

AUTO-ÉVALUATION DES INTERNES EN NEUROCHIRURGIE PAR TESTS DE CONCORDANCE DE SCRIPT (TCS) Processus d'élaboration des tests

F. CAIRE⁽¹⁾, J.-C. SOL⁽²⁾, J.-J. MOREAU⁽¹⁾, P. ISIDORI⁽³⁾, B. CHARLIN⁽⁴⁾

(1) Service de Neurochirurgie, CHU Dupuytren, 87042 Limoges.

(2) Service de Neurochirurgie, CHU de Rangueil, 31400 Toulouse.

(3) DCAM, Université Victor-Segalen - Bordeaux II, 33076 Bordeaux Cedex.

(4) URDESS, Université de Montréal, Faculté de Médecine, Montréal, Québec, Canada H3C 3J7.

SUMMARY: Self-assessment for neurosurgery residents by script concordance test (SCT). The process of test elaboration

F. CAIRE, J.-C. SOL, J.-J. MOREAU, P. ISIDORI, B. CHARLIN (*Neurochirurgie*, 2004, 50, 66-72).

Background. — *The script concordance test (SCT) is a new tool to assess clinical reasoning in a context of uncertainty. It was chosen for computer-based self-assessment training for neurosurgery residents.*

Purpose. — *The aim of this work is to describe the construction of a first series of questions and provide teachers practical information needed to build a SCT.*

Method. — *The theoretical background of the SCT is explained. We review the steps we followed when writing of a series of more than two hundred questions.*

Results. — *We present the different steps of our work, including writing clinical cases, validation of questions and elaboration of scoring systems. We explain the main difficulties we encountered.*

Conclusion. — *The script concordance test is an interesting easy-to-construct assessment tool useful for evaluating clinical competence. This work summarizes our experience in building a SCT, which should be helpful for teachers interested in this new assessment tool, especially in neurosurgery.*

Key-words: education, training assessment, neurosurgery, residency, script concordance test.

RÉSUMÉ

Contexte. — Le test de concordance de script (TCS) est un outil récemment développé qui évalue le raisonnement clinique en contexte d'incertitude. Il a été choisi comme moyen d'auto-évaluation des internes en Neurochirurgie au cours de leur formation.

Objectif. — Ce travail vise à décrire le processus de rédaction d'une banque de questions, afin de fournir aux neurochirurgiens des renseignements pratiques susceptibles de les aider dans la mise en place d'évaluations par TCS.

Méthode. — L'article repose sur l'expérience acquise lors de la création d'une première banque d'items pour le test.

Résultats. — Nous présentons les différents formats de vignettes retenus, les étapes successives du travail de rédaction (définition des situations cliniques et des domaines, rédaction des questions, correction et validation, enfin construction des grilles de réponse), ainsi que les difficultés rencontrées.

Conclusion. — Le respect d'un processus de rédaction bien défini, utilisable en Neurochirurgie mais aussi dans d'autres spécialités, permet la production d'un test de qualité, constituant un bon outil d'auto-évaluation.

L'évaluation du raisonnement clinique est un problème récurrent dans l'enseignement de la Médecine. Les recherches effectuées dans ce do-

maine ont permis d'en identifier les spécificités et de proposer plusieurs instruments [3] : questions à choix multiples (QCM), examen oral, question

Article reçu le 24 décembre 2003. Accepté le 15 janvier 2004.

Tirés à part : F. CAIRE, à l'adresse ci-dessus.
e-mail : fpcaire@yahoo.fr

rédaçtionnelle, examen clinique objectif et structuré (ECOS).

Aucun n'est, toutefois, pleinement satisfaisant. Le test de concordance de script (TCS) est un outil novateur qui vise à évaluer la capacité de raisonnement en contexte d'incertitude [5, 6], c'est-à-dire dans une situation proche d'un problème clinique.

Dans le cadre d'une réforme de l'enseignement du troisième cycle de Neurochirurgie, le Collège de Enseignants a décidé la mise en place d'une évaluation formative portant sur le raisonnement clinique des internes. C'est donc le TCS, informatisé dans le cadre du campus numérique de Neurochirurgie, qui a été retenu comme outil de cette évaluation.

