Évaluation du raisonnement clinique en médecine d'urgence : les tests de concordance des scripts décèlent mieux l'expérience clinique que les questions à choix multiples à contexte riche

Clinical Reasoning Assessment in Emergency Medicine: Script Concordance Tests are More Efficient to Detect Clinical Experience than Rich-Context Multiple Choice Questions.

Jean-Paul FOURNIER[1, 2], Didier THIERCELIN[3], Céline PULCINI[2], Véronique ALUNNI-PERRET[2], Elise GILBERT[4], Jean-Marc MINGUET[5], François BERTRAND[1]

Résumé ***Contexte :*** *La médecine d'urgence est sur le point de devenir une spécialité en France, ce qui implique l'utilisation d'outils de certification adaptés.* ***Buts :*** *Comparer l'aptitude des tests de concordance des scripts (TCS) et des questions à choix multiples à contexte riche (QCM) à identifier le niveau d'expérience des cliniciens en matière de raisonnement clinique en médecine d'urgence.* ***Méthode :*** *60 QCM et 30 TCS ont été rédigés selon les recommandations publiées et administrées à 20 résidents de première année, 16 étudiants de fin de sixième année, et à neuf médecins seniors en poste dans des services d'urgences. L'analyse d'item des QCM a inclus la détermination des indices de difficulté et de discrimination. La fidélité des QCM et des TCS a été mesurée par le coefficient alpha de Cronbach. Les scores obtenus aux deux épreuves par les trois groupes ont été comparés (ANOVA avec correction de Bonferoni-Dunn après vérification de l'homogénéité des variances par un test de Levene). Ils ont été corrélés (test de Pearson) dans le groupe des résidents.* ***Résultats :*** *QCM et TCS ont obtenu des coefficients de fidélité corrects (respectivement 0,85 à 0,95 et 0,92 à 0,96 selon les groupes étudiés). Seuls, les TCS ont pu identifier le niveau d'expérience des différents cliniciens. Les scores des TCS et des QCM variaient dans le même sens, sans corrélation significative. Ce point suggère que QCM et TCS explorent deux aspects complémentaires de la compétence médicale.* ***Conclusion :*** *Le TCS paraît être un outil d'évaluation intéressant en médecine d'urgence, complémentaire d'autres tests d'évaluation.*

Mots-clés Test de concordance de scripts ; question à choix multiples ; médecine d'urgence ; raisonnement clinique ; évaluation.

Abstract ***Context:*** *Emergency medicine will soon be recognized as a medical specialty in France, thus requiring efficient assessment tools.* ***Goal:*** *To compare the ability of Script Concordance Test (SCT) and rich-context Multiple Choice Questions (MCQ) to identify the degree of experience in emergency medicine by clinical reasoning assessment.* ***Methods:*** *60 MCQ and 30 SCT were prepared in respect with published guidelines and were given to 20 residents, 16 sixth year medical students and 9 full-time emergency physicians. Item analysis was performed for MCQ with difficulty and discrimination coefficients. Reliability for both MCQ and SCT tests was determined by a Cronbach coefficient computation. Scores were analysed for all groups by one-way ANOVA with Bonferroni's correction for multiple comparisons after a Levene test of variance homogeneity. Pearson's correlation coefficient was calculated for the group of residents.* ***Results:*** *Both MCQ and SCT were highly reliable, with a Cronbach coefficient values ranging from 0.85 to 0.95 and 0.92 to 0.96, respectively. Degree of experience could only been detected from SCT scores. In the group of residents, SCT and MCQ scores varied in the same way, without any significant correlation. The latter observation suggests that both tests explore complementary competency areas.* ***Conclusion:*** *SCT appears to be a valuable assessment tool in emergency medicine, which may complement other assessment tools.*

Key words Script concordance test; multiple choice questions; emergency medicine; clinical reasoning; assessment.

