

Evaluation de résolution de problèmes mal définis en éthique clinique : variation des scores selon les méthodes de correction et les caractéristiques des jurys

Guy LLORCA*, Pascal ROY**, Benjamin RICHE**

Résumé *Contexte* : La formation médicale évalue facilement les problèmes bien définis posés par la pratique. Les problèmes mal définis sont souvent écartés du fait de leur difficulté d'évaluation. **But** : Étudier les variations liées aux méthodes d'établissement des scores, aux experts et au niveau des étudiants, en matière d'évaluation de la décision médicale. **Matériel et Méthode** : à partir d'une situation clinique aboutissant à une décision complexe, les réponses de 50 étudiants en début d'apprentissage clinique et de 11 internes sont évaluées par 4 jurys différents et un évaluateur isolé. Les 4 jurys ont été constitués par 139 seniors médicaux habitués à la décision complexe. Les tendances des différents jurys, les notations obtenues par deux différents modes de correction (par test de concordance de script ou TCS, et par questions à choix multiples ou QCM) ainsi que leurs corrélations sont comparées. **Résultats** : les professionnels émettent des jugements décisionnels proches. Les jugements des internes sont voisins de ceux des seniors. Les jugements des étudiants débutants sont significativement différents. Les différentes notations sont corrélées, mais celles obtenues par TCS sont plus homogènes. Les notes obtenues par les internes sont significativement plus élevées que celles obtenues par les étudiants, quel que soit le jury. **Conclusions** : le TCS, dans l'évaluation de décisions complexes, comme c'est le cas en matière d'éthique clinique, représente un instrument plus valide qu'une évaluation conduite par un expert isolé, pour mettre en évidence la composante professionnelle des savoirs.

Mots clés Evaluation ; formation médicale ; test de concordance de script ; TCS ; éthique clinique ; méthodes ; cas clinique.

Summary *Aim*: To study variations associated with scoring methods, level of students and background of medical experts in decision making assessment. **Material and method**: 139 medical experts were divided in 4 groups (ethics committee group, medical teacher group, therapeutic teacher group and family physician group). They gave their opinion on a medical case containing complex decision making. These opinions, and the opinion of a unique teacher, were compared with the opinion of 11 fellows and 50 students in their primary year of clinical study. These opinions were then used to produce scores for residents and students. Scores obtained by two different scoring methods (Script concordance test and multiple choice tests) and their correlations were compared. **Results**: The professionals have very similar decision making judgment. The judgment of residents is relatively close to those of professionals. The judgment of students is statistically different. Scores are correlated but scores obtained from SCT are more homogenous. Residents scores are significantly higher than those of students, whatever the expert group used to establish scores. **Conclusion**: In the assessment of difficult decision making, as in clinical ethics, the SCT scoring method represents a more valid way than assessment provided by a unique expert to judge professional knowledge.

Keywords Evaluation; script concordance test; SCT; clinical ethics; medical training; method.

Pédagogie Médicale 2003 ; 4 : 80-8

* Laboratoire de Thérapeutique - 8 avenue Rockefeller - 69373 Lyon cedex 08

** Service de biostatistique - Centre Hospitalier Lyon-Sud - 69495 Pierre-Bénite cedex

Correspondance : Guy Llorca - Service de Rhumatologie - Centre hospitalier Lyon-Sud - 69495 Pierre-Bénite cedex

Tél. : +33 478 86 12 81 - Télécopie : +33 478 86 33 42 - mailto:guy.llorca@chu-lyon.fr

Introduction

Comme certains le font remarquer¹, une formation finit par n'attacher d'importance qu'à ce qu'elle a décidé d'évaluer. La formation médicale n'échappe pas à cette règle², et, de fait, elle a tendance à privilégier ce qui est facilement évaluable, c'est-à-dire, que ce qui appartient aux connaissances factuelles, conceptuelles et procédurales selon la terminologie d'Anderson³ et de Tardif⁴. Les prises de décision en situation complexe sont ainsi souvent écartées de la formation médicale, car elles correspondent à des problèmes mal définis selon la conception de Schön⁵. L'approche de concordance de script, récemment introduite⁶, semble pouvoir apporter une solution à la difficulté d'évaluer la prise de décision en situation mal définie. Cette nouvelle approche repose sur trois principes : 1) l'évaluation porte sur des tâches cliniques authentiques et mal définies ; 2) les questions posées sont des questions que se posent les médecins expérimentés dans ces situations et le format de réponse correspond à celui des réponses possibles dans ce contexte ; 3) les réponses des experts varient dans ce genre de situation. Les jurys sont ainsi constitués de plusieurs personnes expérimentées dans le domaine évalué. Les étudiants reçoivent, pour chaque question, un crédit qui est fonction du nombre de membres du jury qui a répondu comme eux.

