

Élaboration et validation d'un test de concordance de script pour évaluer le raisonnement clinique des infirmières en contexte de détérioration de la condition clinique*

Assessing nurses' clinical reasoning skills in the context of clinical deterioration through a script concordance test

Sarah CÔTÉ¹, Andréanne TANGUAY¹, Robert GAGNON², Bernard CHARLIN²,
Cécile MICHAUD¹

¹ Université de Sherbrooke, 12^e avenue Nord, Sherbrooke Québec, J1H 5N4, Canada

² Université de Montréal, Montréal, Canada

Manuscrit reçu le 29 octobre 2013 ; commentaires éditoriaux soumis aux auteurs le 31 janvier et le 18 février 2014 ; accepté pour publication le 19 février 2014

Mots-clés

Raisonnement clinique ; évaluation du raisonnement clinique ; expertise infirmière ; test de concordance de script ; détérioration clinique ; étude méthodologique

Résumé – Contexte : La complexité des situations cliniques en milieu hospitalier exige, chez les infirmières, un raisonnement clinique solide pour interpréter les signes de détérioration clinique de leurs patients. **But :** Élaborer un test de concordance de script (TCS) évaluant le raisonnement clinique infirmier en contexte de détérioration clinique en milieu hospitalier et évaluer ses qualités psychométriques et son appréciation. **Méthode :** Un TCS de 22 scénarios et de 88 items a été élaboré, puis administré à 47 infirmières novices et 21 expertes. Un coefficient alpha de Cronbach a servi à évaluer la cohérence interne et un test-retest avec 12 participantes, la stabilité temporelle. La validité du contenu a été évaluée à l'aide des indices de la validité du contenu (IVC) et la validité de construit discriminante, avec l'étude des groupes extrêmes. **Résultats :** Après son optimisation, le TCS contient 20 scénarios et 68 items. Son coefficient alpha de Cronbach est de 0,76. L'étude de la stabilité temporelle a révélé une corrélation intra-classe modérée ($r = 0,554$; IC à 95 % = 0,002, 0,847). L'IVC de la pertinence et de la clarté du TCS étaient respectivement de 0,995 et de 0,993. Les scores des novices

* En accord avec les principes de la rédaction épiciène, l'utilisation du genre féminin inclut le masculin en ce qui concerne les participants et les participantes de cette étude. Cela inclut les infirmières et les infirmiers, les étudiants et les étudiantes ainsi que les experts et les expertes. Le terme patient est employé au masculin et englobe les hommes et les femmes, sans préjudice au genre opposé.

étaient moins élevés que ceux des expertes ($p < 0,05$). Il y avait une relation linéaire positive significative entre les scores et l'expertise clinique ($p < 0,05$). Son appréciation est généralement bonne. **Conclusion :** Cette étude a rendu disponible un TCS standardisé pour évaluer le raisonnement clinique d'infirmières en contexte de détérioration clinique en milieu hospitalier.

Keywords

Clinical reasoning; clinical reasoning assessment; nurses' expertise; script concordance test; clinical deterioration; methodological study

Abstract – Background: For nurses, the complexity of clinical situations in hospitals requires solid clinical reasoning to interpret signs of clinical deterioration in patients. **Purpose:** To develop a script concordance test (SCT) assessing nurses' clinical reasoning in context of in-hospital clinical deterioration and evaluate its psychometric properties and appreciation. **Method:** A SCT with 22 scenarios and 88 items was developed then administered to 47 novice and 21 expert nurses. A Cronbach alpha coefficient was used to assess the internal consistency, a test-retest with 12 participants and temporal stability. Content validity was assessed using content validity index (CVI), discriminant validity, with the study of extreme groups. **Results:** After optimization, the SCT had 20 scenarios and 68 items. Cronbach's alpha of TCS was 0.76. The study of the temporal stability showed a moderate ($r = 0.554$; IC at 95% = 0.002, 0.847). The CVI of the SCT relevance and clarity was respectively 0.9952 and 0.9933. Novices' scores were lower than those of experts ($p < 0.05$). There was a significant positive linear relationship between the scores and clinical expertise ($p < 0.05$). Its overall appreciation is good. **Conclusion:** This study made available a standardized SCT to evaluate nurses' clinical reasoning in in-hospital situations of clinical deterioration.