Sa mise en place a débuté en janvier 2002. Nous décrivons ici le processus de rédaction utilisé pour créer une première série d'items portant sur des problèmes cliniques prioritaires ainsi que les difficultés rencontrées à chaque étape. Nous nous efforcerons d'en dégager quelques enseignements à l'intention de futurs rédacteurs de tests en Neurochirurgie ou dans d'autres spécialités médicales.

PRINCIPE DU TEST DE CONCORDANCE DE SCRIPT

Les travaux d'Elstein, puis de G. Bordage [1] ont montré l'importance que revêt le degré d'organisation des connaissances médicales dans la capacité à résoudre des problèmes cliniques. Le terme de « script » désigne les liens, les réseaux associatifs qui se mettent en place entre les connaissances d'un clinicien pour les rendre fonctionnelles en situation clinique. D'après les promoteurs de la théorie des scripts [4, 11], les scripts de connaissances relatives au diagnostic et à la prise en charge apparaissent lorsque les étudiants et résidents commencent à être confrontés à des cas cliniques réels. L'élaboration des scripts conditionne la capacité à agir de façon pertinente dans une situation clinique donnée. Le test de concordance de script (TCS) explore la qualité d'organisation de ces réseaux de connaissances et, de ce fait, la qualité du raisonnement clinique [2, 3, 6].

Par ailleurs, le TCS reflète avant tout la capacité à résoudre des problèmes cliniques dits « mal définis ». Il s'agit de problèmes pour lesquels les données dont on dispose sont incomplètes ou ambiguës [5, 9]. Comme dans une situation clinique réelle, les données de l'interrogatoire et de l'examen clinique initial conduisent à formuler plusieurs hypothèses. Celles-ci sont ensuite testées : le clinicien précise et oriente son examen clinique,

et demande bien souvent des examens complémentaires dans le but de confirmer — ou d'infirmer — les hypothèses de départ. Dans un TCS, l'étudiant doit également considérer plusieurs hypothèses diagnostiques ou options de prise en charge et déterminer l'effet qu'auront de nouveaux éléments cliniques ou paracliniques sur la valeur de ces hypothèses.

On parle *a contrario* de « problème bien défini » lorsque la réponse est univoque (exemple : vous voyez en consultation un homme pour une névralgie cervico-brachiale ; la douleur irradie à la face interne de l'avant-bras jusqu'aux deux derniers doigts : à quelle racine correspond cette topographie ?) : de tels problèmes ne sont pas explorés par les TCS.

Le TCS n'a pas vocation à remplacer les outils déjà existants, comme les QCM, qui s'intéressent à l'évaluation de connaissances dans le cadre de problèmes bien définis. Les deux outils se complètent dans le cadre de l'évaluation du raisonnement clinique.

Le test comprend, en premier lieu, une courte vignette décrivant une situation clinique courante et crédible (figure 1). La description du cas (la vignette) ne contient pas assez d'informations pour que l'on puisse décider de façon certaine du diagnostic ou de la stratégie de prise en charge diagnostique et thérapeutique. Les trois ou quatre

	<i>Si vous pensez à</i>	<i>Et qu'alors vous trouvez à l'examen clinique ou à l'interrogatoire</i>	<i>L'effet sur votre hypothèse diagnostique est le suivant</i>
Item 1	Myélopathie cervicarthrosique	Impressions vertigineuses	-2 -1 0 +1 +2
Item 2	Maladie de Charcot	Radiculalgie C8-D1	-2 -1 0 +1 +2
Item 3	Compression tumorale cervicale	Dysphagie d'évolution progressive	-2 -1 0 +1 +2

Entourez la proposition qui vous semble adéquate (échelle de Likert) :

- 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 pratiquement s'agir que de cette hypothèse.

FIG. 1. — Exemple d'unité diagnostique : unité n° 8;8. M. T..., 58 ans, vous décrit la survenue, lors de ses activités de bricolage, en fin de journée, d'une impression de faiblesse de la main droite et de maladresse gestuelle. On note, dans ses antécédents, une artériopathie chronique des membres inférieurs liée à une intoxication tabagique ancienne.

