

Le test de concordance de script comme outil d'évaluation formative en réanimation médicale

The script concordance test as a formative assessment tool in critical care medicine

Sébastien GIBOT¹, Pierre-Edouard BOLLAERT¹

Résumé *Contexte* : Le test de concordance de script (TCS) vise à comparer l'organisation des connaissances, celle-ci constituant un élément essentiel de la compétence clinique. **Objectifs** : Le but de cette étude exploratoire est d'évaluer si un TCS portant sur des champs particuliers de la réanimation détecte bien les différences d'expérience clinique, si le test est capable de déceler l'effet d'interventions éducatives, telles que le stage, et de comparer ses résultats avec ceux d'un examen standard en France : les épreuves classantes nationales (ECN). **Méthodes** : Un TCS, comportant 26 questions réparties en 7 items, a été administré à 9 réanimateurs médicaux seniors, à 18 étudiants (Deuxième cycle d'études médicales, DCEM, 2-4), et à 13 internes de Diplôme d'études spécialisées (DES). Le test, d'une durée de 20 minutes, a été pratiqué en début puis en fin de stage. Parallèlement, les étudiants ont été soumis à un dossier de type ECN couvrant les champs du TCS et d'une durée de 60 minutes. **Résultats** : Les scores obtenus au TCS par les experts étaient plus élevés que ceux obtenus par les internes ($80,9 \pm 4,8$ vs $64,7 \pm 5,3$) ($p < 0,001$), eux-mêmes supérieurs à ceux obtenus par les étudiants ($56,4 \pm 9,6$) ($p < 0,01$). Une corrélation entre TCS et dossier ECN était notée ($R_s = 0,55$, 95 % IC 0,10-0,81). Au terme de leur stage, tous les étudiants avaient progressé au TCS ($+7,2 \pm 1,5$ points). Une amélioration des performances des internes au TCS était également obtenue en fin de semestre (6 ± 1 points), bien que les valeurs restaient inférieures à celles du groupe d'expert ($p < 0,01$). **Conclusion** : La réalisation d'un TCS spécifique de la réanimation médicale pourrait être envisagée chez les étudiants de deuxième cycle et chez les internes DES. Administré en début de stage, il permettrait de distinguer les étudiants les moins à l'aise en terme de raisonnement clinique ; appliqué en fin de stage, le TCS pourrait constituer un moyen d'évaluer la progression.

Mots clés Test de concordance de script ; réanimation médicale.

Abstract *Background*: The script concordance test (SCT) aims at evaluate knowledge organization, which represents an essential component of the clinical competence. **Aim**: The objectives of this investigative study are to verify the effectiveness of a SCT, which bears on particular fields of intensive care medicine, to differentiate clinical experiences, to evaluate its ability to discern student's improvement following an educational intervention, and to compare its results with the ones obtained from the French standard examination, namely the National Classifying Examination (NCE). **Methods**: A SCT, including 26 questions divided in 7 items, was provided to 9 senior intensive care experts, 18 second year medical students, and 13 residents. The 20 minutes long SCT was completed at the beginning and at the end of training. In addition, students completed a 60 minutes long NCE-like test, which comprised similar questioning themes than the ones included in the SCT. **Results**: SCT scores obtained by experts were higher than those obtained by residents, which were superior to those obtained by students ($80,9 \pm 4,8$ vs $64,7 \pm 5,3$; $p < 0,001$ and $64,7 \pm 5,3$ vs $56,4 \pm 9,6$; $p < 0,01$, respectively). A correlation between the SCT and the NCE-like test was found ($R_s = 0,55$; 95% CI 0,10-0,81). At the end of their training, all students have showed an improvement in their SCT scores ($7,2 \pm 1,5$ points). SCT scores obtained by residents were also higher at the end of the semester (6 ± 1 points), although values were still lower than those obtained by the panel of experts ($p < 0,01$). **Conclusion**: The elaboration of a SCT specific to intensive care medicine could be considered for 2nd year students and residents. The SCT provided at the beginning of training could allow distinguishing students with clinical reasoning difficulties; the SCT completed at the end of training could represent a way to assess students' progression.

Key words Script concordance test; intensive care medicine.

Pédagogie Médicale 2008;9:7-18

Introduction

L'évaluation formative a un rôle important à jouer dans l'acquisition des compétences médicales. Une telle évaluation formative n'est pas régulièrement et systématiquement organisée lors des deuxième et troisième cycles d'études médicales en France. Au contraire, l'évaluation finale des apprentissages est réalisée de manière normative à la fin du deuxième cycle à l'aide d'un examen, les épreuves classantes nationales (ECN), qui recourent à des dossiers cliniques types, pour lesquels les réponses attendues sont très univoques.

Le test de concordance de script (TCS) permet l'évaluation du raisonnement clinique en contexte d'incertitude^{1,2}. Il compare les réponses données par la personne qui passe le test à celles données par un groupe d'experts de référence. Il n'a pas vocation à remplacer des modes d'évaluation traditionnelles déjà existantes comme les questions à choix multiples (QCM), qui explorent bien les connaissances liées aux problèmes cliniques bien définis. Il permet d'enrichir de façon complémentaire l'évaluation des apprentissages, en tant que moyen d'approcher le raisonnement face à des problèmes mal définis, c'est-à-dire lorsque toutes les données nécessaires à la résolution du problème ne sont pas présentes ou lorsqu'il existe des données ambiguës, ce qui est la situation quotidienne en pratique clinique^{1,2}.

