

Intérêt des tests de concordance de script pour évaluer le raisonnement et l'organisation des connaissances des étudiants de quatrième année des études médicales en France

Significance of script concordance tests to assess reasoning and knowledge organization among 4th year medical school students in France

Stéphane JOUNEAU, Régis LURAINÉ et Benoît DESRUES

Service de pneumologie, Hôpital Pontchaillou, Centre hospitalier universitaire de Rennes, Université de Rennes 1, France

Manuscrit reçu le 10 novembre 2011 ; commentaires éditoriaux formulés aux auteurs le 5 avril et le 12 juillet 2012 ; accepté pour publication le 13 juillet 2012

Mots-clés

Tests de concordance de script ; pneumologie ; DCEM2, cas clinique ; grille de correction ; évaluation

Résumé – Contexte : L'examen écrit de pneumologie des étudiants de deuxième année du deuxième cycle des études médicales (DCEM2) à la faculté de médecine de Rennes est constitué d'un cas clinique (CC). **But :** L'introduction des tests de concordance de script (TCS), qui mesurent l'interprétation des données cliniques en contexte d'incertitude, pourrait améliorer l'évaluation du raisonnement clinique des étudiants. **Matériel et méthodes :** Les étudiants du deuxième semestre de DCEM2 ($n = 85$) ont été évalués par un CC (2/3 du score) associé à 20 items de TCS (1/3 du score). La fidélité des TCS a été évaluée par le coefficient α de Cronbach. Les scores des étudiants du CC et des TCS ont été corrélés (test de Spearman). Le degré d'organisation des réponses au CC a été évalué. **Résultats :** Le coefficient α de Cronbach était de 0,35 pour les étudiants. Les scores des étudiants aux TCS étaient significativement corrélés à ceux du CC ($r = 0,304$, $p = 0,005$). Les étudiants avec de bons scores aux TCS avaient des réponses au CC plus hiérarchisées et plus ordonnées que les étudiants avec de mauvais scores aux TCS qui avaient tendance à écrire des réponses plus longues avec une hiérarchisation absente ou plus floue (évaluation qualitative globale). Les étudiants aux réponses les plus hiérarchisées avaient des notes plus élevées comparés aux copies les moins hiérarchisées : 16,3 *versus* 11,8 ($p < 0,0001$) pour les CC et 12,5 *versus* 10,6 ($p = 0,021$) pour les TCS. Les TCS semblent toujours être bons chez les étudiants aux connaissances bien organisées. Les étudiants qui ont effectué leur stage pratique dans le service de pneumologie avaient un score moyen aux TCS supérieur à ceux qui n'étaient pas passés dans le service : 12,6 *versus* 11,3, $p = 0,008$. Cette différence n'était pas retrouvée pour le CC. **Conclusions :** Il semble que l'introduction des TCS améliore l'évaluation du raisonnement clinique des étudiants de DCEM2 lors de l'examen écrit de pneumologie en association avec le CC.

Keywords

Script concordance tests; respiratory medicine; 4th year of medical school; clinical case; assessment grid; clinical reasoning assessment

Abstract – Context: The written examination for respiratory medicine students (DCEM2) in the 4th year at Rennes university medical school is made up of a clinical case (CC). **Goal:** The introduction of script concordance tests (SCT) could enhance the assessment of students' clinical reasoning. **Material and methods:** During the 2nd term, DCEM2 students ($n = 85$) were assessed using a CC (2/3 of score) and 20 SCT items (1/3 of the score). Reliability was determined by a Cronbach coefficient computation. The students' results on CC and SCT were established using a Spearman rank correlation test. The organization of CC answers was assessed. **Results:** Cronbach coefficient was 0.35 for the students. Student scores in SCT were significantly correlated to the results in CC ($r = 0.304$, $p = 0.005$). The students with good scores in SCT had shorter and more organized answers to the CC than students with poor scores in SCT who tended to have longer and less organized answers to the CC (qualitative global assessment). Students with highly organized CC answers had better scores compared to those with poorly organized scripts: 16.3 *versus* 11.8 ($p < 0.0001$) for CC and 12.5 *versus* 10.6 ($p = 0.021$) for SCT. Students with highly organized knowledge seem to always have good results with SCT. Mean SCT score for students who worked in the respiratory department was significantly higher than that of students who had not worked in the ward: 12.6 *versus* 11.3, $p = 0.008$. This difference was not demonstrated with the CC. **Conclusions:** In our opinion, SCT associated to CC seems to enhance the clinical learning assessment of DCEM2 students in their written examination for respiratory medicine.

