developing a LbC tool.  
317 Likewise, while a score is actually calculated after each question or upon completion of the  
318 LbC,<sup>1,27-30</sup> our choice of refraining from using scores also appears to have reinforced the ACEI  
319 component. This choice makes learner self-assessment the only form of evaluation in the  
320 program. We believe this approach is consistent with the choices recommended in competency-  
321 based approaches, and one that should facilitate the inclusion of LbC in competence-based  
322 curricula.<sup>31-34</sup>

323

324 Ultimately, the factors observing a likely development of reflection in this study on a LbC  
325 program on ECG interpretation appear to support the necessary self-regulation of lifelong

351 anonymity of panel members and the absence of scoring seem to have the potential to enhanced  
352 reflection in participants in a LbC training program.

353

#### 354 **Acknowledgements**

355

356 This article received support from the Maison Interuniversitaire des Sciences de l'Homme  
357 d'Alsace (MISHA) and the Excellence Initiative of the University of Strasbourg, France.

358

#### 359 **Disclosure of interest**

360

361 The authors report no conflict of interest.

362

#### 363 **Data availability**

364

365 Verbatim of the focus groups are available on request to the corresponding author.

366

#### 367 **References**

368

369 1. Fernandez N, Foucault A, Dubé S, et al. Learning-by-Concordance (LbC): introducing  
370 undergraduate students to the complexity and uncertainty of clinical practice. *Can Med Educ J*.  
371 2016;7(2):e104-e113.

372 2. Lubarsky S, Durning S, Charlin B. AM last page. The script concordance test: a tool for  
373 assessing clinical data interpretation under conditions of uncertainty. *Acad Med J Assoc Am*  
374 *Med Coll*. 2014;89(7):1089. doi:10.1097/ACM.0000000000000315

375 3. Cooke S, Lemay J-F. Transforming Medical Assessment: Integrating Uncertainty Into

- 376 the Evaluation of Clinical Reasoning in Medical Education. *Acad Med J Assoc Am Med Coll.*  
377 Published online January 31, 2017. doi:10.1097/ACM.0000000000001559
- 378 4. Nguyen QD, Fernandez N, Karsenti T, Charlin B. What is reflection? A conceptual  
379 analysis of major definitions and a proposal of a five-component model. *Med Educ.*  
380 2014;48(12):1176-1189. doi:10.1111/medu.12583
- 381 5. Maudsley G, Strivens J. Promoting professional knowledge, experiential learning and  
382 critical thinking for medical students. *Med Educ.* 2000;34(7):535-544.
- 383 6. Guest CB, Regehr G, Tiberius RG. The life long challenge of expertise. *Med Educ.*  
384 2001;35(1):78-81.
- 385 7. Epstein RM. Mindful practice. *JAMA.* 1999;282(9):833-839.
- 386 8. Mamede S, Schmidt HG. The structure of reflective practice in medicine. *Med Educ.*  
387 2004;38(12):1302-1308. doi:10.1111/j.1365-2929.2004.01917.x
- 388 9. Dewey J. *How We Think: A Restatement of the Relation of Reflective Thinking to the*  
389 *Educative Process.*; 1933.
- 390 10. Schon DA. *The Reflective Practitioner: How Professionals Think In Action.* Basic  
391 Books; 1984.
- 392 11. Nguyen QD, Raymond-Carrier S. Un professionnel de santé qui se pose des questions :  
393 la réflexivité. In: *Comment (Mieux) Former et Évaluer Les Étudiants En Médecine et En*  
394 *Sciences de La Santé ?*. De Boeck supérieur; 2016:45-62.
- 395 12. Goy JJ, Schlaepfer J, Stauffer J-C. Competency in interpretation of 12-lead  
396 electrocardiogram among Swiss doctors. *Swiss Med Wkly.* 2013;143(2122).  
397 doi:10.4414/smw.2013.13806
- 398 13. Jablonover RS, Lundberg E, Zhang Y, Stagnaro-Green A. Competency in  
399 Electrocardiogram Interpretation Among Graduating Medical Students. *Teach Learn Med.*  
400 2014;26(3):279-284. doi:10.1080/10401334.2014.918882

- 401 14. Sibbald M, Sherbino J, Ilgen JS, et al. Debiasing versus knowledge retrieval checklists  
402 to reduce diagnostic error in ECG interpretation. *Adv Health Sci Educ Theory Pract*. Published  
403 online January 29, 2019. doi:10.1007/s10459-019-09875-8
- 404 15. Fent G, Gosai J, Purva M. Teaching the interpretation of electrocardiograms: which  
405 method is best? *J Electrocardiol*. 2015;48(2):190-193. doi:10.1016/j.jelectrocard.2014.12.014
- 406 16. Foucault A, Dubé S, Fernandez N, Gagnon R, Charlin B. Learning medical  
407 professionalism with the online concordance-of-judgment learning tool (CJLT): A pilot study.  
408 *Med Teach*. 2015;37(10):955-960. doi:10.3109/0142159X.2014.970986
- 409 17. Charlin B, van der Vleuten C. Standardized assessment of reasoning in contexts of  
410 uncertainty: the script concordance approach. *Eval Health Prof*. 2004;27(3):304-319.  
411 doi:10.1177/0163278704267043
- 412 18. Lingard L, Kennedy TJ. Qualitative Research Methods in Medical Education. In:  
413 *Understanding Medical Education*. John Wiley & Sons, Ltd; 2010:323-335.  
414 doi:10.1002/9781444320282.ch22
- 415 19. Miles MB, Huberman AM. *Analyse des données qualitatives*. De Boeck Supérieur;  
416 2003.
- 417 20. Pelaccia T, Tardif J. *Comment [mieux] former et évaluer les étudiants en médecine et*  
418 *en sciences de la santé?* De Boeck supérieur; 2016.
- 419 21. Cleary TJ. The use of reflection in medical education: Guide supplement 44.1--  
420 Viewpoint. *Med Teach*. 2011;33(6):500-501. doi:10.3109/0142159X.2010.540596
- 421 22. Jarvis P. Reflective practice and nursing. *Nurse Educ Today*. 1992;12(3):174-181.
- 422 23. Boyd LD. Development of reflective judgement in the pre-doctoral dental clinical  
423 curriculum. *Eur J Dent Educ Off J Assoc Dent Educ Eur*. 2008;12(3):149-158.  
424 doi:10.1111/j.1600-0579.2008.00511.x
- 425 24. Wald HS, Davis SW, Reis SP, Monroe AD, Borkan JM. Reflecting on reflections:

426 enhancement of medical education curriculum with structured field notes and guided feedback.  
427 *Acad Med J Assoc Am Med Coll.* 2009;84(7):830-837. doi:10.1097/ACM.0b013e3181a8592f

428 25. de la Croix A, Veen M. The reflective zombie: Problematizing the conceptual  
429 framework of reflection in medical education. *Perspect Med Educ.* 2018;7(6):394-400.  
430 doi:10.1007/s40037-018-0479-9

431 26. Miller-Kuhlmann R, O'Sullivan PS, Aronson L. Essential steps in developing best  
432 practices to assess reflective skill: A comparison of two rubrics. *Med Teach.* 2016;38(1):75-81.  
433 doi:10.3109/0142159X.2015.1034662

434 27. Foucault A, Dubé S, Fernandez N, Gagnon R, Charlin B. The concordance of judgement  
435 learning tool. *Med Educ.* 2014;48(5):541-542. doi:10.1111/medu.12467

436 28. Cooke S, Lemay J-F, Beran T. Evolutions in clinical reasoning assessment: The  
437 Evolving Script Concordance Test. *Med Teach.* 2017;0(0):1-8.  
438 doi:10.1080/0142159X.2017.1327706

439 29. Dory V, Gagnon R, Vanpee D, Charlin B. How to construct and implement script  
440 concordance tests: insights from a systematic review. *Med Educ.* 2012;46(6):552-563.  
441 doi:10.1111/j.1365-2923.2011.04211.x

442 30. See KC, Tan KL, Lim TK. The script concordance test for clinical reasoning: re-  
443 examining its utility and potential weakness. *Med Educ.* 2014;48(11):1069-1077.  
444 doi:10.1111/medu.12514

445 31. Tardif J. L'évaluation dans le paradigme constructiviste. In: *L'évaluation Des*  
446 *Apprentissages. Réflexions, Nouvelles Tendances et Formation.* CRP; 1993.

447 32. Tardif J. *L'Évaluation Des Compétences- Documenter Le Parcours de Développement.*  
448 Chenelière Education; 2006.

449 33. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR, for the International CBME  
450 Collaborators. The role of assessment in competency-based medical education. *Med Teach.*

451 2010;32(8):676-682. doi:10.3109/0142159X.2010.500704

452 34. Shumway JM, Harden RM. AMEE Guide No. 25: The assessment of learning outcomes  
453 for the competent and reflective physician. *Med Teach.* 2003;25(6):569-584.  
454 doi:10.1080/0142159032000151907

455 35. Jouquan J. L'évaluation des apprentissages des étudiants en formation médicale initiale.  
456 *Pédagogie Médicale.* 2002;3(1):38-52. doi:10.1051/pmed:2002006

457 36. Lindgren S, Karle H. Social accountability of medical education: aspects on global  
458 accreditation. *Med Teach.* 2011;33(8):667-672. doi:10.3109/0142159X.2011.590246

459 37. Wald HS, Borkan JM, Taylor JS, Anthony D, Reis SP. Fostering and evaluating  
460 reflective capacity in medical education: developing the REFLECT rubric for assessing  
461 reflective writing. *Acad Med J Assoc Am Med Coll.* 2012;87(1):41-50.  
462 doi:10.1097/ACM.0b013e31823b55fa

463

464 **Appendices**

465

466 *Interview guide*

467

468 1/ Presentation of the participants and investigators

469

470 2/ How did you hear about this training?

471

472 3/ How did the training go? What involvement did you have?

473

474 4/ What did you think about this training?

475

476 5/ What did you learn?

477 knowledge?

478 Contextual?

479 Behavior?

480 Details about ECG?

481

482

483 6/ Reflection?

484 What did you think about?

485 How did you do this?

486

487 7/ After your training, how did you integrate those learnings into your practice?

488

489 8/ Results: how has your practice changed?

490 Is there an assessment? a reaction of your supervisor?

491

492 9/ What did you think about the uncertainty in this training?

493

494 **Tables**

495

Participants	Age	Sex	Year of postgraduate training
1	26	M	2
2	29	M	3
3	25	F	1
4	26	F	1
5	30	F	3
6	27	F	3
7	26	M	2
8	27	F	2
9	27	F	3
10	26	F	1

496

497 **Table 1: Characteristics of focus group participants**

Deductive coding	
Thematic categories	Representative quotes
<b>The “thoughts and actions” component</b>	
	<i>you realize that there are several possible answers and that it's not because yours is a bit different that you've got it wrong, necessarily” (participant no. 8).</i>
	<i>“About the ECG, since it's normal, I don't do anything, but obviously you reassure [the patient]. Of course, right? Now [during the internship] obviously you always reassure people, you don't even realize, actually” (participant no. 3).</i>
	<i>"you start to think, really thinking, we don't just watch the ECG so we know what are the stakes" (participant no.4)</i>
<b>The ACEI component</b>	
Attentive process	<i>“I would put myself in the situation, I'd tell myself, I'm the doctor, this patient is my patient, and yeah, I was really totally in the case” (participant no. 8)</i>
	<i>"I really did it like I was the GP" (participant no.1)</i>
Critical process	<i>“So, the thing is, it made me ask myself by the end of the 18 cases: have you been dangerous or not?” (participant no.10)</i>
	<i>"I was asking myself about the synthesis: is it really what you should do ?" (participant no.5)</i>
Exploratory process	<i>“You're going to be judged, they'll say : “wait a minute, X wrote this, she missed the infarctus, she didn't send him</i>

	<i>[to the emergency ward] he'd be dead [...]", but the fact that this is completely anonymous [for the participants], I think that for a training program like this it really is very important"</i> (participant no.3)
	<i>"(the formation) allowed me to ask myself some questions"</i> (participant no.1)
Iterative process	<i>"after doing this again and again (by answering questions in the LbC ECG) [...] I understood that in this kind of situation, I could refer to a cardiologist"</i> (participant no.8)
	<i>"Well I adjusted to the answers, anyway – for instance, about reassurance there was a first case where I hadn't written it down, and in the answer, he put reassurance, and so in the last case where it was anxiety and all that, I put reassurance"</i> (participant n°9)
<b>The Self component</b>	
	<i>"it showed me that there is a lot of situations where ECG can help the physician in his practice"</i> (participant no.2)
	<i>"I think it isn't bad, doing it like this, because in the end when I did it I wasn't under any pressure, and so you know, you're not stressed out. When you're being graded or evaluated, you get stressed out, whereas this is really just for us"</i> (participant no. 5)
<b>The change component</b>	

	<i>“My way of reasoning is fairly logical, compared to the others, I’m a little bit reassured, I’ve gained a little bit of trust in myself” (participant no. 3)</i>
	<i>“for instance, you see [atrial fibrillation] [...] I told myself, ‘OK, we have the right to handle that ourselves”” (participant no. 1)</i>
Inductive coding	
Thematic categories	Representative quotes
Panel members identity	<i>“they tell us cardiology specialists, but we don’t know who they are at all” (participant no. 2) – “The thing is, we know cardiologists, who... you know... [do a bad job] [...] so, people we don’t trust completely” (participant no.1)</i>
No scoring	<i>“Now that we are residents, we don’t to be judged [...] we do it because we want to, not because we’re forced to” (participant no.5)</i>

498

499 **Table 2: thematic categories and representative quotes of deductive and inductive coding**

## Synthèse

Ce travail a permis de documenter l'impact d'une formation par concordance sur le développement de la réflexivité chez les futurs médecins généralistes. Cependant, ces résultats sont insuffisants pour conforter l'hypothèse initiale selon laquelle les formations par concordance favoriseraient le développement de la réflexivité.

Comme pour la recherche sur le raisonnement clinique et le burnout, ce travail met en lumière des difficultés méthodologiques dans l'exploration de processus intellectuels complexes. Comment documenter la présence de la réflexivité ? Des entretiens individuels approfondis avec les participants auraient-ils permis de documenter la présence de la composante des cadres conceptuels sous-jacents ? Ce travail sera utile à la communauté des chercheurs pour guider leurs choix méthodologiques sur un tel sujet à l'avenir.

Ces résultats sont toutefois encourageants. Ils démontrent que les participants se sont tout de même engagés dans des interactions critiques, attentives, exploratoires et itératives avec leurs pensées et actions dans une volonté de changement (amélioration). Dans cette perspective, ils confortent en partie la volonté de proposer davantage de formations par concordance aux étudiants de troisième cycle des études médicales ainsi qu'en formation continue. Aussi, ce travail devrait encourager le déploiement des formations par concordance dans de tels milieux de formation.

Le modèle MOT fournit ici encore une grille de lecture intéressante de ces résultats. Les participants exerçaient leur raisonnement clinique dans les cas proposés dans la formation. Ainsi, un parallèle similaire à celui proposé dans l'article sur la grille REFLECT peut être établi entre réflexivité et la procédure "réguler le processus" du modèle MOT. Les données des focus groupes retrouvaient des éléments des trois composantes de cette procédure : analyser ses propres activités cognitives, évaluer son propre processus cognitif et intervenir sur son processus cognitif.

Par exemple, le verbatim "*cela m'a fait me demander à la fin des 18 cas : ai-je été dangereuse ou pas ?*" serait à rattacher à la procédure "évaluer son propre processus cognitif". Le verbatim "*tu commences à réfléchir sur les choses, vraiment réfléchir*" se rattacherait à la procédure "évaluer ses propres activités cognitives". Un verbatim comme "*je me suis dit, OK on a le droit de gérer ça nous même*" se rattacherait à la procédure "intervenir sur son processus cognitif".

Les données de ces focus groupes pourraient appuyer l'hypothèse que la tâche cognitive à l'œuvre dans une formation par concordance comme celle-ci développe la métacognition des

participants. Les mêmes limites méthodologiques seraient cependant à considérer, comme l'impact du questionnement lors du focus groupe sur le processus métacognitif.

## TROISIÈME PARTIE : DÉVELOPPER LE PROFESSIONNALISME EN MÉDECINE GÉNÉRALE

### PROFESSIONNALISME, PROFESSIONNELS ET PROFESSIONNALISATION

Le troisième élément du profil souhaité des futurs médecins généraliste est le professionnalisme. Ce dernier est défini comme une compétence dans le référentiel de formation du troisième cycle de notre discipline (21). Analyser la définition proposée dans ce référentiel est utile : *"assurer l'engagement envers la société et à répondre à ses attentes, à développer une activité professionnelle en privilégiant le bien-être des personnes par une pratique éthique et déontologique, à améliorer ses compétences par une pratique réflexive dans le cadre de la médecine fondée sur des faits probants, à assumer la responsabilité des décisions prises avec le patient"* (13). Nous pouvons discuter cette définition en la découpant en cinq sections.

Le professionnalisme consisterait en premier lieu à *"assurer l'engagement envers la société et à répondre à ses attentes"* (13). Cet engagement envers la société renvoie au concept de contrat social. Développé par Rousseau en 1762 dans son ouvrage éponyme, il peut se résumer à ce que chacun mette *"en commun sa personne et toute sa puissance sous la suprême direction de la volonté générale"* (189). Appliqué à la médecine, le contrat social renvoie au pacte existant entre la profession médicale et la société (190,191). La société reconnaît aux médecins le privilège de l'autoréglementation de leur profession, pourvu qu'ils lui en restent redevables (19,190). Le professionnalisme peut être vu comme ensemble des comportements qui découlent du fait de placer en premier, l'intérêt du patient (191).

La seconde section de la définition stipule que le professionnalisme consiste à *"développer une activité professionnelle en privilégiant le bien-être des personnes par une pratique éthique et déontologique"*. Le professionnalisme ne peut ainsi pas être séparé de la question de l'éthique dans les soins. En effet, si le professionnalisme est un *"un travail engagé qui implicitement demande un savoir et des compétences techniques"*, il demande de manière plus importante encore *"l'application d'une éthique de la vertu"* (191). Le professionnalisme est également ancré dans une pratique déontologique. Apparu en langue française en 1825 selon le conseil de l'Ordre des Médecins (192), le terme de déontologie viendrait des racines grecques *déontos* et *logos* (193). Il s'agirait donc de *"dire si telle ou telle action doit ou ne doit pas être faite"*. Il consiste aujourd'hui en un *"discours sur les devoirs"* des médecins

en France. Les dispositions du code de déontologie médicale s'appliquent à l'ensemble des médecins en activité ainsi qu'aux étudiants selon l'article R.4127-1 du code de la santé publique.

La troisième section de la définition du professionnalisme intègre la réflexivité discutée dans la deuxième partie de ce travail. Le professionnalisme est représenté dans une représentation graphique en forme de fleur du référentiel des compétences en Médecine Générale (la "marguerite des compétences") comme le cercle extérieur, englobant les cinq autres compétences (voir figure n°22 ci-dessous). Le professionnalisme tiendrait donc aussi à l'adoption du pratique réflexive visant à continuellement développer ses compétences.

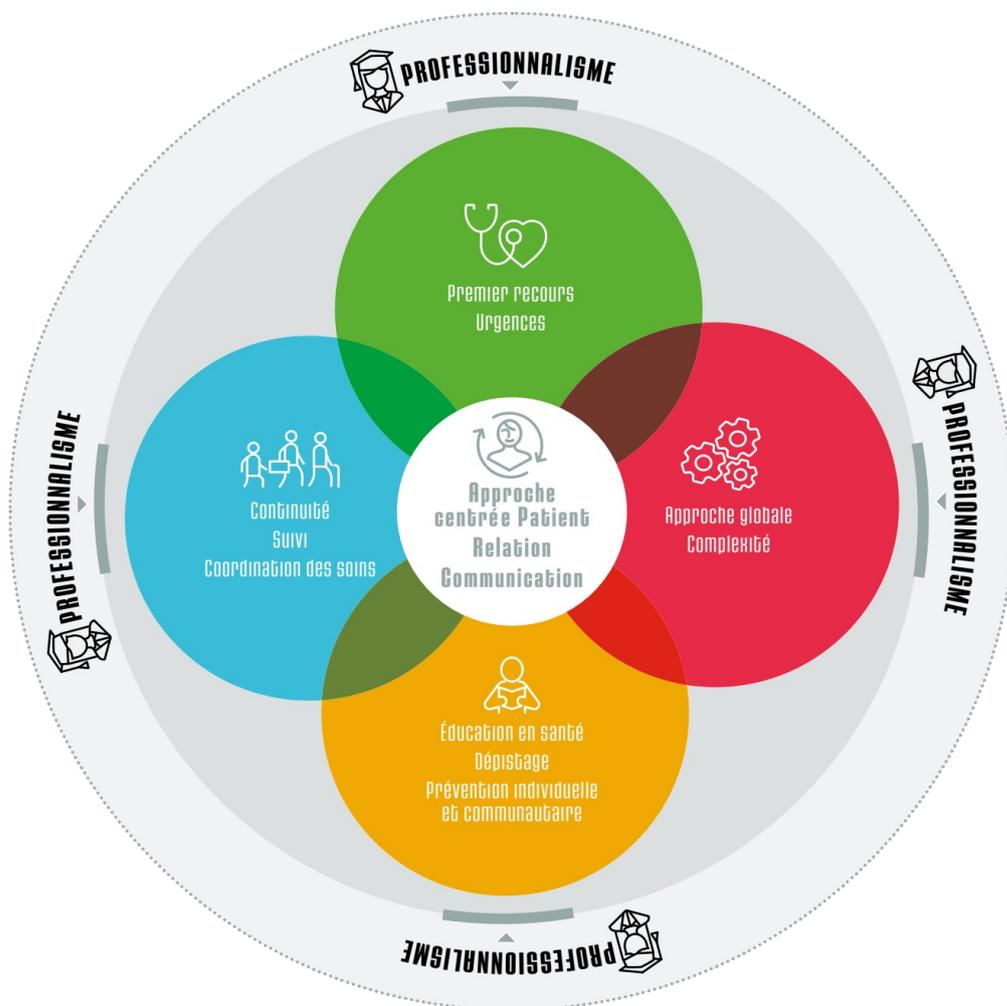


FIGURE 23 : MARGUERITE DES COMPÉTENCES DES MÉDECINS GÉNÉRALISTES

La quatrième section inscrit le respect du cadre de la médecine fondée sur les preuves (*evidence-based medicine*) comme faisant partie du professionnalisme. Ce courant né dans les années 80

en Amérique du Nord (194,195) constitue aujourd'hui le paradigme dominant au niveau mondial dans l'exercice de la médecine. Ce terme désigne *"l'utilisation consciencieuse, explicite et judicieuse des meilleures données disponibles pour la prise de décisions concernant les soins à prodiguer à chaque patient, [...] une pratique d'intégration de chaque expertise clinique aux meilleures données cliniques externes issues de recherches systématiques"* (194,195).

Enfin, le professionnalisme tiendrait à *"assumer la responsabilité des décisions prises avec le patient"*. Cette responsabilité renvoie largement au contrat social dans sa dimension de redevabilité du médecin envers la société(19,190).

