

Utilisation du test de concordance de script au cours du deuxième cycle des études médicales : expérience dans l'enseignement de la neurochirurgie

Use of script concordance tests among fourth- to sixth-year medical students: research in teaching neurosurgery

François CAIRE¹, Benoît MARIN² et Emmanuel CUNY³

1 Service de neurochirurgie, CHU de Limoges, 2 avenue Martin Luther King, 87042 Limoges, France

2 UF Recherche clinique et biostatistiques, CHU de Limoges, 2 avenue Martin Luther King, 87042 Limoges, France

3 Service de neurochirurgie B, CHU de Bordeaux, Place Amélie Raba-Léon, 33076 Bordeaux Cedex et Université Victor Segalen Bordeaux 2, 33076 Bordeaux Cedex, France

Manuscrit reçu le 12 février 2011 ; commentaires éditoriaux formulés aux auteurs le 30 mars 2011 ; accepté pour publication le 28 avril 2011

Mots clés :

Test de concordance de script ; deuxième cycle des études médicales ; raisonnement clinique ; neurochirurgie

Résumé – Contexte. Le test de concordance de script, largement diffusé en troisième cycle et en formation médicale continue, n'est en revanche pas utilisé au cours du deuxième cycle des études médicales. **Objectifs.** Le but de ce travail est d'évaluer si le test de concordance de script peut être utilisé pour classer les étudiants en formation initiale de médecine. **Matériel et méthodes.** Un test constitué de trente items de neurochirurgie est soumis à 75 étudiants à l'issue d'un stage hospitalier de neurochirurgie (43 étudiants de deuxième année – DCEM2 –, 20 de troisième année – DCEM3 – et 12 de quatrième année – DCEM4 – du deuxième cycle des études médicales). **Résultats.** Les scores moyens (sur 100 points) étaient de $51,84 \pm 8,22$, $54,94 \pm 7,30$ et $59,78 \pm 6,08$ respectivement pour les trois groupes d'étudiants. L'analyse de variance mettait en évidence une différence globalement significative entre les groupes ($p = 0,0076$). Les tests *post-hoc* réalisés avec le test *t* de Student n'ont mis en évidence qu'une différence significative 2 à 2 entre les groupes DCEM2 *vs.* DCEM4 ($p = 0,003$). **Conclusion.** Nos résultats suggèrent que le test de concordance de script peut montrer une progression des capacités de raisonnement clinique au cours du deuxième cycle. Cela pose la question de son utilisation éventuelle dans le cadre des épreuves classantes nationales françaises.

Keywords:

Script concordance test; medical students; clinical reasoning; neurosurgery

Abstract – Background. Script concordance tests are widely used to assess clinical reasoning among residents and senior physicians, but not among medical students. **Objective.** The aim of this study was to determine if script concordance tests can be used to rank students during their initial medical training. **Materials and methods.** A 30-item test was completed by a group of 75 medical students (43 fourth-, 20 fifth- and 12 sixth-year students) at the end of a training clerkship. **Results.** The mean scores were respectively: 51.84 ± 8.22 , 54.94 ± 7.30 and

59.78 ± 6.08. The variance analysis identified a significant overall difference between groups ($p = 0.0076$). *Post-hoc* student *t*-tests, using the Bonferroni correction to assess significance ($p < 0.017$), identified only one difference between the DCEM2 and DCEM4 groups ($p = 0.003$). **Conclusion.** Our data suggest that script concordance tests could be used to assess the acquisition of clinical reasoning among fourth- to sixth-year medical students. If our results are confirmed by other studies, script concordance tests could be considered for the French final national ranking examination.