Dans le domaine de l'éthique clinique, les problèmes soulevés sont difficiles à évaluer car ils ne répondent pas au seul rationnel de la médecine fondée sur des preuves⁷, ils doivent impérativement prendre en compte les contextes^{8,9}, et ne peuvent se contenter de recettes toutes faites ou de vérités révélées¹⁰. Dans ces conditions on conçoit la difficulté à établir une légitimité de jury. Il n'en demeure pas moins qu'une professionnalisation existe dans ce domaine, et représente la différence entre le simple bon sens et l'argumentation élaborée (prenant en compte la position des autres et les arguments tirés de références en éthique, de l'expérience clinique, du contexte, de situations semblables etc.). Un système d'évaluation en éthique doit donc permettre de déceler cette progression dans la professionnalisation, tout en tenant compte de l'absence *a priori* de solutions universelles.

Or, les difficultés les plus importantes de la décision médicale sont liées aux contextes^{8,9}, et l'approche proposée par l'éthique clinique, pour évaluer le bien-fondé des décisions en tant que jugement de valeur, est difficile à explorer¹¹. Le test de concordance de script (TCS)⁶, qui permet la prise en compte, non seulement de la démarche cognitive qui aboutit au processus décisionnel, mais aussi des variations d'avis des experts¹², enrichit l'évaluation de la formation médicale, et s'intéresse plus aux savoirs professionnels que techniques, alors que l'évaluation par

questions à choix multiples (QCM) explore certainement mal ces savoirs¹³.

Dans cette étude, nous avons voulu savoir, devant un problème d'éthique clinique, comment se caractérisent les réponses en fonction d'un groupe d'appartenance, comment se situent les scores des étudiants et des internes en fonction des différents jurys logiquement admissibles, et quelle est la fiabilité test-retest du test proposé.

Matériel et méthode

Matériel

Les sujets de l'étude

Deux cents personnes ont participé à l'étude, réparties en 139 experts constituant les membres de jury, 11 internes en cours de spécialisation et 50 étudiants en médecine en début de cursus clinique.

Le stimulus décisionnel et le format de réponse

Il a consisté en un cas clinique, choisi à partir d'une situation réelle observée en rhumatologie, situation soulevant un problème éthique ne nécessitant pas de connaissances médicales spéciales, pour permettre à tous les étudiants de se prononcer. La présentation du cas se terminait par l'exposé d'un problème décisionnel thérapeutique. Six attitudes possibles étaient proposées. Les membres de jury, comme les personnes passant le test (étudiants et internes) devaient exprimer leur opinion à propos des six attitudes possibles proposées devant ce cas. Une radiographie de la hanche du patient était jointe pour ancrer la situation dans le réel et lever toute ambiguïté diagnostique. Les attitudes possibles étaient proposées, selon une grille classique utilisée en TCS⁶ et simplifiée (Figure n° 1).

Constitution des Jurys

Les 139 personnes membres de jury ont été recrutées parmi quatre groupes de « professionnels » seniors choisis pour leurs compétences en pratique médicale, en éthique clinique ou les deux. Le premier jury (CE) était constitué de 15 membres du Comité d'éthique du Centre Hospitalo-Universitaire de Lyon, comportant pour moitié des médecins et pour moitié des non-médecins (juristes, représentants d'association de malades, de cultes, de sciences humaines...). Le second jury (PRE) était constitué de 15 professeurs de médecine particulièrement concernés par l'éthique, exerçant en médecine interne, médecine d'urgence, néphrologie et rhumatologie. Le troisième jury (MG) était représenté par 65 médecins généralistes exerçant en ville. Le quatrième jury (THE) a été

