



Mathieu Lorenzo^{1,2}

1. Département de médecine générale, faculté de médecine de Strasbourg, 4, rue Kirschleger, 67085 Strasbourg

2. Centre de formation et de recherche en pédagogie des sciences de la santé (CFRPS), faculté de médecine de Strasbourg, 4, rue Kirschleger, 67085 Strasbourg

mlorenzo@unistra.fr

exercer 2021;172:185-8.

L'annexe est consultable en ligne à l'adresse : www.exercer.fr

Liens et conflits d'intérêts :

l'auteur déclare n'avoir aucun conflit d'intérêts en relation avec le contenu de cet article.

Les liens d'intérêts éventuels de l'auteur sont disponibles sur le site : www.transparence.sante.gouv.fr

Qu'est-ce que le test de concordance de script ?

Un outil d'évaluation du raisonnement clinique introduit par la réforme du deuxième cycle en France

What is the script concordance test? A tool for assessment of clinical reasoning introduced by the graduate medical school reform in France

LE TCS POUR ÉVALUER LE RAISONNEMENT CLINIQUE

Le test de concordance de script est un outil né il y a une vingtaine d'années au Canada². Il est fondé sur un modèle d'organisation des connaissances des professionnels de santé : les scripts cliniques³. Dans la théorie des scripts cliniques, l'évocation d'une hypothèse diagnostique par le professionnel de santé active un réseau de connaissances. Le professionnel de santé va ensuite interpréter l'ensemble des données cliniques du patient en regard de ces scripts : les données « collent-elles » avec le script activé ? Si oui, l'hypothèse en sort renforcée. Dans le cas contraire, il est probablement nécessaire de considérer une autre hypothèse.

Voici un exemple issu de la pratique de la médecine générale : devant une patiente de 22 ans présentant une douleur abdominale prédominant en fosse iliaque droite, le médecin évoque notamment l'hypothèse de l'appendicite. L'évocation de cette hypothèse active le script « appendicite ». Celui-ci contient les éléments attendus à l'interrogatoire devant une appendicite, les signes cliniques à rechercher à l'examen, les résultats attendus des examens complémentaires, la prise en charge à proposer, etc.

Le médecin demande à la patiente s'il elle a eu de la fièvre. La patiente répond qu'elle a eu 38,7 °C cette nuit avec des frissons. Cette donnée « colle » avec le script « appendicite » : il est fréquent qu'une fièvre accompagne une appendicite aiguë.

À l'examen clinique, le médecin s'attarde particulièrement à palper la fosse iliaque droite. L'absence de défense ou de contracture qu'il constate « ne colle pas » avec son script « appendicite ». Il s'attendait à trouver de tels éléments.

Le principe du TCS est celui-ci : dans une vignette clinique, il est demandé à l'étudiant de juger de l'impact d'une donnée sur le script d'une hypothèse ou d'une option (une « option d'investigation » ou une « option thérapeutique »). L'étudiant répond *via* une échelle de Likert à cinq modalités : ne change rien, impact positif, impact fortement positif, impact négatif, impact fortement négatif. Le **tableau** présente un exemple de questions d'un TCS.

Cet exemple permet d'illustrer plusieurs des difficultés fréquentes rencontrées par les étudiants avec les TCS. En premier lieu, les hypothèses ou options (« si vous pensiez à... ») sont imposées par le format. Il n'est donc pas demandé à l'étudiant de discuter du caractère pertinent ou non de celles-ci au regard de la vignette clinique. Si l'hypothèse ou option ne

INTRODUCTION

La réforme du deuxième cycle des études médicales (R2C) est entrée en vigueur pour les étudiants accédant à la première année du deuxième cycle à la rentrée 2020¹. Cette réforme tente d'inscrire davantage ce deuxième cycle dans une approche par compétences, en visant notamment le développement du raisonnement clinique¹.

En cohérence avec ce dernier objectif, les modalités d'évaluation des apprentissages évoluent en intégrant deux outils relativement méconnus en France : les problèmes à éléments clés (*key feature problems* ou *key feature examinations*) et le test de concordance de script (TCS).