Introduction

L'évaluation de l'apprentissage des étudiants est un des éléments essentiels du processus enseignement-apprentissage. L'évaluation écrite semestrielle des étudiants en deuxième année du deuxième cycle des études médicales (DCEM2) à la faculté de médecine de Rennes (France) est constituée d'un cas clinique (CC), corrigé à l'aide d'une « grille de correction », proche de ceux des épreuves classantes nationales (ECN). Toutefois, cette grille de correction présente clairement des failles. À la lecture de certaines copies, l'organisation des réponses démontre de bonnes connaissances qui sont bien hiérarchisées, allant de l'urgent au moins urgent et du plus important au moins important. À l'inverse, dans d'autres copies, les réponses sont désordonnées et mal hiérarchisées, les réponses des étudiants étant surtout dictées par le but implicite de contenir un maximum des « mots-clés » pris en compte dans la grille. Malheureusement, l'écart des scores attribués respectivement à ces deux types d'étudiants à l'issue de l'application du barème de correction n'est parfois que très faible.

Il serait donc utile d'ajouter à ce CC, une autre méthode d'évaluation du raisonnement clinique. Les

procédés qui permettent d'avoir accès à l'organisation des connaissances et qui font appel à leur réutilisation pertinente plutôt qu'à leur récitation sur commande sont cohérents avec cette intention^[1]. Selon Charlin et al., les outils de mesure du raisonnement clinique existant à l'heure actuelle présentent un défaut majeur : les médecins possédant plusieurs années d'expérience obtiennent des scores à peine meilleurs et parfois inférieurs à ceux des médecins moins expérimentés ou des étudiants dans les simulations de résolutions de problèmes cliniques^[2,3]. Les tests de concordance de script (TCS) sont conçus pour mesurer la capacité à résoudre des problèmes mal définis et à raisonner en contexte d'incertitude^[2]. La théorie des scripts postule qu'en situation clinique, les médecins mobilisent des réseaux de connaissances préétablis, des scripts, qui sont utilisés pour comprendre cette situation et agir en fonction de buts diagnostiques, d'investigation ou de thérapeutique. L'organisation et la hiérarchisation des connaissances sont des éléments importants et le CC classique par sa méthode de cotation, permet mal d'évaluer ces dimensions. Nous avons donc créé des TCS associés à un CC classique pour évaluer les étudiants de DCEM2 lors de l'examen écrit du deuxième semestre. Nos objectifs étaient : 1) de vérifier qu'il était possible

d'introduire des TCS dans les conditions de l'examen écrit de pneumologie des étudiants de DCEM2 dans notre faculté et 2) de montrer que les TCS mesureraient mieux l'impact de l'organisation des connaissances par rapport au CC.

Matériels et méthodes

Population

Les étudiants du deuxième semestre de DCEM2 ($n = 85$) ont satisfait aux exigences d'évaluation des apprentissages requises pour la validation de la séquence de stage au sein du pôle lors d'un examen écrit en mai 2009, comportant notamment un CC de pneumologie (demandé par la faculté) ainsi que 20 questions de pneumologie au format TCS. Les étudiants bénéficiaient d'environ 45 min pour répondre au CC et aux TCS. La pondération des scores était respectivement de 2/3 pour le CC et 1/3 pour les TCS. Cet examen écrit de pneumologie s'intégrait dans une évaluation de fin de semestre du pôle « spécialités médicales », d'une durée totale de trois heures et comprenant aussi trois autres CC – de cardiologie, d'endocrinologie-nutrition et de dermatologie. De ce fait la durée consacrée à l'épreuve de pneumologie pouvait être fluctuante.

Cas clinique (CC)

Il s'agit d'une situation clinique décrite dans un énoncé initial, plus ou moins complexe, suivi de plusieurs questions rédactionnelles, parfois associées à de nouvelles informations, qui évaluent les différentes dimensions des connaissances médicales (diagnostic, choix des examens complémentaires et thérapeutique).

Le thème abordé lors de cet examen était celui de la bronchopneumopathie chronique obstructive (BPCO). Il ne s'agissait pas d'un cas pouvant faire appel à des connaissances dites transversales (c'est-à-dire concernant d'autres spécialités que la pneumologie).

