

# Le test de concordance de script en 20 questions

## *Twenty questions on script concordance tests*

Didier GIET<sup>1,3</sup>, Valérie MASSART<sup>1</sup>, Robert GAGNON<sup>2</sup> et Bernard CHARLIN<sup>2</sup>

<sup>1</sup> Département de médecine générale (DUMG), Université de Liège, Liège, Belgique

<sup>2</sup> Centre de pédagogie appliquée aux sciences de la santé (CPASS), Faculté de médecine, Université de Montréal, Montréal, Canada

<sup>3</sup> Institut de formation et de recherche en enseignement supérieur (IFRES-ULg), Université de Liège, Liège, Belgique

Manuscrit reçu le 26 juillet 2012 ; commentaires éditoriaux formulés aux auteurs le 14 octobre 2012 ; accepté pour publication le 30 octobre 2012

### Mots-clés

Test de concordance de script ; construction ; évaluation ; raisonnement clinique ; médecine générale

### Keywords

Script concordance test; design; assessment; clinical reasoning; family practice

**Résumé – Objectif :** Le présent article se donne comme objectif de répondre, en 20 questions-réponses, aux principales préoccupations que les équipes d'enseignants potentiellement intéressées par le test de concordance de script (TCS) peuvent se poser. Les questions et objections souvent rapportées par les étudiants sont également prises en compte. **Exégèse :** Les principes qui sous-tendent le TCS sont rappelés. Ses atouts et ses limites sont évoqués. Les étapes de construction, de validation, de passation, de calcul de la note sont développées. Les auteurs illustrent le propos en s'appuyant sur une expérience concrète menée au Département de médecine générale de l'Université de Liège.

**Abstract – Objective:** Most teacher concerns about Script Concordance Tests (SCT) are the object of some 20 questions and answers. The questions and objections most commonly reported by students are also taken into account. **Analysis:** The principles of SCT are described. Their strengths and limitations are discussed. The steps in designing, validating, testing and calculating scores are developed. The authors depict the concepts based an experiment led in the Department of Family Practice at the University of Liège.

## Introduction

De nombreux responsables de programmes de formation des médecins ou d'autres professionnels de santé s'interrogent sur la pertinence de leur dispositif d'évaluation des apprentissages des étudiants. Le plus souvent ils souhaitent améliorer la validité ou la fidélité des procédures d'évaluation, avec le souci que ces dispositifs soient équitables et que, grâce à des outils appropriés, ils évaluent autre chose que des connaissances factuelles, en l'occurrence des capacités

complexes telles que le raisonnement clinique qui reste peu exploré spécifiquement. Parfois des considérations logistiques et organisationnelles motivent également la réflexion, notamment en raison de l'accroissement du nombre d'étudiants à évaluer. Ceci peut conduire à souhaiter pouvoir utiliser des procédés d'évaluation dont la correction est automatisable.

Depuis quelques années, le panel des méthodes d'évaluation des apprentissages des étudiants dans le domaine de la santé s'est enrichi du test de concordance de script (TCS). Le TCS est une méthode

innovante d'évaluation du raisonnement clinique, documentée par de nombreuses publications, résumées dans les références citées<sup>[1-6]</sup> et utilisée dans le cadre de plusieurs champs professionnels de la formation en santé (chirurgie, gynécologie, radiologie, médecine d'urgence, maïeutique, art infirmier infirmières, etc.).

Le site du Centre de pédagogie appliquée aux sciences de la santé de la Faculté de médecine de l'Université de Montréal offre un accès aisé aux publications consacrées au TCS démontrant sa fidélité et sa validité (<http://www.cpass.umontreal.ca/tcs.html>).

L'objectif de l'article est de répondre, en 20 questions-réponses, aux principales préoccupations que les équipes d'enseignants potentiellement intéressées par le TCS peuvent se poser. Les objectifs et questions évoqués par les étudiants sont également pris en compte. Les questions retenues abordent successivement les principes qui sous-tendent le TCS, ses atouts et ses limites puis les étapes de construction, de validation, d'administration et de calcul du score. Pour ce faire, les auteurs s'appuient sur une expérience concrète menée au Département de médecine générale de l'Université de Liège (DUMG-ULg).

## Question 1

### Quels sont les principes du test de concordance de script ?

Les situations quotidiennes en pratique clinique sont caractérisées par le fait que certaines informations manquent et qu'il va falloir aller les rechercher ou encore que certaines données disponibles sont ambiguës<sup>[1]</sup>.