L'objectif de cet article est de présenter le principe du TCS et d'en détailler les principales modalités de conception afin de permettre aux généralistes enseignants intervenant dans le deuxième cycle des études médicales de concevoir des questions.



semble pas pertinente pour l'étudiant dans cette situation clinique, il lui sera très difficile de répondre à la question.

Deuxièmement, les questions sont indépendantes. Ainsi, vous ne connaissez pas la température de la patiente dans la question n° 2. Ceci soulève la problématique de la charge cognitive de l'étudiant confronté à une succession de questions indépendantes sur une même vignette⁴. Il peut être difficile pour les étudiants « d'oublier » les informations précédentes.

Enfin, la nuance entre les modalités « impact fortement négatif » et « impact négatif » ou entre « impact fortement positif » et « impact positif » relève d'une certaine finesse dans le raisonnement clinique. Même chez des étudiants en fin de cursus, les déterminants de ce choix sont inconsistants⁵. Certains auteurs proposent alors de n'utiliser que des TCS à trois modalités, notamment en début de formation (impact négatif, ne change rien, impact positif)^{6,7}.

La plus grande originalité des TCS vient de leur mode de notation. Les points sont attribués selon la proximité des réponses des étudiants à celles d'un panel de référence. Pour la question n° 1 de notre exemple, un panel de dix médecins généralistes a répondu. Si huit membres ont répondu « impact positif » (+1), cette réponse vaudra pour les étudiants $8/8 = 1$ point. Si les deux membres

restant du panel ont répondu « impact fortement positif » (+2), cette réponse vaudra $2/8 = 0,25$ point. Les autres réponses vaudront $0/8 = 0$ point. De nombreuses autres méthodes de notation ont été proposées dans la littérature mais celle-ci fait consensus actuellement⁶⁻⁹.

Ce mode de notation a un effet extrêmement intéressant : c'est la proximité entre le raisonnement clinique de l'étudiant et celui d'un panel de référence qui détermine son score. Un étudiant qui obtient un bon score aux TCS est considéré comme ayant un raisonnement clinique proche d'un groupe expert. C'est une manière normative de déterminer ce qui constitue la réussite en matière de raisonnement clinique. Ce qui est fait par la majorité du groupe de référence est considéré comme la « bonne réponse ». C'est aussi un moyen puissant d'illustrer l'incertitude dans la pratique de la médecine : les membres d'un groupe expert ne sont pas toujours d'accord entre eux.

Le seuil de réussite d'un TCS est habituellement fixé à $60/100$ ⁸. Ce seuil a été déterminé à partir de données statistiques et une valeur inférieure correspond à une performance considérée comme « trop éloignée » de celle du groupe de référence pour être acceptable⁸. Il s'agit là aussi d'une différence notable avec le classique seuil de réussite de $10/20$ encore domi-

nant dans les modalités habituelles d'évaluation des apprentissages en médecine.

De ces considérations docimologiques découlent des problèmes d'acceptabilité par les étudiants du TCS. Il leur est souvent difficile d'admettre en début de formation qu'il existe parfois plusieurs réponses « valables » pour un même problème en médecine. Le mode de sélection des membres du panel de référence est également une source d'inquiétude : un TCS n'est valable que si les membres sont réellement experts – s'ils possèdent les « bons » scripts cliniques. Un problème de légitimité fréquemment évoqué par les étudiants concerne le recrutement d'universitaires en médecine, qui n'ont qu'une activité clinique marginale peu représentative. Il semble important pour les enseignants de bien communiquer sur ces considérations et les moyens utilisés pour lutter contre ces menaces afin de favoriser l'acceptabilité du TCS.

Le TCS explore uniquement une sous-partie du raisonnement clinique. Selon une vision hypothético-déductive simple du raisonnement clinique (**figure**), le test de concordance de script explore la phase d'analyse d'adéquation des données cliniques avec des hypothèses diagnostiques. Selon la théorie du double processus, le raisonnement clinique utilise des processus intuitifs (appelés « sys-

Mme Hillinger est une étudiante de 22 ans dans le domaine du textile. Elle se présente à votre cabinet pour des douleurs abdominales prédominant en fosse iliaque droite apparues cette nuit. Elle n'a pas d'antécédents particuliers et ne prend pas de contraception.