Dans un TCS, l'étudiant doit faire la synthèse de ses connaissances dans une situation clinique donnée afin de proposer un diagnostic et/ou une stratégie de prise en charge. Le principe du test, élaboré par Charlin, a déjà été largement décrit³⁻⁵. Nous rappellerons simplement que le TCS est un examen constitué d'un nombre variable d'unités. Chacune de ces unités a pour base une courte vignette décrivant une situation clinique classique susceptible d'être rencontrée en pratique courante par le clinicien. La situation clinique décrite pose un problème, même pour un expert du domaine, dans la mesure où il manque un ou plusieurs éléments pour que l'on puisse proposer de façon certaine un diagnostic ou une stratégie de prise en charge (diagnostique ou thérapeutique). Certaines données peuvent également être ambiguës et difficilement interprétables. Sont ensuite proposées différentes hypothèses de diagnostic ou de prise en charge puis, pour chacune, est

proposé un nouvel élément clinique ou paraclinique, dont on évalue l'impact sur la valeur de l'hypothèse de départ, mesuré sur une échelle de Likert à cinq points.

Différents formats sont possibles. Ils peuvent concerner le choix d'examen complémentaires ou des options de prise en charge thérapeutique. Dans une unité diagnostique, une hypothèse diagnostique est proposée dans la première colonne, tandis qu'une donnée clinique ou paraclinique susceptible de modifier la valeur de cette hypothèse est proposée dans la deuxième colonne.

Le mode de correction du TCS est également assez original puisqu'il n'existe pas de bonne ou de mauvaise réponse unique. La grille de correction est établie par un groupe d'experts qui complètent individuellement le test. Le crédit accordé à chacune des réponses possibles (-2, -1, 0, +1, +2) est calculé selon le nombre d'experts qui l'ont choisie. L'absence de « bonne réponse » tient au fait que le TCS a pour objectif d'évaluer la capacité de l'étudiant à résoudre des problèmes pour lesquelles une telle réponse n'existe pas, plusieurs interprétations étant recevables. Il reflète la capacité de l'étudiant à faire face à une situation courante dans laquelle il existe un certain degré d'incertitude, en formulant des hypothèses et en évaluant l'impact de nouvelles données sur le statut de ces hypothèses.

L'incertitude, aussi bien diagnostique que thérapeutique, est forte en réanimation médicale. Les patients sont en effet le plus souvent « polyopathologiques », ce qui rend l'interprétation des diverses explorations à visées diagnostiques délicates. Par ailleurs, et même si un certain nombre de recommandations internationales existent, la stratégie thérapeutique se doit d'être extrêmement individualisée et de toute façon reste très mal codifiée dans la plupart des situations. Enfin, une particularité supplémentaire en réanimation médicale est que l'état critique du patient justifie sa prise en charge sans délai, bien avant d'avoir en main tous les éléments nécessaires à l'élaboration d'un jugement exhaustivement informé. Tout ceci rend théoriquement séduisant l'utilisation d'un TCS dans l'évaluation de nos étudiants.

Le but de cette étude exploratoire est d'évaluer si un TCS portant sur des champs particuliers de la réanimation détecte bien les différences d'expérience clinique, si le test est capable de déceler l'effet d'interventions éducatives,

1- Service de Réanimation Médicale, Hôpital Central, 29 avenue du Maréchal de Lattre de Tassigny, 54000 Nancy, France

Correspondance : Sébastien Gibot - Service de Réanimation Médicale, Hôpital Central, 29 avenue du Maréchal de Lattre de Tassigny, 54000 Nancy, France - Téléphone : +33 (0)3 83 85 29 70. Mailto:s.gibot@chu-nancy.fr

telles que le stage, et de comparer les résultats qu'il génère par rapport à ceux des ECN, qui représentent, *de facto*, l'examen standard en France.

Méthodes

Construction du TCS

La première étape a consisté en une sélection par un enseignant de réanimation médicale d'une série de situations cliniques représentatives de la pratique en réanimation médicale, en termes de : fréquence, mode de présentation, gravité et possibilités thérapeutiques. Ensuite, pour chacune de ces situations, différents paramètres ont été spécifiés :

- les hypothèses pertinentes de diagnostic, d'investigation ou de traitement ;
- les principales investigations complémentaires à demander pour résoudre le problème ;
- les informations cliniques, positives ou négatives, qui seraient recherchées pour vérifier les hypothèses.

Au cours de la seconde étape, chaque situation clinique a été rédigée sous la forme de scénarios de quelques lignes (vignettes). Pour chaque scénario de cas clinique, formulé de manière concise et précise, plusieurs hypothèses ont été proposées sous la forme d'une série d'items. Le format d'un item dépendait de l'objectif d'évaluation : diagnostic, investigation, thérapeutique. Dans tous les cas, chaque item a comporté trois parties :

- la première partie comprend une hypothèse diagnostique ou une investigation paraclinique ou une option thérapeutique ;
- la deuxième partie présente une information nouvelle (par exemple, une manifestation clinique, un résultat d'exploration complémentaire, etc.) ;
- la troisième partie est une échelle ordinale de type Likert, à cinq points.