Introduction

L'allongement de l'espérance de vie a pour conséquence que les infirmières en milieu hospitalier doivent soigner un nombre croissant de patients^[1] dont la condition de santé est relativement instable^[2]. Les infirmières ont pour rôle de contribuer à la prévention de la maladie ou de complications, au maintien ou au recouvrement de la santé de leurs patients^[3]. Elles ont été formées pour intervenir en ce sens et travaillent en collaboration avec les autres professionnels de la santé. Entre autres fonctions, les infirmières doivent évaluer l'état de santé d'un patient symptomatique et assurer une surveillance clinique de sa condition. Elles entreprennent des actions diagnostiques et thérapeutiques, déterminent le plan de soins et de traitements, appliquent le plan de soins infirmiers et médicaux tout en assurant le suivi de ces personnes^[3]. Ces rôles sont d'une importance majeure pour prévenir la détérioration clinique, qui peut évoluer vers une condition de santé instable. Un raisonnement clinique adéquat est essentiel, particulièrement en contexte de détérioration clinique^[4-5], car il permet de formuler un jugement sur la situation du patient et d'établir les priorités de soins. Le raisonnement clinique consiste

en l'utilisation de la cognition pour analyser et interpréter des données, pour générer des hypothèses et pour concevoir des options de solutions afin de résoudre un problème clinique^[6]. Ce raisonnement varie en fonction de plusieurs caractéristiques de l'infirmière, parmi lesquelles, particulièrement, son expérience clinique^[6-7]. Cela peut notamment s'expliquer au regard de la théorie des scripts^[8]. Cette théorie soutient que l'expertise se développe à partir de réseaux de connaissances théoriques et pratiques, appelés scripts, qui sont organisées et hiérarchisées de manière à ce qu'elles soient rapidement disponibles et accessibles en fonction des tâches cliniques^[8]. Ces scripts sont élaborés et acquis pendant la formation, lors de l'exposition répétée à des situations cliniques rencontrées en pratique et qui fondent l'expérience clinique ; ils favoriseraient une collecte de données plus ciblée et plus pertinente^[7]. Or, les signes de détérioration clinique ne sont pas toujours détectés ou ne sont pas interprétés adéquatement, notamment à cause du haut niveau d'incertitude de la situation, ce qui peut favoriser la survenue de complications cliniques telles que des arrêts cardiorespiratoires, qui rendent nécessaires des transferts vers les unités de soins intensifs, voire qui entraînent des décès évitables^[9]. Afin de mieux soutenir le développement du

raisonnement clinique des infirmières dans ce contexte, il est important de pouvoir l'évaluer. À l'heure actuelle, le raisonnement clinique est évalué chez les infirmières lors de leur formation scolaire, à l'aide d'examens à questions à choix multiples, d'examens oraux, d'examens cliniques objectifs structurés (ECOS) ou de grilles d'évaluation globale. Ces méthodes d'évaluation sont jugées imparfaitement appropriées pour la mesure du raisonnement clinique parce qu'elles ne l'évaluent pas dans sa globalité, parce qu'elles ne sont pas spécifiques du raisonnement clinique et parce qu'elles ne distinguent pas les différents niveaux d'expertise des personnes évaluées^[10]. Son évaluation à l'aide d'une méthode fidèle, valide et spécifique permettrait de déterminer les principales difficultés reliées à la reconnaissance et l'interprétation des signes de détérioration clinique, puis de concevoir une formation pour les diminuer.