C'est ainsi que le TCS s'appuie sur une vignette clinique authentique, qui rend compte d'une situation professionnelle courante dans la discipline.

Chaque cas constitue un problème, même pour un médecin expérimenté, car les données présentées sont insuffisantes, soulèvent une ambiguïté ou encore exposent à des conflits de valeurs<sup>[2]</sup> (un conflit de valeurs en situation professionnelle peut être par exemple illustré par une décision à prendre, face à un

patient, en tenant compte, d'une part, de la situation globale et complexe de ce patient et d'autre part, des recommandations de bonne pratique).

On demande à l'étudiant de poser un micro-jugement en matière diagnostique, d'investigation ou de traitement, lorsque des éléments d'information supplémentaires lui sont communiqués. La tâche consiste donc à envisager l'effet que produirait la découverte d'une nouvelle donnée sur le statut d'une des options pertinentes dans la situation, présentée comme hypothèse au candidat.

Le candidat doit choisir entre les différentes propositions formulées à l'aide d'une échelle de Likert à une variable et à cinq échelons allant de « beaucoup moins probable » à « beaucoup plus probable »<sup>[3]</sup> ou encore de « la moins appropriée » à « la plus appropriée ».

Un exemple de TCS explorant un problème thérapeutique est donné dans le tableau I. Après l'énoncé d'une situation clinique, trois hypothèses indépendantes sont présentées à l'étudiant ; il doit les envisager séparément, l'une après l'autre.

En effet, il s'agit d'hypothèses réalistes possiblement émises par un médecin amené à raisonner à propos d'une telle situation clinique. Dans son raisonnement, ces hypothèses sont envisagées indépendamment les unes des autres.

Le candidat est évalué sur base d'un score qui mesure le **degré de concordance** entre ses réponses et celles d'un panel de référence constitué de médecins expérimentés dans le domaine envisagé.

La grille de réponse des experts constitue la référence pour déterminer le score des candidats. Cette grille peut être optimisée selon différentes méthodes décrites plus loin et proposées dans la littérature<sup>[12,20]</sup>.

Le score du candidat est déterminée par une méthode de pondération des choix de réponse. Pour chaque item, les réponses données par les experts du panel sont prises en compte pour établir le score du candidat. À cet effet, on définit un « crédit » pour chaque niveau de l'échelle de Likert ; celui-ci représente le score qu'un candidat peut obtenir s'il choisit la catégorie correspondante. Dans la méthode habituelle, un crédit de 1 est accordé au(x) niveau(x) de l'échelle de Likert choisi(s) par le plus grand nombre d'experts, les autres niveaux recevant un

**Tableau I.** Exemple de test de concordance de script explorant un problème thérapeutique et dont les données présentées exposent successivement l'apprenant à un conflit de valeurs, à des données insuffisantes et à une ambiguïté.

*Un enfant de 3 ans présente une température mesurée à 39 °C depuis la veille et une toux très importante. Il présente également une allergie aux poils de chien. Toute la famille a passé le week-end en présence du chien d'un ami.*

<i>Si vous pensiez donner</i>	<i>Et que vous apprenez que</i>	<i>L'attitude envisagée devient</i>				
un antipyrétique et un antitussif	son auscultation est peu contributive mais ses parents expriment la crainte d'un foyer infectieux	-2	-1	0	+1	+2
un antipyrétique et un antihistaminique	son auscultation révèle des râles crépitants à la base droite	-2	-1	0	+1	+2
un antibiotique et un antihistaminique	il souffre d'un reflux gastro-œsophagien	-2	-1	0	+1	+2
-2 la moins appropriée -1 moins appropriée	0 ni plus ni moins appropriée (pertinence de la décision envisagée non influencée par la nouvelle donnée)	+1 plus appropriée +2 la plus appropriée				

**Tableau II.** Exemple de test de concordance de script explorant un problème diagnostique.

*Vous êtes appelé à domicile comme médecin de garde car une dame de 54 ans a fait un malaise à la suite d'une douleur épigastrique intense. Elle vous signale souffrir d'ulcères gastriques récidivants depuis des années.*