Chaque item a été bâti afin qu'une réflexion soit nécessaire pour y répondre. De fait, le format des questions a consisté à présenter une hypothèse et à faire préciser quel effet (négatif, neutre ou positif) avait le statut de cette nouvelle hypothèse (diagnostic final, résultat d'un examen biologique ou paraclinique, mais aussi option thérapeutique) qui n'était pas présente dans le scénario clinique initial. Chaque item étant indépendant des autres, les questions ultérieures ont concerné d'autres hypothèses et d'autres données ; en effet, le but de notre travail n'était pas de déterminer l'effet cumulatif d'une série d'informations cliniques, mais de préciser l'impact d'une information clinique, biologique et/ou paraclinique sur une hypothèse diagnostique, une suggestion d'investigation, mais aussi sur une option thérapeutique. Une fois le test construit, il a été soumis pour avis à des réanimateurs médi-

caux expérimentés afin de recueillir leurs appréciations sur la pertinence des questions ainsi que sa faisabilité.

Elaboration d'un dossier clinique de « type ECN »

Parallèlement au TCS, un dossier clinique de type ECN était élaboré, explorant les mêmes contenus et situations cliniques que le TCS. La grille de correction, de 0 à 100, apprécie la conformité des réponses des candidats par rapport à des réponses types définies en tenant compte des recommandations ou des consensus d'experts. Dans ce type de dossier, les situations sont clairement définies et la bonne réponse unique.

Participants

Le TCS a été soumis à un groupe de neuf réanimateurs médicaux (= experts) constituant le panel de référence. Ce groupe d'experts a permis l'élaboration de l'échelle de notation (ci-après).

Il a été secondairement soumis à deux groupes d'étudiants :

- le premier groupe comportant 18 étudiants en médecine de la deuxième à la quatrième année de deuxième cycle des études médicales (DCEM2 à DCEM4) ;
- le deuxième groupe comportant 13 internes inscrits en diplôme d'études spécialisées (DES) (anesthésie-réanimation, cardiologie, médecine interne, et néphrologie) ayant réalisé entre zéro et deux semestres de réanimation médicale.

Administration du test et du dossier clinique de « type ECN »

Le principe du test était expliqué avant administration. Le TCS était administré une première fois en début de stage, ne faisant l'objet d'aucun « *debriefing* » avec les étudiants, puis une nouvelle fois en fin de stage (trois mois pour les étudiants, six mois pour les internes), l'objectif étant d'évaluer la progression. Les étudiants devaient réaliser le test en 20 minutes.

Indépendamment, les étudiants, hors internes, devaient également répondre au dossier clinique ECN, 60 minutes leur étant accordées, conformément au temps imparti pour le traitement d'un tel dossier aux ECN. Ce dossier était également administré en fin de stage au cours duquel un enseignement formel couvrant les champs celui-ci a été délivré.

Système de notation du TCS

Le système de notation est fondé sur le principe selon lequel toute réponse fournie par un expert a une valeur intrinsèque, même si cette réponse ne concorde pas avec celles des autres experts³.

Recherche et Perspectives

Dans ce travail, le groupe de médecins réanimateurs constituait le panel d'experts. Leurs réponses étaient utilisées pour attribuer un score à chaque question et bâtir la grille de correction. Pour chaque item, la réponse donnait droit à un crédit qui correspondait au nombre d'experts qui l'avaient choisie. Tous les items devant offrir le même crédit maximum, les scores bruts étaient transformés proportionnellement pour obtenir un crédit d'un point pour la réponse qui avait été la plus choisie par les membres du panel. Les autres choix recevaient un crédit partiel. Ainsi, pour transformer les scores, tous étaient divisés par le nombre de membres qui avaient donné la réponse la plus choisie. Le score total pour le test était la somme de tous les crédits obtenus à chaque item. Ce score total était ensuite transformé par une simple règle de trois pour obtenir une note exprimée sur 100. Chaque expert était évalué en retirant sa propre grille de réponse du panel et en la soumettant à la nouvelle règle de correction établie par les huit autres experts. Un logiciel de correction automatisée disponible sur Internet a été utilisé pour la notation (<http://www.cme.umontreal.ca/tcs>).

Evaluation qualitative du point de vue des étudiants

La compréhension et l'opinion des étudiants à l'égard du TCS et de ses modalités ont été recueillies de façon informelle lors d'un entretien semi-structuré de *debriefing* à la fin du stage.

Analyses statistiques

Pour chacun des trois groupes (étudiants en médecine, internes de DES et médecins réanimateurs seniors), les notes étaient exprimées en moyennes \pm déviations standards après vérification de la normalité en utilisant le test de Kolmogorov-Smirnoff. Le coefficient de cohérence interne du TCS était apprécié par le test de Cronbach. Les scores moyens recueillis dans chaque groupe étaient comparés par une analyse de variance de type ANOVA, corrigée par le test *post-hoc* de Newman-Keuls. Une corrélation entre TCS et note obtenue au dossier ECN était recherchée en utilisant le test de corrélation de Spearman. Enfin, la progression entre la fin et le début du stage était étudiée par ANOVA pour mesures répétées. Une valeur bilatérale de $p < 0,05$ était considérée comme significative. Le logiciel Graphprism (*Graphpad software*) était employé.