Pédagogie Médicale 2006;7:20-30

Introduction

À l'issue d'une séquence d'enseignement dédiée à l'acquisition de compétences cliniques, l'évaluation des apprentissages des étudiants doit vérifier leur maîtrise d'un corpus de connaissances spécifiques ainsi que leur capacité à les réutiliser dans les conditions prévisibles de leur exercice professionnel, au service du processus de raisonnement clinique face à des problèmes de santé représentatifs de ceux auxquels ils seront exposés [1, 2].
Avec la création du diplôme d'études supérieures complémentaires (DESC), la médecine d'urgence est sur le point de devenir une discipline spécialisée en France, ce qui implique l'utilisation d'outils de certification adaptés.
Les questions à choix multiples (QCM) sont largement utilisées dans cette optique, au point de représenter une part importante des épreuves utilisées pour l'accréditation des candidats par l'*American Board of Emergency Medicine* (ABEM). Il s'agit de QCM à contexte riche, explorant non pas la simple acquisition de connaissances, mais leur application contextualisée [3]. De ce fait, elles exigent de l'étudiant un niveau cognitif élevé [3].
Bien qu'ils soient peu familiers de ce format de questions, il a été montré que les étudiants français atteignaient un niveau de performance à peine inférieur à celui d'étudiants américains de niveau d'études équivalent [4].
Récemment décrit, le test de concordance des scripts (TCS) permet d'explorer le raisonnement clinique en comparant l'utilisation que font des novices d'une information médicale à celle qu'en feraient des experts confrontés au même problème [5, 6]. Surtout, il permet d'analyser le raisonnement en contexte d'incertitude [7, 8], ce qui renforce sa pertinence en médecine d'urgence où le médecin est régulièrement confronté à des situations complexes et floues, à gérer dans un temps court et avec des moyens limités. Il repose sur la théorie des scripts comme modèle d'organisation des connaissances [9]. Il a été utilisé avec succès en chirurgie [10], gynécologie [7], radiologie [11] et urologie [12]. Ces deux aspects de la compétence clinique – application contextualisée des connaissances et raisonnement en situation d'incertitude – sont complémentaires.
De manière générale, un examen permet de transformer une performance observée en score mesurable [1]. À ce titre, il doit être valide, fidèle et réalisable [1]. Explorant le raisonnement clinique, dont la pertinence est indissociable de l'expérience clinique clinique, le TCS devrait pouvoir distinguer les cliniciens en fonction de leur niveau d'expérience. À l'inverse, ni les QCM [13], ni les épreuves rédactionnelles [14] ne permettent de distinguer régulièrement les cliniciens en fonction de leur niveau d'expérience. Enfin, explorant le même domaine de connaissances, ces deux formats devraient conduire à l'obtention de scores positivement corrélés.
Nous avons utilisé le TCS pour évaluer les apprentissages des étudiants à l'issue d'un enseignement optionnel de médecine d'urgence destiné aux résidents. Les objectifs de ce travail étaient de vérifier les caractéristiques intrinsèques du TCS (validité, fidélité, et faisabilité), de confirmer son aptitude à identifier les experts (ce qui serait à mettre au crédit de sa validité [1]) et enfin de vérifier la corrélation des scores obtenus au TCS et aux QCM.

Matériel et méthodes

L'exercice de responsabilités en service d'urgences en France est actuellement subordonné à l'obtention d'un diplôme universitaire. Le service des urgences du Centre Hospitalier Universitaire (CHU) de Nice organise depuis 1995 un enseignement optionnel de médecine d'urgence [Diplôme Universitaire de Médecine d'Urgence (DUMU)] ouvert aux résidents, d'une durée d'une année, constitué d'enseignements théoriques sous forme de séminaires mensuels, d'enseignements pratiques sous forme d'ateliers (gestes techniques, etc.) et de stages dans des unités accréditées.

Préparation des questions

Les questions utilisées pour l'examen organisé à l'issue du DUMU ont été sélectionnées à partir d'un thésaurus de situations cliniques rencontrées dans le service d'accueil des urgences du CHU de Nice, constitué en 2000 et regroupant 9 523 dossiers. Ce service possède deux unités satellites, prenant en charge directement les urgences traumatologiques et psychiatriques. Les enfants et les patientes souffrant d'affections gynéco-obstétricales sont accueillis sur un autre site. Les critères de choix des situations retenues ont été : fréquence et

1- Médecine Générale d'Urgence - Hôpital Saint Roch - CHU Nice, France
2- Département de Pédagogie Médicale - Faculté de Médecine de Nice Sophia Antipolis - France
3- Service d'Aide médicale urgente (SAMU 06) - CHU Nice - France
4- Service d'Accueil des Urgences - Hôpital de Cannes - France
5- Service d'Accueil des Urgences - Hôpital de Draguignan - France
Correspondance : Jean-Paul Fournier - Département de Pédagogie médicale - Faculté de Médecine (Université de Nice Sophia Antipolis)
28 Avenue de Valombrose - 06107 Nice cedex 2 - France - Téléphone : + (4) 93 37 77 77 - Mailto:fournier.jp@chu-nice.fr

gravité, par exemple : insuffisance cardiaque congestive, décompensation de broncho-pneumopathie chronique obstructive pour les situations fréquentes ; dissection aortique, fissuration d'anévrysme de l'aorte abdominale pour les situations graves. Elles ont été sélectionnées par un panel de quatre experts, qui ont privilégié conduites diagnostiques et thérapeutiques de façon équilibrée pour l'élaboration des QCM et du TCS.
Les QCM ont été rédigées selon les recommandations du *National Board of Medical Examiners* (NBME)[3] et n'ont inclus que des questions dites de type C (qui proposent à l'étudiant la meilleure réponse et quatre distracteurs). Elles ont été soumises à trois experts pour validation : validité scientifique, absence d'ambiguïté des intitulés, vérification du caractère incontestable de la réponse retenue. Soixante QCM, toutes indépendantes les unes des autres, ont été rédigées. Toutes avaient le même poids pour l'attribution des indices numériques contributifs au score : 1 si la bonne réponse était retenue, 0 sinon. Un exemple est fourni en *annexe 1*.