Tests de concordance de script (TCS)

– Construction des TCS

Vingt items répartis sur six vignettes de TCS ont donc complété le CC et couvraient de nombreux chapitres de la pneumologie. Ces TCS ont été élaborés comme décrit dans la littérature (figure 1 : exemple d'une vignette de TCS)^[2,3]. Ils ont été établis par un binôme d'enseignant (S.J. et B.D.). Ces TCS analysaient successivement les différents objectifs d'évaluation : diagnostic (2×4 items), investigation (2×3 items) et traitement (2×3 items). 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'impact sur l'hypothèse discutée est appréciée sur une échelle de Likert de -2 à $+2$ (colonne 3). Les vignettes ont été soumises à un panel de 13 experts qui ont attribué individuellement à chaque information un score de -2 à $+2$ en fonction de leur impact respectif sur leur décision médicale^[4]. Ce groupe d'experts était constitué de pneumologues, praticiens du service de pneumologie du CHU de Rennes : 2 professeurs d'université – praticien hospitalier (PU-PH), 6 praticiens hospitaliers (PH), 2 chefs de cliniques – assistants des hôpitaux (CCA), un pneumologue libéral et 2 internes enfin de cursus. 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 score du test correspond à la somme des scores obtenus pour chaque item. Ce principe correspond à l'établissement de scores combinés (*aggregate scoring method*)^[5]. Ces TCS ont été vérifiés en utilisant la grille de qualité adaptée par Caire^[6].

– Séances de préparation

Les étudiants ont bénéficié d'une séance d'une heure, par groupe de 20 à 22. Ils ont été informés par écrit *via* le « guide de l'étudiant » remis à chacun en début de semestre, à la fois de cette modalité particulière d'évaluation pour la pneumologie

Un homme de 55 ans, sans antécédent particulier, consulte pour une fièvre à 38,5 °C, une douleur basithoracique droite et une dyspnée, symptômes évoluant depuis 24 heures.		
<u>Si vous évoquez (option diagnostique)</u>	<u>Et que vous trouvez (nouvelle information)</u>	<u>Cette nouvelle donnée rend votre hypothèse</u>
Une pleuro-pneumopathie bactérienne	Un foyer de râles crépitants en base droite	-2 -1 0 +1 +2
Une pleuro-pneumopathie bactérienne	Une radiographie thoracique normale	-2 -1 0 +1 +2
Une embolie pulmonaire	Des D-Dimers à 0,8 (normale inf. à 0,5)	-2 -1 0 +1 +2
Une embolie pulmonaire	Une expectoration purulente	-2 -1 0 +1 +2
<u>Entourez la proposition qui vous semble adéquate :</u> -2 : quasi improbable -1 : moins probable 0 : l'information n'a aucun effet sur l'hypothèse +1 : plus probable +2 : quasi certaine		

Fig. 1. Exemple de la première vignette diagnostique des tests de concordance de script de l'examen écrit de pneumologie des étudiants en DCEM2.

et aussi de la tenue de ces séances obligatoires d'enseignement dirigé. Cependant la présence à ces séances n'a été contrôlée. Lors de ces séances, les principes et buts de cette méthode d'évaluation étaient expliqués aux étudiants et ils répondaient à six questions (sur deux cas différents) au format TCS en guise « d'entraînement ».

– Validation des TCS

La cohérence interne des TCS a été mesurée par le coefficient α de Cronbach (logiciel SPSS).

– Évaluation du degré d'organisation des réponses

Le degré d'organisation des réponses au CC a d'abord été évalué par un des auteurs (S.J.) en lisant les copies des étudiants ayant obtenu les meilleurs et les plus mauvais scores aux TCS ou au CC ($n = 20$). Cette première évaluation était purement qualitative et globale, ne s'appuyant sur aucun barème. Secondairement, le degré d'organisation des réponses au CC a été mesuré par deux autres évaluateurs (R.L. et B.D.) pour toutes les copies ($n = 85$), en aveugle des scores obtenus au CC et aux TCS et en aveugle des autres évaluateurs. Cette dimension organisation – hiérarchie a

été cotée de 1 (faible) à 4 (forte), ce qui a conduit à déterminer deux catégories : étudiants aux réponses bien organisées (3–4) et étudiants aux réponses mal organisées (1–2). Nous avons comparé les scores des TCS et du CC des étudiants de ces deux catégories.