FIG. 1. — Example of a diagnostic unit, unit 8.8. A 58-year-old patient describes weakness in the right hand and poor control of movements when working on his hobby at the end of the day. The patient has a history of chronic lower limb arteriopathy related to long-term smoking habits.

hypothèses diagnostiques (ou options de prise en charge diagnostique ou thérapeutique) les plus pertinentes sont proposées dans la première colonne. Les questions qui suivent dans la

<i>Si vous pensez à</i>	<i>Et qu'alors vous trouvez ou que vous apprenez</i>	<i>L'effet sur votre hypothèse diagnostique est le suivant</i>
Hypothèse diagnostique	Nouvelle information clinique ou résultat d'un examen complémentaire	-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 pratiquement s'agir que de cette hypothèse.		

FIG. 2. — Format d'une unité diagnostique.

FIG. 2. — *Format of a diagnostic unit.*

<i>Si vous pensiez faire</i>	<i>Et qu'alors vous trouvez</i>	<i>L'effet sur la nécessité de demander ce test est le suivant</i>
Examen complémentaire	Nouvelle information clinique ou résultat d'un autre examen	-2 -1 0 +1 +2
Entourez la proposition qui vous semble adéquate : -2 = absolument contre-indiqué ; -1 = peu utile ou plutôt néfaste ; 0 = non pertinent ; +1 = utile et souhaitable ; +2 = indispensable.		

FIG. 3. — Format d'une unité d'investigation.

FIG. 3. — *Format of an investigation unit.*

<i>Si vous pensiez faire</i>	<i>Et qu'alors vous trouvez</i>	<i>L'effet sur la nécessité de réaliser ce geste est le suivant</i>
Option thérapeutique	Nouvelle information clinique ou résultat d'un autre examen	-2 -1 0 +1 +2
Entourez la proposition qui vous semble adéquate : -2 = absolument contre-indiqué ; -1 = peu utile ou plutôt néfaste ; 0 = non pertinent ; +1 = utile et souhaitable ; +2 = indispensable.		

FIG. 4. — Format d'une unité de thérapeutique.

FIG. 4. — *Format of a therapeutic unit.*

deuxième colonne visent à déterminer l'effet qu'aurait la découverte d'une nouvelle information sur ces options. Les réponses sont portées dans la troisième colonne, sur une échelle de Likert à cinq entrées.

Trois formats de questionnaires sont donc possibles : unités diagnostiques (*figure 2*), unités d'investigation (*figure 3*), unités de thérapeutique (*figure 4*).

Il est important de noter que l'information contenue dans les différents items d'une même vignette n'est pas cumulative. Chaque item demande une réflexion reposant sur son seul contenu, dans le contexte décrit dans la vignette.

Nous appellerons « unité » l'ensemble constitué par une vignette avec ses items. En docimologie, on appelle « item » les éléments d'un instrument de mesure. Dans le contexte des TCS de Neurochirurgie, nous avons choisi de créer 3 items par vignette, chaque item étant constitué par une hypothèse et par la nouvelle information correspondante (*figure 1*). Le terme « test de concordance de script » désigne l'ensemble de l'examen, qui comporte un certain nombre d'unités.

L'établissement de la grille de correction est confiée à un groupe d'experts (idéalement 10 ou plus). Les tests de concordance de script étant destinés à explorer des situations cliniques pour lesquelles les hypothèses diagnostiques et les options de prise en charge ne sont ni univoques ni consensuelles, il n'est pas demandé aux experts de s'entendre sur une « bonne réponse » : toutes les réponses données sont prises en compte dans l'établissement de la grille.

Pour chaque item, la valeur accordée à chacune des réponses sur l'échelle de Likert est fonction du nombre d'experts qui l'ont choisie. Les candidats qui donnent la réponse qui a été choisie par le plus grand nombre d'experts (réponse modale) reçoivent un point, les réponses choisies par un plus petit nombre d'experts reçoivent un crédit qui correspond au nombre d'experts qui ont donné cette réponse divisé par le nombre d'experts de la réponse modale. Les réponses qui n'ont été choisies par aucun expert reçoivent 0. Par exemple, si, pour un item, 8 experts (sur 10) répondent « + 1 » et 2 experts « + 2 » (réponse modale), 1 point sera attribué à la réponse « + 1 » (8/8) et 0,25 point sera attribué à la réponse « + 2 » (2/8). Les autres réponses (-2, -1 ou 0) sont cotées 0.