Le test de concordance de script (TCS) est une méthode d'évaluation adaptée pour la mesure du raisonnement clinique en contexte d'incertitude^[10]. En simulant des conditions de pratique authentique, le TCS permet de vérifier, à l'aide d'un score, le niveau d'organisation et d'élaboration des connaissances et, ainsi, d'évaluer la qualité des scripts^[10]. Il permet aussi d'évaluer les habiletés à interpréter les données collectées dans un contexte caractérisé par l'incertitude^[11]. Il prend la forme de cas – appelés par la suite scénarios, qui décrivent en quelques lignes des situations cliniques de pratique. Même pour un expert, ces scénarios constituent un problème parce que des informations sont manquantes ou ambiguës, ou parce que plusieurs options d'hypothèses diagnostiques, d'investigation ou d'interventions peuvent être évoquées de façon pertinente. Les scénarios sont suivis de plusieurs questions, appelées items. Chaque item est construit en fonction du modèle du raisonnement clinique dit hypothético-déductif^[11]. Ainsi, chaque item se divise en trois parties : 1) une hypothèse, 2) une nouvelle information et 3) une échelle de type Likert qui quantifie l'effet de la nouvelle information sur la probabilité ou la pertinence de l'hypothèse. La réponse formulée à l'aide de cette échelle de Likert en cinq points permet de vérifier l'ampleur de l'effet de la nouvelle information sur l'hypothèse ainsi que sa direction (effet positif ou négatif). Les hypothèses

qui sont proposées correspondent à des hypothèses diagnostiques, d'investigations ou d'interventions, qui seraient habituellement générées lors du processus du raisonnement clinique, en pratique clinique. La nouvelle information est un signe clinique ou un résultat d'examen, qui pourraient être recherchés ou demandés pour confirmer ou infirmer l'hypothèse. Le barème qui permet de coter les réponses est ensuite construit à l'aide de réponses d'expertes par la méthode des scores combinée^[10-11], qui sera décrite plus loin.

L'élaboration des TCS et la vérification de leurs qualités psychométriques sont guidées par des données probantes^[10-13]. Sur la base des données disponibles, le TCS s'est révélé être une méthode d'évaluation du raisonnement clinique valide en médecine. Cependant, la fidélité des TCS a été évaluée de manière incomplète dans la littérature scientifique. En effet, lors de l'évaluation de la fidélité des TCS, seule la cohérence interne est généralement évaluée, sans évaluation de la stabilité temporelle ce qui ne rend pas compte de la fidélité réelle du TCS^[14].

L'élaboration et la validation de TCS ont fait l'objet de deux seules études en sciences infirmières et aucune des deux n'a traité du sujet de la détérioration clinique. Dans la première étude, Deschênes et al.^[15] avaient développé un TCS de 29 scénarios et de 73 items pour évaluer le raisonnement clinique infirmier empreint de *Human caring*. Le TCS avait une bonne cohérence interne globale, avec un coefficient alpha de Cronbach de 0,86. Il y avait une bonne homogénéité à l'intérieur des trois sous-échelles du TCS (représentant chacune une dimension du *Human caring*) et de l'ensemble du TCS avec des corrélations variant de 0,50 à 0,91 entre les différentes dimensions ($p < 0,01$). La comparaison des scores moyens des étudiants (53,3 %) et ceux des experts (61,6 %) à l'aide du test *t* de Student avait démontré qu'ils étaient différents d'une manière statistiquement significative et qu'ils augmentaient avec l'expérience clinique ($p < 0,01$). La deuxième étude est celle de Latreille^[16]. Elle avait pour but de vérifier si le TCS de 15 scénarios et 45 items développé puis administré à 30 étudiantes en sciences infirmières, à 40 infirmières travaillant en pédiatrie et à 15 infirmières expertes dans le domaine de la pédiatrie, était

un instrument de mesure pertinent pour l'évaluation du raisonnement clinique et s'il pouvait distinguer le raisonnement clinique en fonction du niveau d'expertise des participants. Les scores moyens obtenus étaient de 62,3 % pour les étudiants, 75,0 % pour les infirmières et 77,4 %, pour les experts. Il n'y avait pas de différences statistiquement significatives entre les scores des infirmières et des experts et cela pouvait s'expliquer par le fait que des infirmières auraient pu être classées dans le groupe des experts étant donné leur grande expérience clinique. Malgré ce résultat, il y avait une augmentation statistiquement significative des scores selon le niveau d'expertise des participants lorsqu'on comparait les scores des étudiants à ceux des experts ($p < 0,01$). Il y avait aussi une corrélation significative et positive entre l'expérience clinique et les scores élevés au TCS ($r = 0,44$; $p < 0,000$). Cette étude avait pour limite d'avoir une cohérence interne inférieure à ce qui était attendu : après optimisation du TCS, celui-ci comportait 36 items et son coefficient alpha de Cronbach s'élevait à 0,61. Dans les deux études, la stabilité temporelle n'avait pas été évaluée. En somme, ces deux études méthodologiques en sciences infirmières ont permis la construction de TCS valides et ont mené à la conclusion que le raisonnement clinique infirmier peut être évalué grâce au TCS.