Énoncé

- Monsieur Georges L... a 65 ans. Nouvellement retraité de la fonction publique (facteur), il est veuf et vit seul. Il est atteint d'une polyarthrite rhumatoïde avec coxite destructrice. Cette atteinte évolue rapidement : elle est très douloureuse et a détruit le cartilage polaire supérieur, si bien que le patient ne peut pratiquement plus marcher sans aide et plus de 100 mètres. Il ne veut plus continuer ainsi (il parle de suicide).
- Le traitement médical de sa coxite n'a pas apporté de bénéfice suffisant sur la fonction et le traitement de la douleur doit passer par les opiacés pour être efficace.
- Dans son passé, il a fait une pancréatite, il y a 10 ans, dont il conserve une insuffisance pancréatique avec diabète, il a fait un infarctus du myocarde il y a 5 ans, a subi un triple pontage coronarien et conserve une insuffisance cardiaque, il est atteint depuis 2 ans d'un syndrome d'apnée du sommeil appareillé en ventilation non agressive nocturne.
- Il est examiné par un chirurgien orthopédiste qui retient l'indication théorique de prothèse totale de hanche. Lors de la visite pré-anesthésique, l'anesthésiste réfute l'indication pour risques anesthésiques trop élevés.

représenté par 44 enseignants de thérapeutique, membres de l'Association Pédagogique Nationale pour l'Enseignement de la Thérapeutique, et rassemblant l'essentiel des spécialités médicales, dont la médecine d'urgence et la gériatrie. Une 5^e évaluation a été réalisée par un professeur de thérapeutique impliqué en éthique clinique, et isolé pour servir de correcteur unique. Ainsi deux méthodes de correction ont pu être comparées : la première méthode - basé sur une recherche de concordance entre les personnes testées et toutes les réponses des membres des jurys - et la seconde méthode, plus classique, de type QCM, basé sur la réponse du correcteur unique donnant un crédit d'1 point à la bonne réponse et de 0 à tout autre. Cette seconde correction a été réalisée avec deux modalités selon que plusieurs réponses possibles étaient acceptées et pondérées pour chaque attitude (CUM), ou qu'une seule réponse par attitude était retenue (CUU). Ainsi, 6 évaluations permettaient des corrections en deux modalités : soit par test de concordance de script (les 4 jurys CE, PRE, MG et THE qui, regroupés, ont constitué l'appréciation des seniors), soit par QCM (les 2 corrections par correcteur unique CUM et CUU).

Les étudiants

Ils ont été représentés par deux groupes. Le premier était constitué par l'essentiel des étudiants de la promotion de la Faculté de Médecine Lyon-Nord et débutant leur seconde année de deuxième cycle d'études médicales, c'est-à-dire, ayant appris la séméiologie, mais débutant l'étude de la pathologie et les stages d'externe à l'hôpital (Groupe étudiants = 50 personnes). Le second groupe était constitué par l'essentiel de la promotion actuelle des

Figure n° 1 : Grille des jugements aux attitudes proposées.

Attitudes possibles (1)	Jugements (2)				
A - Accepter la décision de l'anesthésiste	-2	-1	0	+1	+2
B - Chercher à faire changer d'avis l'anesthésiste	-2	-1	0	+1	+2
C - Rechercher un anesthésiste d'un avis différent	-2	-1	0	+1	+2
D - Chercher à faire changer d'avis le malade	-2	-1	0	+1	+2
E - Se donner quelques semaines de réflexion avec le malade	-2	-1	0	+1	+2
F - Demander un « avis éthique »	-2	-1	0	+1	+2

Entourez le niveau de vos jugements (colonne 2) pour chaque attitude (une seule réponse par attitude) avec :
 -2 = Totalement inacceptable ; -1 = Difficilement acceptable ;
 0 = Non pertinent dans cette situation ; +1 = Plutôt acceptable ; +2 = Totalement acceptable

internes en Rhumatologie de la région Rhône-Alpes (Groupe internes = 11 personnes). Ce second groupe était supposé posséder une expertise professionnelle supérieure au groupe des étudiants débutants.

Méthode

Les tests

La situation clinique a été proposée individuellement, à chaque participant à l'étude, après avoir obtenu un accord de participation libre. L'énoncé, la radiographie et les attitudes possibles ont été proposés, en projection, aux différents groupes. La grille de réponse individuelle a été proposée en support papier, en demandant à chacun de répondre individuellement. Les réponses devaient porter sur chaque attitude, en ne choisissant qu'une possibilité par attitude, selon une échelle de Likert en 5 points (Figure n° 1). La situation a été proposée une seconde fois, à une demi-journée d'intervalle (pour éviter au maximum un biais de contamination²), au groupe des étudiants débutants, pour s'assurer de la stabilité de leurs réponses.