<i>Si votre hypothèse est</i>	<i>Et que vous apprenez que</i>	<i>La nouvelle information...</i>				
une récurrence d'ulcère gastrique	la douleur était post-prandiale	-2	-1	0	+1	+2
un infarctus du myocarde	la douleur n'est pas apparue à l'effort	-2	-1	0	+1	+2
une lithiase vésiculaire	la patiente présente une hypercholestérolémie	-2	-1	0	+1	+2
-2 : rend l'hypothèse diagnostique beaucoup moins probable (hypothèse pratiquement éliminée)		0 : n'a aucun effet sur l'hypothèse diagnostique		+1 : rend l'hypothèse diagnostique plus probable		
-1 : rend l'hypothèse moins probable		+2 : rend l'hypothèse beaucoup plus probable (il ne peut s'agir pratiquement que de cette hypothèse)				

crédit inférieur en fonction du nombre d'experts les ayant choisis. Le score final du candidat est alors la somme des scores obtenus aux items. Ramenée en pourcentage, cette somme est l'équivalent de la moyenne des scores.

## Question 2

### Qu'évalue le test de concordance de script ?

Le TCS mesure une dimension du raisonnement médical, liée à l'interprétation des données dans des contextes de situations authentiques de la pratique.

Les scores obtenus reflètent donc fondamentalement la capacité d'interprétation des données

cliniques. Cette capacité, qui est au cœur du raisonnement clinique dans son ensemble, est cependant liée à d'autres variables. En tant que composante du jugement clinique, son exercice exige la maîtrise d'une base de connaissances cliniques riche et structurée. Les scores fournissent donc des informations relatives à la fois à la capacité d'interprétation des données cliniques et à ces autres dimensions.

Le TCS permet ainsi d'évaluer le candidat dans un contexte d'incertitude<sup>[1,2,4,5]</sup>. Plus précisément, c'est un micro-jugement, à un moment donné du raisonnement clinique du médecin, qui est évalué. En effet, lorsqu'un médecin est confronté à une situation clinique, il mobilise des réseaux de connaissances préétablis (des scripts), qui sont utilisés pour analyser cette situation et agir en fonction du but qui peut être

l'investigation, diagnostic ou le traitement. Ce sont ces associations, ces liens, qui permettent de prendre des décisions qui s'appuient sur la force ou la faiblesse d'une hypothèse<sup>[4]</sup>. L'objectif d'une question dans le cadre d'un TCS n'est donc pas, pour la situation énoncée, d'aboutir au diagnostic ou à la décision d'investigation ou de traitement : il s'agit d'analyser une étape au sein du processus de raisonnement.

Le TCS étudie donc véritablement chez les candidats une série de micro-jugements, qui font partie intégrante du processus de raisonnement clinique. C'est ainsi la qualité de la séquence d'une multitude de micro-jugements situés au cœur du processus du raisonnement clinique qui est évaluée.

### **Question 3** **Le test de concordance de script** **remplace-t-il d'autres modes** **d'évaluation ?**

Le TCS explore le raisonnement clinique en situation d'incertitude. Il permet d'enrichir de façon complémentaire l'évaluation des apprentissages, en tant que moyen d'approche du raisonnement face à des problèmes mal définis<sup>[1]</sup>.

Le TCS ne remplace donc pas d'autres modes d'évaluation traditionnels comme, par exemple, le questionnaire à choix multiple (QCM) qui explore les connaissances liées à des problèmes cliniques bien définis<sup>[2]</sup>.

À titre d'exemple, le Département de médecine générale a introduit le TCS en complément d'un questionnaire « vrai-faux » (évaluation de connaissances factuelles), de postes d'examen clinique objectif structuré (ECOS) et d'entrevues médicales simulées (EMS) (évaluation de compétences complexes, d'actions prédéfinies).

Plus précisément, le questionnaire « vrai-faux » est composé de 200 questions indépendantes pour lesquelles l'étudiant doit identifier la véracité ou la fausseté de chaque proposition. Des connaissances factuelles telles que l'application de la législation de délivrance de médicaments ou du règlement relatif à

l'interruption du temps de travail, par exemple, sont ainsi évaluées. Des connaissances médicales non discutables sont également évaluées par ce questionnaire.

Les situations d'ECOS et d'EMS consistent en 18 box de consultations par lesquels chaque étudiant passe successivement. Dans les ECOS, en sept minutes, il doit résoudre une partie de consultation face à un patient standardisé (une anamnèse, un examen clinique, la rédaction d'une ordonnance...). Dans les EMS, en 15 min, il doit gérer la prise en charge globale d'un problème soumis par un patient. Des compétences de communicateur, de clinicien, de promoteur de la santé, d'engagement éthique, etc. sont ainsi évaluées, sur base d'une grille d'observation critériée.