Les vignettes du TCS ont été rédigées selon les recommandations de Charlin[5, 6]. Le processus de préparation et d'établissement des scores est illustré en *annexe 2*. Brièvement, chaque vignette présente un scénario clinique court à partir duquel peuvent être formulées plusieurs hypothèses pertinentes, diagnostiques ou thérapeutiques (colonne 1). Une information supplémentaire (clinique, résultat de biologie, …) est alors disponible (colonne 2), dont l'intérêt pour l'hypothèse discutée est apprécié sur une échelle de Likert de -2 à +2 (colonne 3). Les vignettes cliniques correspondaient à des situations retenues par les quatre mêmes experts, qui ont également validé les hypothèses diagnostiques ou thérapeutiques et les informations supplémentaires fournies. Elles ont été soumises à un panel de neuf autres experts qui ont attribué individuellement à chaque information un score de -2 à +2 en fonction de leurs poids respectifs sur leur décision médicale. Ce groupe d'experts était constitué de praticiens, enseignants et/ou occupant des responsabilités dans les services d'urgences de la région depuis au moins cinq ans et était distinct du groupe de seniors défini plus loin. Le score choisi a été transformé en crédit en divisant le nombre d'experts attribuant un même score par le nombre d'experts ayant retenu le score le plus choisi (par convention on attribue un crédit de 1 à ce score). Le processus d'établissement des scores est illustré en *annexe 2*. Le score du test correspond à la somme des scores obtenus à chaque item. Ce principe correspond à l'établissement de scores combinés *(aggregate scoring method)*[15].

Le TCS a été composé de 30 situations cliniques, toutes indépendantes les unes des autres, comportant chacune trois items, soit 90 items en tout. Deux QCM et deux items du TCS ont été réécrits en raison d'une formulation imparfaite des questions. Aucun n'a été contesté pour sa valeur scientifique. Le ratio de questions diagnostiques et thérapeutiques était similaire dans les deux épreuves (0,44 vs 0,53, p = NS). La préparation de l'examen a duré trois semaines environ. Il a été administré en juin 2003. L'épreuve a duré 2h30.

Validation de l'examen

Le test par QCM a été soumis à une analyse d'items avec détermination des indices de difficulté (valeur p : proportion de participants répondant correctement à une question donnée) et de discrimination (r.bis : coefficient de corrélation bisériale de points[16]). Les items possédant un indice de difficulté ou de corrélation insuffisants ont été revus ultérieurement pour validation définitive.
La consistance interne des deux tests a été mesurée par le coefficient α de Cronbach[17] qui doit être au moins égal à 0,80 pour garantir une fidélité correcte[18].

Populations étudiées

Trois groupes de niveaux d'expérience croissante ont été définis et comparés :
– étudiants : il s'agissait de 16 étudiants volontaires venant de valider la sixième année des études médicales et désireux de s'inscrire au DUMU. Ils ont été recrutés lors de l'entretien de pré inscription en septembre 2003 ;
– résidents : il s'agissait des 20 résidents de première année inscrits au DUMU au titre de l'année universitaire 2002 – 2003 ;
– seniors : il s'agissait de neuf médecins volontaires en poste dans le service des urgences de la région ou y effectuant régulièrement des gardes de seniors. Aucun n'appartenait au groupe des experts. Six d'entre eux étaient titulaires du DUMU. Tous avaient au moins cinq années d'expérience professionnelle avec prise de responsabilités. Ils ont été placés dans les mêmes conditions de rédaction que les étudiants et les résidents.

Analyse statistique

Les scores obtenus aux deux tests par les trois catégories de praticiens – étudiants, résidents, seniors – ont été ramenés à 100. Compte tenu de la faiblesse des effectifs, l'homogénéité des variances des scores obtenus aux deux tests a été vérifiée par un test de Levene. Les scores obtenus à chacun des tests par les trois groupes ont été comparés par une analyse de variance (ANOVA) avec

correction *post hoc* par la méthode de Bonferoni-Dunn pour limiter l'effet du hasard.
La corrélation entre les scores obtenus aux QCM et au TCS a été étudiée par le test de Pearson pour les résidents auxquels est destiné cet enseignement. La significativité a été définie pour $p < 0,05$.
Les calculs ont été effectués avec le logiciel StatView 5.0® pour Macintosh®.

Résultats

Résultats du TCS

Les scores obtenus au TCS par les trois groupes étudiés sont résumés sur *le tableau 1*. Les étudiants ont obtenu les scores les plus faibles, les seniors les scores les plus élevés, les résidents se situant entre les deux. Il existait une différence très significative entre les scores obtenus par les étudiants et ceux des résidents (50,26 ± 6,29 *vs* 55,80 ± 5,31, $p = 0,0055$) et entre ceux des étudiants et des seniors (50,26 ± 6,29 *vs* 59,85 ± 4,14, $p = 0,0002$). Etudiants et résidents ont présenté la variabilité des scores la plus importante. Ces résultats sont résumés sur *la figure 1*. La consistance interne du test était très correcte, le coefficient a étant compris entre 0,85 (seniors) et 0,95 (étudiants).