Traitement des réponses

Les données recueillies sous le format de l'échelle de Likert ont été traitées de deux façons. La première a consisté à regrouper les réponses en trois catégories au lieu de cinq, de manière à percevoir plus aisément les tendances globales de réponse dans chaque groupe étudié. Les tendances globales ont été ainsi obtenues en regroupant les réponses en 3 classes de jugements (positive, non pertinente et négative). La seconde façon a consisté à traiter les données en conservant les cinq catégories de réponses pour effectuer une correction selon la méthode du test de concordance de script (TCS), données exploitées en scores bruts transformés pour obtenir un maximum de 60 points, en attribuant le même nombre de point pour chaque attitude (10 points par attitude), puisque chaque attitude était *a priori* recevable^{6,14}.

La correction

Elle a été effectuée pour chacune des 6 possibilités différentes de correction. Pour la correction CUM, 2 ou 3 réponses étaient acceptées comme correctes par attitude (avec 75 % des points attribués à la réponse préférée du correcteur et 25 % des points pour les réponses voisines d'un degré sur l'échelle de Likert).

Pour la correction CUU, seule la réponse du correcteur unique était considérée comme correcte et recevait la totalité des points. Ainsi, 4 corrections ont tenu compte des variations des membres constituant le jury correspon-

dant, une correction a tenu compte des variations proches de l'opinion d'un correcteur unique, et une correction n'a tenu compte que de la seule opinion stricte de ce correcteur.

L'analyse statistique

La reproductibilité intra-observateur des jugements portés par les étudiants a été évaluée en utilisant le coefficient de corrélation simple et, pour mieux évaluer la part du hasard, en calculant le coefficient Kappa de Cohen¹⁵. Les valeurs supérieures à 0,75 représentent un niveau d'accord excellent entre évaluateurs, des valeurs comprises entre 0,40 et 0,75 un bon niveau d'accord, et des valeurs inférieures à 0,40 un faible niveau d'accord¹⁶.

La comparaison des réponses des différents groupes testés a été effectuée pour chacune des questions (A à F). Les tables de contingence des distributions des scores des seniors et des étudiants ont été analysées à l'aide d'un modèle log-linéaire¹⁷. Les scores ont été analysés comme une variable qualitative ordinaire. Le modèle avec interaction entre les variables scores (ordinaire) et groupe (nominale) permet de tester l'hypothèse d'une répartition différente des scores. Ce modèle, construit pour des réponses à 5 et à 3 niveaux, est comparé à un modèle à pente unique (absence d'interaction) à l'aide d'un test du rapport de vraisemblance.

Des valeurs de p inférieures à 5 % sont considérées comme significatives. Le paramètre de pente est le logarithme du rapport des proportions de deux scores successifs. Une valeur négative de ce paramètre caractérise une évaluation négative, une valeur positive une plus forte proportion d'évaluations positives, une valeur nulle une distribution uniforme des scores.

Enfin, les notes obtenues à l'ensemble des réponses par les étudiants et par les internes ont été comparées au moyen d'un test t. Une recherche de corrélation entre les notes attribuées par les différents modes de correction a été effectuée par le calcul du coefficient de corrélation simple.

Résultats

Fiabilité mesurée par test-retest

La fiabilité des réponses des étudiants est bonne. La différence pour chaque attitude est faible et, pour les tendances, une corrélation très hautement significative existe entre les deux tests ($r = 0,95$; $p < 0,001$). L'utilisation du coefficient Kappa s'inscrit dans le même sens (Tableau 1). La valeur la plus basse, observée pour la question F, est liée à l'augmentation des réponses positives lors de la deuxième mesure (on passe de 34 à 41 réponses positives).

Recherche et Perspectives

Analyse des réponses

Globales

Les tendances de l'ensemble des 4 jurys de seniors apparaissent comme nettement différentes de celles des étudiants alors que celles des internes semblent beaucoup plus proches de celles des seniors (Tableau 2). Cette constatation est bien retrouvée par le calcul des paramètres de pente, tant à 5 qu'à 3 niveaux, pour la confrontation des réponses des seniors à celles des étudiants (Tableau 3). La confrontation des réponses des internes n'a pu être réalisée en raison du faible effectif testé. Les étudiants donnent des réponses significativement opposées à celles des seniors pour les attitudes A à D. Il existe une convergence manifeste des réponses aux attitudes E et F qui plaide pour un accord sur ces propositions.