### **Question 4** **Qui rédige les questions d'un test** **de concordance de script ?**

Ce sont des spécialistes de la discipline qui rédigent les questions : des médecins généralistes pour un TCS en médecine générale, des anesthésistes pour un TCS en anesthésie, etc.

Dans l'expérience qui nous occupe, neuf médecins généralistes-enseignants ont rédigé chacun 10 vignettes cliniques. Celles-ci ont été relues (relecture croisée) et amendées par tous les enseignants. Plusieurs vignettes ont été éliminées suite à cette première relecture en fonction des objectifs visés dans le TCS, de la qualité de rédaction et de la prévalence des pathologies.

Un des enseignants médecin généraliste à Liège explique avoir pris son agenda de consultations, identifié les patients rencontrés la semaine précédente au cabinet et rédigé une courte vignette pour quelques-uns d'entre eux : quelles sont les caractéristiques des patients ? Qu'est-ce qui les a amenés au cabinet ? Quelles informations ont été révélées par l'anamnèse ? Quelles ont été les premières hypothèses diagnostiques, thérapeutiques ou d'investigation du médecin suite à l'anamnèse même parcellaire ?

## Question 5

### Comment s'assurer que le questionnaire est représentatif de la discipline ?

Outre les recommandations formulées sur le site de l'Université de Montréal<sup>[13]</sup>, voici quelques actions proposées pour aider à construire un test représentatif de la discipline visée :

- lors de la rédaction de chaque question, les enseignants ciblent des situations courantes, prévalentes, représentatives de leur pratique professionnelle (ils ne ciblent pas une situation exceptionnelle d'un patient).
- Une relecture du questionnaire complet doit être effectuée par au moins un ou deux enseignants. Ils doivent lister les thématiques abordées dans chacune des questions. Ils veilleront à supprimer les redondances éventuelles et vérifieront que les domaines principaux de la discipline soient abordés.

En médecine générale, les enseignants se sont assurés que le test était composé équitablement des différents domaines de la discipline et qu'aucun n'était surreprésenté.

## Question 6

### De combien de questions doit être composé le TCS ?

Les études<sup>[6]</sup> montrent qu'il faut une soixantaine de questions (hypothèses) pour atteindre une bonne fidélité de mesure (coefficient de consistance interne au moins supérieur à 0,70).

Le TCS en médecine générale présente 45 cas cliniques composés chacun de trois hypothèses soit 135 questions (hypothèses diagnostiques, d'investigation ou de traitement).

## Question 7

### Des médecins généralistes chevronnés réussissent-ils mieux le TCS que les étudiants ?

Les études psychométriques publiées sur le TCS à partir de 1998 ont vérifié sa validité de construit<sup>[1]</sup>

(adéquation entre les scores obtenus à un test et le cadre théorique sur lequel il est construit) : ces études confirment l'hypothèse selon laquelle les médecins expérimentés obtiennent des scores supérieurs à ceux des internes et des résidents, qui sont eux-mêmes supérieurs à ceux des externes. Ces observations apportent ainsi des arguments en faveur de la validité de construit.

Les tests usuels, comme un questionnaire à choix multiple, donnent des résultats paradoxaux : « ce ne sont pas les professionnels expérimentés qui obtiennent les meilleurs scores, mais les personnes de niveau intermédiaire : les résidents en fin de formation, proches de leurs examens de certification »<sup>[6]</sup>. Ce phénomène est appelé « effet intermédiaire ».

Or, pour le TCS, cet effet intermédiaire disparaît<sup>[1]</sup> illustrant une acquisition progressive à la faveur de la pratique professionnelle.

## Question 8

### Le test de concordance de script peut-il être utilisé à la fois dans le cadre d'une épreuve certificative et dans celui d'une épreuve formative ?

Le TCS peut être utilisé dans une démarche certificative comme dans une démarche formative<sup>[2]</sup>.

Cet objectif doit être précisé avant la construction du TCS car la méthode d'élaboration du TCS peut varier (voir partie « construction du TCS » ci-dessous).