N°	Si vous pensiez à...	... et qu'alors vous trouvez...	... l'impact sur votre hypothèse ou option est...				
Q1	une appendicite	une température à 38,7 °C	-2 <input type="checkbox"/>	-1 <input type="checkbox"/>	0 <input type="checkbox"/>	+1 <input type="checkbox"/>	+2 <input type="checkbox"/>
Q2	une grossesse extra-utérine	que la patiente ne se souvient plus de la date de ses dernières règles	-2 <input type="checkbox"/>	-1 <input type="checkbox"/>	0 <input type="checkbox"/>	+1 <input type="checkbox"/>	+2 <input type="checkbox"/>
Q3	demandeur une échographie abdominale	que la patiente est enceinte	-2 <input type="checkbox"/>	-1 <input type="checkbox"/>	0 <input type="checkbox"/>	+1 <input type="checkbox"/>	+2 <input type="checkbox"/>

Tableau 1 - Exemple de questions d'un TCS et des réponses du panel

-2 = fortement négatif, -1 = négatif, 0 = ne change rien, +1 = positif, +2 = fortement positif.

Exemple de calcul du barème : pour la question n° 1, huit membres ont répondu « +1, impact positif » et deux membres ont répondu « +2, impact fortement positif ».

La réponse « +1 » vaudra pour les étudiants $8/8 = 1$ point. La réponse « +2 » vaudra $2/8 = 0,25$ point. Les autres réponses vaudront $0/8$ point.

tème 1 ») et des processus analytiques (« système 2 »)¹⁰. Le TCS explore la partie analytique du raisonnement clinique : la phase de raisonnement logique, consciente, qui contrôle les processus intuitifs. Au travers du TCS, c'est l'efficacité de l'organisation des connaissances cliniques de l'étudiant qui est évaluée (ses scripts cliniques).

Il est donc nécessaire de garder à l'esprit que le TCS n'est pas l'outil miracle en évaluation du raisonnement clinique. Son utilisation doit être combinée à d'autres méthodes et outils pour correctement évaluer le raisonnement clinique des étudiants¹¹. Des méthodes comme les examens cliniques objectifs structurés ou les protocoles demandant de réfléchir à voix haute permettent par exemple de documenter d'autres aspects du raisonnement clinique, comme la génération d'hypothèses¹¹.

COMMENT CONCEVOIR UN TCS ?

Comme décrit dans la première partie de cet article, le TCS reflète les microjugements analytiques que tout médecin généraliste effectue constamment face à un patient. Il est donc relativement simple de concevoir des questions de TCS avec la méthode suivante : rappelez-vous une situation que vous avez récemment vécue. Souvenez-vous des hypothèses diagnostiques précoces que vous avez évoquées. Des manœuvres que vous avez considérées pour l'examen clinique. De l'impact du résultat de celles-ci sur vos hypothèses. Avec cette gymnastique d'explicitation de son propre raisonnement clinique, tout professionnel de santé peut rapidement écrire des questions de TCS. Par exemple : « *j'ai reçu hier Mme Hilm pour une toux évoluant depuis une semaine. Alors que j'évoquais une bronchite, j'ai réalisé une auscultation pulmonaire. Quel fut l'impact sur mon hypothèse principale d'entendre des ronchi ? Quel aurait été l'impact d'entendre un foyer de crépitations en base droite ? etc.* ».