Résultats des QCM

Un seul item a été éliminé à l'issue de la revue des items déficients. Le calcul du score a donc porté sur 59 items. Les scores obtenus sont résumés sur le tableau 2. Les résidents ont obtenu le score moyen le plus élevé (59,15 ± 8,93). Il n'y avait cependant aucune différence significative lors de la comparaison des scores obtenus par les trois groupes étudiés. Là encore, ce sont les étudiants et les résidents qui présentaient la plus grande variabilité des scores. Ces résultats sont représentés sur *la figure 2*. La consistance interne du test était très correcte, le coefficient a se situant entre 0,92 et 0,96 *(tableau 2)*.

Corrélation des scores au TCS et aux QCM

Les scores obtenus au TCS et aux QCM par les résidents variaient dans le même sens, sans que l'on puisse mettre en évidence une corrélation significative sur *la figure 3* ($R2 = 0,0164$; $p = 0,5905$).

Tableau 1 :
Scores obtenus par les différentes populations évaluées au test de concordance de scripts

	Effectif	Moyenne	Maximum	Minimum	SD	α
Etudiants	16	50,26	56,76	29,54	6,29	0,95
Résidents	20	55,80	68,83	47,69	5,31	0,87
Seniors	9	59,85	67,38	53,01	4,14	0,85

SD : déviation standard - α : coefficient de Cronbach

Figure 1 :
Scores (sur 100) obtenus par les différentes populations au test de concordance des scripts

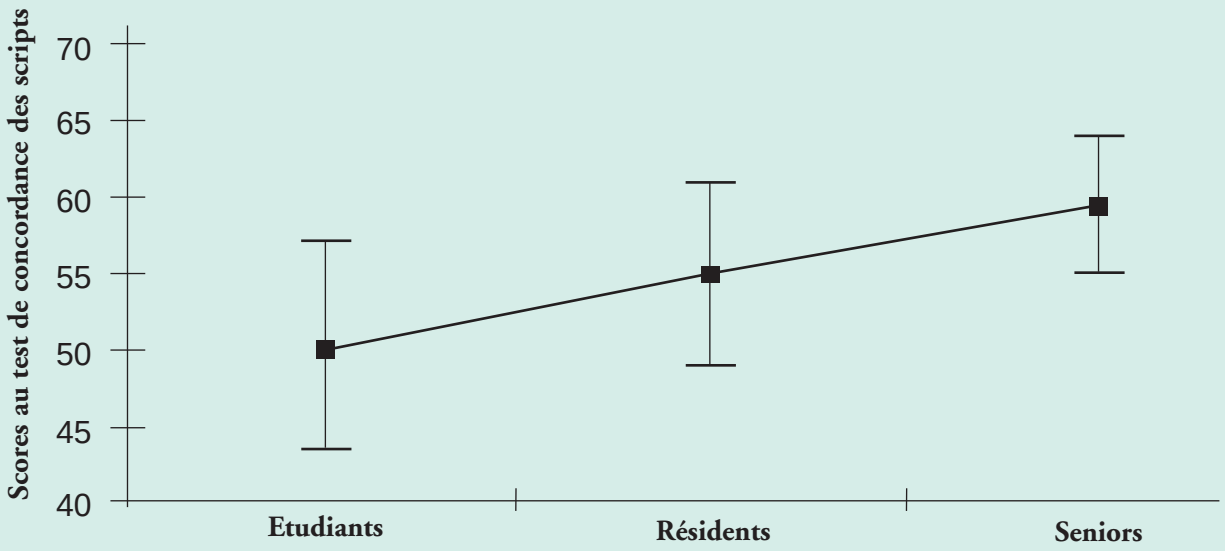


Tableau 2 :
Scores obtenus par les différentes populations évaluées aux questions à choix multiple

	Effectif	Moyenne	Maximum	Minimum	SD	α
Etudiants	16	54,66	69,49	37,29	10,75	0,92
Résidents	20	59,15	74,58	44,07	8,93	0,96
Seniors	9	58,19	64,41	50,85	5,08	0,93

SD : déviation standard - α : coefficient de Cronbach

Figure 2 :
Scores (sur 100) obtenus par les différentes populations aux questions à choix multiple