Dans une démarche formative, l'utilisation de boîtiers de vote peut s'avérer intéressante pour visualiser immédiatement les choix exprimés par l'auditoire. Le débat peut ensuite être ouvert une fois que ceux-ci sont confrontés à l'avis des experts. Ainsi, certains auteurs utilisent le TCS en formation médicale continue<sup>[7,8]</sup>.

L'expérience liégeoise a permis d'identifier des étudiants dont le raisonnement clinique posait problème et de mettre en place secondairement un accompagnement dans la suite de leur formation. Dans ce contexte, l'usage du TCS est ainsi certificatif et formatif.

### Question 9

#### Dans un test de concordance de script certificatif, comment détermine-t-on le niveau de réussite d'un candidat ?

Dans l'hypothèse de l'usage certificatif du TCS, il convient de fixer le niveau en deçà duquel la performance de l'étudiant est jugée problématique.

En fait, la détermination du score de passage est le point problématique du TCS. Il n'y a pas de méthode qui fasse consensus.

Les auteurs du TCS à l'Université de Montréal ont choisi la conduite suivante pour sanctionner les résidents en médecine de famille<sup>[9]</sup> : « Un résident est une personne qui aspire à faire partie de la population des médecins de famille. La moyenne et l'écart-type du panel de référence (ce panel de référence étant assimilé à un échantillon de la population des médecins de famille) sont calculés. Par analogie, avec les méthodes statistiques d'appartenance à un groupe, le candidat qui prétend faire partie de cette population ne peut se situer en-dessous de deux écart-types en dessous de la moyenne du panel ». Cette méthode de détermination de la réussite au TCS est objective, mais le choix de deux écart-types est, lui, arbitraire.

### Question 10

#### Quelle est la bonne réponse à une question de test de concordance de script ?

Il n'y a pas qu'une seule réponse correcte à une question de TCS. C'est là l'intérêt de ce type d'évaluation.

La situation proposée prête à discussion, plusieurs actions sont possibles. Dans une telle situation d'incertitude, il n'existe pas une seule manière d'agir. Ces incertitudes sont illustrées par la répartition des réponses des experts sur différents échelons de l'échelle.

Chaque réponse d'expert est prise en compte et l'étudiant est crédité de points s'il se positionne comme l'a fait au moins un des experts. Ce mode de correction du TCS 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<sup>[1]</sup>.

Une autre méthode de correction consiste à pondérer la réponse de l'étudiant en fonction de la distance qui le sépare de la réponse modale. Un étudiant qui est proche de la réponse modale sera plus crédité qu'un étudiant qui est plus éloigné de la réponse modale<sup>[10]</sup>.

### Question 11

#### Comment détermine-t-on le score attribué à la réponse majoritaire ?

Classiquement, un point complet est attribué à l'étudiant qui choisit l'option prise par la majorité des experts (par exemple, les experts se sont répartis comme suit : 7 experts choisissent la proposition -1 et 4 experts choisissent la réponse -2. L'étudiant qui choisit comme 7 experts (la majorité) recueillent 7/7 point soit 1 point. Un étudiant qui choisit l'option -2 comme 4 experts recueillent 4/7 point soit 0,57 point.

On pourrait imaginer n'attribuer que des points partiels aux étudiants, proportionnellement à la répartition des experts c'est-à-dire qu'il serait envisageable d'attribuer 0,64 point à un étudiant qui a répondu comme 7 experts sur 11 et 0,36 point pour un étudiant qui a répondu comme 4 des experts sur 11. Ce barème pourrait être appliqué s'il est bien annoncé préalablement aux étudiants.

La méthode la plus classiquement employée, comme cela a été le cas au DUMG-ULg, permet d'attribuer un point maximum pour chaque question.

### Question 12

#### Quel serait le score obtenu par un étudiant qui répond au hasard ?

L'étudiant qui répond systématiquement au hasard peut être crédité de points, comme dans toute évaluation dont les questions sont fermées.

Dans un questionnaire « vrai-faux », la probabilité que l'étudiant fournisse une réponse correcte grâce au seul hasard est de un sur deux<sup>[11]</sup>.