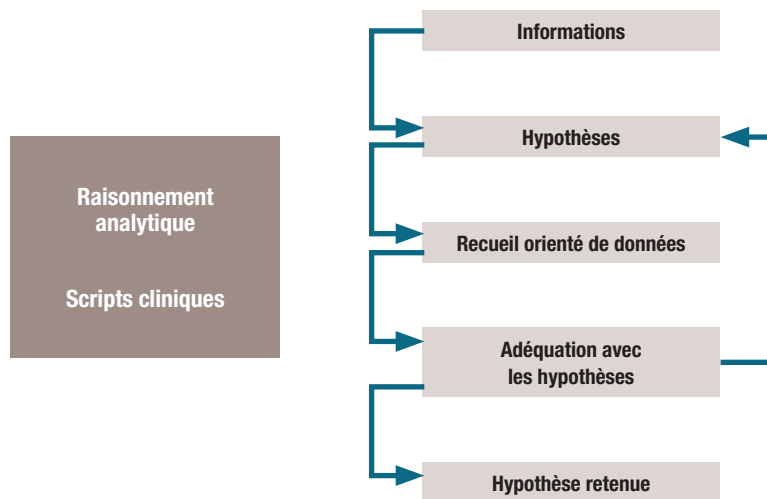


Figure - Le modèle hypothético-déductif du raisonnement clinique

Un conseil pratique est d'éviter de réfléchir à ce que le « panel répondrait » lors de la création des questions. Cette difficulté habituelle chez l'enseignant conduit à écarter un grand nombre de questions jugées comme « trop simples ». L'expérience montre que le panel est souvent plus divisé qu'on ne le pense. Écrivez donc rapidement tous ces microjugements cliniques que vous réalisez dans votre pratique clinique. Vous déciderez seulement dans un deuxième temps de garder ou non ces questions selon les réponses du panel.

Pour faciliter cette phase de rédaction, un masque de saisie de questions de TCS est téléchargeable en **annexe** de cet article.

Il est communément conseillé de créer au maximum trois questions par vignette clinique (scénario)⁸. Ceci pour des raisons de charge cognitive pour les étudiants (et le panel) du fait de l'indépendance des propositions entre elles⁸. Historiquement, les questions de TCS avaient des modalités de réponse spécifiques selon qu'il s'agisse de questions diagnostiques, d'investigation ou de thérapeutiques. L'ancrage universel proposé dans l'exemple (et sur la plateforme SIDES) permet de mélanger des questions des trois dimensions pour une même vignette. Vous pouvez donc, pour une même vignette, interroger un raisonnement portant sur une hypothèse diagnostique, une proposition thérapeutique ou d'investigation. Pour obtenir au final entre 90 et 100 ques-

tions dans un TCS, il est nécessaire de créer entre 120 et 150 questions, car deux étapes ultérieures de sélection sont nécessaires.

Concernant le panel de référence, il doit comprendre entre 10 et 20 membres pour des examens à fort enjeu⁸. Pour des TCS en formatif (pour guider les apprentissages ultérieurs des étudiants), trois à cinq membres suffiront pour refléter l'incertitude sur ces sujets. La désignation des membres du panel pose également question : qui est « expert » de la discipline ? Ce sont habituellement les professionnels désignés comme tels par leurs pairs. Cette question de la légitimité des membres du panel est cruciale pour la validité du TCS et son acceptabilité pour les étudiants. Il est parfois proposé de demander aux professionnels de la discipline de désigner les experts par un vote.

Le nombre de questions d'un TCS dépend aussi de l'usage. Pour un examen sommatif (vérifier que les apprentissages attendus ont bien eu lieu), il faut une centaine de questions. 90 questions semblent un minimum pour des raisons de validité de contenu principalement. En formatif, aucun nombre de questions n'est prescrit.

Le rôle du panel est de permettre la création du barème de notation, mais aussi de sélectionner les bonnes questions à poser aux étudiants. En effet, les questions pour lesquelles le panel est unanime sont davantage à considérer comme des questions à choix



explorant des connaissances. À l'inverse, des questions pour lesquelles les réponses du panel sont très dispersées sont probablement de mauvaises questions : mal formulées, ou traitant d'un domaine trop controversé. Pour sélectionner les meilleures questions, un calculateur en format Excel est disponible gratuitement sur le site de l'université de Montréal : https://www.cpass.umontreal.ca/recherche/axes-de-recherches/concordance/tcs/corriger_tcs/.

Avec cet outil, les questions à privilégier sont celles avec une variance intermédiaire (ni faible ni élevée) dans l'onglet « items stats ». Cette première sélection conduit habituellement à écarter entre 10 et 30 % des questions créées.

Une deuxième étape de sélection des questions est parfois proposée après que les étudiants ont répondu⁸. Il s'agit de comparer le profil de réponse des étudiants et celui du panel. Le calculateur permet d'éva-

luer la corrélation entre les réponses de ces deux groupes. Une « bonne » question devrait avoir globalement des profils de réponse superposables entre les étudiants et les membres du panel. Des profils de réponses très différents (une faible corrélation) peut indiquer que la question n'a pas été comprise de la même manière par les étudiants et les experts. L'outil Excel[®] permet de sélectionner les questions à inclure dans l'analyse finale avec un calcul du coefficient alpha de Cronbach pour augmenter la validité du TCS. Ces considérations docimologiques avancées n'ont de sens que dans une utilisation à fort enjeu. La limite principale pour une appropriation large du TCS dans la communauté universitaire est la relative complexité de ces deux étapes de sélection.