Cette définition dans le référentiel de médecine générale en France diffère d'autres définitions du professionnalisme. Le professionnalisme est une notion interprétée de manière très variée selon les milieux professionnels, les époques, les cultures (20).

Selon la charte du professionnalisme médical pour le nouveau millénaire, le professionnalisme est fondamentalement basé sur les principes de la primauté du bien-être des patient, de l'autonomie des patients et de la justice sociale (196). Le professionnalisme comprend un ensemble d'engagements (196) :

1. Envers la compétence professionnelle
2. Envers l'honnêteté à l'égard des patients
3. Envers la confidentialité des patients
4. Envers l'entretien de rapports convenants avec les patients
5. Envers l'amélioration de la qualité des soins
6. Envers l'amélioration de l'accès aux soins
7. Envers la juste répartition de ressources limitées
8. Envers la connaissance scientifique
9. Envers le maintien de la confiance par la gestion des conflits d'intérêts
10. Envers les responsabilités professionnelles

Selon Robert *et al.*, le professionnalisme se réfère à *"l'agir professionnel , qui conjugue les savoirs spécifiques à chaque profession et le comportement attendu des professionnels concernés vis-à-vis de l'ensemble de la société"* (18). Le *"professionnel"* est défini comme une *"personne qui appartient à un groupe (une profession) et qui possède les connaissances (savoirs), les aptitudes (savoir-faire) et les*

*attitudes (savoir-être) acquises au cours d'une formation, qui sont spécifiques à sa profession et qui sont mises au service des autres membres de la société" (18).*

En cohérence avec cette définition, Bélisle propose de concevoir la professionnalisation des étudiants comme *"le processus de transformation d'un individu en professionnel, qui résulte d'apprentissages relatifs à des compétences spécifiques, une culture professionnelle et une identité professionnelle propre" (197).*

En dehors du milieu spécifique de la médecine ou celui plus large des sciences de la santé, le concept de professionnalisation fait l'objet d'une littérature riche. Pour Wittorski, la professionnalisation serait à la fois une *"mise en mouvement des individus dans des contextes de travail"*, une *"institutionnalisation de l'activité"* et une *"fabrication d'un professionnel par la formation"* (198). Il développe ainsi trois sens différents au terme de professionnalisation : la constitution d'un groupe social autonome (« professionnalisation-profession »), l'accompagnement de la flexibilité du travail (« professionnalisation-efficacité du travail ») et le processus de « fabrication » d'un professionnel par la formation (« professionnalisation-formation ») (198). Le Boterf discute la professionnalisation sous l'angle de l'exigence de confiance et de l'employabilité à l'issue d'une formation (199). Il théorise un modèle de *"navigation professionnelle"* dans l'ingénierie d'une formation professionnalisante (199).

Ces différentes définitions illustrent une partie de la polysémie qui entoure ces termes. Développer le professionnalisme des futurs médecins généralistes correspondrait au processus de professionnalisation selon Marilou Bélisle (197). Il s'agirait de réaliser des apprentissages en termes de compétences, de développer son identité professionnelle et d'acquérir la culture professionnelle correspondante (197). L'identité professionnelle selon Bélisle est considérée comme *"une représentation dynamique de soi en tant que professionnel(le), basée sur le rapport à soi, au groupe professionnel, à la société et à la formation"* (197). La culture professionnelle est vue dans ce modèle comme *"l'ensemble des normes, valeurs, attitudes et symboles partagés par les membres d'une même profession"* (197).

Ce sens de la professionnalisation correspondrait selon Richard Wittorski à la *"fabrication"* du futur professionnel (198). Cette *"professionnalisation-fabrication"* est l'un des trois sens qu'il développe du concept de professionnalisation (198). Pour Guy le Boterf, cette vision de la professionnalisation désigne *"l'accompagnement de la transformation de l'individu en professionnel compétent"* (199).

## COMMENT ACCOMPAGNER LE DÉVELOPPEMENT DU PROFESSIONNALISME ?

Comment favoriser le développement du professionnalisme des futurs médecins généralistes ? La réponse à cette question est grandement liée à la définition de ce qu'est le professionnalisme.

Il ressort des diverses ressources sur la thématique qu'il est nécessaire de commencer par établir une définition commune du professionnalisme à l'échelle de l'institution de formation et rendre explicite dans le programme de formation les attentes en termes de développement du professionnalisme (18,20,200). Un autre point commun est de convenir qu'il ne suffit pas de côtoyer en stage des professionnels pour que se développe "naturellement" le professionnalisme des étudiants (18,20,200). Si ces professionnels côtoyés en stage sont des "modèles de rôle" en termes de professionnalisme pour les étudiants (15,201), il est important d'intégrer en quoi cette exposition fait partie du "curriculum caché" (20,201). Ce curriculum caché en termes de professionnalisme s'oppose ainsi au "curriculum formel" et peut amener les étudiants à reproduire inconsciemment des conduites inappropriées (20,201). Cela souligne l'importance de favoriser l'adoption d'une pratique réflexive tout au long du cursus dans des espaces de parole dédiés visant à travailler le professionnalisme (18).

Les propositions concrètes de Robert *et al.* sont d'organiser des séances en petits groupes où les étudiants pourront échanger autour de vignettes décrivant des comportements non professionnels et de demander aux étudiants de rédiger individuellement des histoires vécues comme difficiles et de les partager avec leurs pairs (18). Robert *et al.* proposent également d'utiliser des outils spécifiquement conçus pour développer le professionnalisme comme les formations par concordance de jugement (FpC-jugement) (18).

La question essentielle est ensuite l'évaluation du professionnalisme. O'Sullivan *et al.* proposent premièrement de documenter le développement du professionnalisme à partir des stages via l'observation directe (20). Cette observation directe peut utiliser des outils validés comme le *professionalism mini-exercice* (P-MEX) (202). Il n'existe cependant pas à ma connaissance de tels outils validés en français.

Une autre possibilité évoquée par O'Sullivan *et al.* est de réaliser des auto-évaluations et des évaluations par les pairs du professionnalisme (20). Des ECOS dédiés, des traces d'apprentissages dans un

portfolio, les sessions d'analyses de vignettes de comportement non professionnels peuvent constituer autant de sources de données supplémentaires pour documenter le développement du professionnalisme (18).

Ces données doivent être analysées à la lumière d'indicateurs du professionnalisme. Des exemples d'indicateurs repris par Robert *et al.* (18) d'après les travaux de Cruess (203), de Symons *et al.* (204) et de Phelan *et al.* (205) sont les suivants :

1. L'étudiant cherche toujours à compléter sa formation et à développer ses compétences. Pour y parvenir, il sollicite du feedback, accepte la critique, se juge objectivement, identifie ses lacunes, reconnaît ses erreurs et prend des mesures pour les corriger.
2. L'étudiant est à l'heure, il est disponible pour les patients et ses collègues, il réalise jusqu'au bout les tâches qui lui sont confiées, utilise un langage adapté, et respecte les règles et les procédures en œuvre dans le service.
3. L'étudiant écoute activement les patients, montre du respect et de l'empathie pour les patients et leur famille, prévient les patients de ce qu'il va faire (par exemple, pendant l'examen clinique) et leur en explique les raisons. Il préserve la confidentialité, fait preuve de franchise, défend les intérêts du patient, discute avec lui des options (tout en lui offrant un choix et en l'aidant à décider), et l'encourage à poser des questions en y répondant de façon claire (en particulier en expliquant les termes techniques). Il maintient des limites appropriées avec les patients et ses collègues.

Dans le cas spécifique des FpC-jugement, l'évaluation peut comporter le calcul d'un score de concordance entre les réponses de l'étudiant et celles du panel ou uniquement se baser sur une auto-évaluation (134,135). J'ai entrepris il y a quelques années de travailler à proposer aux internes en DES de Médecine Générale une FpC-jugement parmi les formations optionnelles.

Dans une optique d'utilisation formative de ces FpC-jugement, la richesse du feedback apporté est primordiale. Dans ce contexte, j'ai réfléchi à une piste pour améliorer celui-ci.

ENRICHIR LA RÉTROACTION : INCLURE DES PATIENTS DANS LES PANELS DE RÉFÉRENCE DES FORMATIONS PAR  
CONCORDANCE DE JUGEMENT

Lors d'un atelier que j'animais en 2017 sur les FpC-jugement au congrès de la Société Internationale Francophone d'Éducation Médicale (SIFEM), un participant m'a interpellé sur la possibilité d'intégrer des patients dans les panels de référence. Cette suggestion m'a immédiatement semblée pertinente. Alors que le concept de professionnalisme met l'accent sur le rapport entre le professionnel et la société, les FpC-J n'utilisent qu'un panel constitué de pairs. Inclure des patients permettrait probablement de mieux refléter ce contrat social dans l'apprentissage du professionnalisme via les FpC-jugement. Il en irait ainsi en quelques sortes d'une amélioration de la validité de construit de l'outil.

Il n'existait pas à l'époque à notre connaissance de travaux ayant étudié l'impact de l'ajout de patients aux panels de FpC-jugement. Dans ce contexte, j'ai travaillé avec un étudiant de troisième cycle de médecine générale à réaliser une étude visant à analyser les profils de réponses de panels de patients versus de médecins généralistes à une série de question d'une FpC-jugement.

#### Contexte de l'étude

Les FpC-jugement sont un des outils proposés pour développer le professionnalisme des étudiants en sciences de la santé (129). Ces FpC-jugement exploitent le principe du modèle de rôle : les membres du panel sont recrutés par un vote des étudiants pour désigner les professionnels qui sont pour eux des modèles de rôle en matière de professionnalisme. Les membres du panel sont en général entre 3 et 5 (129,134,135). Ils répondent aux mêmes questions puis leurs réponses et les justifications de celles-ci servent à constituer le feedback pour les étudiants (129,134,135). Le feedback proposé illustre l'incertitude et la complexité dans les soins : il n'existe pas forcément une réponse unanime.

Considérant la potentielle validité de construit supérieure de l'inclusion de patients dans les panels de FpC-J, il serait intéressant dans cette optique de comparer les profils de réponse des patients et des pairs dans une FpC-jugement. À notre connaissance, il n'existait pas d'étude sur le sujet.

Nous avons réalisé une étude qui visait à comparer les profils de réponses de médecins généralistes et de patients à une FpC-jugement.

#### Matériels et méthodes

Nous avons soumis quinze vignettes d'une formation par concordance de jugement (FpC-J) en ligne à un panel de médecins généralistes et un panel de patients.

Les participants étaient recrutés sur la base du volontariat dans cinq cabinets de médecine générale des régions Grand-Est et Bourgogne en France sur les mois de mai à juin 2019. Les médecins généralistes ont été recrutés par la méthode de proche en proche sur la même période. Les critères d'inclusion étaient d'avoir plus de 20 ans, de disposer d'une connexion internet ainsi que d'un smartphone ou ordinateur ou tablette électronique. Le consentement à participer était recueilli au début du questionnaire.

Pour chaque participant, nous avons recueilli plusieurs variables d'intérêt pour nos analyses : le sexe, la tranche d'âge (par 20 ans), la profession et le temps de réponse au questionnaire. Les participants devaient préciser leur catégorie professionnelle : médecin généraliste, autre professionnel de santé ou autre. Les participants ayant une activité professionnelle « autre » devaient préciser leur profession exacte dans un champ de texte libre. Les participants ayant une activité professionnelle « autre » constituaient le groupe « patients ».

Les réponses à la FpC-jugement ont été recueillies pour chacune des quinze questions. Ces réponses comportaient le choix d'une des quatre modalités de réponse sur l'échelle de Likert ainsi que la justification de cette réponse en texte libre.

Nous avons effectué une analyse statistique des réponses pour rechercher des différences significatives entre les groupes ainsi qu'une analyse thématique des justifications.

Nous avons effectué une régression logistique ordinaire pour calculer des odds-ratio. Ces odds-ratio quantifient la force de l'association entre deux événements. Dans cette étude, un odds-ratio qui n'inclut pas 1 dans l'intervalle de confiance signifie que les répartitions entre les quatre catégories de Likert étaient significativement différentes entre les deux panels. Pour chaque question, nous présentons l'odds-ratio, son intervalle de confiance et la p-value.

## Résultats

327 personnes ont accepté de participer et 132 ont complété la FpC-jugement (40,4%). Nous avons analysé uniquement les réponses complètes. Parmi les 132 réponses complètes, 40 ont été données par des médecins généralistes (30,3%) et 70 par des patients (53,0%). Vingt-deux autres participants appartenaient à des professions de santé autres et ont donc été exclus du champ de notre analyse. Au total, 110 participants ont été inclus dans l'étude.

Il y avait une différence statistiquement significative pour neuf questions (60%) entre les panels. Les détails des résultats sont présentés dans le tableau n°11.



	Médecins généralistes n (%)	Patients n (%)	Odds-ratio [CI95%]	p
--	-----------------------------------	-------------------	--------------------	---

### Question 1

Totalement inacceptable	1 (2.5)	9 (12.86)	0.31 [0.14-0.65]	<b>0.001</b>
Difficilement acceptable	3 (7.5)	19 (27.14)		
Acceptable	23 (57.5)	29 (41.43)		
Totalement acceptable	13 (32.5)	13 (18.57)		

### Question 2

Totalement inacceptable	13 (32.5)	30 (42.86)	0.79 [0.39-1.62]	0.519
Difficilement acceptable	19 (47.5)	25 (35.71)		
Acceptable	6 (15.0)	10 (14.29)		
Totalement acceptable	2 (5.0)	5 (7.14)		

### Question 3

Totalement inacceptable	4 (10)	8 (11.43)	2.67 [1.29-5.68]	<b>0.008</b>
Difficilement acceptable	26 (65)	21 (30)		
Acceptable	6 (15)	27 (38.57)		
Totalement acceptable	4 (10)	14 (20)		

### Question 4

Totalement inacceptable	5 (12.5)	7 (10)	2.94 [1.42-6.18]	<b>0.004</b>
Difficilement acceptable	8 (20)	6 (8.57)		
Acceptable	15 (37.5)	13 (18.57)		
Totalement acceptable	12 (30)	44 (62.86)		

### Question 5

Totalement inacceptable	26 (65)	57 (81.43)		
Difficilement acceptable	11 (27.5)	12 (17.14)		
Acceptable	1 (2.5)	0 (0)	0.41 [0.17-0.98]	<b>0.046</b>
Totalement acceptable	2 (5)	1 (1.43)		

#### Question 6

Totalement inacceptable	2 (5)	3 (4.29)		
Difficilement acceptable	14 (35)	6 (8.57)		
Acceptable	11 (27.5)	16 (22.86)	3.78 [1.79-8.17]	<b>&lt;0.001</b>
Totalement acceptable	13 (32.5)	45 (64.29)		

#### Question 7

Totalement inacceptable	15 (37.5)	21 (30)		
Difficilement acceptable	17 (42.5)	16 (22.86)		
Acceptable	7 (17.5)	22 (31.43)	2.24 [1.11-4.58]	<b>0.024</b>
Totalement acceptable	1 (2.5)	11 (15.71)		

#### Question 8

Totalement inacceptable	2 (5)	18 (25.71)		
Difficilement acceptable	8 (20)	18 (25.71)		
Acceptable	14 (35)	15 (21.43)	0.39 [0.19-0.78]	<b>0.008</b>
Totalement acceptable	16 (40)	19 (27.14)		

#### Question 9

Totalement inacceptable	13 (32.5)	17 (24.29)		
Difficilement acceptable	17 (42.5)	26 (37.14)		
Acceptable	7 (17.5)	19 (27.14)	1.66 [0.82-3.42]	0.159
Totalement acceptable	3 (7.5)	8 (11.43)		

#### Question 10

Totalement inacceptable	29 (72.5)	63 (90)	0.31 [0.10-0.86]	<b>0.024</b>
-------------------------	-----------	---------	------------------	--------------

Difficilement acceptable	10 (25)	5 (7.14)		
Acceptable	1 (2.5)	0 (0)		
Totalement acceptable	0 (0)	2 (2.86)		

### Question 11

Totalement inacceptable	7 (17.5)	36 (51.43)		
Difficilement acceptable	22 (55)	26 (37.14)	0.24 [0.11-0.52]	<b>&lt;0.001</b>
Acceptable	7 (17.5)	7 (10)		
Totalement acceptable	4 (10)	1 (1.43)		

### Question 12

Totalement inacceptable	25 (62.5)	38 (54.29)		
Difficilement acceptable	13 (32.5)	19 (27.14)	1.63 [0.76-3.57]	0.211
Acceptable	1 (2.5)	9 (12.86)		
Totalement acceptable	1 (2.5)	4 (5.71)		

### Question 13

Totalement inacceptable	0 (0)	3 (4.29)		
Difficilement acceptable	0 (0)	5 (7.14)	0.65 [0.29-1.40]	0.270
Acceptable	15 (37.5)	23 (32.86)		
Totalement acceptable	25 (62.5)	39 (55.71)		

### Question 14

Totalement inacceptable	1 (2.5)	13 (18.57)		
Difficilement acceptable	12 (30)	20 (28.57)	0.57 [0.28-1.14]	0.110
Acceptable	13 (32.5)	16 (22.86)		
Totalement acceptable	14 (35)	21 (30)		

### Question 15

Totalement inacceptable	4 (10)	15 (21.43)	0.52 [0.25-1.06]	0.069
Difficilement acceptable	15 (37.5)	30 (42.86)		

Acceptable	17 (42.5)	19 (27.14)
Totalement acceptable	4 (10)	6 (8.57)

**TABLEAU 11 : DÉTAILS DES RÉPARTITIONS DES RÉPONSES DES PANELS À LA FPC-JUGEMENT**

L'analyse thématique des justifications pour ces neuf questions a montré des différences entre les panels. Les patients avaient souvent un jugement plus sévère et si les thèmes étaient parfois communs, ils ajoutaient fréquemment des concepts non évoqués par le panel de médecins généralistes comme la critique du principe de confraternité.

#### Discussion

Ces nombreuses divergences entre les panels montrent que l'ajout de patients aux panels des FpC-jugement a le potentiel d'enrichir fortement le feedback proposé.

Les points de vue "extérieurs" à la profession pourraient permettre de rapprocher le profil des nouveaux diplômés des attentes des patients en matière de professionnalisme. Cela serait de nature à augmenter la validité de construit de cet outil par rapport au concept de professionnalisme et à celui de professionnalisation.

L'impact de cet ajout de patient sur les apprentissages des étudiants devra cependant être documenté dans de futurs travaux.

L'article intégral dans sa version en anglais se trouve ci-dessous.

**Patients' judgments of medical professionalism – Concordance of judgment learning tool scores and justifications between patients and general practitioners**

Mathieu Lorenzo<sup>1,2</sup>, Pierre Bailly<sup>1</sup>

<sup>1</sup> Department of General practice, Faculty of Medicine, Strasbourg, France

<sup>2</sup> Centre for Training and Research in Health Sciences Education (CFRPS), Faculty of Medicine, Strasbourg, France

Correspondence: Dr M. Lorenzo – Department of General practice, Faculty of Medicine, 4 rue Kirschleger, 67000 Strasbourg, France - mlorenzo@unistra.fr

**Authorship**

ML and PB designed the study, collected and analyzed the data, drafted the manuscript and approved the published version.

**Acknowledgements**

Professor Nicolas Fernandez from the university of Montréal, Canada, helped critically revise the manuscript.

**Funding**

None

**Conflict of interest**

None

**Ethical approval**

This project was approved by the Ethics Committee of the Regional Hospital Group of Mulhouse Sud-Alsace (GHRMSA), France, on 19 June 2018.

## **Abstract**

### *Introduction*

Professionalism in medicine can be defined as the range of behaviors resulting from the effort to put the patient's interest first. The Concordance of Judgement Learning Tool (CJLT) tool has been developed for distance, asynchronous learning of professionalism in health sciences students. The learning of professionalism is induced by the participant's comparison of his own response with those of the panel members. Whilst, CJLT programs typically include peers in their panels, we believe that they could include patients. This would ensure better consideration of the social contract inherent in medical professionalism. Accordingly, we have conducted a study aimed at comparing CJLT response patterns of general practitioners (PCPs) and patients within panels.

### *Methods*

We conducted a quantitative prospective study of responses to a CJLT program based on a panel of PCPs and a panel of patients. The main objective was to search for statistically significant differences between the response patterns of the two panels. The secondary objective was to better account for observed divergences by conducting a qualitative analysis of justifications.

### *Results*

110 participants were included in the study: 70 patients and 40 PCPs. We found a significant difference in response patterns between the PCP and patient groups for nine of the fifteen questions (60%). The themes of justifications frequently

differed between groups in questions with significant differences in response patterns.

### *Conclusions*

Including patients in CJLT panels has potential to enrich the feedback offered to students in these online training programs. It can make the notion of social contract inherent in medical professionalism more explicit and allow us to train healthcare professionals that better fit society's expectations. This approach raises numerous questions as to the relevant teaching choices to adopt; these will be the subject of further research.

developed at the University of Montreal for distance, asynchronous learning of professionalism in health sciences students.<sup>5,8</sup> Participants are presented with short text vignettes describing the questionable behavior (in terms of professionalism) of a healthcare professional in a healthcare situation. They are asked to judge this behavior on a Likert scale with four levels, from 'completely unacceptable' to 'totally acceptable',<sup>9</sup> and to justify their choice in writing.

CJLT uses the role model principle: a panel of recognized experts in the given field are asked to answer the same vignettes. In the case of CJLT, panel members are selected by students who vote for the professionals that are role models for them in terms of professionalism. There are, as a rule, three to five panel members. Their answers on the same vignettes and the justifications are used for student feedback. The feedback offered illustrates uncertainty and complexity in healthcare: there isn't necessarily a single, unanimous response. The learning of professionalism is induced by the participant's comparison of his own response with those of the panel members.

Participants obtain a score based on concordance between their choices and the choices of the panel members.<sup>5</sup> This scoring method is similar to script concordance test (SCT).<sup>10,11</sup> For each question, participants can obtain 1 point, a fraction of point or no point.

While the concept of professionalism emphasizes the relationship between the professional and society, CJLT only include health professionals in their panels; a CJLT for primary care physicians (PCP) residents would only include PCPs in the panel. We believe that if patients were included, there would be better consideration of the social contract inherent in medical professionalism. In a social cognitive perspective, including patients could offer a more diversified social interaction than the usual

professional role models. Co-regulated learning models (CRL),<sup>12</sup> place such social transactions at the heart of assessment for learning. To the best of our knowledge, no study so far has examined the effect of including patients in panels of CJLT.

To that effect, we invited patients to become panel members along with PCPs and sought to compare their response patterns in a CJLT.

### **Material and methods**

We conducted a prospective study of responses to a CJLT among a panel of PCPs and a panel of patients.

The main objective was contrasting concordance scores of the CJLT PCPs versus patients and gain deeper understanding of the differences by conducting a thematic analysis of panel member justifications.

We chose a mixt method for this research for two reasons: (i) statistically significant differences between the panels will translate in different scores for students in a CJLT and (ii) qualitative analysis of justifications is the heart of the feedback provided by CJLT in an assessment for learning perspective.