Introduction

Le test de concordance de script (TCS), apparu au début des années 2000^[1] est devenu progressivement un outil d'usage courant en médecine. Son intérêt principal est de permettre d'approcher non plus les seules connaissances techniques mais le raisonnement clinique lui-même, sous l'angle de la capacité de prise de décision en contexte d'incertitude^[2]. À ce titre, il n'est comparable directement à aucun autre outil d'évaluation. Sa validité est maintenant bien établie^[3]. On a notamment pu mettre en évidence une progression des scores selon qu'il est administré respectivement à des étudiants en médecine, à des internes et à des praticiens expérimentés^[4-8]. Les modalités techniques de construction et d'administration d'un test sont également bien connues^[9-14].

En revanche, l'étendue de ses applications reste à explorer. Il est majoritairement utilisé comme instrument d'évaluation formative en troisième cycle (cursus post-gradué), spécialisé^[15] ou non, et en formation médicale continue, mais des portes ont été ouvertes récemment, aussi diverses que l'évaluation de la prise de décision per-opératoire en chirurgie^[16,17] ou l'aide à l'enseignement en premier cycle (partie pré-clinique du cursus pré-gradué). À notre connaissance, l'intérêt du TCS pour le deuxième cycle des études de médecine, c'est-à-dire le cycle d'apprentissage clinique, n'a pas été évalué jusqu'à présent. On peut pourtant supposer qu'il pourrait permettre de mesurer l'acquisition progressive des capacités de raisonnement clinique par des étudiants.

Nous exposons ici les résultats d'une étude préliminaire portant sur les données recueillies de façon

prospective dans un groupe d'étudiants accueillis en stage de neurochirurgie au CHU de Bordeaux. Notre objectif était de savoir si, à l'issue de ce stage, on pouvait observer une différence significative entre les scores obtenus à un TCS par des étudiants de deuxième (DCEM 2), troisième (DCEM 3) et quatrième (DCEM 4) années des études médicales.

Matériel et méthodes

Le test de concordance de script

Depuis le début des années 2000, a été mis en place à l'instigation du collège des enseignants de neurochirurgie une banque de questions de cette discipline, destinée à des étudiants de deuxième et de troisième cycle. La création de cette banque a déjà été décrite en détail^[13]. Elle est constituée exclusivement de tests de concordance de script.

Les questions ont été rédigées par des neurochirurgiens seniors (médecins hospitaliers et hospitalo-universitaires), puis relues et critiquées ; enfin les grilles de notation ont été établies, le plus souvent lors de sessions du collège des enseignants de neurochirurgie. Le panel d'experts était variable, toujours compris entre 7 et 11 personnes.

Le TCS proposé dans cette étude comportait 10 cas issus de la banque existante. Chaque cas était constituée de trois questions (ou items), soit 30 questions au total. Un exemple de cas est illustré par la figure 1. Les cas étaient choisis dans la banque selon leur pertinence au regard du programme du deuxième cycle des études médicales et selon leur niveau de difficulté.

Vous êtes de garde en neurochirurgie. A 4 h du matin, une infirmière des soins intensifs vous appelle pour vous signaler que monsieur H «s'aggrave». Vous connaissez bien ce malade, hospitalisé dans l'après-midi pour une hémorragie méningée de la vallée sylvienne droite, grade III dans la classification de Hunt et Hess avec un discret déficit moteur distal du membre supérieur gauche.

Si vous pensiez à	Et qu'alors vous trouvez	L'effet sur votre hypothèse Diagnostique est le suivant
Hydrocéphalie aiguë	Score de Glasgow à 8. TA à 180/120. hémiparésie gauche complète	-2 -1 0 +1 +2
Récidive hémorragique	Paralysie du VI. Bradycardie. Somnolence sans majoration du déficit	-2 -1 0 +1 +2
Vasospasme	Majoration du déficit gauche, brachio-facial. TA à 110/60	-2 -1 0 +1 +2

Entourez la proposition qui vous semble adéquate :

- 2 l'hypothèse est pratiquement éliminée
- 1 l'hypothèse devient moins probable
- 0 l'information n'a aucun effet sur l'hypothèse
- +1 l'hypothèse devient plus probable
- +2 il ne peut s'agir pratiquement que de cette hypothèse

Fig. 1. Exemple de cas figurant parmi les dix retenus pour le test de concordance de script administré dans l'étude.