Résultats

Description et qualité du TCS

Le TCS comportait 26 questions réparties en sept items. Le test couvrait une question courante en réanimation qui est la prise en charge d'une détresse respiratoire

aiguë puis celle d'un choc septique. Un extrait est présenté en *annexe 1* et la totalité du test est disponible sur demande.

Le contrôle qualité *a priori* était satisfaisant⁷, à l'exception du nombre d'experts requis : 10 étant recommandés, pour seulement neuf ici (*Tableau 1*). La fidélité était satisfaisante comme en témoignait un coefficient α de Cronbach à 0,83.

Le dossier clinique de type ECN

Ce dossier reprenait la thématique développée dans le TCS : 61 items étaient cotés, répartis en six questions. Les questions posées étaient en accord avec le programme national de l'ECN. Le dossier ainsi qu'un extrait de sa grille de correction sont présentés en partie (*Annexe 2*) (dossier complet disponible sur demande).

Performance des différents groupes au TCS

La valeur moyenne obtenue par ce groupe d'experts était de $80,9 \pm 4,8$. La valeur minimum était de 74,5 et la maximum de 87,7.

Les scores obtenus par les experts étaient plus élevés que ceux obtenus par les internes ($64,7 \pm 5,3$) ($p < 0,001$), eux-mêmes supérieurs à ceux obtenus par les étudiants ($56,4 \pm 9,6$) ($p < 0,01$). De manière attendue, les valeurs obtenues par les experts étaient également supérieures à celles obtenues par les étudiants ($p < 0,001$) (*Figure 1*).

Corrélation entre TCS et dossier type ECN

Nous avons ensuite cherché si la performance des étudiants au TCS était corrélée à celle obtenue au dossier de type ECN.

La valeur moyenne obtenue au dossier ECN était de $57,5 \pm 9,7$. Une corrélation assez ferme entre TCS et dossier ECN était notée ($R_s = 0,55$, 95% Intervalle de Confiance (IC) à 95 % : 0,10-0,81) (*Figure 2*).

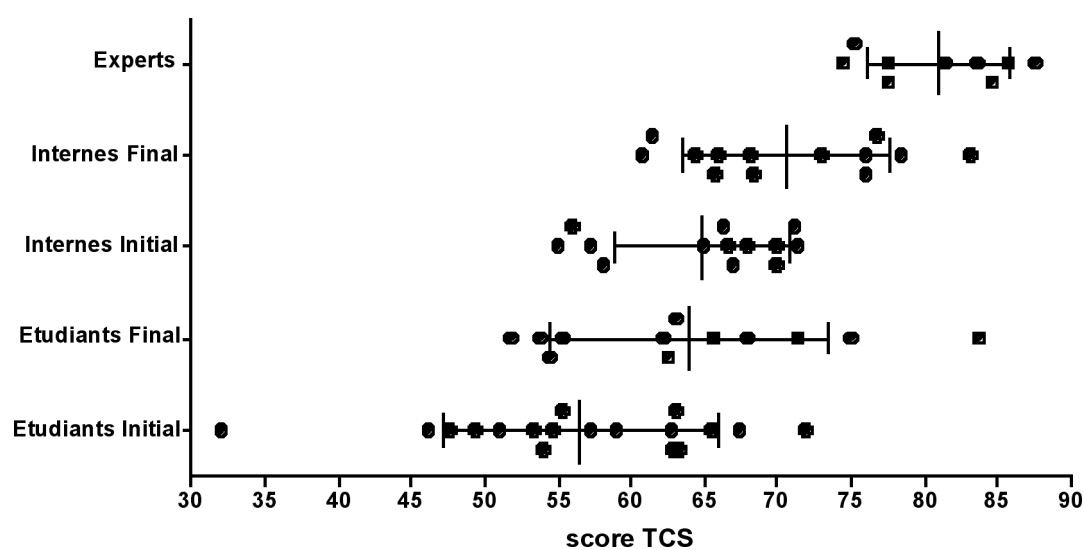
Evolution de la performance au terme du stage

Au terme de leur stage, l'ensemble des étudiants et des internes était à nouveau soumis au TCS (*Figures 1 et 3*). Tous les étudiants, pris individuellement, avaient progressé au terme de leur stage. La moyenne de progression était de $7,2 \pm 1,5$ points. Ainsi, la valeur « finale » était statistiquement supérieure à la valeur « initiale » ($p < 0,05$), et rejoignait la valeur obtenue par les internes au début de leur stage. Ces étudiants progressaient également au dossier ECN en moyenne de $14,6 \pm 6,4$ points. En ce qui concerne les internes, une amélioration des performances au TCS était également obtenue en fin de semestre (6 ± 1 points), bien que les valeurs restaient inférieures à celles du groupe d'experts ($p < 0,01$).