Des seniors

L'évaluation des paramètres de pente des différents groupes de seniors met en évidence une convergence des

réponses aux questions B, C, E et F (Tableau 4). L'analyse globale des réponses à 5 et 3 niveaux ne met pas en évidence de différence décelable ($\chi^2 = 0,4791$, $p = 0,9235$). L'analyse des pentes des réponses à la question A met en évidence une communauté d'opinions négatives, mais bien plus affirmée pour les jurys CE et PRE que pour les jurys THE et MG. La significativité est, là, d'ordre quantitatif. L'analyse des pentes de réponses à la question D met en évidence une opposition manifeste du jury MG par rapport aux autres jurys. La significativité est, pour cette attitude, d'ordre qualitatif.

Des étudiants et des internes

Les tendances des réponses des étudiants diffèrent significativement de celles des seniors pour les attitudes A, B, C et D. Ces tendances diffèrent également de celles des internes. En revanche, les réponses aux attitudes E et F sont en accord avec celles des internes et des seniors (Tableaux 2 et 3).

Les tendances des réponses des internes diffèrent peu de

Tableau 1 : Reproductibilité intra-observateurs évaluée par la méthode test-retest (le coefficient Kappa de Cohen pour chaque attitude proposée est donné, avec, entre parenthèses, l'erreur type de ce coefficient)

Attitudes	A	B	C	D	E	F
Kappa	0,71 (0,10)	0,48 (0,13)	0,45 (0,14)	0,47 (0,12)	0,59 (0,13)	0,39 (0,14)

Tableau 2 : Tendances des réponses des seniors, des étudiants et des internes.

Attitudes	A	B	C	D	E	F
Réponses	(%)	(%)	(%)	(%)	(%)	(%)
Seniors						
Négative	79	32	25	49	11	8
Non pertinente	7	13	19	20	17	24
Positive	14	56	56	31	73	68
Étudiants						
Négative	34	50	70	24	14	0
Non pertinente	14	14	10	16	16	28
Positive	52	36	20	60	70	70
Internes						
Négative	64	27	18	36	9	9
Non pertinente	18	0	9	0	0	36
Positive	18	64	73	64	91	55

Tableau 3 : Évaluation des paramètres de pente des seniors et des étudiants.

Questions	A	B	C	D	E	F
Score à 5 niveaux						
Seniors	- 0,38	+0,06	+0,15	- 0,10	0,41	+0,37
Etudiants	+0,09	- 0,08	- 0,45	+0,32	0,44	+0,40
Interaction	P = 0,0001	P = 0,2269	P < 0,0001	P = 0,0007	P = 0,8370	P = 0,8248
Score à 3 niveaux						
Seniors	- 0,90	+0,23	+0,45	- 0,12	+0,87	+0,77
Etudiants	+0,28	- 0,18	- 0,86	+0,58	+0,90	+1,24
Interaction	P < 0,0001	P = 0,0445	P < 0,0001	P = 0,0009	P = 0,9092	P = 0,0611

Tableau 4 : Évaluation des paramètres de pentes des différents groupes de seniors
(Avec : CE pour Comité d'éthique, PRE pour professeurs compétents en éthique, THE pour enseignants en thérapeutique, MG pour médecins généralistes)

Questions	A	B	C	D	E	F
CE	-2,48	+0,61	+0,61	-0,48	+1,20	+0,47
PRE	-1,47	0,00	+0,47	-0,64	+1,19	+1,10
THE	-0,81	+0,10	+0,21	-0,36	+1,01	+0,67
MG	-0,58	+0,30	+0,58	+0,26	+0,61	+0,61
Interaction	P = 0,0377	P = 0,4747	P = 0,4836	P = 0,0078	P = 0,2988	P = 0,8091

Tableau 5 : Moyennes des notes obtenues avec les différents jurys et méthodes de correction pour les groupes des internes et des étudiants.

(Avec : CE pour Comité d'éthique, PRE pour professeurs compétents en éthique, MG pour médecins généralistes, THE pour enseignants en thérapeutique, CUM pour correcteur unique à réponses multiples et CUU pour correcteur unique à réponse unique.)