#### *Creation and validation of the CJLT*

The text vignettes depicting clinical situations were based on actual clinical situations encountered in primary care and related to medical professionalism. For each vignette, participants were asked to judge the behavior of a third party depicted in the described situation by answering the question: 'For you, this attitude/this behavior would be...'. The Likert scale included the four following options: totally unacceptable, hardly acceptable, acceptable, and totally acceptable. Participants had to add a mandatory

## **Introduction**

Professionalism in medicine can be defined as the range of behaviors resulting from the effort to put the patient's interest first.<sup>1</sup> It is thus closely connected to the question of ethics in healthcare. While professionalism is 'a commitment that implicitly requires technical knowledge and competences', it even more importantly requires 'applying the ethics of virtue'.<sup>1</sup> Professionalism supposes an implicit social contract between the profession and society,<sup>1,2</sup> which creates mutual trust in the doctor-patient relationship. Society allows doctors the privilege to self-regulate their profession, as long as they engage to be accountable to the public.<sup>2,3</sup>

There is no consensus regarding the best way to teach professionalism in medicine.<sup>4,5</sup> In a recent review by Berger *et al.*, it was impossible to identify best practices for teaching professionalism due to the heterogeneity of the studies on this broad concept.<sup>6</sup> However, evident themes in the literature for teaching professionalism are role modelling and personal reflections (with tutoring).<sup>7</sup> Many different methods are described in the literature: reports on critical incidents, interviews of professors, learning programs relying on scenarios of professional dilemmas, professionalism classes, group discussions between professors and students based on scenarios.<sup>7</sup> Most of these methods are time-consuming, which complicates their use in medical education facilities with limited resources.<sup>5</sup> As of the time of writing, the global COVID-19 pandemic is an additional obstacle to teaching professionalism, by considerably limiting classroom teaching.

The innovative Concordance of Judgement Learning Tool (CJLT) tool has been

text justification of their response.

Thirty (30) such vignettes comprising three questions each were created and distributed according to three dimensions of professionalism identified by Van de Camp: ten items each for the individual, institutional and interactional dimensions.<sup>13</sup>

The vignettes and questions were written in everyday language, minimizing medical jargon. They were then corrected and revised by a group of three PCPs affiliated with the primary care department of the University of Strasbourg and a patient. The group then selected a total of fifteen vignettes (five for each dimension of professionalism), for a total of 45 questions. Each vignette had three questions.

A first test questionnaire comprising five vignettes with three questions each was then created using the Limesurvey® software. The goal was to test user-friendliness and feasibility. The questionnaire was submitted online to a pilot group: six patients and five PCPs. Their feedback led to some changes, such as the inclusion of a video explaining how the test works, to facilitate appropriation of the tool.

As the average response time observed for fifteen questions was 30 minutes, we decided to limit the final questionnaire to fifteen questions (one per vignette), to prevent difficulties in enrolling participants.

The CJLT used for the study ultimately comprised fifteen questions pertaining to fifteen different clinical situations, divided evenly according to Van de Camp's three dimensions of professionalism.<sup>13</sup> The list of questions is provided in Appendix no. 1.

As this CLJT was first created for this research project, no reference panel was defined between our two panels. In this study, participants were patients and PCPs. No students were included and no score was calculated.

Participants answered the same questions and gave a justification for every question.

#### *Selection of the participants*

Participants were enrolled on a voluntary basis among patients consulting in five primary care practices in the Grand-Est and Bourgogne regions in France in May-June 2019. PCPs were contacted using the snowball method starting with professional relations of the authors. This was done in person, by email or by phone call. Every participant was asked to suggest potential future participants. Their email addresses were collected for the purpose of sending them the link to the CJLT. These potential participants were then free to transfer the CJLT link to potential future participants in their social circles.

The inclusion criteria for patients were being older than 20, having an internet connection and a smartphone or personal computer or tablet. Exclusion criteria for patients were being a healthcare professional or a PCP. For PCPs, the only inclusion criteria were having an internet connection as well as a smartphone or personal computer or tablet. Consent was obtained at the beginning of the questionnaire.

#### *Collected variables*

We administered the online CJLT to every patient and PCPs who accepted to participate. For each participant, we collected several variables of interest for our analyses: sex, age, profession and questionnaire response time. Age was divided into ten-year brackets from 20 to 100 years old (for instance, 31-40). Participants were asked to specify whether they were a PCP, a healthcare professional, or belonged to a different professional category. Healthcare professionals and participants with an

'other' professional activity were asked to fill in their profession in a free text field. Participants with an 'other' professional activity formed the patient group. Responses to the CJLT were collected for each of the fifteen questions. They comprised the choice of one out of four levels on the Likert scale and the justification of that choice in a free text field.

### *Statistical analysis*

This being a pilot study, we had no data at our disposal to assess the number of individuals to include in each group to reach a power of at least 80%. We empirically decided to include 100 participants, aiming for a minimum proportion of a third of PCPs. The statistical analysis included a descriptive part and an inferential part.

We performed the descriptive statistical analysis of quantitative variables by establishing position parameters (average, median, minimum, maximum, first and third quartiles) and spread parameters (variance, standard deviation, range, IQR) for each variable.

We described qualitative variables by establishing the numbers and proportions of each modality in the sample. Whenever this was useful, cross-tabulations were made, including numbers and proportions by line, proportions by column and proportion compared to the total, for each cell in the table.

The processing of Likert scale results is a debated subject in the medical education literature.<sup>14,15</sup> Likert scales are ordinal-scales but are frequently treated as interval scales. Use of non-parametric tests is encouraged to test percentages of responses in each categories rather than means and standard deviations.<sup>14</sup> For this study, we

performed an ordinal logistic regression to calculate an odds-ratio. These odds-ratio quantifies the strength of the association between two events. In this study, an odds-ratio that do not include 1 in the confidence interval means that the repartitions between the four Likert categories were statistically significant between the two panels. For each question, we present the odds-ratio, its confidence interval and the p-value. The significance threshold was set at 0.05. The quantitative analysis was done using the R software (version 3.5.3).

We deliberately did not subject these multiple comparisons to a statistical correction (like the Holm-Bonferroni method) despite the issue of alpha risk inflation. This issue is taken into consideration in our interpretation of the results.

#### *Qualitative analysis*

A thematic analysis<sup>16</sup> of justifications of the response was performed whenever statistically significant differences between the groups' averages were found. The objective was to identify possible explanations of these differences and assess whether they were relevant to practice. Coding was first done by PB and then revised by ML. Inconsistencies were subjected to negotiation between PB and ML to reach a consensus. Data saturation was not attempted considering the study design. This was PB's first experience of qualitative research, whereas ML is an experienced qualitative researcher. This part was not intended to achieve the expected level of methodological rigor in qualitative research due to the limitations inherent in our study (type of data collected and method of collection).<sup>17</sup>

#### *Ethics and permissions*

This project was approved by the Ethics Committee of the Regional Hospital Group of Mulhouse Sud-Alsace (GHRMSA) on 19 June 2018.

## **Results**

### *Population*

327 individuals accepted to participate and 132 completed the CJLT (40.4%). We analyzed complete responses only. Among the 132 complete responses, 40 were given by PCPs (30.3%) and 70 by patients (53.0%). An additional 22 participants were in health professions other than PCP and were accordingly excluded from the scope of our analysis. In total, 110 participants were included in the study.

The PCP group was younger than the patient group. There was a majority of women in both the PCP group (n=22; 55%) and the patient group (n=44; 62.86%). Participants in the PCP group responded more quickly (28-minute mean response time, vs. 42 minutes for the patient group). Details on participants' characteristics according to groups are provided in Table 1.

### *Main objective*

We found a significant difference in response patterns between the PCP and patient groups for nine questions (60%). Details on the distribution of responses, odds-ratios and p-values are provided in Table 2.

### *Secondary objective*

The themes of justifications frequently differed between groups in the case of questions for which statistically significant differences in response patterns were observed. For reasons of space in this manuscript, we have chosen to present the themes for three questions. Full results of thematic analysis can be found in

#### Appendix 1.

Question no.1 read: *'It is 8.15 pm. A PCP is getting ready to bring in her last patient for the day. She is an hour late and she has an important appointment tonight at 8.30. She gets a phone call from the mother of a 7-month-old patient. She has had a 40°C fever since the morning; the mother asks if she can make an emergency visit. She sounds worried. The doctor is considering referring her to the on-call doctor.*

*In your view, this attitude/behavior is...'*

The justifications given by PCPs (who mostly agreed with the attitude described) emphasized compliance with the organization of continuity of care and privacy: *'You can say yes to everything and leave the office at 11 every night; on-call doctors are there to help us have a life outside of the office; that said, it depends on the 8.30 appointment and the patient's story'* (PCP 27); *'patient's personal life – if need for tests given the hour the emergency department will cover it – ill since the morning but calls after 8'* (PCP 67); *'the on-call doctor is here for that after 8'* (PCP 70)

While most patients' justifications cited similar reasons, some saw the scenario as an example of a lack of professionalism:

*'Poor work ethic'* (patient 96); *'Not very professional'* (patient 90); *'legally acceptable, but less so from a moral and human perspective'* (patient 110).

Question no. 6 read: *'A PCP receives a 34-year-old patient, consulting because she has been presenting with vaginal itching for two days. After interviewing her, he thinks*

*she has a vaginal yeast infection, and asks her to undress so that he can examine her and take a vaginal swab. The patient refuses, because she does not want to be examined by a man. She insists on being given a treatment anyway.*

*The PCP considers refusing to prescribe her a treatment because he was unable to examine the patient.*

*In your view, this attitude/behavior is...'*

PCPs mostly agreed with the physician's attitude in the scenario. They justified it mainly on the grounds that a PCP has the right to deny a prescription if a clinical examination has not been performed, for legal liability reasons: *'He's within his rights, he has no obligation'* (PCP 68); *'No blind treatment, this isn't a grocery store'* (PCP 93); *'[...] any medical diagnosis requires a clinical examination. The PCP could then be faulted for having missed another diagnosis'* (PCP 85); *'[...] prescribing without examining makes you exposed to risks'* (PCP 66).

Conversely, mainly among the participants who found the PCP's attitude 'hardly acceptable', the argument of having to respect the patient's choice dominated: *'No formal indication of a vaginal swab as a first-line step, empirical treatment may be enough, the patient has a right to refuse to be examined'* (PCP 47); *'Possibility of going with a probabilistic treatment'* (PCP 120); *'Prescribing an antifungal cream without performing an examination is common. The patient is free to turn down a pelvic examination if she's not OK with it'* (PCP 99).

A recurring underlying concern for PCPs was the continuity of care: *'[...] consider consulting shortly with a female gynecologist or a female colleague in the office who can handle gynecology if symptoms linger despite treatment'* (PCP 69); *'[...] give out addresses of female colleagues'* (PCP 99).

Patients' justifications were mostly along these same lines: *'The doctor cannot prescribe a drug without examining the patient'* (Patient 34); *'No treatment without diagnosis. The doctor can refer the patient to a female colleague or specialist'* (Patient 43).

Many participants argued that the patient had chosen to consult with a man and therefore could not refuse to be examined by him: *'It's up to the patient to choose a female doctor'* (Patient 28), *'The patient can just go see a woman doctor if she finds being examined by a man intolerable!!!'* (Patient 39), *'The patient had to consult with a doctor that suited her. She should not prevent the doctor from making a proper diagnosis'* (Patient 56).

Such criticisms of the patient's refusal were expressed indifferently by male and female participants.

Question no.11 read: *'A doctor receives the wife of one of his friends for a consultation. She is usually treated by one of his colleagues. In her file, he sees that she has regularly consulted over the past six months for a variety of symptoms and numerous complaints. He notices marks on her arms during the clinical examination. He thinks she might be a victim of spousal abuse, but does not ask her because he doesn't feel capable of handling the situation.'*

*In your view, this attitude/behavior is...'*

In their justifications, PCPs argued that the refusal to address the spousal abuse issue was only acceptable if the continuity of care was ensured: *'He can refer her to the colleague who usually sees her; he could call that colleague'* (PCP 132); *'Acceptable for him not to want to handle it, but offer to refer her to a colleague and to stay in touch if needed'* (PCP 50); *'Complex situation; we have a right to have our limitations in terms*

*of competences, but write it down in her file for her usual PCP' (PCP 67); 'Better refer her to another healthcare professional if you don't feel competent, but this shouldn't delay patient care' (PCP 70).*

Patients were more critical of the PCP's attitude, viewing his failure to address the issue of spousal abuse as professional misconduct: *'Irresponsible' (Patient 19), 'Failure to assist a person in potential danger' (Patient 106), 'He has to learn to handle this kind of situation. This is inadmissible. It's tantamount to "failure to assist a person in danger". This is also part of his duties, of his role' (Patient 45).*

In their view, the PCP's friendship with the possible offender should not be a factor: *'The doctor should play his part as a doctor. Personal considerations must be irrelevant' (Patient 13), 'A doctor should be able to make a difference between his private life and his professional duty' (Patient 60).*

## **Discussion**

### *Main results*

We evidenced statistically significant differences in responses to nine questions (60%) of the CJLT between the PCP group and the patient group.

Comparative thematic analysis of the two groups sheds light on the underlying factors of differences in reasoning in these situations.

### *Comparison with literature data*

To our knowledge, this study is the first to explore the inclusion of patients in CJLT panels. The many differences observed in responses to our questions between the PCP group and the patient group suggests that including patients in CJLT panels may have benefits for learning. In particular, the thematic analysis of justifications supports

this by highlighting the diversity of viewpoints introduced by patients in the judgment of professionalism. As medical students use a variety of sources of information that "facilitate the interpretation of the experience and the construction of knowledge from it", patients are a source of credible "learning cues".<sup>18</sup> In a study by van de Wiel *et al.* on deliberate practice,<sup>19</sup> the patients encountered in clinical practice were valued most for professional development. The inclusion of patients in CJLT panels appears to enrich the feedback offered to learners and strengthen the conceptual validity of that tool.

The social contract inherent in the concept of professionalism becomes more explicit for learners thanks to the patients' justifications of responses. In question 1, patients' justifications on lack of professionalism revealed that patients often expect from their PCP some extra work for emergency visits. Question 11 on spousal abuse was also a great example of these different views on professionalism. As PCPs tend to judge acceptable the attitude if continuity of care was respected, patients demand to "*make a difference between personal life and professional duty*". Both examples highlight the social contract in professionalism: the society expect future healthcare professionals that will meet their needs.

As more than half of the questions of the CJLT had significant differences between patients and PCPs, including patients in references panels could have a big impact on CJLT scores. These scores calculated with both patients and peers choices could give a better feedback on professionalism performance because it will be more coherent with the concept itself. Futures studies might investigate the impact of the introduction of patients in CJLT scores against classical peer-model scores.

The question of the mode of selection for the patient panels is also posed. There are currently no established criteria for this in the literature. We believe that patients who are eager to collaborate in the education of future healthcare professionals, such as expert patients (partner patients), could form a group that would provide more valuable feedback. However, we also believe it is important to diversify patient panels so as not to reflect only what teachers would like patients to be, but rather what they actually are. A random selection process could accordingly be an option.

Regarding the ideal number of panel members, there is no data on the psychometric properties of CJLT according to that factor. In the study by Foucault et al., the peer panel comprised ten members.<sup>1</sup> A conceptual fairly similar tool, the script concordance test (SCT), usually involves ten to twenty members for high-stakes cases.<sup>11</sup> In our view, the CJLT should be used mainly for training purposes. In that context, calculating a score that would allow for such psychometric analyses is of little value. Additionally, the absence of score calculation in these tools could improve the development of evaluative judgment in future healthcare professionals by having them exclusively perform a self-evaluation task on their responses.<sup>20</sup>

The CJLT that include patients in their panels can be one of multiple tools aimed at documenting self-regulation of student learning on professionalism,<sup>12,21–23</sup> with a view to maintaining high standards of professionalism during entire careers. Moreover, contemporary perspective on regulated learning are moving to models of co-regulated learning.<sup>12</sup> co-regulation is defined as "*a transitional process in a learner's acquisition of regulated learning, within which learners and others share a common problem-solving plane*".<sup>24</sup> In these models of co-regulated learning, standards are co-

constructed. Seeing professionalism as a co-constructed standard in health care, patients are needed in CJLT panels.

### *Strengths and weaknesses*

This study is the first to confront the responses of PCPs and patients on CJLT. We have only been able to address a few themes here. Other studies with larger panels and addressing other professionalism-related themes will be necessary to dig deeper into the subject and expand on this first approach.

Out of 327 questionnaires, only 132 were completed. This mediocre completion rate may be explained by several factors: difficulties in using the tool (which is still little used in France), the time needed to complete the training (more than 45 minutes on average – more than in our pilot group), and curiosity, which may have led a majority of participants to open the link simply to discover the tool.

The age of participants can also cause a bias, as 75 per cent of the population of PCPs in this study is under 40, whereas the national average age of PCPs in France is about 51 years old.<sup>25</sup> The low participation of older age brackets may result from lesser ease with online tools, and scarcer internet connections at home.

There is a recruitment bias in our study, which partly relates to the methodological choice of circulating the questionnaire using the snowball method. Our sample is characterized by an over-representation of service sector professionals in the patient group. While we have chosen not to apply a statistical correction to these multiple comparisons, it is possible that at least one statistically significant difference observed in this study is actually related to alpha risk inflation. Still, even in that case, this would not significantly alter our results.

## **Conclusion**

Including patients in CJLT panels has potential to enrich the feedback offered to students in these online training programs. It will make the notion of social contract inherent in medical professionalism more explicit and allow us to train healthcare professionals that better fit society's expectations.

This will in our view be in years to come a valuable tool to further develop in institutions that train healthcare professionals, as a complement to traditional methods for teaching professionalism. Exceptional health crises like the one triggered by COVID-19 call for increased distance training resources on professionalism, as society will not lower its expectations of healthcare professionals. This approach raises numerous questions as to the relevant teaching choices to adopt; these will be the subject of further research.

## References

1. Brody H, Doukas D. Professionalism: a framework to guide medical education. *Med Educ*. 2014;48(10):980-987. doi:10.1111/medu.12520
2. Cruess SR, Cruess L.R., Steinert Y. Linking the teaching of professionalism to the social contract: A call for cultural humility: *Medical Teacher*: Vol 32, No 5. Published January 19, 2010. Accessed March 28, 2018. <https://www.tandfonline-com.scd-rproxy.u-strasbg.fr/doi/full/10.3109/01421591003692722?scroll=top&needAccess=true>
3. Cruess SR, Cruess RL. Professionalism: a contract between medicine and society. *CMAJ Can Med Assoc J*. 2000;162(5):668-669.
4. Cruess RL, Cruess SR. Teaching professionalism: general principles. *Med Teach*. 2006;28(3):205-208. doi:10.1080/01421590600643653
5. Foucault A, Dubé S, Fernandez N, Gagnon R, Charlin B. Learning medical professionalism with the online concordance-of-judgment learning tool (CJLT): A pilot study. *Med Teach*. 2015;37(10):955-960. doi:10.3109/0142159X.2014.970986
6. Berger AS, Niedra E, Brooks SG, Ahmed WS, Ginsburg S. Teaching Professionalism in Postgraduate Medical Education: A Systematic Review. *Acad Med J Assoc Am Med Coll*. 2020;95(6):938-946. doi:10.1097/ACM.0000000000002987
7. Birden H, Glass N, Wilson I, Harrison M, Usherwood T, Nass D. Teaching professionalism in medical education: a Best Evidence Medical Education (BEME) systematic review. BEME Guide No. 25. *Med Teach*. 2013;35(7):e1252-1266. doi:10.3109/0142159X.2013.789132
8. Foucault A, Dubé S, Fernandez N, Gagnon R, Charlin B. The concordance of judgement learning tool. *Med Educ*. 2014;48(5):541-542. doi:10.1111/medu.12467
9. Centre de pédagogie appliquée aux sciences de la santé (CPASS). Guide de formation par concordance de jugement clinique. Santé & Médecine presented at the: 15:17:32 UTC. [https://fr.slideshare.net/CPASS\\_UdeM/guide-de-formation-par-concordance-de-jugement-clinique](https://fr.slideshare.net/CPASS_UdeM/guide-de-formation-par-concordance-de-jugement-clinique)
10. Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten C. The Script Concordance test: a tool to assess the reflective clinician. *Teach Learn Med*. 2000;12(4):189-195. doi:10.1207/S15328015TLM1204\_5
11. Dory V, Gagnon R, Vanpee D, Charlin B. How to construct and implement script concordance tests: insights from a systematic review. *Med Educ*. 2012;46(6):552-563. doi:10.1111/j.1365-2923.2011.04211.x
12. Rich JV. Proposing a Model of Co-Regulated Learning for Graduate Medical Education. *Acad Med J Assoc Am Med Coll*. 2017;92(8):1100-1104. doi:10.1097/ACM.0000000000001583
13. Van De Camp K, Vernooij-Dassen MJFJ, Grol RPTM, Bottema BJAM. How to conceptualize professionalism: a qualitative study. *Med Teach*. 2004;26(8):696-702. doi:10.1080/01421590400019518
14. Jamieson S. Likert scales: how to (ab)use them. *Med Educ*. 2004;38(12):1217-1218. doi:10.1111/j.1365-2929.2004.02012.x
15. Pell G. Use and misuse of Likert scales. *Med Educ*. 2005;39(9):970; author reply 971. doi:10.1111/j.1365-2929.2005.02237.x
16. Kiger ME, Varpio L. Thematic analysis of qualitative data: AMEE Guide No. 131. *Med Teach*. 2020;42(8):846-854. doi:10.1080/0142159X.2020.1755030
17. LaDonna KA, Taylor T, Lingard L. Why Open-Ended Survey Questions Are Unlikely to Support Rigorous Qualitative Insights. *Acad Med J Assoc Am Med Coll*. 2018;93(3):347-349. doi:10.1097/ACM.0000000000002088
18. Watling C, Driessen E, van der Vleuten CPM, Lingard L. Learning from clinical

- work: the roles of learning cues and credibility judgements. *Med Educ.* 2012;46(2):192-200. doi:10.1111/j.1365-2923.2011.04126.x
19. van de Wiel MWJ, Van den Bossche P, Janssen S, Jossberger H. Exploring deliberate practice in medicine: how do physicians learn in the workplace? *Adv Health Sci Educ Theory Pract.* 2011;16(1):81-95. doi:10.1007/s10459-010-9246-3
  20. Tai J, Ajjawi R, Boud D, Dawson P, Panadero E. Developing evaluative judgement: enabling students to make decisions about the quality of work. *High Educ.* 2018;76(3):467-481. doi:10.1007/s10734-017-0220-3
  21. Rovers SFE, Clarebout G, Savelberg HHCM, de Bruin ABH, van Merriënboer JIG. Granularity Matters: Comparing Different Ways of Measuring Self-Regulated Learning. *Metacognition Learn.* 2019;14(1):1-19. doi:10.1007/s11409-019-09188-6
  22. Bransen D, Govaerts MJB. How to conceptualise self-regulated learning: Implications for measurement. *Med Educ.* n/a(n/a). doi:10.1111/medu.14183
  23. Gandomkar R, Yazdani K, Fata L, et al. Using multiple self-regulated learning measures to understand medical students' biomedical science learning. *Med Educ.* Published online February 3, 2020. doi:10.1111/medu.14079
  24. Hadwin A, Oshige M. Self-Regulation, Coregulation, and Socially Shared Regulation: Exploring Perspectives of Social in Self-Regulated Learning Theory. *Teach Coll Rec.* 2011;113(2):240-264.
  25. M.Bachelet; M Anguis ; Les médecins d'ici à 2040 : une population plus jeune, plus féminisée et plus souvent salariée;Drees; Mai 2017;numéro 1011. Accessed November 29, 2019. <https://drees.solidarites-sante.gouv.fr/IMG/pdf/er1011.pdf>