Les étudiants

Le groupe était constitué de 75 étudiants de deuxième cycle effectuant un stage d'externat de quatre mois en neurochirurgie entre 2007 et 2010. Il s'agissait de 43 étudiants de DCEM2, 20 étudiants de DCEM3, 12 de DCEM4. Tous avaient au préalable suivi l'enseignement théorique de neurochirurgie. L'évaluation par TCS était réalisée à l'issue du stage.

Analyse des résultats

Pour chacun des trois groupes, les scores étaient exprimés en moyenne \pm écart type, médiane, intervalle interquartile (IQR) et extrêmes, le score maximal étant de 100 points. Les trois moyennes ont été comparées globalement en utilisant une analyse de variance. La normalité de la distribution des scores dans les trois groupes a été testée au travers du test de Shapiro-Wilk. L'homogénéité des variances a été

testée en utilisant le test de Barlett. Les tests *post-hoc* ont été réalisés en utilisant le test *t* de Student. Une correction de Bonferroni a été utilisée afin de déterminer la significativité des tests *post-hoc*, en évitant l'inflation du risque alpha du fait de la multiplicité des tests, en réduisant le seuil de significativité à 0,017.

Résultats

Scores moyens

Tous les questionnaires ont été remplis intégralement par tous les étudiants. Il n'a pas été signalé *a posteriori* d'erreur ou d'ambiguïté dans la rédaction des questions. Le score maximal étant de 100 points, les scores moyens pour chacun des trois groupes de niveau étaient :

- pour les DCEM2 : 51,84 \pm 8,22 (médiane : 51,36, IQR [46,56–57,03])
- pour les DCEM3 : 54,94 \pm 7,30 (médiane : 53,53 IQR [49,93–59,33])

- pour les DCEM4 : $59,78 \pm 6,08$ (médiane : 60,60 IQR [55,36–63,15]).

La distribution des scores est illustrée par les figures 2 et 3.

Analyse statistique

Les conditions d'application permettant la réalisation d'une analyse de variance étaient remplies : le score suivait une loi normale dans les trois groupes examinés (test de Shapiro-Wilk : $p = 0,92$, $p = 0,41$, $p = 0,86$ pour DCEM2, DCEM3 et DCEM4 respectivement), les variances étaient considérées comme homogènes (test de Barlett : $p = 0,47$). L'analyse de variance mettait en évidence une différence globalement significative entre les groupes ($p = 0,0076$). Les tests *post-hoc* réalisés avec le test *t* de Student et en appliquant la correction de Bonferroni du degré de significativité n'ont mis en évidence qu'une différence significative deux à deux entre les groupes DCEM2 vs. DCEM4 ($p = 0,003$). Les comparaisons des groupes DCEM2 vs. DCEM3 et DCEM3 vs. DCEM4 étaient non statistiquement significatives ($p = 0,15$ et $0,064$, respectivement).

Discussion

Dans ce travail, réalisé dans une population de 75 étudiants de deuxième cycle, nous observons une progression des scores avec le niveau d'étude : $51,84 \pm 8,22$ pour les DCEM2, $54,94 \pm 7,30$ pour les DCEM3, et $59,8 \pm 6,08$ pour les DCEM4. Il existe une différence significative entre les scores des trois groupes ($p = 0,0076$), liée à l'écart entre les étudiants de DCEM4 et les étudiants de DCEM2 ($p = 0,003$). Ce travail suggère donc une progression des capacités d'interprétation des données cliniques, dans un contexte de raisonnement sur des cas clinique, au cours des études de médecine entre la quatrième (DCEM2) et la sixième (DCEM4) années. Ceci tendrait à confirmer la validité de construit de ce test utilisé chez des étudiants en médecine en formation initiale.