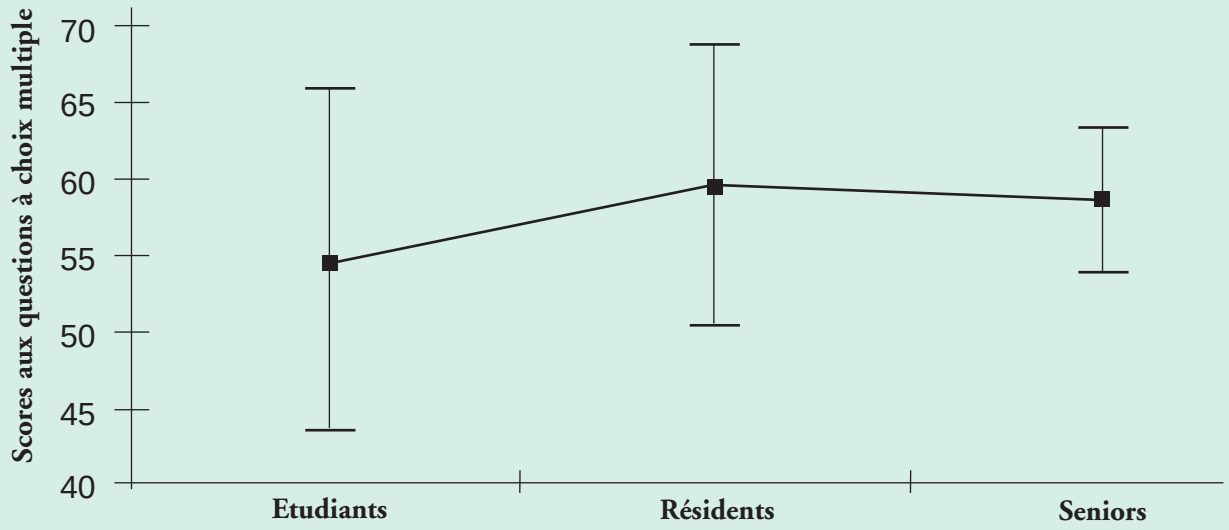
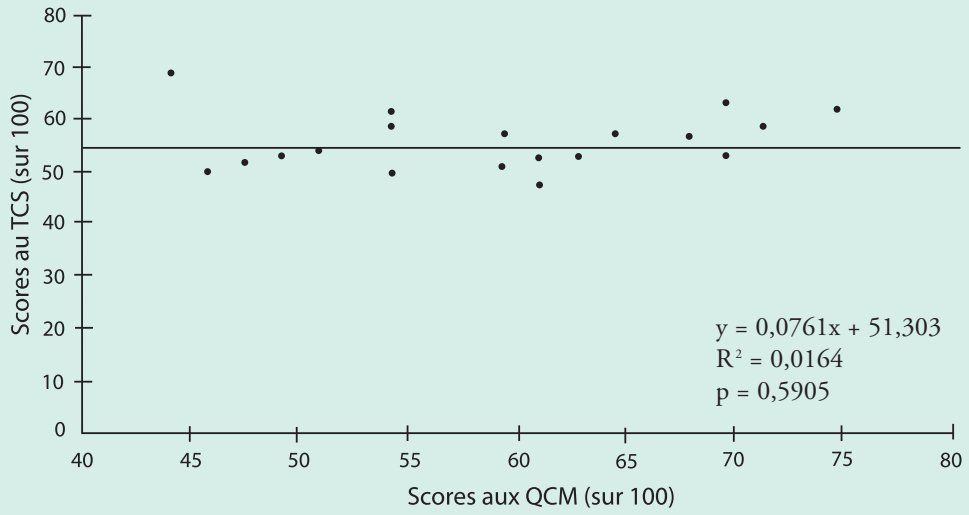


Figure 3 :
Analyse de la corrélation des scores (sur 100) obtenus par les résidents au test de concordance des scripts (TCS) et aux questions à choix multiples (QCM)



Discussion

Les QCM n'ont pu distinguer le niveau d'expérience clinique des trois catégories de cliniciens évalués. Comme on pouvait s'y attendre, ce sont les résidents, qui venaient de bénéficier de l'enseignement, qui ont obtenu les scores les plus élevés. Les étudiants ont eu les scores les plus faibles, sans que l'on puisse mettre en évidence une différence significative entre leurs résultats et ceux des résidents et des seniors. Étudiants et résidents présentaient la variabilité la plus élevée des scores. Ce point, joint à la faiblesse des effectifs, peut expliquer que ces différences n'aient pas été significatives. Les seniors ont obtenu des scores intermédiaires. Là encore, ces résultats étaient attendus, les QCM[13], comme d'ailleurs les questions rédactionnelles[14], pouvant ne pas distinguer les seniors des juniors (étudiants, résidents).

À l'inverse, le TCS a pu distinguer les trois catégories de cliniciens, en fonction de leur degré d'expérience, les étudiants obtenant les scores les plus faibles, les seniors les scores les plus élevés et les résidents des scores intermédiaires. Il existait une différence significative entre les scores obtenus d'une part, par les étudiants et les résidents et, d'autre part, entre ceux des résidents et des seniors. Ces résultats confirment les résultats observés dans d'autres disciplines spécialisées[7, 10 – 12].