## Tables

Table 1: Details on participants' characteristics according to groups

	PCP n (%)	Patients n (%)	Total n (%)
Women	22 (55)	44 (62,86)	66 (60)
Men	18 (45)	26 (37.14)	44 (40)
<b>Age</b>			
20-30	13 (32.5)	7 (10)	20 (18.18)
31-40	17 (42.5)	15 (21.43)	32 (29.09)
41-50	1 (2.5)	10 (14.29)	11 (10)
51-60	7 (17.5)	10 (14.29)	17 (15.45)
61-70	2 (5)	23 (32.86)	25 (22.73)
71-80	0	1 (1.43)	1 (0.91)
81-90	0	3 (4.29)	3 (2.73)
91-100	0	1 (1.43)	1 (0.91)
Median response time	28 min	42 min	38 min

Table 2 : Details on the distribution of responses and statistical analysis results between PCPs and patients groups

	PCPs n (%)	Patients n (%)	Odds-ratio [CI95%]	p
<b>Question 1</b>				
Totally unacceptable	1 (2.5)	9 (12.86)	0.31[0.14-0.65]	<b>0.001</b>
Hardly acceptable	3 (7.5)	19 (27.14)		
Acceptable	23 (57.5)	29 (41.43)		
Totally acceptable	13 (32.5)	13 (18.57)		
<b>Question 2</b>				
Totally unacceptable	13 (32.5)	30 (42.86)	0.79 [0.39-1.62]	0.519
Hardly acceptable	19 (47.5)	25 (35.71)		
Acceptable	6 (15.0)	10 (14.29)		
Totally acceptable	2 (5.0)	5 (7.14)		
<b>Question3</b>				
Totally unacceptable	4 (10)	8 (11.43)	2.67 [1.29-5.68]	<b>0.008</b>
Hardly acceptable	26 (65)	21 (30)		
Acceptable	6 (15)	27 (38.57)		
Totally acceptable	4 (10)	14 (20)		
<b>Question 4</b>				
Totally unacceptable	5 (12.5)	7 (10)	2.94 [1.42-6.18]	<b>0.004</b>
Hardly acceptable	8 (20)	6 (8.57)		
Acceptable	15 (37.5)	13 (18.57)		
Totally acceptable	12 (30)	44 (62.86)		
<b>Question 5</b>				
Totally unacceptable	26 (65)	57 (81.43)	0.41 [0.17-0.98]	<b>0.046</b>
Hardly acceptable	11 (27.5)	12 (17.14)		
Acceptable	1 (2.5)	0 (0)		
Totally acceptable	2 (5)	1 (1.43)		
<b>Question 6</b>				
Totally unacceptable	2 (5)	3 (4.29)	3.78 [1.79-8.17]	<b>&lt;0.001</b>
Hardly acceptable	14 (35)	6 (8.57)		
Acceptable	11 (27.5)	16 (22.86)		
Totally acceptable	13 (32.5)	45 (64.29)		
<b>Question 7</b>				
Totally unacceptable	15 (37.5)	21 (30)	2.24 [1.11-4.58]	<b>0.024</b>
Hardly acceptable	17 (42.5)	16 (22.86)		
Acceptable	7 (17.5)	22 (31.43)		
Totally acceptable	1 (2.5)	11 (15.71)		
<b>Question 8</b>				
Totally unacceptable	2 (5)	18 (25.71)	0.39 [0.19-0.78]	<b>0.008</b>

Hardly acceptable	8 (20)	18 (25.71)		
Acceptable	14 (35)	15 (21.43)		
Totally acceptable	16 (40)	19 (27.14)		

#### Question 9

Totally unacceptable	13 (32.5)	17 (24.29)		
Hardly acceptable	17 (42.5)	26 (37.14)	1.66 [0.82-3.42]	0.159
Acceptable	7 (17.5)	19 (27.14)		
Totally acceptable	3 (7.5)	8 (11.43)		

#### Question 10

Totally unacceptable	29 (72.5)	63 (90)		
Hardly acceptable	10 (25)	5 (7.14)	0.31 [0.10-0.86]	<b>0.024</b>
Acceptable	1 (2.5)	0 (0)		
Totally acceptable	0 (0)	2 (2.86)		

#### Question 11

Totally unacceptable	7 (17.5)	36 (51.43)		
Hardly acceptable	22 (55)	26 (37.14)	0.24 [0.11-0.52]	<b>&lt;0.001</b>
Acceptable	7 (17.5)	7 (10)		
Totally acceptable	4 (10)	1 (1.43)		

#### Question 12

Totally unacceptable	25 (62.5)	38 (54.29)		
Hardly acceptable	13 (32.5)	19 (27.14)	1.63 [0.76-3.57]	0.211
Acceptable	1 (2.5)	9 (12.86)		
Totally acceptable	1 (2.5)	4 (5.71)		

#### Question 13

Totally unacceptable	0 (0)	3 (4.29)		
Hardly acceptable	0 (0)	5 (7.14)	0.65 [0.29-1.40]	0.270
Acceptable	15 (37.5)	23 (32.86)		
Totally acceptable	25 (62.5)	39 (55.71)		

#### Question 14

Totally unacceptable	1 (2.5)	13 (18.57)		
Hardly acceptable	12 (30)	20 (28.57)	0.57 [0.28-1.14]	0.110
Acceptable	13 (32.5)	16 (22.86)		
Totally acceptable	14 (35)	21 (30)		

#### Question 15

Totally unacceptable	4 (10)	15 (21.43)		
Hardly acceptable	15 (37.5)	30 (42.86)	0.52 [0.25-1.06]	0.069
Acceptable	17 (42.5)	19 (27.14)		
Totally acceptable	4 (10)	6 (8.57)		

## Synthèse

Ce travail a permis d'explorer une suggestion pédagogique reçue lors d'un congrès de pédagogie. Il s'agit en ce sens d'une démarche scientifique de vérification de la pertinence d'une idée nouvelle. Il s'agit d'un travail novateur car aucun article n'a jamais été publié sur l'inclusion de patients dans les panels de FpC-jugement.

Si cet article présente une méthodologie mixte (quantitative et qualitative), c'est bien la partie quantitative qui doit être mise en avant. Comme énoncé dans la section sur les formations par concordance et le développement de la réflexivité, l'analyse qualitative des justifications de nos participants n'est pas robuste méthodologiquement. Il s'agit ici d'un travail de "défrichage" du sujet. Des travaux ultérieurs employant une méthodologie qualitative robuste devront approfondir cette thématique.

Concrètement, ce travail va permettre de déployer la première FpC-jugement intégrant des patients dans les panels pour les étudiants de troisième cycle de médecine à l'Université de Strasbourg. L'impact de cette FpC-jugement sur les apprentissages des étudiants en matière de professionnalisme sera à documenter dans des travaux ultérieurs.

En appliquant le langage de la MOT à ce travail, il est possible d'analyser les justifications des participants selon les types de connaissances. Les participants raisonnaient en maniant des connaissances déclaratives concrètes (exemples) mais aussi des connaissances déclaratives abstraites (concepts). Néanmoins, ce sont surtout des connaissances stratégiques abstraites (principes) qui étaient mises en avant. Les participants évoquaient ainsi des principes tels que la confraternité, l'humanité, l'empathie. Bien que la tâche cognitive à l'œuvre dans une FpC-jugement ne soit pas de l'ordre du raisonnement clinique, des points communs notables peuvent être soulignés. Les vignettes et les justifications que faisaient les participants de leurs jugements révélaient des influences d'origine émotionnelle et relationnelles comme dans la procédure "réguler le processus" du modèle MOT. Les participants tenaient compte dans leurs jugements de la perspective du patient (et notamment de ses craintes) comme dans le principe "le clinicien tient compte, tout au long du processus, de la perspective du patient à l'égard de son problème : ses croyances, ses craintes, ses attentes, ses choix, ses émotions et l'impact du problème sur sa vie".

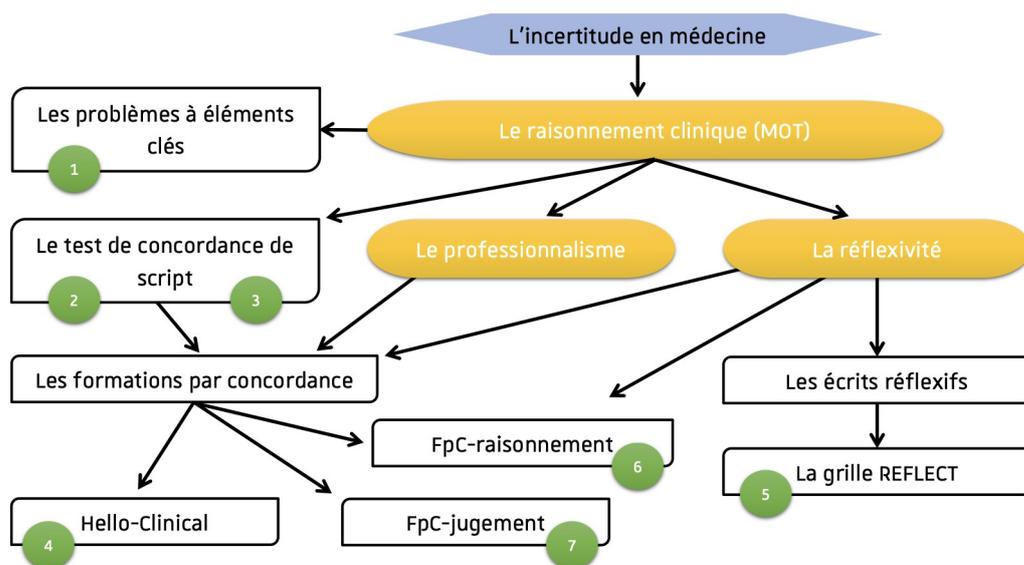
Aussi, l'utilisation de patients dans les panels de FpC-jugement pourraient indirectement favoriser le développement du raisonnement clinique en illustrant l'importance de ces influences cognitives, émotionnelles et de la prise en compte de la perspective du patient dans les soins.

## SYNTHÈSE ET PERSPECTIVES

Au travers de ce document, j'ai retracé mon parcours de formation et de recherche en Pédagogie des Sciences de la Santé débuté lors de mon master en 2013. A partir de problématiques vécues en formation initiale en tant qu'étudiant puis de problématiques de formateur dans un second temps, je me suis approprié un certain nombre de concepts en pédagogie des sciences de la santé.

Mon questionnement sur l'incertitude dans l'exercice de la médecine m'a amené à m'intéresser au raisonnement clinique. J'ai découvert des outils d'évaluation du raisonnement clinique comme les problèmes à éléments clés et le TCS. Le champ du raisonnement clinique m'a amené à vers les concepts de réflexivité et de professionnalisme. J'ai découvert le groupe des formations par concordance et je me suis intéressé à son application à un outil utilisant des patients virtuels. J'ai considéré l'impact des écrits réflexifs et de leur évaluation sur le développement de la réflexivité.

Plusieurs travaux de recherche originaux ont découlé de ce cheminement intellectuel en pédagogie des sciences de la santé. J'ai tout d'abord montré que le burnout n'avait pas d'impact sur la performance du raisonnement clinique évalué par un TCS dans une population d'étudiants en troisième cycle de médecine générale. J'ai ensuite montré via l'outil Hello-Clinical que le nombre d'années d'expérience en médecine générale semble être associé aux prescriptions complémentaires en situation d'incertitude. J'ai traduit et adapté en langue française un outil validé d'évaluation de la réflexivité des traces écrites. J'ai documenté l'impact d'une formation par concordance sur le développement de la réflexivité. J'ai enfin montré que l'inclusion de patients dans les panels de FpC-jugement avait le potentiel pour enrichir la qualité de la rétro-information apportée aux étudiants. Je reprends ci-dessous la figure n°1 présentée en préambule qui illustre ce cheminement et fait apparaître le "fil conducteur" principal de cette thèse. Les numéros correspondent aux différents articles présentés précédemment.



#### REPRISE DE LA FIGURE N°1

L'évaluation des apprentissages en pédagogie des sciences de la santé est le fil conducteur principal de cette thèse. La modélisation par objets typés et le modèle MOT sont les fils conducteurs secondaires. Ainsi, ces différents travaux forment un ensemble cohérent. Ils visent à permettre de mieux former les professionnels de santé en général, les étudiants de troisième cycle de médecine générale en particulier. Les deux articles didactiques sur les problèmes à éléments clés et les TCS serviront à favoriser l'appropriation de ces outils relativement novateurs en France dans le cadre de la réforme du deuxième cycle des études médicales. Si les universitaires de Médecine Générale créent des problèmes à éléments clés et des questions de TCS en lien avec cette discipline, cela influencera la nature et la qualité des apprentissages réalisés par les étudiants en amont de leur troisième cycle. En proposant au cours de ce troisième cycle en médecine générale diverses formations par concordance, il sera possible de favoriser le développement du raisonnement clinique, du professionnalisme et de la réflexivité chez ces étudiants. L'utilisation de la grille REFLECT dans l'analyse des écrits réflexifs pourrait également favoriser le développement de la réflexivité chez ces mêmes étudiants.

L'évaluation des apprentissages en pédagogie des sciences de la santé est désormais ma thématique de recherche principale. Ces différents travaux m'ont donné davantage de légitimité et de confiance pour animer des formations sur cette thématique.

De ces travaux découlent de nombreuses perspectives. L'outil Hello-Clinical finalisera son développement en 2022. Il fera l'objet d'une étude ambitieuse visant à rechercher une corrélation entre les difficultés de raisonnement clinique documentées sur des patients virtuels à celles constatées en stage chez des étudiants en médecine. Des formations à destination des internes de médecine générale de la Faculté de Médecine de Strasbourg verront le jour via Hello-Clinical dans les prochaines années. Des travaux ultérieurs viseront à documenter l'impact de l'utilisation des patients virtuels en formation sur les apprentissages des étudiants pour mieux cerner la place de ce genre d'outil dans la formation des futurs professionnels de santé. Des formations seront également proposées dans le cadre du développement professionnel continu (DPC) via Hello-Clinical aux médecins généralistes en exercice dans les prochaines années. Il sera possible à l'avenir d'étendre ces formations à divers professionnels de santé en formation initiale et continue : infirmiers, kinésithérapeutes, pharmaciens, sages-femmes, aides-soignants, ambulanciers, orthophonistes.

A la suite de la publication de la version française de la grille REFLECT, son utilisation pourra être encouragée pour aider au développement de la réflexivité des internes de Médecine Générale de la Faculté de Médecine de Strasbourg mais aussi plus largement au niveau national. Cette grille pourrait également se révéler une aide dans le cadre de la réforme du deuxième cycle des études médicales pour optimiser le développement de la réflexivité de plus jeunes étudiants. De nouveaux travaux pourront être menés pour documenter les propriétés psychométriques de la version française de la grille REFLECT dans des études de validation de grande ampleur.

De nouvelles FpC-raisonnement sont en cours de finalisation sur les pathologies cutanées en médecine générale et commencent à être proposées aux internes de Strasbourg. Ces formations feront l'objet de déclinaisons en DPC à destination des médecins généralistes en exercice.

Une FpC-jugement est en cours de recrutement d'un panel mixte patients/médecins généralistes pour être proposée aux internes Strasbourgeois. Des travaux ultérieurs pourront déterminer l'impact de l'inclusion de pairs étudiants dans ces panels de référence.

## BIBLIOGRAPHIE

1. Pelaccia T. Comment les médecins urgentologues raisonnent-ils au regard des spécificités de leur cadre et de leur mode d'exercice ? [Internet] [These de doctorat]. Strasbourg; 2014 [cité 22 nov 2021]. Disponible sur: <https://www.theses.fr/2014STRAG034>
2. Pelaccia T, Tribby E. La pédagogie médicale est-elle une discipline ? *Pédagogie Médicale*. mai 2011;12(2):121-32.
3. Medical education. In: Wikipedia [Internet]. 2021 [cité 21 nov 2021]. Disponible sur: [https://en.wikipedia.org/w/index.php?title=Medical\\_education&oldid=1052153512](https://en.wikipedia.org/w/index.php?title=Medical_education&oldid=1052153512)
4. M. Harden JG Graham Buckley, IR Hart, R. BEME Guide No. 1: Best Evidence Medical Education. *Med Teach*. 1 janv 1999;21(6):553-62.
5. Albert M. Understanding the debate on medical education research: a sociological perspective. *Acad Med J Assoc Am Med Coll*. oct 2004;79(10):948-54.
6. Albert M, Hodges B, Regehr G. La recherche en éducation médicale : entre le service et la science. *Pédagogie Médicale*. mai 2006;7(2):73-81.
7. Bunniss S, Kelly DR. Research paradigms in medical education research. *Med Educ*. 1 avr 2010;44(4):358-66.
8. Arrêté du 2 septembre 2020 portant modification de diverses dispositions relatives au régime des études en vue du premier et du deuxième cycle des études médicales et à l'organisation des épreuves classantes nationales. ESRS2018628A sept 10, 2020.
9. Arrêté du 8 avril 2013 relatif au régime des études en vue du premier et du deuxième cycle des études médicales.
10. Tardif J. Développer un programme par compétences : de l'intention à la mise en œuvre. *Pédagogie Collégiale*. mars 2003;16(3):36-44.
11. Parent F, Jouquan J, De Ketele J-M. CanMEDS and other « competency and outcome-based approaches » in medical education: clarifying the ongoing ambiguity. *Adv Health Sci Educ Theory Pract*. mars 2013;18(1):115-22.
12. Arrêté du 12 avril 2017 portant organisation du troisième cycle des études de médecine.

13. Compagnon L, Bail P, Huez J, Stalnikiewicz B, Ghasarossian C, Zerbib Y, et al. Définitions et descriptions des compétences en médecine générale. *Exercer*. 2013;(108):148-55.
14. Compagnon L, Bail P, Huez J, Stalnikiewicz B, Ghasarossian C, Zerbib Y, et al. Les niveaux de compétences. *Exercer*. 2013;(108):156-64.
15. Parent F, Jouquan J. Inscrire une formation dans une approche par compétences. In: Comment (mieux) former et évaluer les étudiants en médecine et en sciences de la santé ? De Boeck Supérieur; 2016. p. 107-23.
16. Faucher C, Pelaccia T, Nendaz M, Audétat M-C, Charlin B. Un professionnel de santé qui résout efficacement les problèmes : le raisonnement clinique. In: Comment (mieux) former et évaluer les étudiants en médecine et en sciences de la santé ? De Boeck Supérieur; 2016. p. 33-44.
17. Nguyen QD, Raymond-Carrier S. Un professionnel de santé qui se pose des questions : la réflexivité. In: Comment (mieux) former et évaluer les étudiants en médecine et en sciences de la santé ? De Boeck Supérieur; 2016.
18. Robert D, Payot A, Lajeunesse Y. Un professionnel de santé qui se comporte de façon professionnelle et éthique : le professionnalisme. In: Comment (mieux) former et évaluer les étudiants en médecine et en sciences de la santé ? De Boeck Supérieur; 2016.
19. Cruess SR, Cruess RL. Professionalism: a contract between medicine and society. *CMAJ Can Med Assoc J*. 7 mars 2000;162(5):668-9.
20. O'Sullivan H, van Mook W, Fewtrell R, Wass V. Integrating professionalism into the curriculum: AMEE Guide No. 61. *Med Teach*. 2012;34(2):e64-77.
21. Arrêté du 21 avril 2017 relatif aux connaissances, aux compétences et aux maquettes de formation des diplômés d'études spécialisées et fixant la liste de ces diplômés et des options et formations spécialisées transversales du troisième cycle des études de médecine - Légifrance [Internet]. Disponible sur: <https://www.legifrance.gouv.fr/loda/id/LEGITEXT000034508221/2018-04-25/>
22. Dutch College of General Practitioners. Medically unexplained symptoms. 2013.
23. Van Der Weijden T, van Velsen M, Dinant G-J, van Hasselt CM, Grol R. Unexplained complaints in general practice: prevalence, patients' expectations, and professionals' test-ordering behavior. *Med Decis Mak Int J Soc Med Decis Mak*. juin 2003;23(3):226-31.
24. Biehn J. Managing uncertainty in family practice. *Can Med Assoc J*. 15 avr 1982;126(8):915-7.

25. Thomson GH. Tolerating uncertainty in family medicine. *J R Coll Gen Pract.* juin 1978;28(191):343-6.
26. Bloy G. L'incertitude en médecine générale : sources, formes et accommodements possibles. *Sci Soc Santé.* 2008;26(1):67-91.
27. Fox RC. L'incertitude médicale. Paris, France, Belgique: L'Harmattan; 1988. 202 p.
28. Freeston MH, Rhéaume J, Letarte H, Dugas MJ, Ladouceur R. Why do people worry? *Personal Individ Differ.* 1 déc 1994;17(6):791-802.
29. Birrell J, Meares K, Wilkinson A, Freeston M. Toward a definition of intolerance of uncertainty: A review of factor analytical studies of the Intolerance of Uncertainty Scale. *Clin Psychol Rev.* nov 2011;31(7):1198-208.
30. Anderson KG, Dugas MJ, Koerner N, Radomsky AS, Savard P, Turcotte J. Interpretive style and intolerance of uncertainty in individuals with anxiety disorders: a focus on generalized anxiety disorder. *J Anxiety Disord.* déc 2012;26(8):823-32.
31. Fergus TA, Wu KD. The intolerance of uncertainty scale: measurement invariance, population heterogeneity, and its relation with worry among self-identifying White and Black respondents. *Assessment.* oct 2013;20(5):555-64.
32. Oglesby ME, Medley AN, Norr AM, Capron DW, Korte KJ, Schmidt NB. Intolerance of uncertainty as a vulnerability factor for hoarding behaviors. *J Affect Disord.* 20 févr 2013;145(2):227-31.
33. Lieu TA, Schroeder SA, Altman DF. Specialty choices at one medical school: recent trends and analysis of predictive factors. *Acad Med J Assoc Am Med Coll.* oct 1989;64(10):622-9.
34. Cooke GPE, Doust JA, Steele MC. A survey of resilience, burnout, and tolerance of uncertainty in Australian general practice registrars. *BMC Med Educ.* 2013;13:2.
35. Cooke G, Tapley A, Holliday E, Morgan S, Henderson K, Ball J, et al. Responses to clinical uncertainty in Australian general practice trainees: a cross-sectional analysis. *Med Educ.* déc 2017;51(12):1277-88.
36. Simpkin AL, Schwartzstein RM. Tolerating Uncertainty — The Next Medical Revolution? *N Engl J Med.* 3 nov 2016;375(18):1713-5.
37. Nevalainen M, Kuikka L, Pitkälä K. Medical errors and uncertainty in primary healthcare: a comparative study of coping strategies among young and experienced GPs. *Scand J Prim Health Care.* juin

2014;32(2):84-9.