Un tel résultat peut paraître surprenant, dans la mesure où les étudiants qui ont participé à l'étude avaient tous suivi le même enseignement théorique de neurochirurgie au préalable ; leurs connaissances dans ce domaine étaient donc théoriquement les mêmes. En outre, ils terminaient tous un stage d'externat dans le service ; leur expérience clinique de la neurochirurgie était donc globalement homogène. L'existence d'une différence entre les scores des étudiants de DCEM2 et DCEM4 tendrait donc à montrer que la spécificité de contenu dans le champ du raisonnement clinique n'est pas absolue. Ainsi, l'expérience acquise par les étudiants au cours de leur externat dans les différentes disciplines rencontrées pourrait être en partie transférable à l'étude de nouvelles disciplines. Il est en effet probable que les connaissances acquises en radiologie, en stage de réanimation, de neurologie ou de rhumatologie sont partiellement utilisables dans le champ de la neurochirurgie et pourraient au moins permettre aux étudiants de créer facilement des liens opérationnels avec des connaissances antérieures. Les étudiants s'approprieraient progressivement une « culture générale » médicale.

Le coefficient de Cronbach pour ce TCS et pour cet échantillon était de 0,32. Cette valeur faible est probablement liée au caractère très court du test proposé à nos étudiants, en l'occurrence 10 cas, soit un total de trente questions (ou items). Nous souhaitons en effet proposer à nos étudiants un test pour lequel le temps de réponse n'excédait pas 30 minutes. Par ailleurs, les étudiants étaient pour la première fois confrontés au TCS.

À l'heure actuelle, la mise en œuvre du TCS reste classiquement limitée au troisième cycle et à la formation médicale continue. Son utilisation au cours des études de médecine est marginale. Par définition, le TCS ne peut s'adresser qu'à des étudiants qui sont en train d'acquérir progressivement des connaissances cliniques : en premier cycle, il ne peut donc être utilisé que comme outil d'aide à l'enseignement^[18] et non comme moyen d'évaluation, fut-elle formative. En deuxième cycle en revanche, il pourrait être utilisé pour mettre en évidence une

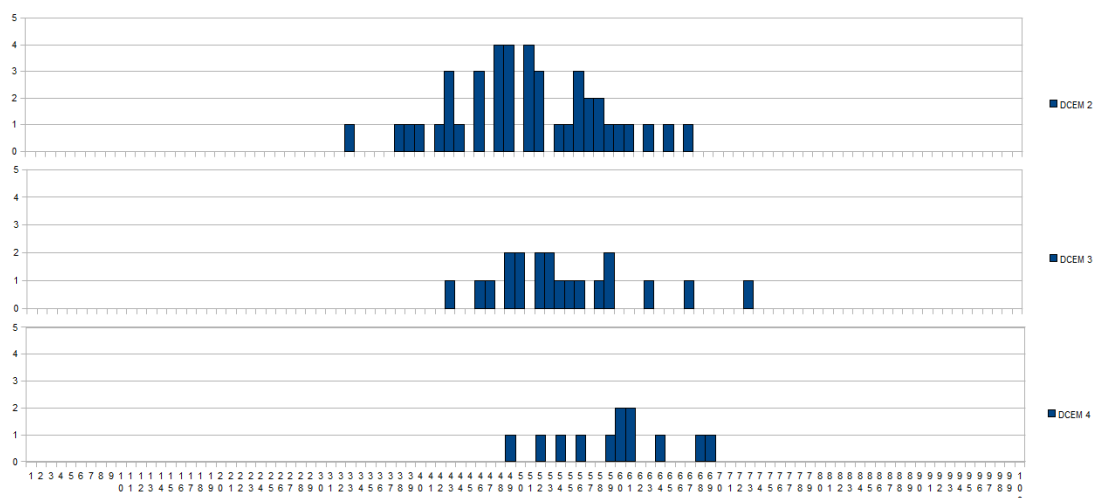


Fig. 2. Répartition des scores (score maximal = 100) pour les trois groupes d'étudiants : DCEM2 ($n = 43$), DCEM3 ($n = 20$) et DCEM4 ($n = 12$). DCEM : deuxième cycle des études médicales.