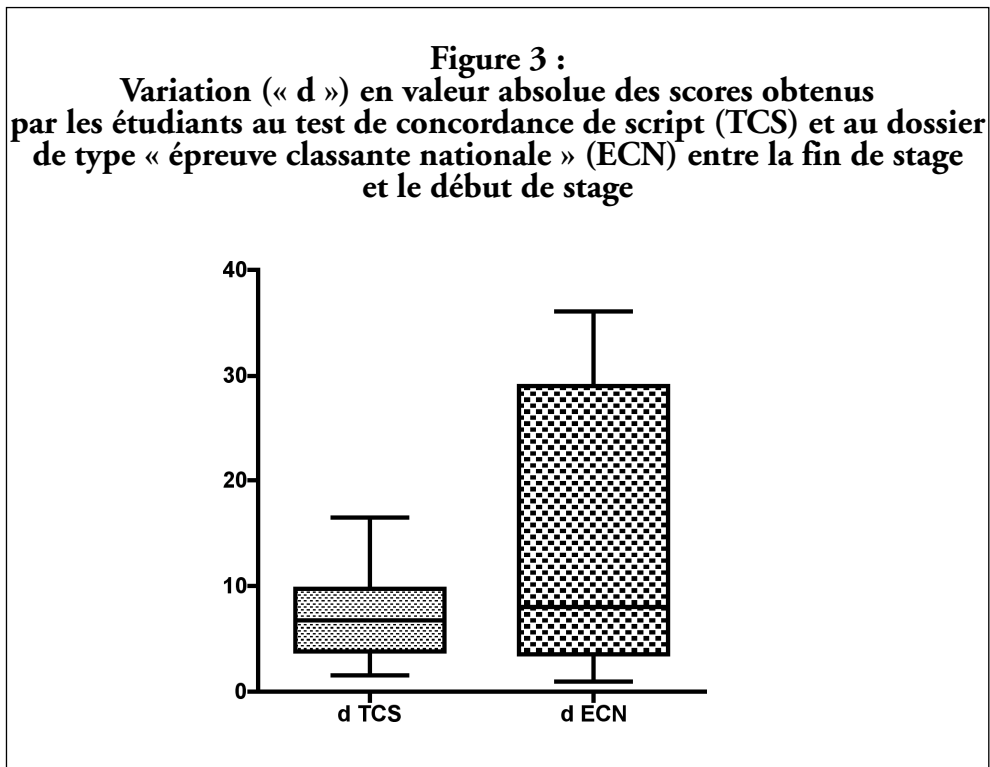
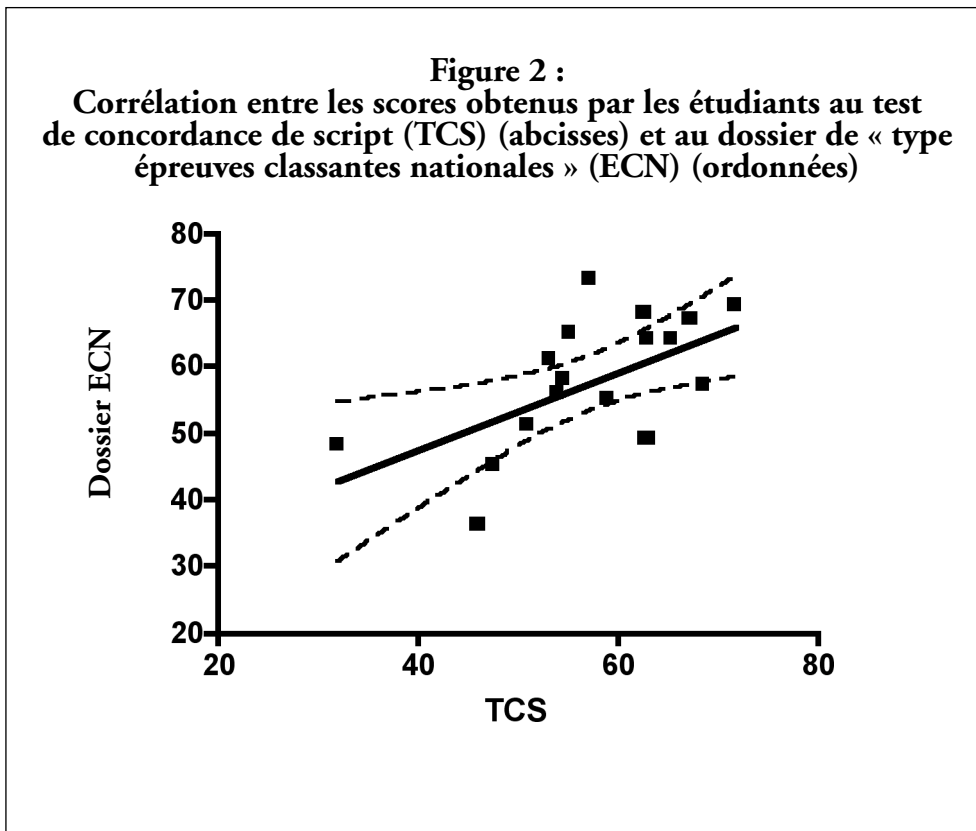
**Tableau 1 :
Contrôle de la qualité *a priori* du test de concordance de script**