Dans le cadre d'un TCS en médecine générale, Vanbelle<sup>[12]</sup> a calculé les scores qu'obtiendrait un étudiant s'il choisissait systématiquement un échelon de l'échelle. À titre d'exemple, si un étudiant répond systématiquement -2, -1, 0, +1 ou +2, Vanbelle<sup>[12]</sup> a calculé qu'il récolterait respectivement 0,28 ; 0,40 ; 0,64 ; 0,43 ou 0,14 point. Cet étudiant qui répondrait au hasard (c'est-à-dire qu'il a une chance sur cinq de choisir chaque niveau de l'échelle de Likert), obtiendrait un score de 0,38 point (soit 38 %).

Ce score donne la note minimale que peut obtenir un étudiant qui ne connaîtrait absolument rien et qui répondrait toujours au hasard. Il faut noter que si on établit les scores de passage à partir des écarts-types du panel, sachant que le score moyen des panels est très souvent autour de 80 et que l'écart-type est de l'ordre de 5, il n'y a aucune chance pour qu'un étudiant réussisse en répondant strictement au hasard.

### **Question 13** **Qui compose le panel de référence et combien de membres doit-il comprendre ?**

Il s'agit de spécialistes de la discipline qui exercent une « pratique validée ». Le choix de la composition du panel de référence dépend en effet de l'objet de l'évaluation<sup>[1]</sup>. Il convient de se demander quelle spécialité est confrontée dans sa pratique quotidienne aux cas présentés dans le TCS.

Dans l'expérience en médecine générale, il s'agissait de médecins généralistes maîtres de stage. Ces médecins reçoivent régulièrement des étudiants et futurs médecins généralistes en stage au cabinet. La plupart travaillent en groupe et rendent visible leur activité professionnelle. Les membres du panel ont une expérience professionnelle d'au moins 10 à 15 années. Il s'agit de médecins généralistes considérés comme chevronnés et attentifs aux recommandations de bonne pratique.

Gagnon et al. ont montré qu'un panel de référence devait comporter au moins une quinzaine de membres pour des TCS dont l'enjeu est certificatif<sup>[12,14]</sup>.

Pour les TCS à visée formative, sans enjeu de classement ou de certification, le panel peut être composé de 8 à 10 membres.

Le nombre de membres du panel de référence doit pouvoir refléter les variations d'opinions entre médecins expérimentés<sup>[1]</sup>.

### **Question 14** **Comment préparer les experts du panel de référence ?**

Avant de faire passer le TCS au panel d'experts, il est important de les réunir et de leur présenter la méthode.

Le traitement anonyme des données doit leur être garanti avant d'obtenir leur consentement. Il est important de spécifier que le TCS n'est pas une méthode de consensus et que toutes les réponses seront prises en compte. Lors de l'administration du test, il convient d'annoncer que les réponses sont individuelles et qu'aucun débat ou discussion ne doit avoir lieu.

Dans l'expérience en médecine générale, une quinzaine d'experts ont été réunis et les principes et objectifs du TCS leur ont été présentés. Une discussion a suivi. Des exemples de questions ont été proposés en groupe et chacun a pu s'essayer à la démarche particulière. En fin de réunion, le TCS a alors été soumis à l'ensemble des experts réunis dans un local commun, sans interaction entre eux.

### **Question 15** **Comment optimiser les réponses du panel d'experts ?**

Quand toutes les questions ont été relues et validées par les enseignants, elles sont soumises au panel d'experts.

Au terme de cette phase, en fonction des résultats enregistrés, certaines questions pour lesquelles les réponses d'experts divergent totalement peuvent être éliminées. Selon Charlin<sup>[6]</sup>, il s'agit vraisemblablement de questions mal construites. Elles sont donc à reformuler ou à éliminer.

Duggan<sup>[15]</sup>, lors de la construction d'un TCS en gynécologie à l'Université d'Adélaïde, a optimisé le panel des experts en conservant les questions dont les réponses traduisaient une concordance raisonnable entre experts, c'est-à-dire une distribution des réponses sur l'échelle de Likert autour de la réponse modale après suppression des valeurs visiblement aberrantes. Les questions considérées comme « discordantes » sont celles dont les valeurs sont bimodales, soit réparties sur toute l'échelle de Likert, soit centrées aux deux extrémités de l'échelle<sup>[16]</sup>. Ces questions ne sont pas considérées comme appropriées pour leur utilisation dans un examen certificatif ou dont l'objet est la recherche. C'est aussi le cas lorsqu'un des échelons de l'échelle de Likert est choisi par un seul expert.