Dans cette optique, notre étude avait pour but la construction d'un TCS pour évaluer le raisonnement clinique d'infirmières prenant soin de patients présentant des signes de détérioration clinique (TCS-DC) en milieu hospitalier, dans des unités de soins de médecine et de chirurgie, et d'en évaluer la validité, la fidélité, incluant sa cohérence interne ainsi que sa stabilité temporelle.

Méthodes

Contexte de l'étude

L'étude a été réalisée au Québec où la formation des infirmières peut prendre différents cheminements^[3]. Ainsi, le cheminement collégial peut consister soit en une formation professionnelle de trois ans (diplôme

d'études collégiales ; DEC en soins infirmiers), qui permet de se présenter à l'examen de l'Ordre des infirmières et infirmiers du Québec (OIIQ), soit en une formation scientifique générale de deux ans (DEC en sciences de la nature). Dans les deux cas, des programmes de formation universitaire en sciences infirmières de niveau baccalauréat – licence dans certains pays, tels qu'en France – (BScInf) sont possibles et varient en durée. La formation collégiale professionnelle et le BScInf permettent à la diplômée de se présenter à l'examen de l'OIIQ, dont la réussite autorise l'exercice de la profession infirmière. Après la réussite de leur formation et avant la réussite de l'examen de l'OIIQ, les diplômées en soins infirmiers ou en sciences infirmières portent le titre de candidates à l'exercice de la profession infirmière (CEPI)^[3]. Il est important de mentionner qu'un baccalauréat en sciences infirmières peut donner accès à des études supérieures, de niveau maîtrise, puis de niveau doctorat.

Participant·es

Les participant·es de cette étude peuvent être classées selon qu'elles sont des expertes ou des novices en sciences infirmières. Les expertes, selon les recommandations de Petrucci^[16], devaient avoir une expérience propre au domaine d'intérêt. En l'occurrence, les participant·es de l'étude devaient donc avoir une expérience auprès de patients à risque de détérioration clinique. Les expertes recrutées avaient au moins cinq ans d'expérience clinique en soins critiques ou elles étaient détentrices d'un diplôme universitaire en sciences infirmières et travaillaient activement dans une unité de soins critiques à titre d'infirmières et de monitrices cliniques (professeures de stage) pour une université canadienne. Elles ne devaient pas avoir quitté les soins directs aux patients depuis plus d'un an, à moins d'enseigner activement en soins critiques. Les novices étaient des étudiantes en baccalauréat de la même université, candidates à l'exercice de la profession infirmière ou des infirmières ayant moins de trois ans d'expérience clinique. Toutes les participant·es ont été recrutées sur une base volontaire (échantillonnage de convenance).

Étapes préliminaires à la construction et à la validation du TCS-DC

Avant de commencer la construction du TCS-DC, il a fallu déterminer, selon les recommandations tirées des écrits scientifiques^[10, 18], la population à étudier (les infirmières œuvrant en médecine ou en chirurgie dans les milieux hospitaliers), le domaine à évaluer (la détérioration clinique) et le contexte d'utilisation du TCS (évaluer le développement professionnel ou les acquis à la fin d'une activité éducative).

Par la suite, des grilles et des guides ont été conçus par l'étudiante-chercheuse, en collaboration avec ses directrices de maîtrise, afin de faciliter le processus de validation du TCS-DC. Premièrement, une « grille de correction pour l'élaboration du TCS-DC » a été élaborée. Celle-ci a été entre autres adaptée de celle disponible sur le site du Centre de pédagogie appliquée aux sciences de la santé (CPASS) de l'Université de Montréal^[18]. Elle contient des critères pour corriger les scénarios et les items du TCS-DC afin de vérifier qu'aucun élément n'est manquant. Deuxièmement, afin de vérifier que tout le contenu pertinent en détérioration clinique était évalué par le TCS-DC, un « Tableau du contenu » a été développé. Celui-ci présente d'un côté, les principaux troubles physiologiques et pathologiques rencontrés en contexte de détérioration clinique et, d'un autre côté, les rôles infirmiers impliqués dans ce contexte qui devaient être évalués. Finalement, deux guides pour faciliter l'évaluation de la validité apparente et du contenu du TCS-DC ont été conçus. Le premier, le « Guide de validation du contenu du test de concordance de script en contexte de détérioration clinique : pertinence et clarté » a servi pour évaluer la validité apparente et du contenu du TCS-DC. Le second, le « Guide d'évaluation du test de concordance de script en contexte de détérioration clinique : pré-test » a été utilisé par l'étudiante-chercheuse lors d'entrevues pour évaluer la ressemblance des scénarios du TCS-DC avec la pratique clinique sur les unités de médecine et de chirurgie, en milieu hospitalier et pour valider les instructions pour accéder au TCS et le remplir.