– Statistiques

Les scores sur 20 sont présentés sous forme de moyenne \pm écarts-type suivis du score minimal et maximal entre crochets. La distribution des scores des étudiants au CC comme aux TCS n'étant pas normale (test de Kolmogorov-Smirnov), nous avons donc utilisé des tests non paramétriques pour nos analyses. Pour comparer les scores obtenus aux TCS par les experts et ceux obtenus par les étudiants, ainsi que les scores d'« organisation-hiérarchie » des étudiants, nous avons utilisé un test de Mann et Whitney. Pour rechercher une corrélation entre les scores obtenus au CC et les scores des TCS, nous avons utilisé le test de Spearman. Un $p < 0,05$ était considéré comme statistiquement significatif. Les calculs ont été effectués avec le logiciel GraphPad Prism 5[®].

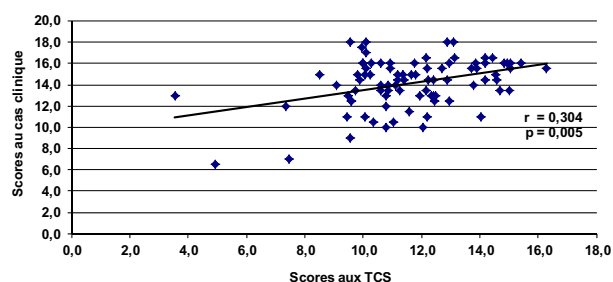


Fig. 2. Les scores obtenus aux tests de concordance de script (TCS) étaient significativement corrélés à ceux obtenus au cas clinique (CC ; $r = 0,304$, $p = 0,005$).

Résultats

Le coefficient α de Cronbach était de 0,35 pour les TCS des étudiants.

La moyenne obtenue par les étudiants de DCEM2 au CC était de $14,2 \pm 2,3$ [6,5–18,0]. La moyenne des scores obtenus aux TCS par les experts était de $16,7 \pm 1,6$ [13,5–18,9]. La moyenne obtenue aux TCS par les étudiants était de $11,6 \pm 2,2$ [3,6–16,3]. Aucun étudiant n'a failli à l'examen par le seul résultat aux TCS.

Pour deux items de TCS, le panel d'expert a répondu de manière unanime mais les étudiants n'ont eu que respectivement 50 %, 30 % et 60 % de bonnes réponses. Quatre autres items retrouvaient seulement un des experts en désaccord avec le reste du panel, mais il s'agissait d'un expert différent à chaque item.

Les scores des étudiants obtenus aux TCS étaient significativement corrélés à ceux obtenus au CC ($r = 0,304$, $p = 0,005$, figure 2).

Nous avons constaté que certains étudiants avaient de bons scores au CC mais de mauvais aux TCS et inversement. Nous avons donc relu (évaluation globale) les copies des étudiants ayant obtenu les meilleurs et les plus mauvais scores aux TCS ou au CC ($n = 20$). Pour les étudiants ayant eu de bons scores au CC et des scores nettement moins bons aux TCS, on constatait que ces étudiants avaient tendance à remplir au maximum l'espace de réponse. Pour les étudiants qui avaient de bons scores aux deux types d'évaluations on remarquait qu'ils avaient des réponses au CC très hiérarchisées. Pour

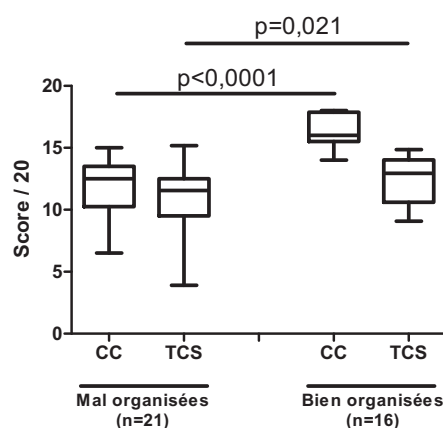


Fig. 3. Comparaison des résultats au cas clinique (CC) et aux tests de concordance de script (TCS) des étudiants avec les moins organisées (score moyen d'organisation – hiérarchie < 2 , $n = 21$) aux étudiants avec les copies les plus organisées (score moyen d'organisation – hiérarchie > 3 , $n = 16$).

les étudiants ayant eu de mauvais scores aux deux examens, les étudiants n'ont pas cherché à remplir la place libre par le maximum de mots. Par contre, on notait une tendance à la récitation de parties de leurs cours, non adaptée à la question posée.