Par ailleurs, une étude a comparé l'effet de différentes stratégies d'optimisation du panel sur la fiabilité des scores. Les auteurs concluent à un effet très négligeable du retrait de réponses aberrantes<sup>[20]</sup>.

### **Question 16** **Que faire d'une question pour laquelle les avis d'experts se concentrent sur un seul échelon de l'échelle de Likert ?**

Si tous les membres du panel donnent la même réponse, la question devient proche d'une question à choix multiple<sup>[6]</sup>. Il ne faut pas systématiquement éliminer ce type de questions. Il faut cependant veiller à ne pas multiplier leur nombre dans un TCS.

### **Question 17** **Combien de temps faut-il pour construire un test de concordance de script ?**

Si la construction d'un TCS relève d'un travail d'équipe, un temps d'appropriation de la méthode doit être prévu. L'élaboration du test peut prendre plusieurs semaines à plusieurs mois.

Dans le contexte très particulier du Département de médecine générale de Liège où les enseignants

partagent leur temps professionnel entre enseignement et pratique, l'expérience a montré que six à neuf mois étaient nécessaires pour construire un test. Un temps important a en effet été consacré à l'appropriation et à l'acceptation des principes du TCS par chacun.

Des experts du TCS peuvent construire, eux, un test de qualité en quelques jours à peine<sup>[17]</sup>.

### **Question 18** **Comment préparer les étudiants ?**

Avant d'administrer le TCS aux étudiants, une séance d'échange de vues sur les principes et modalités de l'épreuve, de une à deux heures, doit être organisée, afin de leur permettre de se familiariser au format singulier de l'épreuve.

En médecine générale, un mois environ avant le TCS, les étudiants sont réunis pour une présentation du cadre de l'évaluation et des fondements du TCS. Des exercices au moyen des boîtiers de vote sont proposés pour faciliter la discussion et le débat à propos de cette nouvelle méthode.

### **Question 19** **Quelles sont les objections exprimées par les étudiants à l'égard du test de concordance de script ?**

Les étudiants expriment souvent une difficulté à raisonner sur base de l'hypothèse qui leur est fournie. Celle-ci n'est effectivement pas toujours la plus évidente à leur esprit ou peut leur paraître inappropriée. Le candidat doit donc admettre l'hypothèse qui lui est présentée, la faire sienne et y confronter la nouvelle information.

Les étudiants vivent parfois difficilement l'obligation de trancher dans un contexte qui correspond à une « zone grise » pour laquelle il s'avère malaisé d'appliquer des règles de bonne pratique ou issues de la médecine factuelle.

Les étudiants rencontrent quelquefois des difficultés à interpréter la consigne relative au niveau

« zéro » de l'échelle de Likert. Choisir cette réponse ne constitue pas une modalité d'abstention mais signifie que la nouvelle donnée n'a pas d'effet sur l'hypothèse. Il est important de le leur rappeler dans les instructions précédant la passation du TCS.

Les étudiants expriment parfois leur réticence à être comparés à un panel d'experts praticiens de terrain d'une discipline. Ils s'interrogent sur ce qui lui confère une valeur d'exemple ou d'étalon. Il convient alors d'expliquer les critères de composition du panel, évoqués dans cet article.

## Question 20

### Combien de temps faut-il aux étudiants pour répondre à un test de concordance de script ?

D'une faculté à l'autre, le temps d'examen peut varier entre une et trois heures<sup>[18]</sup>. Il faut compter à peu près une heure pour répondre à un test composé de 75 questions.

Dans un contexte d'évaluation certificative, il faut une heure à une heure et demie aux étudiants pour répondre au TCS de médecine générale composé de 45 cas à trois questions chacun (aucune limite de temps ne leur est imposée). Nous observons une certaine lassitude après une bonne heure d'examen en raison de l'important niveau de concentration exigé par le TCS.

Par ailleurs, il ne semble pas souhaitable de multiplier les situations cliniques. Chaque nouvelle situation clinique nécessite effectivement un nouvel effort d'analyse du contexte. Ainsi, générer deux à trois hypothèses (au lieu d'une seule) par situation clinique épargne l'attention demandée aux étudiants<sup>[19]</sup>.

## Contributions

Valérie Massart et Didier Giet ont collationné les questions qui sont fréquemment posées au sujet du TCS par les enseignants et les étudiants. Les réponses ont été construites conjointement par l'ensemble des auteurs. La version finale du manuscrit a été approuvée par tous les auteurs.