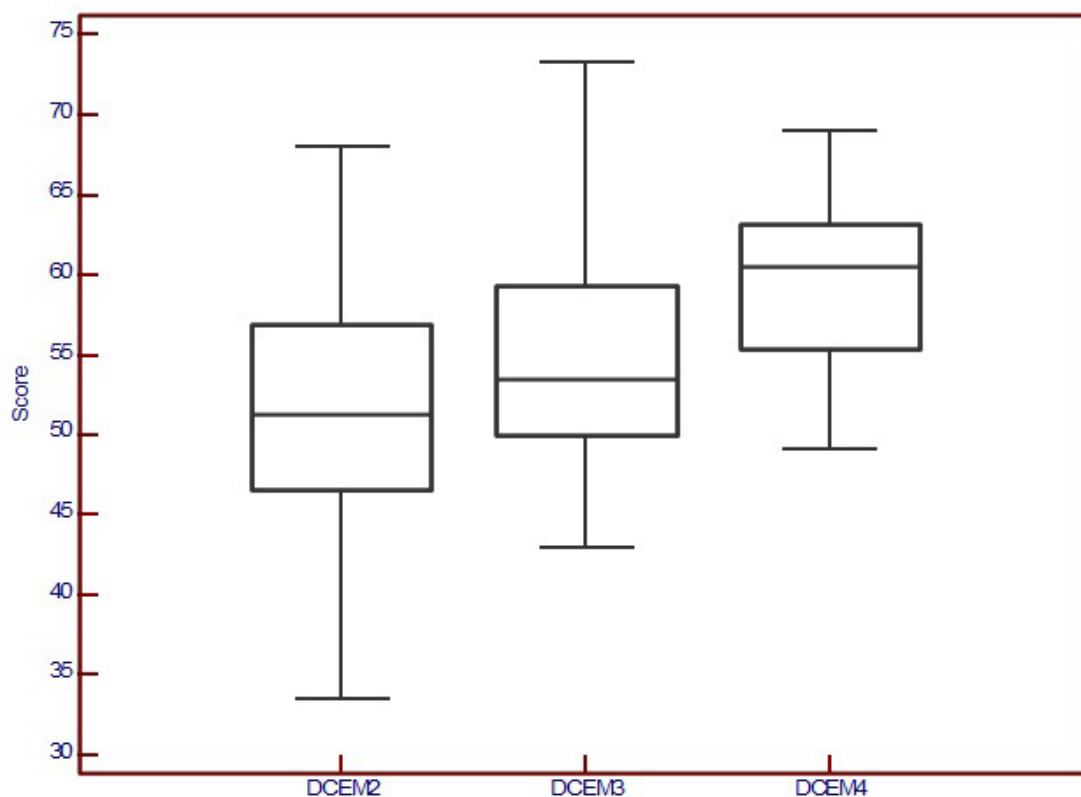


Fig. 3. Scores sur 100 points obtenus par des étudiants de DCEM2, 3 et 4 au test de concordance de script : « Boîte à moustache » représentant la médiane, l'intervalle inter-quartile et les extrêmes pour les trois groupes d'étudiants. DCEM : deuxième cycle des études médicales.

progression dans les capacités de raisonnement clinique des étudiants.

En contexte français, il a d'ailleurs été envisagé d'inclure le TCS parmi les outils d'évaluation dans le cadre des épreuves classantes nationales (ECN) en fin de deuxième cycle. Les étudiants semblent très favorables à cette idée, à laquelle s'est intéressée notamment l'Association nationale des étudiants en médecine de France (ANEMF).