L'examen a été soigneusement construit, en termes de validités et de fidélité. Il a été préparé en trois semaines environ par un seul enseignant. Les QCM utilisées dans ce travail ont été rédigées à partir d'un corpus de situations qu'auront à gérer les médecins en poste dans un service d'urgences. Elles ont été rédigées selon les recommandations du NBME[3], revues par trois experts pour leur contenu scientifique et la précision de leur intitulé. Enfin, elles ont fait l'objet d'une analyse d'item à l'issue de l'épreuve. De fait, elles étaient conformes aux normes du NBME[19]. Ce point garantit leur validité apparente[1]. Le nombre des questions assure la validité de contenu de l'épreuve[1]. Enfin, elles sont conformes à l'exercice futur des médecins formés, ce qui conditionne la validité de construit de l'épreuve[1, 20]. Les coefficients alpha de Cronbach (0,92 et 0,96 selon les groupes) garantissent la consistance interne de l'épreuve[18]. Les questions retenues pour le TCS ont été construites à partir du même corpus de situations cliniques, sélectionnées par les mêmes experts. Ils ont revu les vignettes cliniques et les items retenus pour en garantir la validité scientifique et l'absence d'ambiguïté. Les différents niveaux d'ancrage des échelles de Likert (*voir annexe 2*) ont été tirés de l'article de Charlin[6]. De ce fait, la validité apparente de cette épreuve est atteinte. La validité de contenu est garantie par le nombre de questions utilisées[1]. Enfin, les coefficients alpha de Cronbach obtenus (0,85 à 0,95 selon les groupes) garantissent la consistance interne de l'épreuve[18]. Les deux épreuves apparaissaient donc valides et fidèles. Elles exploraient le même corpus de connaissances ; il paraissait donc intéressant de les comparer.

Les scores obtenus aux deux épreuves ont été comparés chez les résidents auxquels cet enseignement est destiné. Il aurait été logique d'observer une corrélation, même faible, entre les scores obtenus aux deux épreuves. En effet, utilisation de connaissances factuelles en situation clinique (QCM) et raisonnement clinique (TCS) portant sur le même corpus de situations sont *a priori* liés. Les scores variaient dans le même sens, mais il n'a pas été possible de mettre en évidence la corrélation statistiquement significative attendue entre leurs valeurs. L'effectif limité du groupe peut certes expliquer ces résultats. Surtout, la revue des questions utilisées pour les deux épreuves a montré que, si elle étaient bien issues du même corpus de situations cliniques, avec la même répartition entre situations de prise de décisions diagnostiques ou thérapeutiques, elles n'étaient pas identiques en termes de répartition d'une spécialité à l'autre (métabolique, gastroentérologie, etc.). Il est possible que, si les questions avaient exploré exactement le même champ de connaissances, la corrélation ait été statistiquement significative.

Le TCS permet de mesurer le degré d'organisation des connaissances et leur niveau d'élaboration. Il vise à mesurer l'adéquation des liens au sein des connaissances cliniques bien plus que la simple présence d'éléments de connaissances[6, 9]. Un des avantages fondamentaux du TCS est qu'il intègre le contexte d'incertitude dans la prise de décision, qu'elle soit diagnostique ou thérapeutique[7, 8]. L'appréciation de l'incertitude ou de l'absence de consensus est rendue possible par le mode de calcul des scores faisant appel aux scores composites[7]. Le principe de tels scores est de pondérer la réponse (ce qui correspond ici à l'échelle de Likert) et d'obtenir le score final en additionnant les scores de chaque item[15]. Une telle approche permet en outre de limiter l'erreur de mesure[15] et de mieux distinguer experts et novices[7].

La gestion de l'incertitude est commune en médecine d'urgence où l'on traite régulièrement des situations floues, correspondant souvent à des syndromes ou des symptômes (douleur thoracique, dyspnée,etc.) plutôt que des entités cliniques bien définies (infarctus du myocarde, pneumonie communautaire, etc.), dans un temps et avec des moyens limités. La gestion de l'incertitude en clinique explique en partie le peu d'entrain des cliniciens à appliquer les recommandations de bonne pratique clinique[21]. Cette situation n'est pas exceptionnelle

en médecine d'urgence où l'on observe par exemple près de 30 % d'inadéquation avec les recommandations de prise en charge diagnostique de l'embolie pulmonaire[22] ou plus de 40 % d'inadéquation de la prescription de l'antibiothérapie lors de la prise en charge des pneumonies communautaires[23]. À ce titre, le TCS constitue donc un outil d'évaluation conceptuellement séduisant en médecine d'urgence et, au-delà, aux autres domaines médicaux.

Une autre limite importante de ce travail est représentée par la taille et la composition des groupes de seniors, de résidents et d'étudiants. Le groupe des résidents est, sans doute, représentatif des candidats à cet enseignement qui sont tous des résidents de première année, depuis que cet enseignement existe. En revanche, les groupes de seniors et d'étudiants ont été constitués de volontaires qui, eux, ne sont pas forcément représentatifs des seniors exerçant des responsabilités en service d'urgence ou des étudiants en fin de cursus. Le type même du recrutement du service d'urgence du CHU de Nice a limité les questions portant sur des situations et problèmes psychiatriques, traumatologiques, pédiatriques ou gynécologiques, pas ou peu représentés dans le thésaurus qui a servi de base à la rédaction des questions utilisées pour les deux épreuves.