38. Nevalainen M, Kuikka L, Sjöberg L, Eriksson J, Pitkala K. Tolerance of uncertainty and fears of making mistakes among fifth-year medical students. *Fam Med.* avr 2012;44(4):240-6.

39. Lorenzo M. Tolérance à l'incertitude en médecine générale: une démarche de théorisation ancrée à partir de 14 entretiens [Thèse d'exercice]. [France]: Université de Strasbourg (2009-....). Faculté de médecine; 2013.

40. Young M, Thomas A, Lubarsky S, Ballard T, Gordon D, Gruppen LD, et al. Drawing Boundaries: The Difficulty in Defining Clinical Reasoning. *Acad Med J Assoc Am Med Coll.* juill 2018;93(7):990-5.

41. Barrows HS. Problem-solving learning. *Med Educ.* mai 1980;14(3):242-3.

42. Pelaccia T, Tardif J, Tribby E, Charlin B. An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory. *Med Educ Online.* 2011;16.

43. Bleakley A. Re-visioning clinical reasoning, or stepping out from the skull. *Med Teach.* 3 avr 2021;43(4):456-62.

44. Daniel M, Rencic J, Durning SJ, Holmboe E, Santen SA, Lang V, et al. Clinical Reasoning Assessment Methods: A Scoping Review and Practical Guidance. *Acad Med.* juin 2019;94(6):902-12.

45. Charlin B, Lubarsky S, Millette B, Crevier F, Audétat M-C, Charbonneau A, et al. Clinical reasoning processes: unravelling complexity through graphical representation. *Med Educ.* mai 2012;46(5):454-63.

46. Charlin B, Tardif J, Boshuizen HP. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Acad Med J Assoc Am Med Coll.* févr 2000;75(2):182-90.

47. Higgs J, Jensen GM, Loftus S, Christensen N. *Clinical reasoning in the health professions.* Edinburgh: ElsevierHealth Sciences; 2018. 532 p.

48. Référentiel CanMEDS :: Le Collège royal des médecins et chirurgiens du Canada [Internet]. [cité 27 janv 2021]. Disponible sur: <https://www.royalcollege.ca/rcsite/canmeds/canmeds-framework-f>

49. Schumacher D. Competency 6. Make informed diagnostic and therapeutic decisions that result in optimal clinical judgment. *Acad Pediatr.* 1 mars 2014;14(2):S21-3.

50. ACGME. Family Medicine Milestones [Internet]. 2019. Disponible sur: <https://www.acgme.org/Portals/0/PDFs/Milestones/FamilyMedicineMilestones.pdf?ver=2020-09-01-150203-400>

51. Tuning Learning Outcomes/Competences for Undergraduate Medical Education in Europe [Internet]. Disponible sur: [http://www.unideusto.org/tuningeu/images/stories/Summary\\_of\\_outcomes\\_TN/Learning\\_Outcomes\\_Competences\\_for\\_Undergraduate\\_Medical\\_Education\\_in\\_Europe.pdf](http://www.unideusto.org/tuningeu/images/stories/Summary_of_outcomes_TN/Learning_Outcomes_Competences_for_Undergraduate_Medical_Education_in_Europe.pdf)
52. Arrêté du 31 juillet 2009 relatif au diplôme d'Etat d'infirmier.
53. Association des infirmières et infirmiers du Canada. Cadre des compétences de base des infirmières et infirmiers praticiens du Canada [Internet]. 2010. Disponible sur: [https://www.cna-aiic.ca/-/media/cna/page-content/pdf-fr/competency\\_framework\\_2010\\_f.pdf](https://www.cna-aiic.ca/-/media/cna/page-content/pdf-fr/competency_framework_2010_f.pdf)
54. Décret n° 2013-798 du 30 août 2013 relatif au régime des études en vue du certificat de capacité d'orthophoniste. 2013-798 août 30, 2013.
55. Alliance canadienne des organismes de réglementation en orthophonie et en audiologie. Profil de compétences nationales pour l'orthophonie [Internet]. 2018. Disponible sur: <https://caaspr.ca/sites/default/files/2019-08/Profil-de-compe%CC%81tences-nationales-pour-lorthophonie.pdf>
56. Hege I, Adler M, Durning S, Edelbring S, Huwendiek S, Kononowicz A, et al. Development of a longitudinal interprofessional clinical reasoning curriculum – an international project. Jahrestag Ges Für Med Ausbild GMA [Internet]. 17 nov 2020 [cité 6 juin 2021]; Disponible sur: <https://www.egms.de/en/meetings/gma2020/20gma003.shtml>
57. Wilson RM, Harrison BT, Gibberd RW, Hamilton JD. An analysis of the causes of adverse events from the Quality in Australian Health Care Study. Med J Aust. 3 mai 1999;170(9):411-5.
58. Berner ES, Graber ML. Overconfidence as a Cause of Diagnostic Error in Medicine. Am J Med. 1 mai 2008;121(5, Supplement):S2-23.
59. Hauer KE, Ciccone A, Henzel TR, Katsufakis P, Miller SH, Norcross WA, et al. Remediation of the deficiencies of physicians across the continuum from medical school to practice: a thematic review of the literature. Acad Med J Assoc Am Med Coll. déc 2009;84(12):1822-32.
60. Yates J, James D. Predicting the « strugglers »: a case-control study of students at Nottingham University Medical School. BMJ. 29 avr 2006;332(7548):1009-13.
61. Audétat M-C, Dory V, Nendaz M, Vanpee D, Pestiaux D, Junod Perron N, et al. What is so difficult about managing clinical reasoning difficulties? Med Educ. févr 2012;46(2):216-27.
62. Hicks PJ, Cox SM, Espey EL, Goepfert AR, Bienstock JL, Erickson SS, et al. To the point: medical

education reviews--dealing with student difficulties in the clinical setting. *Am J Obstet Gynecol.* déc 2005;193(6):1915-22.

63. Frellsen SL, Baker EA, Papp KK, Durning SJ. Medical school policies regarding struggling medical students during the internal medicine clerkships: results of a national survey. *Acad Med J Assoc Am Med Coll.* sept 2008;83(9):876-81.

64. Gruppen LD, Frohna AZ. Clinical reasoning. In: *International Handbook of Research in Medical Education.*

65. Croskerry P. A universal model of diagnostic reasoning. *Acad Med J Assoc Am Med Coll.* août 2009;84(8):1022-8.

66. Hogarth RM. *Educating Intuition.* 1 edition. Chicago: University of Chicago Press; 2001. 312 p.

67. Norman G. Research in clinical reasoning: past history and current trends. *Med Educ.* avr 2005;39(4):418-27.

68. Deciding Analytically or Trusting your Intuition? The Advantages and Disadvantages of Analytic and Intuitive Thought by Robin M. Hogarth :: SSRN [Internet]. [cité 23 juill 2016]. Disponible sur: [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=394920](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=394920)

69. Kahneman D. A perspective on judgment and choice: mapping bounded rationality. *Am Psychol.* sept 2003;58(9):697-720.

70. Croskerry P. Context is everything or how could I have been that stupid? *Healthc Q Tor Ont.* 2009;12:e171-176.

71. Norman GR, Eva KW. Diagnostic error and clinical reasoning. *Med Educ.* janv 2010;44(1):94-100.

72. Wikipedia contributors. Intuition. In: *Wikipédia* [Internet]. Wikimedia Foundation, Inc.; 2012 [cité 27 août 2012]. Disponible sur: <http://fr.wikipedia.org/w/index.php?title=Intuition&oldid=80635719>

73. Nendaz M, Charlin B, Leblanc V, Bordage G. Le raisonnement clinique: données issues de la recherche et implications pour l'enseignement. *Pédagogie Médicale.* nov 2005;6(4):235-54.

74. Pelaccia T, Tardif J, Tribby E, Ammirati C, Bertrand C, Dory V, et al. How and When Do Expert Emergency Physicians Generate and Evaluate Diagnostic Hypotheses? A Qualitative Study Using Head-Mounted Video Cued-Recall Interviews. *Ann Emerg Med.* 29 mai 2014;

75. Moulton CE, Regehr G, Mylopoulos M, MacRae HM. Slowing down when you should: a new model of expert judgment. *Acad Med J Assoc Am Med Coll.* oct 2007;82(10 Suppl):S109-116.

76. Croskerry P, Cosby KS, Schenkel SM, Wears RL. Patient Safety in Emergency Medicine. Philadelphia: Lippincott Williams and Wilkins; 2008. 448 p.
77. Charlin B, Boshuizen HPA, Custers EJ, Feltovich PJ. Scripts and clinical reasoning. *Med Educ.* déc 2007;41(12):1178-84.
78. Epstein S. Integration of the cognitive and the psychodynamic unconscious. *Am Psychol.* 1994;49(8):709-24.
79. Hammond KR. Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice. Oxford University Press; 2000. 449 p.
80. Kahneman D, Klein G. Conditions for intuitive expertise: a failure to disagree. *Am Psychol.* sept 2009;64(6):515-26.
81. Schmidt HG, Rikers RMJP. How expertise develops in medicine: knowledge encapsulation and illness script formation. *Med Educ.* déc 2007;41(12):1133-9.
82. Elstein AS, Schwartz A, Schwarz A. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ.* 23 mars 2002;324(7339):729-32.
83. Merkebu J, Battistone M, McMains K, McOwen K, Witkop C, Konopasky A, et al. Situativity: a family of social cognitive theories for understanding clinical reasoning and diagnostic error. *Diagn Berl Ger.* 27 août 2020;7(3):169-76.
84. Durning SJ, Artino AR. Situativity theory: a perspective on how participants and the environment can interact: AMEE Guide no. 52. *Med Teach.* 2011;33(3):188-99.
85. Chang RW, Bordage G, Connell KJ. The importance of early problem representation during case presentations. *Acad Med J Assoc Am Med Coll.* oct 1998;73(10 Suppl):S109-111.
86. Bordage G. [Decision management in medicine: some mental processes and practical consultations]. *Rev Med Interne.* juin 2005;26 Suppl 1:S14-17.
87. Bowen JL. Educational strategies to promote clinical diagnostic reasoning. *N Engl J Med.* 23 nov 2006;355(21):2217-25.
88. Paquette G. Modélisation des connaissances et des compétences [Internet]. Presse de l'université du Québec. 2002 [cité 23 nov 2014]. 388 p. (Hors collection). Disponible sur: <http://www.puq.ca/catalogue/livres/modelisation-des-connaissances-des-competences-172.html>
89. Tardif J. L'évaluation dans le paradigme constructiviste. In: L'évaluation des apprentissages:

réflexions, nouvelles tendances et formation. Sherbrooke (Québec), Canada: éd. du CRP, Faculté d'éducation; 1993. p. 27-56.

90. Bertrand C, Dory V, Pelaccia T, Durand E, Vaux J, Charlin B, et al. Choisir un outil d'évaluation. In: Comment (mieux) former et évaluer les étudiants en médecine et en sciences de la santé ? De Boeck Supérieur; 2016. p. 33-44.

91. Lorenzo M. Les problèmes à éléments clés - Un nouvel outil d'évaluation introduit par la réforme du deuxième cycle. *exercer*. oct 2021;(176):375-9.

92. Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten C. The Script Concordance test: a tool to assess the reflective clinician. *Teach Learn Med*. 2000;12(4):189-95.

93. Lorenzo M. Qu'est-ce que le test de concordance de script ? Un outil d'évaluation du raisonnement clinique introduit par la réforme du deuxième cycle en France. *exercer*. avr 2021;(172):185-8.

94. Sweller J. Cognitive Load During Problem Solving: Effects on Learning. *Cogn Sci*. 1988;12(2):257-85.

95. Gawad N, Wood TJ, Cowley L, Raiche I. The cognitive process of test takers when using the script concordance test rating scale. *Med Educ*. 2020;54(4):337-47.

96. Bland AC, Kreiter CD, Gordon JA. The psychometric properties of five scoring methods applied to the script concordance test. *Acad Med J Assoc Am Med Coll*. avr 2005;80(4):395-9.

97. Wilson AB, Pike GR, Humbert AJ. Analyzing script concordance test scoring methods and items by difficulty and type. *Teach Learn Med*. 2014;26(2):135-45.

98. 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.

99. Lubarsky S, Dory V, Duggan P, Gagnon R, Charlin B. Script concordance testing: from theory to practice: AMEE guide no. 75. *Med Teach*. 2013;35(3):184-93.

100. Freudemberger HJ. Staff Burn-Out. *J Soc Issues*. mars 1974;30(1):159-65.

101. Maslach C, Jackson SE. The measurement of experienced burnout. *J Organ Behav*. 1 avr 1981;2(2):99-113.

102. Maslach C, Schaufeli WB, Leiter MP. Job burnout. *Annu Rev Psychol*. 2001;52:397-422.

103. Ishak WW, Lederer S, Mandili C, Nikravesh R, Seligman L, Vasa M, et al. Burnout during residency training: a literature review. *J Grad Med Educ.* déc 2009;1(2):236-42.
104. Galam E, Komly V, Tourneur AL, Jund J. Burnout among French GPs in training: a cross-sectional study. *Br J Gen Pract.* mars 2013;63(608):e217.
105. Dyrbye L, Shanafelt T. A narrative review on burnout experienced by medical students and residents. *Med Educ.* janv 2016;50(1):132-49.
106. Erschens R, Keifenheim KE, Herrmann-Werner A, Loda T, Schwille-Kiuntke J, Bugaj TJ, et al. Professional burnout among medical students: Systematic literature review and meta-analysis. *Med Teach.* 1 févr 2019;41(2):172-83.
107. Deligkaris P, Panagopoulou E, Montgomery AJ, Masoura E. Job burnout and cognitive functioning: A systematic review. *Work Stress.* avr 2014;28(2):107-23.
108. Durning SJ, Costanzo M, Artino AR, Dyrbye LN, Beckman TJ, Schuwirth L, et al. Functional Neuroimaging Correlates of Burnout among Internal Medicine Residents and Faculty Members. *Front Psychiatry.* 2013;4:131:doi:10.3389/fpsy.2013.00131.
109. Brunsberg KA, Landrigan CP, Garcia BM, Petty CR, Sectish TC, Simpkin AL, et al. Association of Pediatric Resident Physician Depression and Burnout With Harmful Medical Errors on Inpatient Services. *Acad Med J Assoc Am Med Coll.* 30 avr 2019;
110. McBee E, Ratcliffe T, Picho K, Schuwirth L, Artino AR, Yepes-Rios AM, et al. Contextual factors and clinical reasoning: differences in diagnostic and therapeutic reasoning in board certified versus resident physicians. *BMC Med Educ.* 15 nov 2017;17(1):211.
111. Kwah J, Weintraub J, Fallar R, Ripp J. The Effect of Burnout on Medical Errors and Professionalism in First-Year Internal Medicine Residents. *J Grad Med Educ.* oct 2016;8(4):597-600.
112. Maslach C, Jackson SE, Leiter MP. Maslach Burnout Inventory Manual. In: 3rd edition. Palo Alto, CA: Consulting Psychologists Press; 1996. p. 191-218.
113. Rotenstein LS, Torre M, Ramos MA, Rosales RC, Guille C, Sen S, et al. Prevalence of Burnout Among Physicians: A Systematic Review. *JAMA.* 18 2018;320(11):1131-50.
114. Brenninkmeijer V, VanYperen N. How to conduct research on burnout: advantages and disadvantages of a unidimensional approach in burnout research. *Occup Environ Med.* 1 juin 2003;60(suppl 1):i16-20.

115. Ernst M. Le syndrome de burnout des internes en médecine générale à la Faculté de Médecine de Strasbourg: prévalence et analyse d'entretiens. [France]: Université de Strasbourg. Faculté de médecine; 2009.
116. Schaufeli WB, Van Dierendonck D. A cautionary note about the cross-national and clinical validity of cut-off points for the Maslach Burnout Inventory. *Psychol Rep.* juin 1995;76(3 Pt 2):1083-90.
117. Kovess-Masfety V, Saunder L. Le burnout : historique, mesures et controverses. *Arch Mal Prof Environ.* 1 févr 2017;78(1):16-23.
118. Cooke S, Lemay J-F. Transforming Medical Assessment: Integrating Uncertainty Into the Evaluation of Clinical Reasoning in Medical Education. *Acad Med.* juin 2017;92(6):746-51.
119. Attali C, Huez J-F, Valette T, Lehr-Drylewicz A-M. Les grandes familles de situations cliniques. *Exercer.* 2013;24(108):165-9.
120. Dory V, Gagnon R, Vanpee D, Charlin B. How to construct and implement script concordance tests: insights from a systematic review. *Med Educ.* 1 juin 2012;46(6):552-63.
121. McConnell MM, Eva KW. The role of emotion in the learning and transfer of clinical skills and knowledge. *Acad Med J Assoc Am Med Coll.* oct 2012;87(10):1316-22.
122. Kozlowski D, Hutchinson M, Hurley J, Rowley J, Sutherland J. The role of emotion in clinical decision making: an integrative literature review. *BMC Med Educ.* 15 déc 2017;17(1):255.
123. Croskerry P, Norman G. Overconfidence in clinical decision making. *Am J Med.* mai 2008;121(5 Suppl):S24-29.
124. Croskerry P, Abbass AA, Wu AW. How doctors feel: affective issues in patients' safety. *Lancet Lond Engl.* 4 oct 2008;372(9645):1205-6.
125. Croskerry P. A universal model of diagnostic reasoning. *Acad Med J Assoc Am Med Coll.* août 2009;84(8):1022-8.
126. Guillou P, Pelaccia T, Bacqué M-F, Lorenzo M. Does burnout affect clinical reasoning? An observational study among residents in general practice. *BMC Med Educ.* déc 2021;21(1):35.
127. 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.* 1 févr 2011;16(5):601-8.
128. 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.

129. Fernandez N, Foucault A, Dubé S, Robert D, Lafond C, Vincent A-M, et al. Learning-by-Concordance (Lbc): introducing undergraduate students to the complexity and uncertainty of clinical practice. *Can Med Educ J.* oct 2016;7(2):e104-13.

130. 3.1 - Formations par concordance (FpC) - Centre de pédagogie appliquée aux sciences de la santé (CPASS) - Faculté de Médecine - Université de Montréal [Internet]. Centre de pédagogie appliquée aux sciences de la santé (CPASS). [cité 27 janv 2021]. Disponible sur: <https://www.cpass.umontreal.ca/recherche/groupe-de-recherche-cpass/axes-de-recherches/concordance/formation-et-evaluation-par-concordance/fpc/>

131. Ericsson KA. Deliberate Practice and the Acquisition and Maintenance of Expert Performance in Medicine and Related Domains. *Acad Med.* oct 2004;79(10):S70.

132. Ericsson KA, Hoffman RR, Kozbelt A, Williams AM. *The Cambridge Handbook of Expertise and Expert Performance.* Cambridge University Press; 2018. 985 p.

133. Peyrony O, Hutin A, Truchot J, Borie R, Calvet D, Albaladejo A, et al. Impact of panelists' experience on script concordance test scores of medical students. *BMC Med Educ.* déc 2020;20(1):1-8.

134. Foucault A, Dubé S, Fernandez N, Gagnon R, Charlin B. Learning medical professionalism with the online concordance-of-judgment learning tool (CJLT): A pilot study. *Med Teach.* 2015;37(10):955-60.

135. Foucault A, Dubé S, Fernandez N, Gagnon R, Charlin B. The concordance of judgement learning tool. *Med Educ.* mai 2014;48(5):541-2.

136. Charlin B, Gagnon R, Sibert L, Vleuten CV der. Le test de concordance de script, un instrument d'évaluation du raisonnement clinique. *Pédagogie Médicale.* 1 août 2002;3(3):135-44.

137. Cooke S, Lemay J-F. Transforming Medical Assessment: Integrating Uncertainty Into the Evaluation of Clinical Reasoning in Medical Education. *Acad Med J Assoc Am Med Coll.* 31 janv 2017;

138. Bates J, Schrewe B, Ellaway RH, Teunissen PW, Watling C. Embracing standardisation and contextualisation in medical education. *Med Educ* [Internet]. [cité 1 déc 2018];0(0). Disponible sur: <https://onlinelibrary.wiley.com/doi/abs/10.1111/medu.13740>

139. Logan RL, Scott PJ. Uncertainty in clinical practice: implications for quality and costs of health care. *Lancet Lond Engl.* 2 mars 1996;347(9001):595-8.

140. van der Weijden T, van Velsen M, Dinant G-J, van Hasselt CM, Grol R. Unexplained Complaints in

General Practice: Prevalence, Patients' Expectations, and Professionals' Test-Ordering Behavior. *Med Decis Making*. 1 mai 2003;23(3):226-31.

141. Kassirer JP. Our stubborn quest for diagnostic certainty. A cause of excessive testing. *N Engl J Med*. 1 juin 1989;320(22):1489-91.

142. Lorenzo M. Tolérance à l'incertitude en médecine générale: une démarche de théorisation ancrée à partir de 14 entretiens [Thèse d'exercice]. [France]: Université de Strasbourg (2009-....). Faculté de médecine; 2013.

143. Bugter-Maessen AM, Winkens RA, Grol RP, Knottnerus JA, Kester AD, Beusmans GH, et al. Factors predicting differences among general practitioners in test ordering behaviour and in the response to feedback on test requests. *Fam Pract*. juin 1996;13(3):254-8.

144. van der Weijden T, van Bokhoven MA, Dinant G-J, van Hasselt CM, Grol RPTM. Understanding laboratory testing in diagnostic uncertainty: a qualitative study in general practice. *Br J Gen Pract*. déc 2002;52(485):974-80.

145. Allison JJ, Kiefe CI, Cook EF, Gerrity MS, Orav EJ, Centor R. The association of physician attitudes about uncertainty and risk taking with resource use in a Medicare HMO. *Med Decis Mak Int J Soc Med Decis Mak*. sept 1998;18(3):320-9.

146. Kristiansen IS, Hjortdahl P. The general practitioner and laboratory utilization: why does it vary? *Fam Pract*. mars 1992;9(1):22-7.

147. Spielberger CD, Gorsuch RL, Lushene RE. *Manual for the State-Trait Anxiety Inventory*. 1970 [cité 12 juin 2020]; Disponible sur: <http://ubir.buffalo.edu/xmlui/handle/10477/2895>

148. Gerrity MS, White KP, DeVellis RF, Dittus RS. Physicians' Reactions to Uncertainty: Refining the constructs and scales. *Motiv Emot*. 1 sept 1995;19(3):175-91.

149. Le Roux V. Validation culturelle et psychométrique du PRU, Physicians' Reactions to Uncertainty scales, questionnaire américain de mesure des réactions affectives des médecins face à l'incertitude. Université de Brest - Bretagne Occidentale; 2015.