Jury	Moyennes ± SD/60		Maximum Minimum		Coefficients de variation		p
	Étudiants	Internes	Étudiants	Internes	Étudiants	Internes	
CE	29,6 ± 11,7	40,0 ± 11,1	55,1 – 9,80	56,8 – 22,7	0,39	0,27	< 0,0001
PRE	30,4 ± 10,5	39,9 ± 12,9	54,1 – 10,1	58,5 – 10,1	0,35	0,34	< 0,0001
MG	38,5 ± 6,90	45,6 ± 7,50	59,4 – 23,4	57,4 – 35,0	0,18	0,16	< 0,0001
THE	38,9 ± 7,10	43,2 ± 6,60	52,1 – 21,6	52,6 – 31,5	0,18	0,15	= 0,010
CUM	21,9 ± 9,50	32,1 ± 11,1	50,0 – 3,30	46,7 – 16,7	0,44	0,34	< 0,0001
CUU	16,8 ± 10,4	24,1 ± 8,30	50,0 – 0,00	50,0 – 0,00	0,62	0,61	< 0,0001

celles des experts, à l'exception de l'attitude D (Tableau 2) où ils se situent en position intermédiaire entre les seniors et les étudiants. Là encore, les attitudes E et F paraissent en accord avec les seniors et les étudiants.

Analyse des notations

Analyse des notations des étudiants

Les notes obtenues par les étudiants ne sont véritablement supérieures à la moyenne qu'avec les jurys MG et THE (Tableau 5). Les différences des moyennes attribuées par chaque jury sont toujours très hautement significatives (t compris entre $-5,338$ et $13,694$ et p toujours $< 0,0001$), à l'exception des différences entre les jurys CE et PR ($t = 0,700$, $p = 0,4871$) et MG et THE ($t = -0,499$, $p = 0,6203$). La note zéro n'est possible qu'avec le correcteur isolé CUU (QCM à réponse unique).

L'analyse des rangs des étudiants, met en évidence des corrélations hautement significatives entre les différents jurys (r compris entre $0,50$ et $0,94$, p compris entre $0,05$ et $0,0001$), à l'exception du correcteur unique par rapport au jury THE et quelque soit le mode de correction ($r = 0,24$ et $0,19$ selon le mode de correction ; p non significatif).

Analyse des notations des internes

Les notes obtenues par les internes sont toujours supérieures à la moyenne avec tous les jurys et tous les modes de correction (Tableau 5).

Les différences des moyennes attribuées par chaque jury ne sont statistiquement significatives que pour les modalités utilisant le correcteur unique (t compris entre $8,836$ et $9,963$ et $p < 0,0001$). La note zéro n'est possible qu'avec le jury CUU (QCM à réponse unique). L'analyse des rangs des internes met en évidence d'excellentes corrélations, là encore, entre les différents jurys (r compris entre $0,70$ et $0,88$, p compris entre $0,01$ et $0,0001$), à l'exception du correcteur unique qui attribue des rangs différents de ceux attribués par les jurys MG et THE (r respectivement à $0,47$ et $0,40$; p non significatif).

Comparaison des notations entre les internes et les étudiants

Les moyennes obtenues par les internes sont toujours significativement plus élevées que celles obtenues par les étudiants, quel que soit le jury ou le mode de correction (Tableau 5).

Les corrélations entre les différents modes de correction sont très proches pour les internes et les étudiants, si l'on tient compte de la différence d'effectif de ces deux populations étudiées.

Discussion

Cette étude porte sur l'analyse des comportements en condition de décision complexe. En effet, les problèmes simples de la pratique médicale sont liés à des connaissances factuelles à niveau de preuve généralement élevé, et répondent à l'analyse conditionnelle simple, que l'on peut qualifier de 1° ordre (« Si...Alors... »). Les évaluations par QCM à contexte simple se prêtent bien à ce type de connaissances.

Dans la réalité de la décision médicale quotidienne, les problèmes sont plus complexes. Ils peuvent être liés à un enchaînement conditionnel de 2° ordre (« Si... et Si...Alors... »), et sont valablement abordés par les QCM à contexte riche.

L'introduction de la démarche fondée sur des preuves, et la différenciation entre savoirs techniques et savoirs professionnels, obligent à considérer le degré de définition des problèmes traités, car ils peuvent répondre à un autre enchaînement conditionnel, que l'on peut qualifier de 3° ordre (« Si... et Si... Mais... Alors... »), lorsque les différentes réponses possibles correspondent à celles que des experts auraient pu proposer. Le TCS se prête mieux à l'exploration de ces savoirs et attitudes⁶.

Dans le cas de problèmes mal définis devant aboutir à des décisions complexes, les propositions d'attitudes possibles sont toutes *a priori* valables, et leur choix préférentiel dépend du niveau de preuve, bien entendu, mais, de manière aussi importante, du contexte^{8,9,18}. C'est le domaine de l'éthique clinique en tant que jugement de valeur des attitudes possibles¹⁰.