À ces restrictions prés, nos résultats suggèrent une complémentarité entre QCM à contexte riche et TCS pour l'évaluation d'un enseignement de médecine d'urgence. Ce point est particulièrement intéressant au moment où la médecine d'urgence devient une discipline spécialisée autonome en France. D'autres tests comme les stations d'examen clinique objectif structuré (ECOS) permettent d'évaluer des aspects plus techniques ou relationnels de la compétence clinique[2]. Ils sont également complémentaires des TCS[10], dans une approche globale de l'évaluation de l'acquisition de la compétence clinique en enseignement médical[24].

Remerciements

Nous remercions le Pr Bernard Charlin d'avoir accepté la relecture de la première version du manuscrit. Nous remercions les étudiants et les médecins qui ont participé à cette étude.

Contributions

Jean-Paul Fournier a conçu l'étude, rédigé les questions, corrigé l'examen et rédigé le manuscrit. Didier Thiercelin, Élise Gilbert, Jean-Marc Minguet et François Bertrand ont sélectionné les situations cliniques, revu et critiqué les questions. Céline Pulcini et Véronique Alunni-Perret ont revu le manuscrit.

Annexe 1 : Exemple de QCM à contexte riche

Une femme de 24 ans se présente aux urgences pour céphalées. Elles sont apparues brutalement, de siège rétro-orbitaire, avec des nausées et un épisode de vomissement. Le bruit et la lumière sont très pénibles. Elle a pris sans succès 6 gélules de paracétamol (Dafalgan®). Elle vient de commencer un traitement par gestodène, éthinylestradiol (Méliane®). Ses seuls antécédents notables sont constitués par des sinusites maxillaires à répétition et une allergie au triméthoprime sulfaméthoxazole (Bactrim®). Elle est actuellement traitée par roxythromycine (Rulid®). La température est à 37,8°C. L'examen neurologique est normal. La nuque est souple. Quelle est la prochaine étape de la prise en charge ?

A- Injection de kétoprofène (Profénid®)
B- Injection de morphine
C- Injection de paracétamol (Perfalgan®)
D- Perfusion de dihydro ergotamine
E- Pulvérisation nasale de sumatriptan (Imigrane®)

Annexe 2 :
Réalisation d'un test de concordance de scripts

Rédaction de la vignette clinique et des items (étape 1), soumission à neuf experts qui attribuent une note de -2 à +2 à la nouvelle information (seconde colonne) en fonction du poids qu'ils lui accordent (étape 2). La note est transformée en crédit pour l'item en divisant le nombre d'experts ayant choisi une valeur particulière par le nombre d'experts ayant choisi la valeur la plus sélectionnée (étape 3) : pour I1 : -2 est la valeur la plus choisie, par 4 experts : -2 donne un crédit de 1 (1 = 4/4), -1 est choisie par 3 experts, elle donne un crédit de 0,75 (0,75 = 3/4), 0 est choisie par 2 experts, elle donne un crédit de 0,5 (0,5 = 2/4). 1 et 2 n'ont pas été choisies et donnent donc un crédit de 0. Les étudiants (R1 à R8) répondent aux mêmes questions : leurs choix de valeurs (étape 4) est transformé en score (étape 5) à partir du score établi par les experts. Le score total est la somme du score de chaque item.

Étape 1 : rédaction de la vignette clinique et des trois items (I1 à I3) :
Une patiente de 65 ans est adressée par une clinique pour suspicion d'embolie pulmonaire. Elle a des antécédents de diabète, d'insuffisance cardiaque et de broncho-pneumopathie chronique obstructive.

Si vous pensiez faire *(option d'examen complémentaire)*	**Et qu'alors vous trouvez** *(nouvelle information obtenue par examen clinique ou exam en complémentaire)*	**L'effet sur la nécessité de demander ce test est le suivant**
Un angioscanner thoracique (I1)	Un traitement par metformine (Glucophage retard®)	-2 -1 0 +1 +2
Un dosage des D dimères (I2)	Un cancer de l'ovaire en cours de traitement	-2 -1 0 +1 +2
Un écho-Doppler veineux (I3)	Un signe de Homans	-2 -1 0 +1 +2
-2 : absolument contre-indiqué *-1 : peu utile ou plutôt néfaste*	*0 : non pertinent dans cette situation*	*+1 : utile et souhaitable* *+2 : indispensable*

Étape 2 : attribution des scores par les experts : nombre d'experts ayant respectivement attribué chacune des notes pour chaque item

		-2	-1	0	+1	+2
Q1	I1	4	3	2	0	0
	I2	0	1	4	3	1
	I3	0	1	4	3	1

Étape 3 : attribution du crédit pour l'item

		-2	-1	0	1	2
Q1	I1	1	0,75	0,5	0	0
	I2	0	0,25	1	0,75	0,25
	I3	0	0,25	1	0,75	0,25

Étape 4 : établissement du score des candidats (1)