150. Feddock CA. The lost art of clinical skills. *Am J Med*. avr 2007;120(4):374-8.

151. Chamberland M. Comment exploiter les résultats de la recherche pour améliorer nos activités d'enseignement et d'apprentissage du raisonnement clinique ? *Pédagogie Médicale*. 1 nov 2005;6(4):197-9.

152. Saint-Pierre L. La métacognition, qu'en est-il? *Rev Sci Léducation*. 1994;20(3):529-45.
153. Noel B, Cartier SC, Tardif J. De la métacognition à l'apprentissage autorégulé. *De Boeck Supérieur*; 2016. 227 p.
154. Flavell JH. Metacognitive aspects of problem solving. In: *The nature of intelligence*. Hillsdale, NJ: Lawrence Erlbaum; 1976. p. 231-5.
155. Noël B. *La métacognition*. Bruxelles: De Bœck-Wesmaël; 1991.
156. Dewey J. *How we think* [Internet]. Boston : D.C. Heath & Co.; 1910 [cité 7 févr 2021]. 242 p. Disponible sur: <http://archive.org/details/howwethink00deweiala>
157. Kolb DA, Fry R. Toward an applied theory of experiential learning. In: *Studies of group process*. C. Cooper. New York: Wiley; 1975. p. 33-57.
158. Kolb DA. *Experiential Learning: Experience as the Source of Learning and Development*. 1er édition. Englewood Cliffs, N.J: Financial Times/ Prentice Hall; 1983. 272 p.
159. Boud D, Keogh R, Walker D. *Reflection: Turning Experience into Learning*. Routledge; 2013. 171 p.
160. Schön DA. *Educating the Reflective Practitioner: Toward a New Design for Teaching and Learning in the Professions*. 1st edition. San Francisco, Calif.: Jossey-Bass; 1990. 384 p.
161. Schön DA. *The reflective practitioner: how professionals think in action*. Arena; 1995. 388 p.
162. Nguyen QD, Fernandez N, Karsenti T, Charlin B. What is reflection? A conceptual analysis of major definitions and a proposal of a five-component model. *Med Educ*. déc 2014;48(12):1176-89.
163. CNGE. Grille d'évaluation des RSCA.
164. Taha A. *RSCA: récit de situation complexe authentique : de l'idée à la réalisation*. 2018.
165. Wald HS, Borkan JM, Taylor JS, Anthony D, Reis SP. Fostering and evaluating reflective capacity in medical education: developing the REFLECT rubric for assessing reflective writing. *Acad Med J Assoc Am Med Coll*. janv 2012;87(1):41-50.
166. Sandars J. The use of reflection in medical education: AMEE Guide No. 44. *Med Teach*. août 2009;31(8):685-95.
167. Schön DA. *Le praticien réflexif: à la recherche du savoir caché dans l'agir professionnel*. Montréal, Canada: les Éd. Logiques, DL 1994; 1994. 418 p.
168. Mann K, Gordon J, MacLeod A. Reflection and reflective practice in health professions education: a systematic review. *Adv Health Sci Educ Theory Pract*. oct 2009;14(4):595-621.

169. Lundgren H, Poell RF. On Critical Reflection: A Review of Mezirow's Theory and Its Operationalization. *Hum Resour Dev Rev.* 1 mars 2016;15(1):3-28.
170. Miller-Kuhlmann R, O'Sullivan PS, Aronson L. Essential steps in developing best practices to assess reflective skill: A comparison of two rubrics. *Med Teach.* 2 janv 2016;38(1):75-81.
171. Aukes LC, Geertsma J, Cohen-Schotanus J, Zwierstra RP, Slaets JPJ. The development of a scale to measure personal reflection in medical practice and education. *Med Teach.* mars 2007;29(2-3):177-82.
172. Andersen NB, O'Neill L, Gormsen LK, Hvidberg L, Morcke AM. A validation study of the psychometric properties of the Groningen Reflection Ability Scale. *BMC Med Educ.* 10 oct 2014;14:214.
173. Boenink AD, Oderwald AK, De Jonge P, Van Tilburg W, Smal JA. Assessing student reflection in medical practice. The development of an observer-rated instrument: reliability, validity and initial experiences. *Med Educ.* avr 2004;38(4):368-77.
174. Koole S, Dornan T, Aper L, De Wever B, Scherpbier A, Valcke M, et al. Using video-cases to assess student reflection: development and validation of an instrument. *BMC Med Educ.* 20 avr 2012;12:22.
175. Michaud C. Le portfolio : un en-(je)u de formation et de développement professionnel. [Internet] [phdthesis]. Université Claude Bernard - Lyon I; 2010 [cité 28 nov 2020]. Disponible sur: <https://tel.archives-ouvertes.fr/tel-00558922>
176. Michaud C. Le portfolio, quel lien entre les écrits réflexifs et les compétences ? *Mes Éval En Éducation.* 2012;35(2):9.
177. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline: Validation of instruments or scales. *J Eval Clin Pract.* avr 2011;17(2):268-74.
178. Onaisi R, Severac F, Lorenzo M. Évaluer la réflexivité à travers les traces écrites d'apprentissage des étudiants en santé : traduction et adaptation interculturelle de la grille REFLECT. *Pédagogie Médicale.* 2021;22(1):15-26.
179. Cleary TJ. The use of reflection in medical education: Guide supplement 44.1--Viewpoint. *Med Teach.* 2011;33(6):500-1.
180. Chaffey LJ, de Leeuw EJJ, Finnigan GA. Facilitating students' reflective practice in a medical course: literature review. *Educ Health Abingdon Engl.* déc 2012;25(3):198-203.
181. Lafortune L. L'accompagnement et l'évaluation de la réflexivité en santé. *Applications en*

éducation et en formation. Presses de l'Université du Québec. 2015. 256 p.

182. Uygur J, Stuart E, De Paor M, Wallace E, Duffy S, O'Shea M, et al. A Best Evidence in Medical Education systematic review to determine the most effective teaching methods that develop reflection in medical students: BEME Guide No. 51. *Med Teach*. janv 2019;41(1):3-16.

183. LaDonna KA, Taylor T, Lingard L. Why Open-Ended Survey Questions Are Unlikely to Support Rigorous Qualitative Insights. *Acad Med J Assoc Am Med Coll*. mars 2018;93(3):347-9.

184. Pelaccia T, Tardif J. Comment [mieux] former et évaluer les étudiants en médecine et en sciences de la santé? Louvain-la-Neuve: De Boeck supérieur; 2016.

185. Jarvis P. Reflective practice and nursing. *Nurse Educ Today*. juin 1992;12(3):174-81.

186. Boyd LD. Development of reflective judgement in the pre-doctoral dental clinical curriculum. *Eur J Dent Educ Off J Assoc Dent Educ Eur*. août 2008;12(3):149-58.

187. Wald HS, Davis SW, Reis SP, Monroe AD, Borkan JM. Reflecting on reflections: enhancement of medical education curriculum with structured field notes and guided feedback. *Acad Med J Assoc Am Med Coll*. juill 2009;84(7):830-7.

188. de la Croix A, Veen M. The reflective zombie: Problematizing the conceptual framework of reflection in medical education. *Perspect Med Educ*. déc 2018;7(6):394-400.

189. Rousseau J-J. *Le contrat social, ou, Principes du droit politique*. P. Pourrat; 1839. 528 p.

190. Cruess SR, Cruess RL, Steinert Y. Linking the teaching of professionalism to the social contract: A call for cultural humility. *Med Teach*. 1 janv 2010;32(5):357-9.

191. Brody H, Doukas D. Professionalism: a framework to guide medical education. *Med Educ*. 2014;48(10):980-7.

192. Introduction aux commentaires du code de déontologie [Internet]. Conseil National de l'Ordre des Médecins. 2019 [cité 12 févr 2021]. Disponible sur: <https://www.conseil-national.medecin.fr/code-deontologie/introduction-commentaires-code-deontologie-art-1/introduction-commentaires-code>

193. Bentham J, Bentham G. *Essai sur la nomenclature et la classification des principales branches d'art-et-science: ouvrage extrait du Chrestomathia de Jérémie Bentham*. Bossange frères; 1823. 266 p.

194. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 13 janv 1996;312(7023):71-2.

195. Davidoff F, Haynes B, Sackett D, Smith R. Evidence based medicine. *BMJ*. 29 avr

1995;310(6987):1085-6.

196. Sereni D. Le professionnalisme médical pour le nouveau millénaire : une charte pour les praticiens. *Pédagogie Médicale*. 1 févr 2004;5(1):43-5.

197. Bélisle M. Perceptions de diplômés universitaires quant aux effets d'un programme professionnalisant et innovant sur leur professionnalisation en contexte de formation initiale [Internet]. Université de Sherbrooke; 2011 [cité 11 mars 2020]. Disponible sur: <https://savoirs.usherbrooke.ca/handle/11143/960>

198. Wittorski R. La professionnalisation. *Savoirs*. 2008;n° 17(2):9-36.

199. Boterf GL. *Professionnaliser: Construire des parcours personnalisés de professionnalisation*. Editions Eyrolles; 2016. 171 p.

200. Barrier JH, Brazeau-Lamontagne L, Colin R, Quinton A, Llorca G, Ehua FS. La formation au professionnalisme des futurs médecins. Recommandations du Conseil Pédagogique de la CIDMEF. *Pédagogie Médicale*. 1 mai 2004;5(2):75-81.

201. Chamberland M, Hivon R. Les compétences de l'enseignant clinicien et le modèle de rôle en formation clinique. *Pédagogie Médicale*. 1 mai 2005;6(2):98-111.

202. Bajwa NM, Nendaz MR, Galetto-Lacour A, Posfay-Barbe K, Yudkowsky R, Park YS. Can Professionalism Mini-Evaluation Exercise Scores Predict Medical Residency Performance? Validity Evidence Across Five Longitudinal Cohorts. *Acad Med* [Internet]. 27 août 2019 [cité 3 sept 2019]; Publish Ahead of Print. Disponible sur: [https://journals.lww.com/academicmedicine/Abstract/publishahead/Can\\_Professionalism\\_Mini\\_Evaluation\\_Exercise.97503.aspx](https://journals.lww.com/academicmedicine/Abstract/publishahead/Can_Professionalism_Mini_Evaluation_Exercise.97503.aspx)

203. Cruess RL, Cruess SR. Teaching professionalism: general principles. *Med Teach*. 1 janv 2006;28(3):205-8.

204. Symons AB, Swanson A, McGuigan D, Orrange S, Akl EA. A tool for self-assessment of communication skills and professionalism in residents. *BMC Med Educ*. déc 2009;9(1):1-7.

205. Phelan S, Obenshain SS, Galey WR. Evaluation of the noncognitive professional traits of medical students. *Acad Med J Assoc Am Med Coll*. oct 1993;68(10):799-803.

## ANNEXES

### ANNEXE N°1 : INDICATEURS DES NIVEAUX DE DÉVELOPPEMENT DES COMPÉTENCES EN MÉDECINE GÉNÉRALE

		L'interne...
<b>PREMIER RECOURS, URGENCES</b>	<b>Novice</b>	<ul style="list-style-type: none"> <li>➤ Prend conscience de l'amplitude du champ d'activités possible en exercice ambulatoire et s'interroge sur ses capacités à y faire face.</li> </ul>
	<b>Intermédiaire</b>	<ul style="list-style-type: none"> <li>➤ Recueille, accepte sans rejeter et analyse les demandes explicites les plus fréquentes, tente de les gérer en repérant la demande réelle en essayant de les hiérarchiser dans une vision centrée maladie plus que centrée patient et tenant compte des prévalences liées au contexte. Fait des tentatives pour repérer la demande réelle derrière la plainte alléguée, en essayant d'intégrer les antécédents et le contexte de vie du patient ; a encore du mal à élargir sa vision centrée maladie et a besoin de soutien sous la forme de supervision pour se centrer patient.</li> <li>➤ Décide sans avoir systématiquement obtenu un diagnostic de maladie et accepte d'en parler au patient.</li> <li>➤ Accepte l'idée que les demandes urgentes recouvrent aussi des urgences ressenties.</li> <li>➤ Étend peu à peu le champ de ses capacités interventionnelles et en fait bénéficier de façon pertinente les problèmes ou situations de patients qui nécessitent une intervention extérieure.</li> </ul>
	<b>Compétent</b>	<ul style="list-style-type: none"> <li>➤ Fait face aux plaintes les plus prévalentes de premier recours en mobilisant des ressources internes et externes permettant leurs résolutions</li> <li>➤ S'organise pour faire face aux plaintes les plus prévalentes de premier recours en participant aussi à la permanence de soins</li> <li>➤ Fait des diagnostics de situation</li> <li>➤ Évoque les stades précoces des maladies et en dehors des situations d'urgence se donne le temps (cf. suivi)</li> <li>➤ Évoque la possibilité de symptômes biomédicalement inexpliqués (SBI)</li> <li>➤ Élargit le champ de la consultation aux autres dimensions de la consultation et aux autres problèmes de santé en programmant éventuellement des actions de prévention en accord avec le patient</li> <li>➤ Gère les urgences ressenties par le patient</li> <li>➤ Collabore avec les autres intervenants et assume ses responsabilités</li> </ul>

		L'interne...
<b>RELATION, COMMUNICATION, APPROCHE CENTREE PATIENT</b>	<b>Novice</b>	<ul style="list-style-type: none"> <li>➤ Identifie les difficultés inhérentes à la mise en pratique des habiletés relationnelles et communicationnelles et repère que l'acquisition des compétences et des capacités dans le domaine de la relation et de la communication doivent faire l'objet d'une formation.</li> <li>➤ Utilise principalement une communication verbale.</li> <li>➤ Explique les décisions et espère obtenir l'adhésion du patient.</li> <li>➤ Communique avec l'entourage des patients à partir de ce qu'il pense être important pour le patient.</li> <li>➤ Communique avec différents intervenants en utilisant différents médias.</li> </ul>
	<b>Intermédiaire</b>	<ul style="list-style-type: none"> <li>➤ Identifie les données communicationnelles et relationnelles qui participent à la démarche décisionnelle.</li> <li>➤ Dans l'analyse d'une consultation peut utiliser certaines notions de psychologie médicale afin de mieux comprendre le patient et le sens de ses réactions.</li> <li>➤ Dans les situations courantes, construit une relation en s'appliquant à utiliser les habiletés d'une communication centrée patient.</li> <li>➤ Accepte l'idée que l'on ne peut tout aborder et tout régler dans le temps d'une seule consultation.</li> <li>➤ Repère et exprime ses difficultés relationnelles et communicationnelles.</li> </ul>
	<b>Compétent</b>	<ul style="list-style-type: none"> <li>➤ En dehors des situations très complexes, mène en autonomie un entretien centré patient et structure ce dernier</li> <li>➤ Dans les conditions habituelles, de complexité modérée, est capable de gérer les émotions, de rester empathique et respectueux</li> <li>➤ Dans les conditions habituelles, de complexité modérée, est capable de mettre en place une relation de soutien, à effet psychothérapeutique bénéfique que pour le patient</li> <li>➤ Lors de situations et/ou de relations qui posent problème (agressivité, séduction, sympathie, rejet, etc.) construit et tente de maintenir la relation tout en se questionnant sur la nature de celle-ci</li> <li>➤ Communique sur ses erreurs en tenant compte de l'avis du patient et en acceptant d'être remis en cause</li> <li>➤ Dans les conditions habituelles, réfléchit à sa capacité communicationnelle avec le patient et son entourage</li> <li>➤ Met en œuvre, avec les intervenants médicaux, médico-sociaux et l'entourage du patient, une relation opérationnelle dans l'intérêt du patient</li> </ul>

		L'interne...
<b>APPROCHE GLOBALE, PRISE EN COMPTE DE LA COMPLEXITE</b>	<b>Novice</b>	<ul style="list-style-type: none"> <li>➤ Entrevoit qu'il existe des données psycho sociales, culturelles, éthiques, juridiques et administratives dans la démarche décisionnelle et qu'il est nécessaire de les prendre en compte.</li> <li>➤ Accepte l'idée que s'occuper du patient ne se réduit pas à se centrer sur sa maladie mais que cette démarche n'est pas évidente pour lui.</li> </ul>
	<b>Intermédiaire</b>	<ul style="list-style-type: none"> <li>➤ A conscience qu'une situation clinique ne peut pas se réduire au diagnostic médical et qu'il est nécessaire d'intégrer d'autres aspects pour comprendre et gérer cette situation clinique. Tente de passer du diagnostic médical à un diagnostic qui intègre une partie du contexte sans pour autant qu'il s'agisse d'un diagnostic de situation.</li> <li>➤ A conscience qu'il existe des temporalités différentes entre le médecin et le patient dans toutes les situations, en particulier en cas de discordance (temps nécessaire à chaque patient).</li> <li>➤ Accepte l'idée qu'il existe plusieurs réponses acceptables en fonction des différentes analyses possibles. De ce fait il prend en compte une partie de la complexité en situation. Reconnaît la place de l'incertitude dans la démarche décisionnelle.</li> </ul>
	<b>Compétent</b>	<ul style="list-style-type: none"> <li>➤ Dans les situations habituelles, tient compte des données émanant de plusieurs champs et de plusieurs sources, tente de les intégrer dans une décision centrée patient</li> <li>➤ Prend le temps nécessaire et suffisant pour explorer une situation. Laisse le temps au patient de métaboliser</li> <li>➤ Gère simultanément plusieurs problèmes de nature différente en les hiérarchisant</li> <li>➤ Utilise le temps comme allié, comme une aide à la décision en adéquation avec la situation du patient</li> <li>➤ En fonction des situations, est en mesure de modifier sa posture</li> <li>➤ Dans les conditions habituelles, de complexité modérée, est capable de mettre en place une relation de soutien, à effet psychothérapeutique bénéfique pour le patient</li> <li>➤ Fait la différence entre incertitude personnelle et incertitude professionnelle</li> </ul>

	L'interne...
EDUCATION, PREVENTION, DEPISTAGE, SANTE INDIVIDUELLE ET COMMUNAUTAIRE	<p><b>Novice</b></p> <ul style="list-style-type: none"> <li>➤ Accepte la place et l'importance des différentes composantes de cette grande compétence dans l'activité du généraliste.</li> <li>➤ Définit ce que recouvrent les 3 niveaux de prévention de l'OMS, primaire, secondaire et tertiaire.</li> <li>➤ Possède des notions vagues de ce que recouvre l'éducation du patient avec ses 3 niveaux d'activité, du plus général au plus spécifique : l'éducation pour la santé du patient, l'éducation du patient à sa maladie et l'éducation thérapeutique du patient.</li> <li>➤ Se sent responsable de la gestion de la santé du patient.</li> <li>➤ Argumente ses propositions dans le but d'obtenir l'adhésion du patient, par une approche logique centrée sur son propre raisonnement et sans tenir compte des représentations du patient.</li> </ul>
	<p><b>Intermédiaire</b></p> <ul style="list-style-type: none"> <li>➤ Réalise des consultations dédiées à la prévention en les intégrant aux soins à partir de la demande du patient et de ses contraintes de médecin</li> <li>➤ Réalise les démarches et gestes de prévention dans les situations les plus simples</li> <li>➤ Repère et exprime ses difficultés à changer de posture de soignant</li> <li>➤ Cherche la collaboration et le soutien de l'entourage familial pour aider le patient</li> <li>➤ Travaille avec d'autres intervenants impliqués dans la prévention et l'éducation du patient</li> </ul>
	<p><b>Compétent</b></p> <ul style="list-style-type: none"> <li>➤ Intègre couramment dans son activité de soins et dans la durée des moments dédiés à la prévention individuelle, au dépistage organisé et à l'éducation du patient</li> <li>➤ Accompagne le patient dans une démarche d'éducation à sa santé (posture d'éducateur)</li> <li>➤ Clarifie les tensions entre enjeux individuels et collectifs de la prévention pour rechercher l'adhésion du patient</li> <li>➤ Collabore activement avec d'autres intervenants impliqués dans la prévention et l'éducation du patient</li> </ul>

	L'interne...
CONTINUITE, SUIVI, COORDINATION DES SOINS AUTOUR DU PATIENT	<p><b>Novice</b></p> <ul style="list-style-type: none"> <li>➤ Accepte l'idée qu'il va être amené à revoir les patients</li> <li>➤ Utilise le dossier médical</li> <li>➤ Fait volontiers appel à d'autres intervenants sur des critères décisionnels centrés sur le biomédical et le médecin</li> <li>➤ Transmet les informations nécessaires à la continuité des soins</li> <li>➤ Utilise et prend en compte les informations des autres intervenants</li> <li>➤ Met en place une relation médecin malade basée sur une posture expert « haute » et perçoit ses limites</li> <li>➤ Se rend disponible pour la permanence des soins</li> </ul>
	<p><b>Intermédiaire</b></p> <ul style="list-style-type: none"> <li>➤ Utilise le temps dans la démarche décisionnelle dans certaines situations</li> <li>➤ Prend en compte les problèmes et les plaintes afin d'organiser le suivi</li> <li>➤ Construit une relation dans le temps en essayant de faire participer le patient à la décision et à la démarche</li> <li>➤ Utilise et renseigne le dossier médical dans une optique de suivi</li> <li>➤ Fait le lien entre les différents moments ponctuels de recours</li> <li>➤ Met en œuvre une relation avec les intervenants, en particulier paramédicaux et médicosociaux, en adaptant les moyens de communication et d'information à la situation et à l'intervenant avec lequel il communique (cf. communication)</li> </ul>
	<p><b>Compétent</b></p> <ul style="list-style-type: none"> <li>➤ Conçoit que le patient a une histoire personnelle et une vie qui déterminent ses traits de caractère et qui influencent le type de suivi</li> <li>➤ Hiérarchise les plaintes et les problèmes et établit un suivi centré patient</li> <li>➤ Utilise le temps comme allié, comme une aide à la décision en adéquation avec la situation du patient</li> <li>➤ Choisit les intervenants en accord avec le patient selon des critères bio-psycho-sociaux</li> <li>➤ Analyse les avis des différents intervenants, les synthétise pour prendre une décision centrée patient</li> <li>➤ Utilise le dossier médical pour programmer un suivi dans une perspective de promotion de la santé au niveau individuel et collectif, de prévention et de dépistage</li> <li>➤ Collabore à la continuité et la coordination du maintien à domicile</li> <li>➤ Participe à l'organisation de l'accessibilité aux soins, y compris lors de ses absences</li> </ul>

		L'interne...
<b>PROFESSIONNALISME</b>	<b>Novice</b>	<ul style="list-style-type: none"> <li>➤ Entrevoit un projet professionnel</li> <li>➤ Accepte son rôle d'interne</li> <li>➤ Présente un engagement altruiste envers le patient</li> <li>➤ Participe aux formations théoriques et pratiques du DES</li> </ul>
	<b>Intermédiaire</b>	<ul style="list-style-type: none"> <li>➤ Manifeste un engagement pour la médecine générale</li> <li>➤ S'occupe du patient avec altruisme, honnêteté, dans le respect des règles déontologiques</li> <li>➤ Prend conscience du besoin d'acquisition permanente de nouvelles connaissances afin d'améliorer ses compétences</li> <li>➤ Définit des objectifs de formation en fonction de son projet professionnel</li> <li>➤ Organise son temps de travail</li> </ul>
	<b>Compétent</b>	<ul style="list-style-type: none"> <li>➤ Assume sa responsabilité envers le patient et la société</li> <li>➤ Collabore avec les autres soignants</li> <li>➤ Organise son outil et son temps de travail</li> <li>➤ Améliore ses compétences</li> <li>➤ Prend en charge le patient avec altruisme</li> <li>➤ S'implique dans le rayonnement de la discipline</li> </ul>

**Mathieu Lorenzo**

Département de médecine générale, et centre de formation et de recherche en pédagogie des sciences de la santé (CFRPS), faculté de médecine de Strasbourg, 4, rue Kirschleger, 67085 Strasbourg

mlorenzo@unistra.fr

*exercer* 2021;176:375-9.