Construction et validation du TCS-DC

La première étape de construction du TCS-DC a consisté à sélectionner des situations de détérioration clinique au cours desquelles une infirmière devait utiliser son raisonnement clinique pour les résoudre. La seconde étape a consisté à transcrire les situations retenues sous la forme de scénarios et d'items. Les scénarios décrivaient des situations cliniques où de la détérioration clinique était probable. Ils ont entre autres été élaborés à l'aide de plusieurs versions du *Modified Early Warning Score*, un outil d'aide au dépistage contenant plusieurs signes de détérioration reconnus comme prédictifs d'un arrêt cardiaque^[4]. Parmi ces signes, on retrouve entre autres des altérations de la pression artérielle, de la fréquence cardiaque, de la fréquence respiratoire, de la température, tant à la hausse qu'à la baisse, ainsi qu'une modification de l'état de conscience^[4]. Chaque item, adapté à la réalité infirmière, a été constitué d'abord d'une hypothèse (diagnostic infirmier, impression de diagnostic médical, option de surveillance, d'évaluation ou d'intervention infirmière), ensuite, d'une nouvelle information (un signe clinique collecté par une infirmière, un résultat d'examen) et finalement d'une échelle de type Likert de cinq niveaux, allant de -2 à +2, avec des libellés en lien avec l'hypothèse. Chaque scénario et chaque item étaient indépendants les uns des autres. Comme la capacité à résoudre un problème clinique ne prédit pas la capacité à résoudre un autre problème, il a été préférable de générer un nombre plus élevé de problèmes plutôt que d'évaluer en profondeur une même situation clinique^[18]. Par ailleurs, Gagnon et al.^[19] recommandaient de construire de 15 à 25 scénarios suivis de trois à cinq items chacun pour assurer une fidélité élevée. Cette version préliminaire du TCS-DC comportait donc 22 scénarios suivis de quatre items. Elle a ensuite été soumise à des expertes évaluatrices afin de réaliser une partie de la troisième étape, consistant à valider le TCS-DC, puis à des novices pour un pré-test. Un extrait d'un scénario et de ses items est présenté sur la figure 1.

Madame Poirier, 68 ans, est hospitalisée sur une unité de soins généraux pour un ulcère artériel au membre inférieur droit qui s'est infecté. Elle n'a pas d'autres antécédents. Vous entrez dans la chambre à 8h30 et Madame Poirier est en diaphorèse et elle vomit. Elle vous avise qu'elle a des palpitations et une douleur épigastrique.

Si vous pensiez	Et qu'alors vous trouvez que	Cette intervention est
1. Contrôler sa glycémie	Madame Poirier est pâle	<input type="checkbox"/> -2 <input type="checkbox"/> -1 <input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 :
2. Surveiller cette patiente toutes les 15 minutes pendant la prochaine heure	Sans intervention, Madame Poirier ne vomit plus, sa douleur épigastrique est passée de 7/10 à 5/10 et elle n'est plus nauséuse. Ses signes vitaux sont : <ul style="list-style-type: none"> - Fréquence cardiaque à 115/min - Fréquence respiratoire à 19/min - Pression artérielle à 128/68 - Saturation en oxygène à 92 % 	<input type="checkbox"/> -2 <input type="checkbox"/> -1 <input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 :
3. Aviser le médecin le plus tôt possible des signes vitaux et des symptômes de Madame Poirier	Après cinq minutes, Madame Poirier ne vomit plus. Sa douleur est toujours à 7/10. Ses signes vitaux sont : <ul style="list-style-type: none"> - Fréquence cardiaque à 120/min - Fréquence respiratoire à 22/min - Pression artérielle à 128/68 - Saturation en oxygène à 92 % 	<input type="checkbox"/> -2 <input type="checkbox"/> -1 <input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 :

- 2 : inutile
-1 : peu utile
0 : ni plus ni moins utile
1 : est utile
2 : très utile

Fig. 1. Exemple d'un scénario de test de concordance de script et de ses items.