		R1	R2	R3	R4	R5	R6	R7	R8
Q1	I1	-2	2	-1	0	-2	-2	-2	2
	I2	2	1	1	1	2	2	0	2
	I3	2	2	1	1	-1	0	2	2

Étape 5 : établissement du score des candidats (2)

		R1	R2	R3	R4	R5	R6	R7	R8
Q1	I1	1	0	0,75	0,5	1	1	1	0
	I2	0,25	0,75	0,75	0,75	0,25	0,25	1	0,25
	I3	0,25	0,25	0,75	0,75	0,25	1	0,25	0,25
Total		1,5	1,0	2,25	2,0	1,50	2,25	2,25	0,5

Références

1. *Jean P, Des Marchais J, Delorme J. Apprendre à enseigner les sciences de la santé. Cahier 4, Sherbrooke: Université De Sherbrooke, 1993.*
2. *Epstein RM, Hundert EM. Defining and assessing professional competence. JAMA 2002;287:226-35.*
3. *Case SM, Swanson DB. Constructing written test questions for basic and clinical sciences 3rd edition, Philadelphia: The National Board of Medical Examiners, 2001.*
4. *Fournier JP, De Champlain AF, Benchimol D, et al. Intérêt de transposer un examen américain dans le cadre du futur examen national classant validant français. Ann Med Interne 2003;154:148-56.*
5. *Charlin B, Roy L, Brailowsky C, Goulet F, Van Der Vleuten C. The script concordance test : a tool to assess the reflective clinician. Teach Learn Med 2000;4:189-95.*
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. *Charlin B, Desaulniers M, Gagnon R, Blouin D, Van Der Vleuten C. Comparison of an aggregate scoring method with a consensus scoring method in a measure of clinical reasoning capacity. Teach Learn Med 2002;14:150-6.*

8. *Charlin B, Van Der Vleuten C. Standardized assessment in context of uncertainty : the script concordance approach. Evaluation and the Health Professions 2004;27:304-19.*
9. *Charlin B, Tardif J, Boshuizen HPA. Scripts and medical knowledge : theory and applications for clinical reasoning instruction and research. Acad Med 2000;75:182-90.*
10. *Brailowsky C, Charlin B, Beausoleil S, Cote S, Van Der Vleuten C. Measurement of clinical reflective capacity early in training as a predictor of clinical reasoning performance at the end of residency : an experimental study on the script concordance test. Med Educ 2001;35:430-6.*
11. *Brazeau-Lamontagne L, Charlin B, Gagnon R, Samson L, Van Der Vleuten C. Measurement of perception and interpretation skills along radiology training : utility of the script concordance approach. Med Teach 2004;26:326-32.*
12. *Sibert L, Charlin B, Corcos J, Gagnon R, Grise P, Van Der Vleuten C. Stability of clinical reasoning assessment results with the script concordance test accross two different linguistic, cultural and learning environments. Med Teach 2002;24:522-7.*
13. *Van Der Vleuten C. PM The assessment of professional competence : development, research, and practical implications. Adv Health Sci Educ Theory Pract 1996;1:41-67.*
14. *Newble DI, Hoare J, Baxter A. Patient management problems: issues of validity. Med Educ 1982;16:137-42.*
15. *Norcini JJ, Shea JA, Day SC. The use of aggregate scoring for a recertifying examination. Evaluation and the Health Professions 1990;13:241-51.*
16. *Crocker L, Algina J. Introduction to classical and modern test theory. New York: Holt Reinhard and Winston, 1986:3-39.*
17. *Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297-334.*
18. *Swanson DB, Case SM. Trends in written assessment : a strangely biased perspective. In : Harden RM, Hart IR (Eds.) Approaches of the assessment of clinical competence. Norwich: Page Brothers, 1992:38-53.*
19. *Swanson DB, Case SM. Variation of item difficulties and discrimination by item format on part I (basic sciences) and part II (clinical sciences) of Us Licensing Examination. In Rothman Ai, Cohen R Eds. Proceedings of the 6th Ottawa Conference on Medical Education, Toronto, University of Toronto Bookstore Custom Publishing, 1995:285-7.*
20. *Laduca A. Validation of professional licensure examination. profession theory, test design, and construct validity. Evaluation and The Health Professions 1994;17:178-97.*
21. *Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PAC, Rubin HR. Why don't physician follow clinical practice guidelines ? JAMA 1999;282:1458-65.*
22. *Chagnon I, Bounameaux H, Aujesky D, et al. Comparison of two clinical prediction rules and implicit assessment among patients with suspected pulmonary embolism. Am J Med 2002;113:269-75.*
23. *Halm EA, Atlas SJ, Borowsky LH, et al. Understanding physicians adherence with a pneumonia practice guideline. Arch Intern Med 2000;160:98-104.*
24. *Jouquan J. L'évaluation des apprentissages des étudiants en formation médicale initiale. Pédagogie Médicale 2002;3:38-52.*

Manuscrit reçu le 13 juillet 2005 ; commentaires éditoriaux formulés aux auteurs le 9 novembre 2005 ; accepté pour publication le 3 janvier 2006.