Liens et conflits d'intérêts : les auteurs déclarent n'avoir aucun conflit d'intérêts en relation avec le contenu de cet article.

Les liens d'intérêts éventuels de chacun des auteurs sont disponibles sur le site : [www.transparence.sante.gouv.fr](http://www.transparence.sante.gouv.fr)

## Les problèmes à éléments clés

### Un nouvel outil d'évaluation introduit par la réforme du deuxième cycle

*The key features examinations. A new assessment tool introduced by the graduate medical school reform*

#### INTRODUCTION

La réforme du deuxième cycle des études médicales (R2C) est entrée en vigueur pour les étudiants accédant à la première année du deuxième cycle à la rentrée 2020<sup>1</sup>. Cette réforme tente d'inscrire davantage ce deuxième cycle dans une approche par compétences en visant notamment le développement du raisonnement clinique<sup>1</sup>.

En cohérence avec ce dernier objectif, les modalités d'évaluation des apprentissages évoluent en intégrant deux outils relativement méconnus en France : les problèmes à éléments clés et le test de concordance de script (TCS).

Les problèmes à éléments clés ont été inventés dans les années 1980 à partir d'une volonté d'améliorer l'évaluation de la prise de décision des futurs médecins lors de l'examen qualifiant national du Canada<sup>2,3</sup>. Ils semblent avoir atteint cet objectif avec une corrélation importante entre les scores aux problèmes à éléments clés et la qualité ultérieure de la pratique clinique<sup>4</sup>. De manière générale, les données de la littérature suggèrent que les problèmes à éléments clés sont un outil valide d'évaluation du raisonnement clinique<sup>4,5</sup>. Ils sont largement utilisés dans les facultés de médecine en Amérique du Nord, aux Pays-Bas ou en Australie depuis plus de vingt-cinq ans<sup>5</sup>. Ils sont souvent désignés par leurs acronymes anglais KFE (pour *key features examinations*) ou KFP (pour *key features problems*).

Les problèmes à éléments clés explorent une sous-partie du processus de raisonnement clinique : la prise de décision clinique. Ils se présentent habituellement sous la forme de vignettes cliniques suivies d'une ou plusieurs questions. Les questions portent sur les éléments essentiels nécessaires pour diagnostiquer ou résoudre une situation clinique. Ce sont les « éléments clés » d'une situation clinique. Les problèmes à éléments clés, combinés dans le cadre de la R2C aux TCS, à des dossiers cliniques, des examens cliniques objectifs structurés (ECOS) et de la supervision en stage pourraient permettre une meilleure évaluation du processus de raisonnement clinique.

La discipline Médecine générale est officiellement en charge de certains des items du deuxième cycle. Le premier volume de l'ouvrage *Référence en médecine générale pour le 2<sup>e</sup> cycle* rédigé sous la direction de Xavier Gocko propose 25 items abordés sous l'angle de la médecine générale<sup>6</sup>. Pour ces items en priorité, mais plus largement pour intégrer la vision de la médecine générale dans les évaluations du deuxième cycle, l'appropriation par les enseignants de la discipline des problèmes à éléments clés est nécessaire.

À la suite de l'article portant sur le TCS, l'objectif de cet article était de présenter le principe des problèmes à éléments clés et d'en détailler les principales modalités de conception.

#### PRINCIPES DES PROBLÈMES À ÉLÉMENTS CLÉS

La littérature en sciences de l'éducation montre que les experts n'utilisent que très peu d'éléments cliniques pour raisonner face à un patient<sup>2,7-9</sup>. Ces quelques éléments clés utilisés par les experts sont contextuels : ils dépendent de la situation. En cohérence avec ces données, les problèmes à éléments clés se focalisent sur la capacité des étudiants à distinguer les éléments clés d'une situation clinique.

Dans un problème à éléments clés classique, une vignette textuelle détaillée est présentée à l'étudiant<sup>5</sup>. Celle-ci contient à la fois des éléments pertinents et non pertinents pour la résolution du cas<sup>5</sup>. Fréquemment, une première question va recueillir les hypothèses diagnostiques de l'étudiant face à la situation clinique<sup>5</sup>. Ce recueil s'effectue sous la forme d'une zone de texte libre en limitant fortement le nombre d'items possibles (ex : trois hypothèses au maximum). Ce recueil est peut-être hiérarchisé : quelle est l'hypothèse n° 1 ? la n° 2 ? etc.

D'autres types de questions peuvent demander à l'étudiant d'indiquer quelles sont les éléments clés de la vignette qui sous-tendent son hypothèse diagnostique principale<sup>5</sup>. L'étudiant choisit ainsi dans une longue liste des éléments présentés dans la vignette. Cette liste est souvent bien plus longue que dans une question à choix classique pour réduire les chances de réponses au hasard et les biais d'indicage<sup>5</sup>.



Mme Haltenbach est une comptable retraitée de 68 ans. Elle se présente ce jour en consultation urgente à votre cabinet. Elle est très fatiguée depuis la veille et trouve qu'elle respire mal.

Vous la connaissez depuis trois ans. Elle a comme antécédents une fracture du poignet droit survenue il y a quatre ans, une appendicite opérée à l'âge de 28 ans, une hypertension artérielle depuis douze ans, une pneumopathie bactérienne du lobe supérieur droit il y a dix ans, une tendinopathie de la coiffe des rotateurs à droite à l'âge de 43 ans, une ablation d'un carcinome basocellulaire de l'oreille gauche il y a deux ans.

Son traitement habituel comprend de l'amlodipine 5 mg 1/j, du cholécalférol 100 000 UI (une ampoule tous les trois mois), du paracétamol 1 g en cas de douleurs.

À l'examen, sa température est à 38,5 °C, sa pression artérielle assise est à 162/88 mmHg au bras droit et à 159/84 mmHg au bras gauche. Sa saturation en oxygène est à 93 % en air ambiant, sa fréquence cardiaque à 96 batt/min. L'auscultation cardiaque retrouve un souffle systolique aortique à 2/6, l'auscultation pulmonaire retrouve des crépitations des deux bases pulmonaires. Il n'y a pas de sibilants. Son abdomen est pléthorique, la palpation ne retrouve pas de défense ni de contracture. Il n'y a pas de masse abdominale.

Question 1 : Quelles sont vos trois hypothèses diagnostiques principales à ce stade ? Hiérarchisez.

1. \_\_\_\_\_
2. \_\_\_\_\_
3. \_\_\_\_\_

Question 2 : Quels sont les quatre éléments de l'énoncé les plus pertinents pour justifier votre hypothèse diagnostique principale parmi les propositions ci-dessous ?

Le sexe féminin	L'âge de 68 ans	La consultation en urgence	La fréquence cardiaque à 96 batt/min	La pression artérielle à 162/88 mmHg au bras droit
La pression artérielle à 159/84 mmHg au bras gauche	Les crépitations des deux bases pulmonaires	L'antécédent de fracture du poignet	L'antécédent d'appendicite	L'antécédent d'hypertension artérielle
L'antécédent de pneumopathie bactérienne	Le souffle systolique	L'antécédent de carcinome basocellulaire	Le traitement par amlodipine	Le traitement par cholécalférol
Le traitement par paracétamol	La température à 38,5 °C	L'absence de masse abdominale	La saturation à 93 %	Etc.

Question 3 : Quelles sont les éléments que vous voudriez particulièrement recueillir à l'interrogatoire en regard de votre hypothèse diagnostique principale ? Sélectionnez-en au maximum cinq.

Présence de frissons	Consommation quotidienne d'alcool	Expectorations purulentes	Hémoptysie	Palpitations
Suivi dermatologique de son basocellulaire	Présence de diarrhées	Perte d'odorat	Perte de goût	Insomnie
Présence d'acouphènes	Acuité visuelle habituelle	Douleurs dans les membres inférieurs	Heure précise de début des symptômes	Tabagisme
Douleur thoracique	Douleur pharyngée	Présence d'un rash cutané	Dysphagie associée	Etc.

Tableau 1 - Exemple de problème à éléments clés

Dans la même optique, les étudiants peuvent devoir indiquer dans une liste qui peut contenir des centaines de possibilités les quelques questions les plus pertinentes pour confirmer leur hypothèse diagnostique principale<sup>5</sup>. Il en est de même pour les manœuvres cliniques les plus pertinentes à proposer, les examens complémentaires pertinents, le traitement médicamenteux, etc. L'ensemble « vignettes + questions » est nommé « cas ». Un exemple de cas de problème à éléments clés selon ces principes est présenté dans le **tableau 1**.

Ce focus sur les éléments clés permettant de résoudre un problème est la différence principale entre les problèmes à éléments clés et les dossiers cliniques classiques. Ces derniers vont souvent évaluer des connaissances déclaratives, et non les éléments clés utilisés pour raisonner en situation de soins<sup>5</sup>. À titre d'exemple, une question du type « Quelles sont les molécules ayant l'indication pour le traitement de la thrombose veineuse profonde ? » explore un savoir théorique superficiel (demande de restituer des connaissances déclaratives).

À ce titre, des exemples de problèmes à éléments clés proposés par Montjotin-Gayton *et al.* dans le numéro 162 de cette revue sont questionnables<sup>10</sup>. Dans leur problème à éléments clés de l'item 47 « Puberté normale et pathologique », la première question est « À partir de quel âge parle-t-on de retard pubertaire chez le garçon ? ». Cette question est un exemple typique d'évaluation de connaissance déclarative. Il n'y a besoin d'aucun contexte clinique pour y répondre car la réponse n'en dépend pas. À l'inverse, pour l'item 44 « Suivi d'un nourrisson et d'un enfant normal », la question « Quel diagnostic le plus probable concernant l'œil gauche de Y suspectez-vous ? » suivant une vignette traitant d'une anomalie de l'œil chez un nourrisson de 12 mois semble un bon exemple de question de problème à éléments clés.

Une limite importante des problèmes à éléments clés est qu'ils ne

peuvent pas évaluer seuls l'intégralité du processus de raisonnement clinique<sup>11,12</sup>. Les problèmes à éléments clés n'évaluent notamment pas correctement la justification diagnostique ou la représentation du problème<sup>12</sup>. Seule une combinaison de plusieurs outils et méthodes d'évaluation du raisonnement clinique permet de correctement l'évaluer<sup>12</sup>.

### MODALITÉS DE CONCEPTION DES PROBLÈMES À ÉLÉMENTS CLÉS

La rédaction des problèmes à éléments à clés nécessite plusieurs heures et plusieurs rédacteurs<sup>12</sup>. En effet, il est conseillé de réunir au minimum trois experts du domaine pour déterminer par consensus quels sont les éléments clés de la situation<sup>12</sup>.

La première étape est de déterminer les objectifs de l'évaluation. S'agit-il d'une évaluation formative (visant à guider les apprentissages ultérieurs) ou d'une évaluation sommative (visant à vérifier que les apprentissages attendus ont eu lieu) ? Dans ce dernier cas, partir des objectifs d'apprentissage du référentiel de formation puis utiliser une table de spécification semble utile pour s'assurer d'une bonne validité de contenu des problèmes à éléments clés<sup>11</sup>. Une table de spécification est un tableau croisant les objectifs d'apprentissage avec le contenu d'une évaluation<sup>13</sup>. Elle sert à s'assurer que l'évaluation couvre bien tous les éléments importants du programme (c'est la validité de contenu). Un exemple de table de spécification est donné dans le **tableau 2**. Le lien avec le rang des connaissances de la R2C (A, B et C)

n'est pas établi à ce jour. Dans la R2C, les connaissances de rang A sont à acquérir par tous les étudiants de 2<sup>e</sup> cycle. Les connaissances de rang B sont des connaissances d'une spécialité exigibles au premier jour du DES. Les connaissances de rang C sont des éléments de spécialité uniquement enseignées en troisième cycle.

Les problèmes à éléments clés sont une des modalités de l'évaluation des connaissances tout en visant à documenter le développement d'une compétence (le raisonnement clinique). Il faudra choisir une des situations de départ (SDD) du nouveau référentiel et faire le lien avec les rangs de connaissances à l'aide des fiches LISA (livret de suivi des apprentissages). Les fiches SDD et LISA sont en cours de finalisation au niveau national à ce jour. Dans le cas d'une évaluation formative, ces considérations de validité de contenu sont moins importantes, et l'accent est mis sur la rétro-information offerte à l'étudiant à la suite de la réalisation des problèmes à éléments clés<sup>13</sup>. C'est cette rétro-information qui permet à l'étudiant de cerner les apprentissages à réaliser pour la suite. Il semble donc utile de prendre un temps dédié visant à expliciter le processus de raisonnement clinique de l'expert sous-tendant la résolution des problèmes à éléments clés pour modéliser celui de l'étudiant<sup>14</sup>.

Pour écrire les cas, le plus pertinent est de partir d'une situation réelle vécue par l'un des membres du comité de rédaction. Dans cette étape, il est utile de contextualiser la vignette<sup>15</sup>. Le fait de donner un nom (fictif) au patient, d'ajouter une profession, des détails personnels, tend à majorer l'engagement des étudiants dans une

utilisation formative des problèmes à éléments clés<sup>15</sup>. Il est dans tous les cas indispensable de présenter les valeurs brutes (sans interprétation) dans la vignette et dans les questions<sup>4,5,15</sup>. Il est conseillé d'éviter d'écrire « il a de la fièvre » mais plutôt « sa température est à 38,5 °C ». Comme dans la vraie vie, cette étape d'interprétation des données permet de mieux discriminer les étudiants en difficulté de raisonnement clinique<sup>14,16</sup>. La vignette présentée en exemple dans le **tableau 1** reprend ces principes.

Selon les objectifs de l'évaluation, le comité de rédaction propose de s'intéresser à la génération d'hypothèses diagnostiques, à l'investigation clinique, à la thérapeutique.

Vient ensuite le choix de la modalité de réponse : question ouverte à réponse courte ? Question à choix multiple ? Choisissez le nombre de réponses attendues pour chaque question et le nombre maximal de propositions possibles (ce nombre peut être légèrement supérieur au nombre de réponses attendues). Dans l'exemple présenté dans le **tableau 1**, il est demandé à l'étudiant d'indiquer trois hypothèses diagnostiques. Trois hypothèses étaient attendues et valaient des points. Une autre possibilité aurait été d'indiquer à l'étudiant de renseigner « trois hypothèses au maximum » et de n'attribuer des points qu'à deux hypothèses. Pour les questions à choix multiple, déterminez le nombre de modalités que vous allez proposer au choix. Il est habituel que ce nombre soit au moins égal au double, voire au triple, du nombre maximal de réponses possibles pour limiter les réponses au hasard<sup>5,12</sup>.

	Reconnaître les symptômes et signes cliniques [...]	Prescrire les examens adaptés [...]	Prendre en charge les motifs de consultation les plus fréquents [...]	Donner les informations adaptées [...]	Etc.
Nombre de cas	3	2	4	1	...
Nombre de questions	4	4	5	2	...

**Tableau 2** - Exemple d'une table de spécification de problèmes à éléments clés pour l'item 22 « Grossesse normale »



La phase cruciale est celle de la détermination par consensus des éléments clés. Quelles sont les hypothèses diagnostiques les plus probables ? Quels sont les éléments de l'énoncé les plus contributifs ? etc. C'est cette phase qui prend le plus de temps.

Dans les questions à choix multiple, il faut choisir les distracteurs (les options de réponses proposées en sus des éléments clés attendus). Les meilleurs distracteurs sont les plus plausibles. Une méthode efficace semble être d'utiliser les erreurs les plus fréquentes des étudiants. Pour cela, une question sous la forme de question ouverte courte peut être posée préalablement à un groupe d'étudiants. Utilisez ensuite les réponses erronées pour créer vos distracteurs et transformer cette question en question à choix pour le groupe d'étudiants suivant.

Pour les questions ouvertes courtes, il est conseillé de déterminer par consensus une grille de correction comprenant les éléments clés attendus, leurs synonymes et équivalences acceptés. La fidélité semble meilleure en recourant à une correction automatisée par ordinateur pour ces questions<sup>4</sup>.

Le nombre de cas à créer dépend des objectifs de l'évaluation. Dans une évaluation sommative ou certificative à fort enjeu, il faut une quarantaine de cas pour atteindre une fidélité satisfaisante<sup>5,12</sup>.

Un tel examen de 40 cas aura une durée d'environ quatre heures<sup>2</sup>. Il semble plus pertinent dans un même temps de limiter à deux ou trois le nombre de questions par cas pour poser plus de cas<sup>17</sup>. La performance en termes de raisonnement clinique étant contextuelle (être expert dans une situation ne préjuge pas d'être expert dans une autre), son évaluation est davantage valide avec une large variété de cas traitant de multiples situations cliniques différentes<sup>17,18</sup>. Dans le cas d'un examen à fort enjeu comme le sera l'examen national de fin de deuxième cycle, ces recommandations de nombre de cas et de nombre de questions par cas devraient logiquement s'appliquer.

Le mode de notation présente également des spécificités. Les recherches ont montré qu'il était préférable d'attribuer le même nombre de points à chaque cas au sein d'un ensemble, à chaque question au sein d'un cas et à chaque élément de réponse au sein d'une question<sup>4,5</sup>. Le calcul du score est une succession de calculs de moyennes (**encadré**).

1. L'étudiant a indiqué 2/3 des hypothèses diagnostiques attendus à la question n° 1. Il a correctement renseigné 4/4 éléments à la question n° 2 et 3/5 à la question n° 3.  
Son score à ce cas est  $(2/3 + 4/4 + 3/5)/3 = 75\%$

2. Sur un ensemble de quatre cas, ses scores sont 75 %, 86 %, 52 % et 45 %.  
Son score global est  $(0,75 + 0,86 + 0,52 + 0,45) / 4 = 64,5\%$ .

**Encadré** - Exemple de calcul de scores dans un problème à éléments clés

Des modes de calcul plus complexes sont décrits comme des attributions de points négatifs, ou des pénalités en cas de nombre trop important de réponses sont également décrites sans que leurs valeurs ajoutées soient évidentes<sup>4,5</sup>.

## CONCLUSION

L'introduction des problèmes à éléments clés dans la R2C est une opportunité de mieux évaluer le raisonnement clinique des étudiants en médecine. Cet article pourrait favoriser l'appropriation des problèmes à éléments clés par les enseignants universitaires de médecine générale. Ainsi, une nouvelle dynamique pourrait s'enclencher afin d'alimenter les bases de questions nationales avec des problèmes à éléments clés en lien avec les soins premiers. ♦

### Résumé

La réforme du deuxième cycle des études médicales en France tente d'inscrire davantage celui-ci dans une approche par compétences en visant notamment le développement du raisonnement clinique. En cohérence avec ce dernier objectif, un nouvel outil d'évaluation a été introduit : les problèmes à éléments clés. L'objectif de cet article est de présenter le principe des problèmes à éléments clés et d'en détailler les principales modalités de conception.

Les problèmes à éléments clés explorent une sous-partie du processus de raisonnement clinique : la prise de décision clinique. Ils se présentent habituellement sous la forme de vignettes cliniques suivies d'une ou plusieurs questions. Les questions portent sur les éléments essentiels nécessaires pour diagnostiquer ou résoudre une situation clinique. Ce sont les « éléments clés » d'une situation clinique. Ce travail propose une méthode pour créer des problèmes à éléments clés à partir d'une situation vécue. Il est conseillé de se réunir à trois experts du domaine pour déterminer par consensus les éléments clés de la situation. L'exploration de la phase précoce du raisonnement clinique au travers de la génération initiale d'hypothèses diagnostiques est pertinente dans un problème à éléments clés.

→ **Mots-clés MESH** : éducation ; enseignement spécialisé en médecine ; enseignement ; pédagogie.

### Summary

The reform of the graduate medical school in France tries to align more with a competency-based approach by focusing in particular on the development of clinical reasoning. In line with this last objective, a new assessment tool is introduced: the key features problems. The objective of this article is to present the principle of key features problems and to detail how to design it.

Key features problems explore a sub-part of the clinical reasoning process: decision making. They usually take the form of clinical vignettes followed by one or more questions. The questions focus on the essential elements needed to diagnose or resolve a clinical situation. These are the «key features» of a clinical situation. Determining by consensus of three or more people the key features of a situation from the trainers' clinical experience is recommended. Exploring the early phase of clinical reasoning and especially the generation of diagnostic hypotheses is often a good option.

→ **Keywords MESH**: education; education; medical; graduate; teaching; pedagogy.

## Références

1. République française. Arrêté du 2 septembre 2020 portant modification de diverses dispositions relatives au régime des études en vue du premier et du deuxième cycle des études médicales et à l'organisation des épreuves classantes nationales.
2. Page G, Bordage G. The medical council of Canada's key features project: a more valid written examination of clinical decision-making skills. *Acad Med J* 1995;70:104-10.
3. Hatala R, Norman GR. Adapting the key features examination for a clinical clerkship. *Med Educ* 2002;36:160-5.
4. Bordage G, Page G. The key-features approach to assess clinical decisions: validity evidence to date. *Adv Health Sci Educ* 2018;1-32.
5. Hrynchak P, Takahashi SG, Nayer M. Key-feature questions for assessment of clinical reasoning: a literature review. *Med Educ* 2014;48:870-83.
6. Gocko X. Références en médecine générale pour le deuxième cycle - 25 Items. Paris : CNGE, 2019. Disponible sur : <https://www.exercer.fr/librairie/produits/produit/70> [consulté le 1 septembre 2021].
7. Pelaccia T, Tardif J, Tribby E, Charlin B. An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory. *Med Educ Online* 2011;16.
8. Pelaccia T, Tardif J, Tribby E, et al. How and when do expert emergency physicians generate and evaluate diagnostic hypotheses? A qualitative study using head-mounted video cued-recall interviews. *Ann Emerg Med* 2014.
9. Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ* 2005;39:98-106.
10. Montjotin-Gayton L, Rodrigues Sousa Magnin M, Plotton C, Gocko X. Nouvelle docimologie pour le deuxième cycle des études médicales. *exercer* 2020;185-6.
11. Bertrand C, Dory V, Pelaccia T, et al. Choisir un outil d'évaluation. In: Comment (mieux) former et évaluer les étudiants en médecine et en sciences de la santé ? Bruxelles : De Boeck Supérieur, 2016:33-44.
12. Daniel M, Rencic J, Durning SJ, et al. Clinical reasoning assessment methods: a scoping review and practical guidance. *Acad Med* 2019;94:902-12.
13. Bertrand C, Dory V, Pelaccia T, Charlin B, Hodges B. Les principes généraux de l'évaluation. In: Comment (mieux) former et évaluer les étudiants en médecine et en sciences de la santé ? Bruxelles : De Boeck Supérieur, 2016:343-55.
14. Faucher C, Pelaccia T, Nendaz M, Audétat MC, Charlin B. Un professionnel de santé qui résout efficacement les problèmes : le raisonnement clinique. In: Comment (mieux) former et évaluer les étudiants en médecine et en sciences de la santé ? Bruxelles : De Boeck Supérieur, 2016:33-44.
15. Cohen DA, Newman LR, Fishman LN. Twelve tips on writing a discussion case that facilitates teaching and engages learners. *Med Teach* 2017;39:147-52.
16. Audétat MC, Laurin S. Supervision of clinical reasoning: methods and a tool to support and promote clinical reasoning. *Can Fam Physician* 2010;56:e127-129, 294-6.
17. Norman G, Bordage G, Page G, Keane D. How specific is case specificity? *Med Educ* 2006;40:618-23.
18. Croskerry P. Context is everything or how could I have been that stupid? *Healthc Q Tor Ont* 2009;12:e171-176.