Notre étude a porté sur un problème de décision complexe issu du réel. L'énoncé de la situation est bien de 3° ordre, avec un apport factuel répétitif, et un apport contextuel déterminant la difficulté de choix des réponses possibles, représenté par une vulnérabilité attachée à un risque aggravé dont l'estimation est scientifiquement impossible.

Dans ces conditions, la préférence du malade et l'expérience du médecin deviennent déterminantes, si les attitudes proposées correspondent bien à des solutions acceptables, sans introduire de leurre, et sont bien dictées par des comportements répondant à des principes ou des règles consensuelles. Dans la réalité, l'attitude adoptée par l'équipe a été l'attitude « E », et le patient, après un mois de réflexion, a décidé de renoncer à l'intervention. Quelques semaines plus tard, à l'occasion d'une chute, il a été victime d'une fracture du col fémoral et a été opéré en urgence sans difficulté particulière !

La méthode utilisée est relativement complexe mais cherche à répondre à la nécessité d'analyser avec suffisamment de performance des données qualitatives concernant des groupes inhomogènes particulièrement en matière d'effectifs. Le regroupement en 3 niveaux de

réponses ne semble pas affecter les significations statistiques (Tableau 3). Il est justifié, ici, par le fait que l'intention décisionnelle représente le critère d'évaluation principal.

L'attitude A, qui représente la perte de chance, est l'élément clé du raisonnement dans la situation proposée. Parmi les seniors, les jurys CE et PRE s'inscrivent davantage contre cette attitude pour des raisons éthiques, alors que les jurys THE et MG s'y opposent avec moins de force pour des raisons vraisemblablement plus pratiques. Les étudiants cautionnent plus cette attitude vraisemblablement par respect de l'avis d'un spécialiste plus compétent qu'eux. Les réponses aux autres attitudes sont logiquement en accord avec la position adoptée devant le refus de l'anesthésiste, refus qui rend complexe cette situation mal définie.

Cette étude met bien en évidence la différence des tendances de jeunes étudiants inexpérimentés par rapport à celles de professionnels aguerris et d'internes qui sont plus proches, dans leurs réponses, des professionnels que des étudiants. Cette différence mesure vraisemblablement le chemin que chaque étudiant doit effectuer pour passer de connaissances techniques à un savoir professionnel.

Si toutes les méthodes de correction sont corrélées, la force de cette corrélation est différente, et notre démarche adaptée du TCS permet d'obtenir des corrélations supérieures à celles observées avec les méthodes de type QCM. Ces différences sont déjà manifestes avec un seul cas et six réponses possibles. Dans la réalité pédagogique, l'utilisation de plusieurs cas est certainement susceptible d'accroître encore les performances de cette démarche.

Dans le cas, comme ici, de problèmes complexes à forte composante contextuelle, la question essentielle est celle de la valeur des différents jurys d'experts possibles, puisqu'il s'agit d'un jugement de valeur nettement lié aux sensibilités des acteurs décisionnels^{18,19,20}. Cette étude, qui confronte des jurys « logiques » dans de telles situations, s'inscrit pour une équivalence des Comités d'éthique pluridisciplinaires et des professeurs de médecine entraînés à la réflexion éthique. Le jury des enseignants de thérapeutique, reflète plus une diversité liée à sa multidisciplinarité intrinsèque. Le jury des médecins généralistes peut être également recevable, mais il reflète plus une vision technique et pratique de la situation. Le correcteur unique, s'il est issu des professeurs entraînés à la réflexion éthique, est également recevable, mais représente une asymétrie autoritaire, et la corrélation obtenue avec lui est de bien moins bonne qualité. Cette constatation indique que la correction par un correcteur unique est certainement moins valide dans de telles situations complexes ; elle doit, en tout cas, tenir compte des opinions voisines des siennes pour éviter l'arbitraire, et oblige à rejeter les réponses

uniques non adaptées à ces situations complexes où n'existe aucune vérité absolue ou révélée.

Cette étude apporte un premier élément de réponse aux questions que pose l'évaluation des savoirs éthiques en médecine¹¹. Elle s'inscrit pour l'utilisation de cette démarche, adaptée du TCS, dans l'apprentissage du raisonnement éthique (ARE). Son utilisation pratique soulève cependant quelques difficultés²⁰ qui sont représentées par l'élaboration de situations pertinentes, la constitution de jury adaptés à ces situations, et la connaissance des sensibilités contextuelles où ces situations sont observées^{19,20}. Cette approche peut autant servir la formation initiale que la formation postdoctorale¹⁰, en mettant en évidence les écarts entre étudiants et professionnels, en situant chaque individu par rapport à un groupe de référence, ou en établissant, à l'occasion d'ateliers d'aide à la décision²¹, des microconsensus susceptibles également d'être comparés à des groupes de référence.