LE TEST DE CONCORDANCE DE SCRIPT EN NEUROCHIRURGIE

Les internes inscrits en DES de Neurochirurgie sont actuellement peu nombreux : cinquante environ. À la suite d'une étude réalisée à partir

d'un questionnaire anonyme portant sur les besoins ressentis par les internes, le Collège des Enseignants de Neurochirurgie a proposé la mise en place d'une auto-évaluation des étudiants au cours de leur cursus [10]. En effet, il est apparu, à la lecture des résultats de ce questionnaire, que les internes ressentaient, du fait de leur petit nombre et de leur éloignement géographique, le besoin de mesurer au cours de leur formation leur niveau d'expertise, tant par rapport aux autres internes que par rapport au niveau de connaissances requis en fin d'internat.

L'intérêt d'une évaluation formative, c'est-à-dire une évaluation qui vise à donner aux étudiants des informations quant à leur progression en cours de cursus est aujourd'hui largement reconnu [7, 12] : l'étudiant peut (sous la supervision de son tuteur dans le cas de la Neurochirurgie) reconnaître ses points forts, mais aussi ses faiblesses, et recentrer si nécessaire ses apprentissages. En Neurochirurgie, le TCS est utilisé comme outil d'une évaluation formative, et non sanctionnelle. Par ailleurs, l'architecture du test permet son utilisation dans le cadre d'une auto-évaluation donnée sur Internet.

En Neurochirurgie, il a été décidé, pour des questions de faisabilité, de limiter la taille de chaque test de concordance de script à 30 unités, soit 90 items, auxquelles l'interne doit répondre en une heure au maximum. Le passage du test se fait par Internet, à partir du site du campus numérique de Neurochirurgie (www.campus-neurochirurgie.org). Un logiciel a été développé au préalable par le département communication, audiovisuel et multimédia (DCAM) de l'Université Bordeaux-2. Le responsable du test choisit initialement, dans la banque de questions, les 30 unités qu'il retiendra pour le test. Ce choix peut se faire par présentation clinique, par domaine de la spécialité, ou de façon aléatoire.

L'étudiant accède au site Internet en utilisant un identifiant propre. Le score obtenu lui est indiqué à l'issue du test. Le test devant être passé par l'ensemble des internes, le site reste accessible pendant une période de temps donnée (une semaine). Chacun peut se connecter, quand il le souhaite, pendant cette période. Cette possibilité de passer le test sur une période d'une semaine seulement a été décidée afin de permettre de présenter les résultats de chaque interne par rapport à l'ensemble du groupe, ainsi qu'à ceux des autres internes plus ou moins avancés dans leur formation.

LES ÉTAPES DE LA RÉALISATION

PREMIÈRE ÉTAPE : DÉFINITION DES DIFFÉRENTS DOMAINES DE LA SPÉCIALITÉ

La population-cible et le domaine à évaluer étant connus, la première étape consistait en la

TABLEAU I. — Liste des 18 domaines de Neurochirurgie.

TABLE I. — *The 18 domains of neurosurgery.*

1. Traumatologie : crâne
2. Traumatologie : rachis
3. Tumoral : crâne
4. Tumoral : rachis
5. Vasculaire : crâne
6. Vasculaire : rachis
7. Dégénératif
8. Fonctionnel
9. Pédiatrie
10. Pathologie nerveuse périphérique
11. Malformatif
12. Hydrocéphalie
13. Infections et parasitoses : crâne
14. Infections et parasitoses : rachis
15. Complications de la chirurgie
16. Diagnostics non chirurgicaux
17. Investigation
18. Localisation

définition des différents domaines de la spécialité, c'est-à-dire des grandes catégories diagnostiques qui permettent de classer les items par étiologie (nosographie). Ce travail a été effectué par le bureau du Collège, en janvier 2002. Quinze domaines ont initialement été retenus. Nous avons été amenés, dans un second temps, à ajouter trois nouveaux domaines (*tableau I*), ne rentrant dans aucun domaine de la spécialité mais fréquemment retrouvés en clinique :

— diagnostics non chirurgicaux (n° 16) : neurologiques ou rhumatologiques, la plupart du temps,

— investigation (n° 17) : il s'agit de questions portant sur la stratégie d'investigation permettant d'arriver au diagnostic,

— localisation (n° 18) : la localisation d'une lésion peut ne faire référence à aucun diagnostic.

DEUXIÈME ÉTAPE : DÉFINITION DES PRÉSENTATIONS CLINIQUES

Nous nous sommes attachés à définir les présentations cliniques considérées comme les plus représentatives de la Neurochirurgie.