Scénario	<ul style="list-style-type: none"> - Décrit une situation problématique, y compris pour les experts. - Décrit une situation appropriée au niveau de compétences des étudiants. - Le scénario est suffisant pour comprendre la question et préciser le contexte. - La présentation clinique est typique. - Le scénario est correctement rédigé. 	<ul style="list-style-type: none"> x oui <input type="checkbox"/> non x oui <input type="checkbox"/> non x oui <input type="checkbox"/> non x oui <input type="checkbox"/> non x oui <input type="checkbox"/> non
Questions	<ul style="list-style-type: none"> - Les questions focalisent sur les points importants. - Les hypothèses sont pertinentes pour les experts. - Les mêmes options ne sont pas retrouvées dans 2 questions consécutives. - La nouvelle information (2^e colonne) est logique, compte tenu de l'hypothèse de la 1^{re} colonne. - Les questions sont construites de façon à disperser les réponses sur tous les points de l'échelle de Likert. - Les questions sont construites pour respecter un équilibre entre basse et haute variabilité. 	<ul style="list-style-type: none"> x oui <input type="checkbox"/> non x oui <input type="checkbox"/> non x oui <input type="checkbox"/> non x oui <input type="checkbox"/> non x oui <input type="checkbox"/> non x oui <input type="checkbox"/> non
Panel d'experts	<ul style="list-style-type: none"> - Entre 10 et 20 - Le panel d'experts est composé de praticiens expérimentés dont la présence dans le panel est légitime, compte tenu du niveau d'expérience des étudiants. - Les experts passent le test dans les mêmes conditions que les étudiants. 	<ul style="list-style-type: none"> <input type="checkbox"/> oui x non x oui <input type="checkbox"/> non x oui <input type="checkbox"/> non

**Figure 1 :
Scores obtenus par les différents groupes au test de concordance de script**



« Initial » : évaluation effectuée en début de stage
« Final » : évaluation effectuée en fin de stage



Opinion des étudiants à l'égard du TCS

L'acceptabilité par les étudiants et les internes est bonne ; ils ont été réceptifs au test et l'ont jugé proche de la réalité professionnelle. Plusieurs étudiants ont cependant rapporté avoir été déstabilisés par la situation – nouvelle pour eux – consistant à être soumis à un examen acceptant plusieurs « bonnes réponses » et par la prise de conscience que les avis d'experts n'étaient pas forcément consensuels.

Discussion

Les principaux résultats de ce travail peuvent être résumés de la façon suivante : 1) le TCS permet effectivement de différencier les étudiants selon leur niveau d'expérience en réanimation médicale ; 2) il existe une bonne corrélation entre les performances obtenues au TCS et au dossier de type ECN par les étudiants en médecine et 3) la progression des étudiants au terme du stage en réanimation est clairement mise en évidence par l'amélioration des résultats au TCS.

Notre travail confirme que le TCS permet de différencier des groupes d'étudiants de niveaux de compétence différents en réanimation médicale. Ainsi, nous avons constaté que les étudiants en médecine (DCEM2, DCEM3 et DCEM4) obtenaient un score inférieur à celui des internes DES, lui-même inférieur au score des experts. Ce résultat attendu reflète le fait que le TCS évalue davantage que la maîtrise des seules connaissances déclaratives mais également le processus de leur mobilisation dans le cadre du raisonnement clinique et il est en accord avec différentes études réalisées dans diverses autres spécialités médicales ou chirurgicales⁸⁻¹⁰. Par exemple, une série comportant 17 étudiants de DCEM 4, neuf internes du DES de médecine générale, cinq internes inscrits au DES de médecine interne, et sept médecins internistes confirmés, avait montré une bonne discrimination du niveau d'expérience clinique par le TCS⁸.

Notre travail confirme également l'excellente faisabilité du TCS. Ce type de méthode d'évaluation est relativement facile à construire et à utiliser. Son acceptabilité par les étudiants et les internes est bonne ; en effet, dans notre étude, les étudiants étaient réceptifs au test et l'avaient jugé proche de la réalité professionnelle.

Il a été montré que la capacité d'un candidat à résoudre un problème clinique ne permet pas de prédire sa capacité à résoudre un autre problème. Ainsi, il est préférable d'éviter toute évaluation longue concernant une même situation clinique et de privilégier les évaluations recensant un nombre plus important de problèmes¹¹. A cet égard, le

TCS, qui consiste en une sélection de situations cliniques variées dans une discipline donnée, permet d'évaluer de manière préférentielle la compétence dans la spécialité concernée. Par ailleurs, le système de notation du TCS, qui permet d'incorporer la variabilité des réponses d'experts du domaine, en fait dès lors un examen utile pour évaluer les participants sur des problèmes mal définis, pour lesquels il n'y a pas nécessairement de consensus entre les experts, ce qui est particulièrement le cas en réanimation médicale.

Il a été mentionné que les étudiants adaptent leur mode d'apprentissage en fonction des dispositifs d'évaluation utilisés. Il est permis de formuler l'hypothèse que la diffusion de l'utilisation du TCS, qui a pour objectif d'évaluer le raisonnement clinique, puisse permettre d'optimiser les activités d'apprentissage et ramener les étudiants « au lit du patient ».