La question de l'utilisation du TCS en deuxième cycle est peu traitée dans la littérature. La plupart des auteurs comparent les scores obtenus par des étudiants aux scores obtenus par des internes d'une part, des praticiens « seniors » d'autre part. Ce type d'étude a notamment été réalisé en médecine interne^[4], en réanimation médicale^[5], en urologie^[6], en radiologie^[8] ou en radiothérapie^[7]. Les groupes d'étudiants peuvent être homogènes ou non et regrouper selon les cas : des étudiants de sixième année seuls^[4], de quatrième année^[7], de la quatrième à la sixième année^[5]. Malheureusement, les résultats de ces travaux ne permettent pas de comparer les scores des étudiants selon leur niveau d'étude.

Le travail réalisé en réanimation par Gibot et Bollaert^[5] est toutefois intéressant. Il compare les scores de trois groupes : des étudiants de DCEM2 à DCEM4, des internes et des réanimateurs « seniors ». Pour les deux premiers groupes, le test est appliqué à deux reprises, en début et en fin de stage de réanimation (trois mois et six mois). Si nous nous intéressons seulement au groupe des étudiants, nous constatons une progression significative des résultats entre les deux passages du test ($p < 0,05$), parallèlement à une progression des scores au dossier de format ECN administré en parallèle. Ces résultats vont donc dans le sens d'une progression des capacités de raisonnement clinique à l'issue du stage mais l'effectif réduit (18 étudiants) ne permet pas de comparaison entre les étudiants de DCEM2, DCEM3 et DCEM4. On notera par ailleurs que, dans cette étude, il existe une corrélation entre les scores obtenus au TCS et à l'épreuve de dossier de type ECN, ce qui appuie la validité externe du TCS.

L'article publié par Sqalli Houssaini *et al.*^[19] est à notre connaissance le seul à s'intéresser spécifiquement au deuxième cycle des études médicales. Il compare deux groupes d'étudiants ayant un niveau de connaissances de base comparable en fin de quatrième année, à l'issue de l'enseignement universitaire de néphrologie. En sixième année, la réalisation d'un nouveau TCS montre que les scores sont significativement plus élevés dans le groupe d'étudiants qui effectuent un stage d'externat en néphrologie en cinquième année que dans le groupe d'étudiants n'ayant pas bénéficié de ce stage. Ce travail montre donc une progression des scores entre la quatrième et la sixième année mais il s'agit surtout pour les auteurs d'illustrer l'intérêt du stage clinique dans l'apprentissage. Ces résultats ne sont donc pas généralisables.

Au total, les données de la littérature ne nous permettent donc pas, en l'état actuel des choses, de comparer nos observations à celles d'autres équipes. Nos résultats doivent être confirmés dans des disciplines différentes. Il serait également intéressant de préciser jusqu'à quel point le TCS peut être discriminant et distinguer des étudiants de DCEM2, DCEM3 et DCEM4.

Conclusion

Notre travail suggère qu'il pourrait exister une progression des capacités de raisonnement clinique, telles que mesurées par le TCS, au cours des études de médecine. Ces résultats préliminaires doivent être confirmés par un travail plus large et transversal, mené de façon prospective. Cela pourrait conduire à poser la question de l'utilisation du TCS au cours des évaluations de fin de deuxième cycle des études médicales, telles qu'elles sont organisées par exemple en France dans le cadre des ECN.

Contributions

François Caire a participé à la rédaction initiale de l'article, à l'analyse et à l'interprétation des résultats. Benoît Marin a participé à la révision critique

de l'article, à l'analyse et à l'interprétation des résultats. Emmanuel Cuny a participé à la conception du projet, à l'analyse et à l'interprétation des résultats, à la révision critique de l'article.

Remerciements. Nous remercions les Drs. et Prs. Bousquet, Frerebeau, Irthum, Kehrl, Lonjon, Mourier, Moreau, Paquis, Pirotte et Sol, qui ont participé à la rédaction et à l'élaboration des items composant le TCS utilisé dans ce travail. Nous remercions le Pr. B. Charlin pour ses conseils et pour la relecture du manuscrit.