## Déclaration d'intérêts

Aucun auteur ne déclare de conflit d'intérêts relatif à l'objet de cet article.

## Références

1. Gibot S, Bollaert P-E. Le test de concordance de script comme outil d'évaluation formative en réanimation médicale. *Pédagogie Médicale* 2008;9:7-18.
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;3:135-44.
3. Fournier JP, Demeester A, Charlin B. Script Concordance test: Guidelines for Construction. *BMC Med Inform Decis Mak* 2008;8:18.
4. Charlin B. and Van der Vleuten C. Standardized assessment in context of uncertainty: The script concordance approach. *Eval Health Prof* 2004;27:304-19.
5. Llorcas G, Roy P, Riche B. Évaluation de résolution de problèmes mal définis en éthique clinique : variation des scores selon les méthodes de correction et les caractéristiques des jurys. *Pédagogie Médicale* 2003;4:26-32.
6. Charlin B, Gagnon R, Kazi-Tani D, Thivierge R. Le test concordance comme outil d'évaluation en ligne du raisonnement des professionnels en situation d'incertitude. *Revue internationale des technologies en pédagogie universitaire* 2005;2:22-7.
7. Petrella R, Davis P. Improving management of musculoskeletal disorders in primary care: the Joint Adventures Program. *Clin Rheumatol* 2007; 26:1061-6.
8. Labelle M, Gagnon R, Thivierge RL, Laprise R, Ste-Marie L-G, Charlin B. Formation continue en petits groupes sur l'ostéoporose : comparaison d'un atelier basé sur le test de concordance de scripts (TCS) et d'un atelier classique. *Pédagogie Médicale* 2003;4:145-53.
9. Charlin B, Gagnon R, Lubarsky S, Lambert C, Meterissian S, Chalk C, Goudreaut 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.
10. Vanbelle S. Agreement between raters and groups of raters, Thèse de Doctorat, Université de Liège, 2009.

- [On-line]. Disponible sur : <http://hdl.handle.net/2268/39575>.
11. Leclercq D. Edumétrie et docimologie pour les praticiens chercheurs. Liège : Éditions de l'Université de Liège 1999.
  12. Vanbelle S, Massart V, Giet D, Albert A. Test de concordance de script : un nouveau mode d'établissement des scores limitant l'effet du hasard. *Pédagogie Médicale* 2007;8:71-81.
  13. Centre de pédagogie appliquée aux sciences de la santé. Construire un test de concordance de script. [On-line]. Disponible sur : <http://www.cpass.umontreal.ca/recherche-et-developpement/test-de-concordance-de-script-/construire-un-tcs.html> ; consulté le 3 juillet 2012.
  14. Gagnon R, Charlin B, Coletti M, Sauvé 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.
  15. Duggan P. Development of a script concordance test using an electronic voting system. *Ergo* 2008;1:35-41.
  16. Charlin B, Gagnon R, Pelletier J, Coletti M, Abi-Rizk G, Nazr C, Sauvé É, van der Vleuten C. Assessment of clinical reasoning in the context of uncertainty: the effect of variability within the reference panel. *Med Educ* 2006;40:848-854.
  17. Kania RE, Verillaud B, Tran H, Gagnon R, Kazitani D, Huy PT, Herman P, Charlin B. Online script concordance test for clinical reasoning assessment in otorhinolaryngology: the association between performance and clinical experience. *Arch Otolaryngol Head Neck Surg* 2011;137:751-5.
  18. Gagnon R., Charlin B, Roy L, St-Martin M, Sauvé É, Boshuizen H and Van der Vleuten C. The Cognitive Validity of the Script Concordance Test: A Processing Time Study. *Teach Learning Med* 2006;18:22-7.
  19. Norman G, Bordage G, Pages G, Keane. How specific is case specificity? *Med Educ* 2006;40:618-23.
  20. Gagnon R, Lambert C, Charlin B, Lubarsky. Optimization of answer keys for script concordance testing: should we exclude deviant respondents, deviant responses or none? *Adv Health Sci Educ Theory Pract* 2011;16:601-8.

---

Correspondance et offprints : Didier Giet, Université de Liège, Département de médecine générale (DUMG), Institut de Formation et de Recherche en Enseignement Supérieur (IFRES), 1 avenue de l'Hôpital, B23, 4000 Liège, Belgique.  
Mailto : D.Giet@ulg.ac.be