Si la corrélation entre performance à un TCS et à des QCM a déjà été montrée, notre travail relève, pour la première fois à notre connaissance, un lien entre TCS et dossier clinique de type ECN. Dans le cadre du présent travail, nous nous sommes attachés à construire un dossier ECN couvrant les champs d'investigation également explorés par le TCS. La corrélation entre les performances à ces deux types de test souligne le fait que même si l'objectif du TCS est d'évaluer le raisonnement clinique, celui-ci ne s'envisage qu'avec une base de connaissances solide. Cependant, la valeur intermédiaire du coefficient de corrélation calculé suggère que si le TCS mesure des éléments en partie communs avec l'ECN, il explore d'autres aspects de la compétence clinique et que ces deux tests ne sont pas redondants.

Enfin, la réalisation répétée du TCS en fin de stage a clairement pu mettre en évidence une progression de tous les étudiants. Il est bien sûr impossible de distinguer si cette amélioration était liée à l'enseignement théorique réalisé en cours de stage par les médecins du service (sous forme de cours théoriques classiques) ou bien au « compagnonnage » et à l'imprégnation des pratiques d'exercice observées au cours des visites quotidiennes au lit du patient.

Ainsi, le TCS évaluant à la fois les connaissances et leur usage pratique et s'avérant assez sensible dans l'analyse de la progression, il serait théoriquement élégant d'élargir son emploi dans l'évaluation aussi bien en fin de second cycle qu'en troisième cycle.

Si le test de concordance de script peut apprécier de manière simple et reproductible le raisonnement et la compétence clinique, cette technique comporte cependant certaines limites^{6,12}. Le choix des experts peut s'avérer

Recherche et Perspectives

délicat et doit être effectué de façon cohérente par rapport au champ des compétences à évaluer (par exemple médecine hospitalière *versus* médecine de soins primaires). Bien entendu, le TCS étant un examen écrit, il ne permet pas d'évaluer l'ensemble des domaines de la compétence clinique comme les habiletés de recueil des données à l'interrogatoire et à l'examen physique ou encore les habiletés relationnelles.

Un point additionnel mérite d'être abordé : si l'acceptation du TCS a été excellente, le « *debriefing* » de ce test en fin de stage (après l'administration finale) s'est avéré problématique. Il a été en effet difficile de faire admettre à des étudiants, habitués au dogme de la bonne réponse unique et ayant en point de mire le « Graal » de l'ECN et de ses grilles de cotation très tranchées, que plusieurs réponses restaient possibles et que les experts n'étaient pas forcément consensuels. Dans le contexte académique culturel qui caractérise l'évaluation des apprentissages au cours des études de médecine en France, une telle expérience permet d'illustrer auprès des étudiants tous les aléas de la pratique du raisonnement en situation d'incertitude mais peut aussi s'avérer délicate, dans la mesure où elle peut induire un conflit cognitif chez des étudiants majoritairement convaincus que les connaissances médiatisées par les recommandations de bonne pratique (*guidelines*) ou par les conclusions des conférences de consensus ont une valeur normative et univoque. Ce point mérite indiscutablement d'être exploité pédagogiquement d'une manière judicieuse.

En conclusion, la réalisation d'un TCS spécifique dans le champ de la réanimation médicale pourrait être envisagée à la fois chez les étudiants de deuxième cycle mais également chez les internes. Administré une première fois en début de stage, il permettrait de distinguer les étudiants les moins à l'aise en termes de raisonnement clinique, qui pourraient alors bénéficier d'un enseignement plus focalisé. Appliqué en fin de stage, le TCS pourrait constituer un excellent moyen d'évaluer la progression et donc l'utilité réelle pour l'étudiant de son stage.

Contributions

Sébastien Gibot a participé à la conception et à la réalisation pratique de l'étude. Pierre-Edouard Bollaert a participé à la conception de l'étude. Les deux auteurs ont solidairement contribué aux versions successives du manuscrit.

**Annexe 1 :
Le test de concordance de script : exemples d'items**

Une femme de 21 ans, sans antécédent, se présente au service d'accueil des urgences en raison de la survenue brutale d'une dyspnée associée à une douleur basi-thoracique droite. L'interrogatoire vous apprend que cette patiente est sous œstroprogestatifs et qu'elle fume environ 10 cigarettes par jour depuis 4 ans. A l'examen clinique, la température est à 38.3°C, la fréquence respiratoire à 28/min, la fréquence cardiaque à 110/min et la pression artérielle à 90/48 mmHg.

Item n° 1		
Si vous pensiez à	Et qu'alors vous disposiez de	Cette nouvelle donnée rendrait l'hypothèse diagnostique
Embolie pulmonaire	Gazométrie artérielle : PaO ₂ =55 mmHg, PaCO ₂ =28 mmHg	-2 -1 0 +1 +2
	Radiographie pulmonaire : discret épanchement pleural droit	-2 -1 0 +1 +2
	Electrocardiogramme : Bloc de branche droit	-2 -1 0 +1 +2
	Dosage de D-dimères : 2800 µg/L	-2 -1 0 +1 +2
-2 : Quasi improbable -1 : Moins probable 0 : Ni plus ni moins probable +1 : Plus probable +2 : Quasi certaine		

Recherche et Perspectives

Annexe 1 (suite) Le test de concordance de script : exemples d'items

Une femme de 21 ans, sans antécédent, se présente au service d'accueil des urgences en raison de la survenue brutale d'une dyspnée associée à une douleur basi-thoracique droite. L'interrogatoire vous apprend que cette patiente est sous œstroprogestatifs et qu'elle fume environ 10 cigarettes par jour depuis 4 ans. A l'examen clinique, la température est à 38.3°C, la fréquence respiratoire à 28/min et la fréquence cardiaque à 110/min.