Les présentations cliniques qui devaient être retenues ont été sélectionnées selon la méthode « PUIGE » : prévalence — urgence — intervention thérapeutique possible — gravité — exemplarité (c'est-à-dire valeur didactique).

Trente-deux présentations ont été retenues initialement ; deux ont été ajoutées secondairement (*tableau II*).

TABLEAU II. — Liste des présentations cliniques retenues.

TABLE II. — *Clinical presentations retained.*

1. Mouvements anormaux
2. Troubles cognitifs
3. Syndrome confusionnel
4. Troubles de la vigilance et coma
5. Main douloureuse insomnante
6. Lombalgies et dorsalgies chroniques
7. Lombo-radiculalgies
8. Névralgie cervico-brachiale
9. Algie faciale
10. Céphalées
11. Syndrome méningé
12. Traumatisme crânien de l'enfant
13. Traumatisme crânien de l'adulte
14. Traumatisme rachidien
15. Complications post-opératoires chez un opéré crânien ou rachidien
16. Paralyse faciale
17. Hémiplégie
18. Monoplégie
19. Paraplégie et tétraplégie (hors traumatismes)
20. Suivi d'une tumeur cérébrale
21. Découverte fortuite d'une image au scanner
22. Crise convulsive inaugurale
23. « Un signe, une image » : corrélation radio-clinique
24. Troubles de la vision et exophtalmie
25. Surdit� unilatérale et acouphènes
26. Déformation crânio-faciale
27. Anomalie du périmètre crânien
28. Écoulement liquidien par le nez ou les oreilles
29. Déformation rachidienne douloureuse
30. Dysraphisme et anomalie de charnière
31. Troubles sphinctériens
32. Troubles de la marche et de la statique
33. Tumeurs de la voûte du crâne
34. Anosmie

Les 4 à 5 hypothèses diagnostiques les plus pertinentes pour chacune des présentations cliniques ont ensuite été définies. La définition des gestes cliniques, de la stratégie d'investigation et de la stratégie thérapeutique pour chaque présentation a été réalisée ultérieurement.

TROISIÈME ÉTAPE : RÉDACTION DES TCS

La rédaction des vignettes cliniques et des items correspondants a été réalisée par deux des auteurs entre mai et octobre 2002. Ce travail a été réalisé au cours d'un préceptorat de pédagogie à l'Unité de Recherche et de Développement en Éducation des Sciences de la Santé (URDESS) de l'Université de Montréal.

Pour chaque présentation clinique, ont été définis, dans la continuité des travaux du Collège des Enseignants, les signes cliniques (confirmant ou infirmant le diagnostic) qui doivent être recherchés à l'interrogatoire et à l'examen clinique, la stratégie d'investigation ainsi que les résultats

attendus des examens complémentaires, enfin les schémas thérapeutiques envisageables.

Le test de concordance de script est un outil en devenir ; il est donc apparu nécessaire de formaliser autant que possible les règles à suivre pour la rédaction des unités. Ainsi, il a été décidé de travailler exclusivement sur la base des trois formats présentés plus haut (vignettes diagnostiques, vignettes d'investigation, vignettes de traitement). Le nombre d'items a également été fixé à trois pour les différentes unités, alors que la plupart des travaux réalisés jusqu'alors utilisaient un nombre variable d'items. En dehors d'un souci évident de lisibilité, nous espérons ainsi faciliter le travail d'interprétation des résultats du test.

Nous avons relevé quatre difficultés particulières au cours de la rédaction des vignettes et des items :

1) Le TCS est utilisable pour l'exploration de *problèmes cliniques mal définis*. Il est intéressant dans des cas où existe un certain degré d'incertitude, c'est-à-dire lorsque les données à la disposition du clinicien sont incomplètes ou ambiguës. Cela exclut toutes les questions (notamment cliniques) auxquelles on peut répondre par « oui » ou « non ». De telles questions sont du ressort d'autres modes d'évaluation, notamment les QCM. Tous les cas pour lesquels il existait une solution objective et consensuelle ont donc été éliminés.