Pour les deux étudiants ayant obtenu les scores les plus bas aux TCS (3,6 et 4,95/20), ils n'avaient répondu qu'à la moitié des TCS. On notait que les items non répondus étaient les 10 derniers.

Nous avons ensuite comparé les scores obtenus au CC et aux TCS des étudiants aux réponses mal organisées (score moyen d'organisation-hiérarchie < 2 , $n = 21$) aux étudiants aux réponses bien organisées (score moyen d'organisation-hiérarchie > 3 , $n = 16$). Les étudiants aux réponses bien organisées avaient un meilleur score aux deux types d'évaluations comparés aux copies mal organisées (figure 3) : 16,3 *versus* 11,8 ($p < 0,0001$) pour les CC et 12,5 *versus* 10,6 ($p = 0,021$) pour les TCS.

Enfin, les étudiants qui étaient passés en stage pratique dans le service de pneumologie ($n = 28$) avaient un score moyen aux TCS supérieur à ceux qui n'étaient pas passés dans le service ($n = 57$) : 12,6 *versus* 11,3, $p = 0,008$ (figure 4). Cette différence n'était pas retrouvée pour le CC.

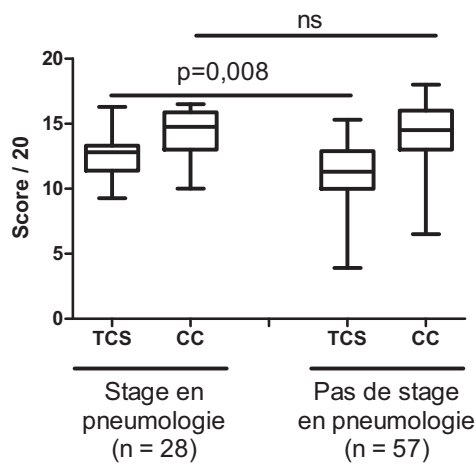


Fig. 4. Comparaison des résultats au cas clinique (CC) et aux tests de concordance de script (TCS). Les étudiants qui sont passés en stage pratique dans le service de pneumologie ($n = 28$) ont un score moyen aux TCS supérieur à ceux qui ne sont pas passés dans le service ($n = 57$). Cette différence n'est pas retrouvée pour le CC.

Discussion

Dans le but d'améliorer l'évaluation du raisonnement clinique de nos étudiants, nous avons introduit des TCS lors de l'examen écrit de pneumologie des étudiants de DCEM2, en association avec un CC. Les scores entre le CC et les TCS sont significativement corrélés. À la relecture des copies des étudiants ayant obtenu les meilleurs et les plus mauvais scores aux TCS ou au CC en évaluation globale, il semblerait que les TCS soient plus sensibles pour discriminer les étudiants avec une bonne organisation de leurs connaissances. Toutefois, ce bénéfice des TCS sur le CC n'apparaissait pas lors de la mesure par deux évaluateurs indépendants de la dimension « organisation – hiérarchie » de toutes les copies de CC. La différence d'organisation des connaissances, évaluée par l'organisation des réponses au CC, pourrait expliquer les résultats de certains étudiants : mauvaises notes aux TCS mais bonne note au CC.

Les TCS que nous avons établis sont conformes aux recommandations^[2,3]. Toutefois, on peut regretter que certaines questions aient manqué d'incertitude. En effet, deux items étaient trop évidents,

comme démontré par le panel d'expert qui a répondu de manière unanime. Il s'agissait plutôt de « questions de cours ». Cependant, on remarque également que les étudiants n'ont pas trouvé ces questions évidentes avec respectivement 50 %, 30 % et 60 % de bonnes réponses. Ces étudiants étaient certainement en situation d'incertitude du fait d'un manque de connaissances ce qui devrait relativiser cette notion dans l'élaboration des TCS. D'autres items ($n = 4$) retrouvaient seulement un des experts en désaccord avec le reste du panel, mais il s'agissait d'un expert différent à chaque item et ces experts n'étaient pas ceux avec l'expérience clinique la moins importante. Ces six items de TCS seront à remplacer lors des prochaines évaluations pour rester dans le contexte de l'incertitude.