Mathieu Lorenzo<sup>1,2</sup>

1. Département de médecine générale, faculté de médecine de Strasbourg, 4, rue Kirschleger, 67085 Strasbourg

2. Centre de formation et de recherche en pédagogie des sciences de la santé (CFRPS), faculté de médecine de Strasbourg, 4, rue Kirschleger, 67085 Strasbourg

mlorenzo@unistra.fr

exercer 2021;172:185-8.

L'annexe est consultable en ligne à l'adresse : [www.exercer.fr](http://www.exercer.fr)

Liens et conflits d'intérêts : l'auteur déclare n'avoir aucun conflit d'intérêts en relation avec le contenu de cet article. Les liens d'intérêts éventuels de l'auteur sont disponibles sur le site : [www.transparence.sante.gouv.fr](http://www.transparence.sante.gouv.fr)

## Qu'est-ce que le test de concordance de script ?

### Un outil d'évaluation du raisonnement clinique introduit par la réforme du deuxième cycle en France

*What is the script concordance test? A tool for assessment of clinical reasoning introduced by the graduate medical school reform in France*

#### LE TCS POUR ÉVALUER LE RAISONNEMENT CLINIQUE

Le test de concordance de script est un outil né il y a vingtaine d'années au Canada<sup>2</sup>. Il est fondé sur un modèle d'organisation des connaissances des professionnels de santé : les scripts cliniques<sup>3</sup>. Dans la théorie des scripts cliniques, l'évocation d'une hypothèse diagnostique par le professionnel de santé active un réseau de connaissances. Le professionnel de santé va ensuite interpréter l'ensemble des données cliniques du patient en regard de ces scripts : les données « collent-elles » avec le script activé ? Si oui, l'hypothèse en sort renforcée. Dans le cas contraire, il est probablement nécessaire de considérer une autre hypothèse.

Voici un exemple issu de la pratique de la médecine générale : devant une patiente de 22 ans présentant une douleur abdominale prédominant en fosse iliaque droite, le médecin évoque notamment l'hypothèse de l'appendicite. L'évocation de cette hypothèse active le script « appendicite ». Celui-ci contient les éléments attendus à l'interrogatoire devant une appendicite, les signes cliniques à rechercher à l'examen, les résultats attendus des examens complémentaires, la prise en charge à proposer, etc.

Le médecin demande à la patiente s'il elle a eu de la fièvre. La patiente répond qu'elle a eu 38,7 °C cette nuit avec des frissons. Cette donnée « colle » avec le script « appendicite » : il est fréquent qu'une fièvre accompagne une appendicite aiguë.

À l'examen clinique, le médecin s'attarde particulièrement à palper la fosse iliaque droite. L'absence de défense ou de contracture qu'il constate « ne colle pas » avec son script « appendicite ». Il s'attendait à trouver de tels éléments.

Le principe du TCS est celui-ci : dans une vignette clinique, il est demandé à l'étudiant de juger de l'impact d'une donnée sur le script d'une hypothèse ou d'une option (une « option d'investigation » ou une « option thérapeutique »). L'étudiant répond *via* une échelle de Likert à cinq modalités : ne change rien, impact positif, impact fortement positif, impact négatif, impact fortement négatif. Le tableau présente un exemple de questions d'un TCS.

Cet exemple permet d'illustrer plusieurs des difficultés fréquentes rencontrées par les étudiants avec les TCS. En premier lieu, les hypothèses ou options (« si vous pensiez à... ») sont imposées par le format. Il n'est donc pas demandé à l'étudiant de discuter du caractère pertinent ou non de celles-ci au regard de la vignette clinique. Si l'hypothèse ou option ne

#### INTRODUCTION

La réforme du deuxième cycle des études médicales (R2C) est entrée en vigueur pour les étudiants accédant à la première année du deuxième cycle à la rentrée 2020<sup>1</sup>. Cette réforme tente d'inscrire davantage ce deuxième cycle dans une approche par compétences, en visant notamment le développement du raisonnement clinique<sup>1</sup>.

En cohérence avec ce dernier objectif, les modalités d'évaluation des apprentissages évoluent en intégrant deux outils relativement méconnus en France : les problèmes à éléments clés (*key feature problems* ou *key feature examinations*) et le test de concordance de script (TCS).

L'objectif de cet article est de présenter le principe du TCS et d'en détailler les principales modalités de conception afin de permettre aux généralistes enseignants intervenant dans le deuxième cycle des études médicales de concevoir des questions.



semble pas pertinente pour l'étudiant dans cette situation clinique, il lui sera très difficile de répondre à la question.

Deuxièmement, les questions sont indépendantes. Ainsi, vous ne connaissez pas la température de la patiente dans la question n° 2. Ceci soulève la problématique de la charge cognitive de l'étudiant confronté à une succession de questions indépendantes sur une même vignette<sup>4</sup>. Il peut être difficile pour les étudiants « d'oublier » les informations précédentes.

Enfin, la nuance entre les modalités « impact fortement négatif » et « impact négatif » ou entre « impact fortement positif » et « impact positif » relève d'une certaine finesse dans le raisonnement clinique. Même chez des étudiants en fin de cursus, les déterminants de ce choix sont inconstants<sup>5</sup>. Certains auteurs proposent alors de n'utiliser que des TCS à trois modalités, notamment en début de formation (impact négatif, ne change rien, impact positif)<sup>6,7</sup>.

La plus grande originalité des TCS vient de leur mode de notation. Les points sont attribués selon la proximité des réponses des étudiants à celles d'un panel de référence. Pour la question n° 1 de notre exemple, un panel de dix médecins généralistes a répondu. Si huit membres ont répondu « impact positif » (+1), cette réponse vaudra pour les étudiants  $8/8 = 1$  point. Si les deux membres

restant du panel ont répondu « impact fortement positif » (+2), cette réponse vaudra  $2/8 = 0,25$  point. Les autres réponses vaudront  $0/8 = 0$  point. De nombreuses autres méthodes de notation ont été proposées dans la littérature mais celle-ci fait consensus actuellement<sup>6-9</sup>.

Ce mode de notation a un effet extrêmement intéressant : c'est la proximité entre le raisonnement clinique de l'étudiant et celui d'un panel de référence qui détermine son score. Un étudiant qui obtient un bon score aux TCS est considéré comme ayant un raisonnement clinique proche d'un groupe expert. C'est une manière normative de déterminer ce qui constitue la réussite en matière de raisonnement clinique. Ce qui est fait par la majorité du groupe de référence est considéré comme la « bonne réponse ». C'est aussi un moyen puissant d'illustrer l'incertitude dans la pratique de la médecine : les membres d'un groupe expert ne sont pas toujours d'accord entre eux.

Le seuil de réussite d'un TCS est habituellement fixé à  $60/100^8$ . Ce seuil a été déterminé à partir de données statistiques et une valeur inférieure correspond à une performance considérée comme « trop éloignée » de celle du groupe de référence pour être acceptable<sup>8</sup>. Il s'agit là aussi d'une différence notable avec le classique seuil de réussite de  $10/20$  encore domi-

nant dans les modalités habituelles d'évaluation des apprentissages en médecine.

De ces considérations docimologiques découlent des problèmes d'acceptabilité par les étudiants du TCS. Il leur est souvent difficile d'admettre en début de formation qu'il existe parfois plusieurs réponses « valables » pour un même problème en médecine. Le mode de sélection des membres du panel de référence est également une source d'inquiétude : un TCS n'est valable que si les membres sont réellement experts – s'ils possèdent les « bons » scripts cliniques. Un problème de légitimité fréquemment évoqué par les étudiants concerne le recrutement d'universitaires en médecine, qui n'ont qu'une activité clinique marginale peu représentative. Il semble important pour les enseignants de bien communiquer sur ces considérations et les moyens utilisés pour lutter contre ces menaces afin de favoriser l'acceptabilité du TCS.

Le TCS explore uniquement une sous-partie du raisonnement clinique. Selon une vision hypothético-déductive simple du raisonnement clinique (figure), le test de concordance de script explore la phase d'analyse d'adéquation des données cliniques avec des hypothèses diagnostiques. Selon la théorie du double processus, le raisonnement clinique utilise des processus intuitifs (appelés « sys-

Mme Hillinger est une étudiante de 22 ans dans le domaine du textile. Elle se présente à votre cabinet pour des douleurs abdominales prédominant en fosse iliaque droite apparues cette nuit. Elle n'a pas d'antécédents particuliers et ne prend pas de contraception.

N°	Si vous pensiez à...	... et qu'alors vous trouvez...	... l'impact sur votre hypothèse ou option est...				
Q1	une appendicite	une température à 38,7 °C	-2	-1	0	+1	+2
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q2	une grossesse extra-utérine	que la patiente ne se souvient plus de la date de ses dernières règles	-2	-1	0	+1	+2
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Q3	demander une échographie abdominale	que la patiente est enceinte	-2	-1	0	+1	+2
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Tableau 1 - Exemple de questions d'un TCS et des réponses du panel

-2 = fortement négatif, -1 = négatif, 0 = ne change rien, +1 = positif, +2 = fortement positif.

Exemple de calcul du barème : pour la question n° 1, huit membres ont répondu « +1, impact positif » et deux membres ont répondu « +2, impact fortement positif ».

La réponse « +1 » vaudra pour les étudiants  $8/8 = 1$  point. La réponse « +2 » vaudra  $2/8 = 0,25$  point. Les autres réponses vaudront  $0/8$  point.

tème 1 ») et des processus analytiques (« système 2 »)<sup>10</sup>. Le TCS explore la partie analytique du raisonnement clinique : la phase de raisonnement logique, consciente, qui contrôle les processus intuitifs. Au travers du TCS, c'est l'efficacité de l'organisation des connaissances cliniques de l'étudiant qui est évaluée (ses scripts cliniques).

Il est donc nécessaire de garder à l'esprit que le TCS n'est pas l'outil miracle en évaluation du raisonnement clinique. Son utilisation doit être combinée à d'autres méthodes et outils pour correctement évaluer le raisonnement clinique des étudiants<sup>11</sup>. Des méthodes comme les examens cliniques objectifs structurés ou les protocoles demandant de réfléchir à voix haute permettent par exemple de documenter d'autres aspects du raisonnement clinique, comme la génération d'hypothèses<sup>11</sup>.

## COMMENT CONCEVOIR UN TCS ?

Comme décrit dans la première partie de cet article, le TCS reflète les microjugements analytiques que tout médecin généraliste effectue constamment face à un patient. Il est donc relativement simple de concevoir des questions de TCS avec la méthode suivante : rappelez-vous une situation que vous avez récemment vécue. Souvenez-vous des hypothèses diagnostiques précoces que vous avez évoquées. Des manœuvres que vous avez considérées pour l'examen clinique. De l'impact du résultat de celles-ci sur vos hypothèses. Avec cette gymnastique d'explicitation de son propre raisonnement clinique, tout professionnel de santé peut rapidement écrire des questions de TCS. Par exemple : « *J'ai reçu hier Mme Hilm pour une toux évoluant depuis une semaine. Alors que j'évoquais une bronchite, j'ai réalisé une auscultation pulmonaire. Quel fut l'impact sur mon hypothèse principale d'entendre des ronchi ? Quel aurait été l'impact d'entendre un foyer de crépitations en base droite ? etc.* ».

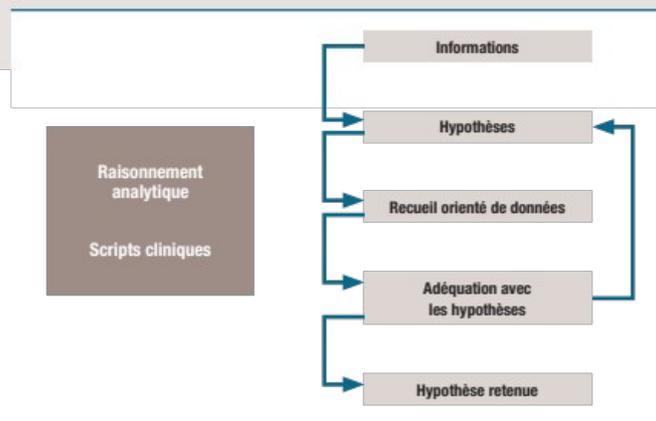


Figure - Le modèle hypothético-déductif du raisonnement clinique

Un conseil pratique est d'éviter de réfléchir à ce que le « panel répondrait » lors de la création des questions. Cette difficulté habituelle chez l'enseignant conduit à écarter un grand nombre de questions jugées comme « trop simples ». L'expérience montre que le panel est souvent plus divisé qu'on ne le pense. Écrivez donc rapidement tous ces microjugements cliniques que vous réalisez dans votre pratique clinique. Vous déciderez seulement dans un deuxième temps de garder ou non ces questions selon les réponses du panel.

Pour faciliter cette phase de rédaction, un masque de saisie de questions de TCS est téléchargeable en **annexe** de cet article.

Il est communément conseillé de créer au maximum trois questions par vignette clinique (scénario)<sup>8</sup>. Ceci pour des raisons de charge cognitive pour les étudiants (et le panel) du fait de l'indépendance des propositions entre elles<sup>8</sup>. Historiquement, les questions de TCS avaient des modalités de réponse spécifiques selon qu'il s'agisse de questions diagnostiques, d'investigation ou de thérapeutiques. L'ancrage universel proposé dans l'exemple (et sur la plateforme SIDES) permet de mélanger des questions des trois dimensions pour une même vignette. Vous pouvez donc, pour une même vignette, interroger un raisonnement portant sur une hypothèse diagnostique, une proposition thérapeutique ou d'investigation. Pour obtenir au final entre 90 et 100 ques-

tions dans un TCS, il est nécessaire de créer entre 120 et 150 questions, car deux étapes ultérieures de sélection sont nécessaires.

Concernant le panel de référence, il doit comprendre entre 10 et 20 membres pour des examens à fort enjeu<sup>8</sup>. Pour des TCS en formatif (pour guider les apprentissages ultérieurs des étudiants), trois à cinq membres suffiront pour refléter l'incertitude sur ces sujets. La désignation des membres du panel pose également question : qui est « expert » de la discipline ? Ce sont habituellement les professionnels désignés comme tels par leurs pairs. Cette question de la légitimité des membres du panel est cruciale pour la validité du TCS et son acceptabilité pour les étudiants. Il est parfois proposé de demander aux professionnels de la discipline de désigner les experts par un vote.

Le nombre de questions d'un TCS dépend aussi de l'usage. Pour un examen sommatif (vérifier que les apprentissages attendus ont bien eu lieu), il faut une centaine de questions. 90 questions semblent un minimum pour des raisons de validité de contenu principalement. En formatif, aucun nombre de questions n'est prescrit.

Le rôle du panel est de permettre la création du barème de notation, mais aussi de sélectionner les bonnes questions à poser aux étudiants. En effet, les questions pour lesquelles le panel est unanime sont davantage à considérer comme des questions à choix



explorant des connaissances. À l'inverse, des questions pour lesquelles les réponses du panel sont très dispersées sont probablement de mauvaises questions : mal formulées, ou traitant d'un domaine trop controversé. Pour sélectionner les meilleures questions, un calculateur en format Excel est disponible gratuitement sur le site de l'université de Montréal : [https://www.cpass.umontreal.ca/recherche/axes-de-recherches/concordance/tcs/corriger\\_tcs/](https://www.cpass.umontreal.ca/recherche/axes-de-recherches/concordance/tcs/corriger_tcs/).

Avec cet outil, les questions à privilégier sont celles avec une variance intermédiaire (ni faible ni élevée) dans l'onglet « items stats ». Cette première sélection conduit habituellement à écarter entre 10 et 30 % des questions créées.

Une deuxième étape de sélection des questions est parfois proposée après que les étudiants ont répondu<sup>8</sup>. Il s'agit de comparer le profil de réponse des étudiants et celui du panel. Le calculateur permet d'éva-

luer la corrélation entre les réponses de ces deux groupes. Une « bonne » question devrait avoir globalement des profils de réponse superposables entre les étudiants et les membres du panel. Des profils de réponses très différents (une faible corrélation) peut indiquer que la question n'a pas été comprise de la même manière par les étudiants et les experts. L'outil Excel<sup>®</sup> permet de sélectionner les questions à inclure dans l'analyse finale avec un calcul du coefficient alpha de Cronbach pour augmenter la validité du TCS. Ces considérations docimologiques avancées n'ont de sens que dans une utilisation à fort enjeu. La limite principale pour une appropriation large du TCS dans la communauté universitaire est la relative complexité de ces deux étapes de sélection.

Pour finir, le temps de passation habituel est compris entre 30 et 60 secondes par question. Il faudra donc fréquemment une heure aux membres du panel de référence pour

compléter leurs réponses. Utiliser des outils numériques tels que la plateforme SIDES ou des questionnaires type LimeSurvey<sup>®</sup> ou Google Forms<sup>®</sup> fait gagner du temps si plusieurs milliers de réponses du panel doivent être traitées.

## CONCLUSION

Le TCS est un outil d'évaluation d'une partie du raisonnement clinique. Les particularités en termes de tâche demandée et de mode de notation nécessitent un temps d'appropriation pour les enseignants et les étudiants.

Une fois ce travail effectué, la possibilité de refléter l'incertitude dans les soins en fait un outil de choix pour l'enseignement de la médecine générale. Il est important que les enseignants de la discipline alimentent les bases de questions de TCS pour que l'importance des soins premiers transparaissent dans les modalités évaluatives en deuxième cycle. ♦

### Résumé

La réforme du deuxième cycle des études médicales va introduire le test de concordance de script (TCS) parmi les outils d'évaluation du raisonnement clinique. L'objectif de cet article didactique est de présenter le principe du TCS et d'en détailler les principales modalités de conception afin de permettre aux généralistes enseignants intervenant dans le deuxième cycle des études médicales d'en concevoir des questions. Le TCS explore une sous-partie du raisonnement clinique : les processus analytiques. La particularité principale du TCS est que le score est attribué en fonction de la proximité de la réponse de l'étudiant avec celles d'un panel de référence.

La deuxième partie de l'article apporte des conseils pratiques quant à la rédaction de questions, à la sélection de celles-ci et à l'utilisation du TCS.

→ Mots-clés : éducation ; enseignement spécialisé en médecine ; enseignement ; pédagogie.

### Summary

The reform of graduate medical school in France will introduce the Script Concordance Test (SCT) as a tool for assessing clinical reasoning. The objective of this didactic article is to present the principle of the SCT and to detail its main design modalities in order to allow general practitioner teachers in the second cycle of medical studies to create SCT questions. The Script Concordance Test explores a subpart of clinical reasoning: analytical processes. The main feature of the SCT is that the score is assigned according to the closeness of the student's response to those of a reference panel.

The second part of the article provides practical advice on writing questions, selecting questions and using the SCT.

→ Keywords: education; education medical graduate; teaching; pedagogy.

### Références

1. République française. Arrêté du 2 septembre 2020 portant modification de diverses dispositions relatives au régime des études en vue du premier et du deuxième cycle des études médicales et à l'organisation des épreuves classantes nationales.
2. Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten C. The script concordance test: a tool to assess the reflective clinician. *Teach Learn Med* 2000;12:189-95.
3. Charlin B, Tardif J, Boshuizen HP. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Acad Med* 2000;75:182-90.
4. Sweller J. Cognitive load during problem solving: effects on learning. *Cogn Sci* 1988;12:257-85.
5. Gawad N, Wood TJ, Cowley L, Raiche I. The cognitive process of test takers when using the script concordance test rating scale. *Med Educ* 2020;54:337-47.
6. Bland AC, Kreiter CD, Gordon JA. The psychometric properties of five scoring methods applied to the script concordance test. *Acad Med* 2005;80:395-9.
7. Wilson AB, Pike GR, Humbert AJ. Analyzing script concordance test scoring methods and items by difficulty and type. *Teach Learn Med* 2014;26:135-45.
8. Dory V, Gagnon R, Vanpee D, Charlin B. How to construct and implement script concordance tests: insights from a systematic review. *Med Educ* 2012;46:552-63.
9. Lubarsky S, Dory V, Duggan P, Gagnon R, Charlin B. Script concordance testing: from theory to practice. *AMEE guide no 75. Med Teach* 2013;35:184-93.
10. Pelaccia T, Tardif J, Tribby E, Charlin B. An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory. *Med Educ Online* 2011;16.
11. Daniel M, Rencic J, Durning SJ, et al. Clinical Reasoning Assessment Methods: a scoping review and practical guidance. *Acad Med* 2019;94:902-12.



Mathieu Lorenzo

## Développer le raisonnement clinique, la réflexivité et le professionnalisme des futurs médecins généralistes en France



### Résumé

A partir de travaux sur le raisonnement clinique, la réflexivité et le professionnalisme, cette thèse décrit l'utilisation possible en formation initiale en Médecine Générale et en recherche en éducation médicale des patients virtuels, des formations par concordance et de la grille REFLECT.

Cinq articles originaux sont présentés au sein d'un document relatant le cheminement scientifique de l'auteur en donnant une cohérence à l'ensemble des travaux réalisés.

Des pistes de recherches originales pour l'avenir sont proposées et les perspectives en matière de formation des futurs médecins généralistes sont décrites.

Mots clés : Pédagogie, éducation médicale, raisonnement clinique, réflexivité, professionnalisme, formation par concordance, patients virtuels, Hello-Clinical

### Abstract

Based on work on clinical reasoning, reflexivity and professionalism, this thesis describes the possible use in initial training in General Medicine and in research in medical education of virtual patients, training by concordance and the REFLECT tool.

Five original articles are presented in a document relating the scientific path of the author, giving coherence to all the work carried out.

Original avenues of research for the future are proposed and the perspectives in terms of training of future general practitioners are described.

Keywords: pedagogy, medical education, clinical reasoning, reflection, professionalism, learning-by-concordance, virtual patients, Hello-Clinical