2) Le TCS ne permet pas d'explorer des *présentations cliniques très inhabituelles ou atypiques* : le clinicien, n'ayant pas rencontré auparavant de situation similaire, ne peut pas faire appel à des scripts, c'est-à-dire à des réseaux de connaissances déjà organisées [4]. On sait d'ailleurs que, quel que soit le mode d'évaluation, les questions les plus difficiles ne sont pas les plus discriminantes.

3) Nous avons relevé une relative difficulté d'*écriture des vignettes diagnostiques* faisant appel à des éléments cliniques strictement neurochirurgicaux. En effet, ces vignettes font intervenir très largement des diagnostics différentiels neurologiques et rhumatologiques. Un diagnostic n'est qu'exceptionnellement neurochirurgical d'emblée, et une vignette d'orientation diagnostique peut très facilement ne comporter que des diagnostics neurologiques. Ce problème est toutefois probablement spécifique de la Neurochirurgie, et non transposable à d'autres spécialités.

4) *L'estimation du niveau de difficulté* des items au cours de leur rédaction reste un des problèmes majeurs. Il est en effet délicat d'estimer *a priori* quel est le degré d'expertise requis pour résoudre tel ou tel item. Le risque récurrent est de rédiger un item non discriminant parce que trop facile ou trop difficile. Seule l'analyse docimologique *a posteriori* permet d'éliminer de tels items.

TABLEAU III. — Tableau de spécification, reprenant l'exemple de l'unité 8;8 (tableau II).

TABLE III. — *Specification table, recapitulating the example in unit 8.8 (table II).*

Identification de la vignette	Champ professionnel	Domaines
<p>8; 8</p> <p>8^e vignette de la série</p> <p>Présentation clinique n° 8 (névralgie cervico-brachiale)</p>	<p>— examen clinique et interrogatoire</p>	<p>7; 16; 4</p> <p>1^{er} item (dégénératif)</p> <p>2^e item (diagnostics non chirurgicaux)</p> <p>3^e item (tumoral rachis)</p>

QUATRIÈME ÉTAPE : MISE AU POINT DU TABLEAU DE SPÉCIFICATION

Le modèle de tableau de spécification retenu (*tableau III*) classe les unités selon le contenu des items reliés.

Dans la première colonne, les vignettes sont désignées par deux numéros : le premier est celui de la présentation clinique ; le second identifie les différentes vignettes correspondant à cette présentation clinique.

Dans la deuxième colonne est indiqué le « champ professionnel ». Celui-ci désigne le format de la vignette, et donc le domaine de compétence exploré. Nous distinguerons donc les formats suivants :

— questionnaires cliniques : l'information apportée dans la deuxième colonne peut être un élément clinique ou un résultat d'examen complémentaire,

— questionnaires d'investigation (*figure 3*),

— questionnaires de thérapeutique (*figure 4*).

Enfin, dans la troisième colonne sont indiqués successivement les numéros de domaine correspondant à chacun des trois items (*tableau I*).

Le contenu des vignettes a pu être équilibré entre les différents domaines en fonction des données du tableau de spécification. Le nombre total de vignettes rédigées est de 234, soit 702 items.

CINQUIÈME ÉTAPE : CORRECTION ET VALIDATION DES UNITÉS

Les vignettes et les items rédigés ont été envoyés pour correction aux membres du bureau du Collège des Enseignants, plusieurs réunions de formation pédagogique ayant été organisées au préalable afin d'aider les correcteurs.

Une grille de correction a par ailleurs été proposée (*tableau IV*) comme support de cette tâche.

TABLEAU IV. — Grille de correction.

TABLE IV. — *Scoring chart.*

La vignette clinique (énoncé du problème)	La vignette est-elle correctement rédigée ? (incorections, longueur, confusion)
	La vignette correspond-elle à une situation clinique crédible ?
	S'agit-il d'une présentation clinique usuelle ? (la situation ne doit pas être un piège)
Les items (individuellement)	Les hypothèses diagnostiques (ou les options de prise en charge) qui sont proposées sont-elles pertinentes ?
	Les items correspondent-ils à des questions que l'on se pose réellement en situation clinique ?
	Le niveau de difficulté correspond-il à ce qu'un interne devrait savoir ? (item trop facile ou trop difficile)

Lors de cette étape de relecture des vignettes et des items par des experts du domaine, les correcteurs doivent vérifier que la vignette est compréhensible et correctement formulée, que la situation décrite est une présentation classique. Ils doivent ensuite vérifier la pertinence des hypothèses diagnostiques ou des options de prise en charge qui sont formulées dans les items. Il est à noter que les correcteurs doivent non seulement souligner ce qui leur paraît inexact, contestable ou mal formulé, mais aussi proposer un élément qui leur semble plus adapté, afin d'éviter une seconde correction.