Un des biais de notre étude est le fait que le CC était centré sur la BPCO alors que les TCS couvraient la pneumologie dans son ensemble. Cela peut rendre la comparaison des deux types d'examen plus compliquée si on estime que certains étudiants auraient pu faire « une impasse » sur la BPCO ou au contraire travailler tout particulièrement ce sujet.

Une critique de notre travail concerne la validité de l'étude avec un coefficient α de Cronbach bas. Nous avons été très surpris par les résultats de validité de nos TCS et avons identifié de possibles explications. Tout d'abord, notre panel d'experts était de 13, ce qui est conforme aux recommandations^[4], mais peut-être que 15 à 20 experts auraient permis d'améliorer la validité interne de notre étude^[3]. De même, les deux rédacteurs des TCS ont fait partie du panel d'experts. Nous avons bien respecté la règle des trois items en moyenne par vignettes mais n'avons au total que 20 items dans nos TCS. Il semblerait qu'il faille plus de 60 items pour obtenir un coefficient α de Cronbach $> 0,75$, considéré comme valide^[3]. Dans de nombreux articles, les auteurs ont utilisé des TCS comprenant 90 items, ce qui n'aurait pas été possible dans le cadre de l'examen écrit de pneumologie comme organisé actuellement^[7,8]. Toutefois, on retrouve certains articles où malgré un nombre d'items de TCS conséquents ($n = 120$) le coefficient α de Cronbach n'était que de 0,69 et dans un autre travail, un coefficient α de Cronbach égal à 0,83 avec seulement 26 items^[9,10]. En utilisant la méthode de

Spearman et Brown, on peut penser qu'avec 85 questions de TCS, et donc plus de temps pour cette évaluation, nous aurions pu obtenir un coefficient α de Cronbach de 0,8. Par ailleurs, on notait la présence de corrélation négative entre certaines questions du TCS et la totalité du test. Il semble également que certaines questions mesuraient plus d'un construit ou d'une dimension.

Dans notre travail, nous retrouvons une corrélation entre les scores obtenus aux TCS et ceux obtenus au CC. Néanmoins, on remarque que certains étudiants ayant eu 18/20 ($n = 4$) au CC n'ont obtenu que 9,5–10,1–12,9 et 13,1/20 aux TCS. De même, la partie centrale de la droite de régression (figure 2) montre que pour des scores de TCS égaux à 10 ± 2 , les scores obtenus au CC varient énormément (de 18 à 9/20). Les réponses au CC sont en partie l'évaluation de connaissances déclaratives alors que les TCS mesurent l'interprétation des données cliniques en contexte d'incertitude. On note tout de même que globalement les étudiants ayant de bons scores au CC ont de bons scores aux TCS et qu'inversement, les étudiants ayant de mauvais scores au cas clinique avaient également de mauvais scores aux TCS. Il faut également préciser que la présence des étudiants lors des séances d'entraînement aux TCS n'a pas été vérifiée. Certains étudiants n'ont peut-être pas assisté à ces séances, expliquant leur mauvaise note au TCS. Enfin, les étudiants avec de très mauvais scores aux TCS n'avaient pas répondu à tous les TCS, probablement par faute de temps car les items non répondus sont les 10 derniers.

Les étudiants aux réponses les plus organisées ont des scores aux TCS supérieurs aux étudiants aux copies les moins organisées (évaluation qualitative globale [sans barème] et normative) ; ils ont aussi de meilleurs scores au CC. En outre, on note la faible dispersion des scores obtenus aux TCS par les étudiants aux réponses bien organisées comparativement aux étudiants aux copies mal organisées (figure 3). Aucun des étudiants bien organisés n'a eu de mauvais score aux TCS ; les TCS semblent toujours être bons chez les étudiants aux connaissances bien organisées. L'évaluation globale nous a permis de montrer un caractère plus discriminant des TCS par rapport au CC lors de l'analyse des copies aux notes extrêmes

($n = 20$). En revanche, la mesure de la dimension « organisation – hiérarchie » par deux évaluateurs indépendants montrait simplement que les étudiants aux réponses les plus organisées avaient de meilleurs scores au CC et TCS comparativement aux étudiants aux réponses les moins organisées.