Conclusion

Cette étude met en évidence l'apport d'une adaptation du TCS à l'évaluation de situations complexes conduisant à des décisions médicales difficiles. Si l'on prend la précaution de définir correctement les groupes d'experts de référence, cette approche est beaucoup plus riche que l'approche traditionnelle, puisqu'elle permet d'intégrer les données de la médecine fondée sur des preuves et celles issues de l'analyse contextuelle fondée sur le réel. Elle peut être utile tant en formation initiale que continue, et s'inscrit plus dans le cadre de l'évolution des sensibilités et des tendances actuelles.

Remerciements

Nous tenons à remercier tout particulièrement le Docteur Bernard Charlin pour ses conseils avisés, ainsi que toutes les personnes qui ont bien voulu se prêter aimablement à cette étude.

Références

1. Meirieu P. *Dictionnaire encyclopédique de l'éducation et de la formation*. Éditions Nathan, Paris, 1998.
2. Llorca G. *La Formation Médicale (Aspects conceptuels)*. Méditations, Lyon, 1999.
3. Anderson JR. *The architecture of cognition*. Cambridge MA : Harvard University Press, 1985
4. Tardif J. *Pour un enseignement stratégique. L'apport de la psychologie cognitive*. Les éditions Logiques, Montréal, 1997.
5. Schön DA. *The Reflective Practitioner : How Professionals Think in Action*. New York : Basic Books, 1993.
6. Charlin B, Gagnon R, Sibert L, Van der Vleuten C. *Le test de concordance de script, un instrument d'évaluation du raisonnement clinique*. *Pédagogie Médicale*, 2002 ; 3 : 135-44.
7. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. *Evidence-based medicine*. Churchill Livingstone editions, Edinburg, 1998.
8. Llorca G. *Médecine factuelle et décision pratique. Contexte et définitions*. *Pratiques Médicales et Thérapeutiques*, 2002, n° 22, 24-5.
9. Llorca G. *Médecine factuelle et décision pratique. La médecine factuelle à l'épreuve du quotidien*. *Pratiques Médicales et Thérapeutiques*, 2002, n° 23, 25-7.
10. Llorca G. *Éthique et formation médicale. Révision des concepts*. *Pédagogie Médicale*, 2002 ; 3 : 169-79.
11. Brazeau-Lamontagne L. *L'évaluation de l'éthique médicale des résidents*. *Pédagogie Médicale*, 2002 ; 3 : 152-8.
12. Grant J, Mardsen P. *Primary Knowledge. Medical Education and consultant expertise*. *Med Educ* 1998 ; 22 : 173-9.
13. McGaghie WC. *Evaluating Competence for Professional Practice*. In : Curry L, Wergin JF (Eds) *Educating professionals. Responding to new expectations competence and accountability*. San Francisco (Ca) Jossey-bass, 1993.
14. Norcini JJ, Shea JA, Day SC. *The use of the aggregate scoring for a recertification. examination, evaluation and the Health Professions*, 1990 ; 13 : 241-51.
15. Cohen J. *A coefficient of agreement for nominal scales*. *Educational and Psychological Measurements*, 1960, 20 : 37-46.
16. Fleiss JL. *Statistical Measures for Rates and Proportions*, 2nd Edition, Wiley, New York, 1981.
17. Agresti A. *Categorical data Analysis*. Wiley, New York, 1990.
18. Jonsen AR, Siegler M, Winslade WJ. *Clinical ethics*. McGraw-Hill ed., New York, 4° ed., 1998.
19. Roupie E, Santin A, Boulme R, Wartel JS, Lepage E, Lemaire F, Lejonc J-L, Montagne O. *Patient's preferences concerning medical information and surrogacy : results of a prospective study in a French emergency department*. *Intensive Care med*, 2000 ; 26 : 52-6.
20. Llorca G. *Du raisonnement médical à la décision partagée (Introduction à l'éthique en médecine)*. Éditions Med-line, Paris, 2003.
21. Llorca G. *Rôle des Comités d'éthique dans la formation médicale*. *Pédagogie Médicale*. 2001 ; 2 : 157-62.