Item n° 4		
Si vous pensiez à	Et qu'alors	Cette nouvelle donnée rendrait votre attitude
Hospitaliser cette patiente dans un service de médecine	La pression artérielle était à 74/34 mmHg	-2 -1 0 +1 +2
	La saturation périphérique en O ₂ était à 88 % en air ambiant	-2 -1 0 +1 +2
	Des marbrures des membres inférieurs étaient présentes	-2 -1 0 +1 +2
	Il existait une obnubilation	-2 -1 0 +1 +2
-2 : Très inappropriée -1 : Inappropriée 0 : Ni plus ni moins appropriée +1 : Appropriée +2 : Indispensable		

Annexe 2
**Le dossier clinique de type « épreuves classantes nationales » :
exemples de questions et de leurs grilles de correction**

Une femme de 21 ans, sans antécédent, se présente au service d'accueil des urgences en raison de la survenue brutale d'une dyspnée associée à une douleur basi-thoracique droite. L'interrogatoire vous apprend que cette patiente est sous œstroprogestatifs et qu'elle fume environ 10 cigarettes par jour depuis 4 ans. A l'examen clinique, la température est à 38.3°C, la fréquence respiratoire à 28/min, la fréquence cardiaque à 110/min et la pression artérielle à 90/48 mmHg.

Question 1 : Enumérez en les hiérarchisant vos hypothèses diagnostiques et en les justifiant brièvement.

Question 2 : Quels examens paracliniques prescrivez-vous de première intention et qu'en attendez-vous ?

Question 1	31 points	
	Embolie pulmonaire (4) : femme (1) dyspnée (1) brutale (1) douleur (1) tabac (1) œstroprogestatif (1) tachycardie (1) fièvre (1)	
	Pneumopathie (4) dyspnée (1) brutale (1) douleur (1) tachycardie (1) fébricule (1)	
	Pleurésie (2) douleur (1) fébricule (1)	
	Pneumothorax (2) dyspnée (1) brutale (1) douleur (1) tachycardie (1)	
Question 2	18 points	
	ECG (2) bloc de branche droit et/ou aspect S1Q3 (1)	
	RP (2) épanchement aérique ou liquidien (1) foyer alvéolo-interstitiel (1) hypertrophie artère pulmonaire (1)	
	Gazométrie artérielle (2) hypoxémie (1) capnie (1) statut acido-basique (1)	
	Echocardiographie transthoracique (2) aspect de cœur pulmonaire aigu (1)	
	Hémocultures (1)	
	Antigénurie legionella / pneumocoque (1)	

Références

1. Charlin B, van der Vleuten C. Standardized assessment of reasoning in contexts of uncertainty: the script concordance approach. *Eval Health Prof* 2004;3:304-19.
2. Llorca G, Roy P, Riche B. Evaluation de la résolution de problèmes cliniques mal définis en éthique clinique : variation des scores selon les méthodes de correction et les caractéristiques des jurys. *Pédagogie Médicale* 2003;4:80-8.
3. Charlin B, Boshizen HPA, Custers EJFM, Feltovich PJ. Scripts and clinical reasoning. *Med Educ* 2007;41:1178-84.
4. Charlin B, Roy L, Brailovsky CA, Van der Vleuten C. The script concordance test: a tool to assess the reflective clinician. *Teach Learn Med* 2000;12:189-95.
5. 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.
6. Sibert L, Charlin B, Corcos J, Gagnon R, Grise P, Van der Vleuten C. Stability of clinical reasoning assessment of competence in urology with the Script Concordance Test across two sites from different linguistic, cultural and learning environments. *Med Teach* 2002;24:522-7.
7. Caire F, Sol JC, Moreau JJ, Isidori P, Charlin B. Auto-évaluation des internes en neurochirurgie par des tests de concordance de script (TCS) : processus d'élaboration des tests. *Neurochirurgie* 2004;50:66-72.
8. Marie I, Sibert L, Roussel F, Hellot MF, Lechevallier J, Weber J. Le test de concordance de script : un nouvel outil d'évaluation du raisonnement et de la compétence clinique en médecine interne ? *Rev Med Int* 2005;26:501-7.
9. 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.
10. Caire F, Sol JC, Charlin B, Isidori P, Moreau JJ. Le test de concordance de script (TCS) comme outil d'évaluation formative des internes en neurochirurgie: implantation du test sur Internet à l'échelle nationale. *Pédagogie Médicale* 2004;5:87-94.
11. Charlin B, Bordage G, Van der Vleuten C. L'évaluation du raisonnement clinique. *Pédagogie Médicale* 2003;4:42-52.
12. Sibert L, Charlin B, Gagnon R, Corcos J, Khalaf A, Grise P. Évaluation du raisonnement clinique en urologie : l'apport du Test de Concordance de script. *Prog Urol* 2001;11:1213-9.

Manuscrit reçu le 28 septembre 2007 ; commentaires éditoriaux formulés aux auteurs le 7 janvier 2008 ; accepté pour publication le 9 janvier 2008