Les étudiants qui avaient effectué leur stage pratique dans le service de pneumologie avaient de meilleurs scores aux TCS que ceux qui n'étaient pas passés dans le service. Cette différence n'était pas retrouvée avec le CC. Cet élément irait plutôt dans le sens d'une meilleure discrimination des connaissances d'action par les TCS. En effet, on peut supposer que le stage pratique de pneumologie a permis aux étudiants de mieux comprendre leurs cours théoriques et d'améliorer leur raisonnement médical par la mise en pratique au lit du patient.

Il aurait été intéressant de recueillir l'opinion et le ressenti des étudiants par un questionnaire ou mieux, des entretiens semi-directifs sur un groupe d'étudiants tiré au sort.

Enfin, il est possible que les modalités des ECN évoluent avec l'introduction de TCS dans les années à venir. Nous devons donc préparer nos étudiants à cette nouvelle technique d'évaluation.

Conclusion

L'introduction des TCS dans l'examen écrit des étudiants de DCEM2 validant la pneumologie a été possible et semble plus discriminante pour l'évaluation du raisonnement clinique que le simple CC avec sa grille de correction. Bien entendu, un seul instrument d'évaluation ne suffit pas^[1]. L'évaluation authentique et formative faite dans les services est complétée par une évaluation « au lit du malade ». Pour la première fois à Rennes, nous avons introduit un nouvel instrument d'évaluation, les TCS, lors de l'examen écrit de fin de pôle « spécialités médicales » des étudiants en DCEM2 en association au CC. Ces TCS permettent d'évaluer les prises de décision en contexte d'incertitude. Le manque de cohérence interne – fiabilité de nos TCS doit nous amener à pratiquer des améliorations : augmentation

du panel d'experts (>15), et surtout augmentation du nombre de vignettes des TCS (>60). Cet ajout d'un outil d'évaluation supplémentaire nous paraît utile, intéressant et prometteur, ce d'autant que les modalités des ECN pourraient évoluer dans les années à venir et intégrer les TCS. L'extension aux autres disciplines, notamment de notre pôle « spécialités médicales » (cardiologie, endocrinologie et dermatologie) est à envisager, ce qui nous permettrait de mettre un nombre suffisant de vignettes de TCS et un temps suffisant lors de l'examen de « fin de pôle ».

Bibliographie

1. Jouquan J. L'évaluation des apprentissages des étudiants en formation médicale initiale. *Pédagogie Médicale* 2002;31:38-52.
2. 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;33:135-44.
3. Fournier JP, Demeester A, Charlin B. Script concordance tests: guidelines for construction. *BMC medical informatics and decision making* 2008;8:18.
4. Gagnon R, Charlin B, Coletti M, Sauve E, van der Vleuten C. Assessment in the context of uncertainty: how many members are needed on the panel of reference of a script concordance test? *Med Educ* 2005;39:284-91.
5. Norcini JJ, Shea JA, Day SC. The use of aggregate scoring for a recertifying examination. *Evaluation and the Health Professions* 1990;13:241-51.
6. Caire F, Sol JC, Moreau JJ, Isidori P, Charlin B. Self-assessment for neurosurgery residents by script concordance test (SCT). The process of test elaboration. *Neuro-Chirurgie* 2004;50:66-72.
7. Fournier JP, Thiercelin D, Pulcini C, Alunni-Perret V, Gilbert E, Minguet J-M, Bertrand F. 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. *Pédagogie Médicale* 2006;7:20-30.
8. Lambert C, Gagnon R, Nguyen D, Charlin B. The script concordance test in radiation oncology: validation study of a new tool to assess clinical reasoning. *Radiation oncology (London, England)* 2009;4:7.
9. Gibot S, Bollaert P. Le test de concordance de script comme outil d'évaluation formative en réanimation médicale. *Pédagogie Médicale* 2008;1:7-18.
10. Haddad F, Gaspard D, Nasr M, Abou Jaoude S, Nemr E. Étude exploratoire de l'utilisation du test de concordance de script pour l'évaluation des apprentissages de résidents de médecine interne au Liban. *Pédagogie Médicale* 2008;9:135-40.

Correspondance et offprints : Stéphane Jouneau, Service de pneumologie, Hôpital Pontchaillou, 2 rue Henri Le Guilloux, 35033 Rennes Cedex, France.
Mailto : Stephane.jouneau@chu-rennes.fr