Selon le nombre de vignettes, ce travail de correction peut être considérable. La répartition a permis de ne confier à chaque correcteur qu'un nombre limité d'unités.

SIXIÈME ÉTAPE : CONSTRUCTION DES GRILLES DE RÉPONSE

Cette dernière étape a pu être réalisée lors du colloque annuel du Collège des Enseignants de Neurochirurgie, en janvier 2003. Dix-neuf experts ont été répartis en trois groupes, chacun ayant en charge la cotation de 78 unités. Les experts ont passé les tests dans les mêmes conditions que les étudiants, leurs réponses permettant d'établir la grille de correction selon le modèle décrit plus haut.

CONCLUSION

Le test de concordance de script est un outil docimologique novateur, permettant d'évaluer le raisonnement clinique en contexte d'incertitude, c'est-à-dire lorsque le clinicien ne dispose pas de toutes les données permettant la résolution du problème. En Neurochirurgie, le TCS est utilisé comme instrument d'évaluation formative, destiné aux internes DES et DIS. Il permet à l'étudiant de se situer par rapport à la référence que constitue le panel d'experts, mais aussi par rapport aux autres étudiants, de même niveau ou de niveaux différents. Il permet surtout à chacun de suivre sa propre évolution au cours de son cursus, pour les différents domaines de la spécialité. Son informatisation sur le site du campus numérique de Neurochirurgie lui confère, par ailleurs, une grande souplesse de mise en œuvre et d'utilisation.

Le travail qui a été réalisé a permis la mise en place du cadre de cette nouvelle évaluation, ainsi que la rédaction d'une première série d'unités. Un premier test de concordance de script composé de 30 unités portant sur trois domaines différents a ainsi pu être organisé au cours d'une Journée Nationale d'Enseignement.

L'implication de la communauté neurochirurgicale dans ce projet est indispensable, pour en assurer la pérennité et la diffusion, mais aussi la qualité, en permettant la rédaction de nouvelles questions.

Remerciements. — Deux des auteurs ont bénéficié de bourses du Collège de Enseignants de Neurochirurgie pour réaliser ce travail.

RÉFÉRENCES

- [1] BORDAGE G. Elaborated knowledge : a key to successful diagnostic thinking. *Acad Med* 1994 ; 69 : 883-5.
- [2] BRAILOVSKI 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.
- [3] CHARLIN B, BORDAGE G, VAN DER VLEUTEN C. L'évaluation du raisonnement clinique. *Pédagogie Médicale* 2003 ; 4 : 42-52.
- [4] CHARLIN B, TARDIF J, BOSHUIZEN HP. Scripts and medical diagnostic knowledge : theory and applications for clinical reasoning instruction and research. *Acad Med* 2000 ; 75 : 182-90.
- [5] CHARLIN B, VAN DER VLEUTEN C. Standardized assessment in context of uncertainty : the script concordance approach. Evaluation and the health professions. Submitted, July 2003.
- [6] CHARLIN B, GAGNON R, SIBERT L, VAN DER VLEUTEN C. Le test de concordance de script, un instrument d'évaluation de raisonnement clinique. *Pédagogie Médicale* 2002 ; 3(3) : 136-144.
- [7] CROSSLEY J, HUMPHRIS G, JOLLY B. Assessing health professionals. *Med Educ* 2002 ; 36 : 800-5.
- [8] LEGENDRE R. Dictionnaire actuel de l'éducation. 1^{re} édition. Paris : Larousse, 1988.
- [9] 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(2) : 80-88.
- [10] MOREAU JJ. Améliorer la formation en Neurochirurgie. *Neurochirurgie* 2002 ; 48 : 307-8.
- [11] SCHMIDT HG, NORMAN GR, BOSHUIZEN HP. A cognitive perspective on medical expertise : theory and implication. *Acad Med* 1990 ; 65 : 611-21.
- [12] WASS V, VAN DER VLEUTEN C, SHATZER J, JONES R. Assessment of clinical competence. *Lancet* 2001 ; 357 : 945-9.