Pour finir, le temps de passage habituel est compris entre 30 et 60 secondes par question. Il faudra donc fréquemment une heure aux membres du panel de référence pour

compléter leurs réponses. Utiliser des outils numériques tels que la plateforme SIDES ou des questionnaires type *LimeSurvey*[®] ou *Google Forms*[®] fait gagner du temps si plusieurs milliers de réponses du panel doivent être traitées.

CONCLUSION

Le TCS est un outil d'évaluation d'une partie du raisonnement clinique. Les particularités en termes de tâche demandée et de mode de notation nécessitent un temps d'appropriation pour les enseignants et les étudiants.

Une fois ce travail effectué, la possibilité de refléter l'incertitude dans les soins en fait un outil de choix pour l'enseignement de la médecine générale. Il est important que les enseignants de la discipline alimentent les bases de questions de TCS pour que l'importance des soins premiers transparaissent dans les modalités évaluatives en deuxième cycle. ♦

Résumé

La réforme du deuxième cycle des études médicales va introduire le test de concordance de script (TCS) parmi les outils d'évaluation du raisonnement clinique. L'objectif de cet article didactique est de présenter le principe du TCS et d'en détailler les principales modalités de conception afin de permettre aux généralistes enseignants intervenant dans le deuxième cycle des études médicales d'en concevoir des questions. Le TCS explore une sous-partie du raisonnement clinique : les processus analytiques. La particularité principale du TCS est que le score est attribué en fonction de la proximité de la réponse de l'étudiant avec celles d'un panel de référence.

La deuxième partie de l'article apporte des conseils pratiques quant à la rédaction de questions, à la sélection de celles-ci et à l'utilisation du TCS.

→ **Mots-clés** : éducation ; enseignement spécialisé en médecine ; enseignement ; pédagogie.

Summary

The reform of graduate medical school in France will introduce the Script Concordance Test (SCT) as a tool for assessing clinical reasoning. The objective of this didactic article is to present the principle of the SCT and to detail its main design modalities in order to allow general practitioner teachers in the second cycle of medical studies to create SCT questions. The Script Concordance Test explores a subpart of clinical reasoning: analytical processes. The main feature of the SCT is that the score is assigned according to the closeness of the student's response to those of a reference panel.

The second part of the article provides practical advice on writing questions, selecting questions and using the SCT.

→ **Keywords**: education; education medical graduate; teaching; pedagogy.

Références

1. République française. Arrêté du 2 septembre 2020 portant modification de diverses dispositions relatives au régime des études en vue du premier et du deuxième cycle des études médicales et à l'organisation des épreuves classantes nationales.
2. 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.
3. 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.

4. Sweller J. Cognitive load during problem solving: effects on learning. *Cogn Sci* 1988;12:257-85.
5. Gawad N, Wood TJ, Cowley L, Raiche I. The cognitive process of test takers when using the script concordance test rating scale. *Med Educ* 2020;54:337-47.
6. Bland AC, Kreiter CD, Gordon JA. The psychometric properties of five scoring methods applied to the script concordance test. *Acad Med* 2005;80:395-9.
7. Wilson AB, Pike GR, Humbert AJ. Analyzing script concordance test scoring methods and items by difficulty and type. *Teach Learn Med* 2014;26:135-45.

8. Dory V, Gagnon R, Vanpee D, Charlin B. How to construct and implement script concordance tests: insights from a systematic review. *Med Educ* 2012;46:552-63.
9. Lubarsky S, Dory V, Duggan P, Gagnon R, Charlin B. Script concordance testing: from theory to practice: AMEE guide no 75. *Med Teach* 2013;35:184-93.
10. Pelaccia T, Tardif J, Tribby E, Charlin B. An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory. *Med Educ Online* 2011;16.
11. Daniel M, Rencic J, Durning SJ, et al. Clinical Reasoning Assessment Methods: a scoping review and practical guidance. *Acad Med* 2019;94:902-12.