Références

- Charlin B, Roy L, Brailovsky C, Goulet F, Van der Vleuten C. The Script Concordance Test, a Tool to Assess the Reflective Clinician. *Teach Learn Med* 2000;12:189-95.
- Charlin B. Évaluer la dimension d'incertitude du raisonnement clinique, Assessing the uncertainty component of clinical reasoning. *Pédagogie Médicale* 2006;7:5-6.
- Lubarsky S, Charlin B, Cook DA, Chalk C, van der Vleuten CPM. Script concordance method: a review of published validity evidence. *Med Educ* 2011;45:329-38
- Marie I, Sibert L, Roussel F, Hellot M-F, 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.
- Gibot S et Bollaert P-E. Le test de concordance de script comme outil d'évaluation formative en réanimation médicale. *Pédagogie Médicale* 2008;9:7-18
- Sibert L, Charlin B, Corcos J, Gagnon R, Khalaf A, Grise P. Évaluation du raisonnement clinique en urologie. L'apport du test de concordance de script. *Prog Urol* 2001;11:1213-9.
- 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. *Radiat Oncol* 2009; 4:7.
- 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.
- Fournier J-P, Demeester A, Charlin B. Script Concordance Tests: Guidelines for Construction. *BMC Med Inform and Decis Mak* 2008; 8:18.
- Gagnon R, Charlin B, Coletti M, Sauvé E, Van der Vleuten C. Assessment in context of uncertainty: How many members are needed on the panel of reference of a script concordance test? *Med Educ* 2005;39:204-91.
- Gagnon R, Charlin B, Lambert C, Carrière B, der Vleuten V. Script Concordance Testing: More Cases or More Questions? *Adv Health Sci Educ* 2009;14:1382-96
- Charlin B, Gagnon R, Lubarsky S, Lambert C, Meterissian S, Chalk C, Goudreau J, van der Vleuten C. Assessment in the context of uncertainty using the script concordance test: More meaning for scores. *Teach Learn Med* 2010;22:180-6.
- 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.
- Sibert L, Darmoni S, Dahamna B, Hellot M-F, Weber J, Charlin B. On line clinical reasoning assessment with Script Concordance test in urology; results of a French pilot study. *BMC Medical Education. BMC Med Educ* 2006;6:45
- Carrière B, Gagnon R, Charlin B, Downing S, Bordage G. Assessing clinical reasoning in pediatric emergency medicine: validity evidence for a script concordance test. *Ann Emerg Med* 2009;53:647-52.
- Meterissian S, Zabolotny B., Gagnon R. and Charlin B. Is the script concordance test a valid instrument for assessment of intraoperative decision-making skills? *Am J Sur* 2007;193:248-51
- Park AJ, Barber MD, Bent AE, Dooley YT, Dancz C, Sutkin G, Jelovsek JE. Assessment of intraoperative judgment during gynecologic surgery using the script concordance test. *Am J Obstet Gynecol* 2010;203:240-6.
- Hoff L, Bestawros A, Kassis J, Charlin B. Le test de concordance de script comme outil d'enseignement et d'apprentissage: un projet-pilote pour les étudiants en première année de médecine. *Pédagogie Médicale* 2010;10:51-6.
- Sqalli Houssaini T, Bono W, Tachfouti N, Maillard D. Pertinence d'un TCS dans l'évaluation des compétences en néphrologie des étudiants du deuxième cycle de la faculté de médecine de Fès. *Annales de médecine et de thérapeutique* 2009;1:4-10.

Correspondance et offprints : Emmanuel Cuny, Service de neurochirurgie B, Groupe hospitalier Pellegrin-Tripode, Place Amélie Raba-Léon, 33076 Bordeaux Cedex, France
Mailto : emmanuel.cuny@chu-bordeaux.fr