Administration du TCS

Les principes du TCS ont été expliqués à toutes les participantes potentielles à l'aide d'une lettre d'introduction qui leur a été envoyée par message électronique. Les étudiantes ont également été rencontrées et ont reçu des explications sur le fonctionnement du TCS-DC, ainsi que sur leur participation attendue,

consistant à répondre à un TCS à deux reprises dans un intervalle de deux semaines pour un test-retest. Celles qui désiraient participer se connectaient ensuite sur le site internet indiqué dans le message électronique (Moodle), où des instructions étaient à nouveau présentées avant qu'elles ne répondent au TCS. Les participantes pouvaient remplir le TCS-DC à leur convenance : il n'y avait pas de limite de

Tableau I. Participantes aux différentes étapes d'élaboration et de validation du TCS.

	Expertes			Novices		
	<i>n</i>	F	Exp.	<i>n</i>	F	Exp.
Rédaction des situations	3	3	3 inf. : > 10 ans	–	–	–
Validité du contenu (IVC de la pertinence et du contenu clarté)	5	4	2 inf. : > 5 ans 2 inf. : avec expertises particulières 1 inhalo. : avec expertises particulières	–	–	–
Pré-test	–	–	–	19	16	2 CEPI : 0 an 17 inf. : < 3 ans
Validité de construit discriminante Cohérence interne Appréciation	21	15	8 inf. : 3 et 4 ans 6 inf. : 5 à 9 ans 7 inf. : > 10 ans	47	42	2 CEPI : 0 an 45 inf. : < 3 ans
Stabilité temporelle	–	–	–	12	11	12 inf. : < 6 ans

n = nombre, F = femme, Exp. = nombre d'années d'expérience en situation de détérioration clinique, inf. = infirmière, inhalo. = inhalothérapeute, CEPI = candidate à l'exercice de la profession infirmière.

temps, ni de limite d'accès (ils pouvaient le remplir en plusieurs temps). Cette absence de limite visait à favoriser la participation des étudiantes, le but de leur participation étant de valider le TCS et non pas d'évaluer leur performance. Des rappels étaient formulés aux étudiantes toutes les deux semaines afin d'encourager leur participation pour une première fois ou pour le retest.

Description des participantes

Le tableau I résume le nombre de participantes selon leur rôle et leur expertise. Les expertes qui ont contribué à la construction et à la validation du TCS-DC présentaient les caractéristiques suivantes. En tout, trois expertes ont collaboré à la sélection et à la rédaction des scénarios du TCS-DC. Ensuite, cinq expertes évaluatrices ont collaboré à la validation du contenu du TCS-DC. Quatre étaient des infirmières et une était inhalothérapeute (professionnelle de la santé cardio-respiratoire). Parmi les infirmières, une faisait partie d'une équipe d'intervention rapide (équipe appelée sur les unités de soins lorsqu'un patient se détériore) depuis deux ans. L'inhalothérapeute œuvrait en soins généraux et en soins critiques depuis trois ans. Le

TCS-DC a aussi été administré à 23 infirmières expertes faisant partie du personnel enseignant. Celles-ci étaient âgées de 24 à 56 ans. L'âge moyen était de 32 ans. Elles avaient entre trois et 37 années d'expérience clinique, le nombre d'années d'expérience moyen étant de 11,7 ans. Deux d'entre elles n'avaient pas entrepris d'études universitaires ; cinq avait une maîtrise en sciences infirmières, quatre étudiaient en maîtrise et deux en doctorat. Conformément aux recommandations de la littérature scientifique^[20], deux expertes, respectivement de 25 ans et 37 ans d'expérience clinique, ont été retirées du panel de référence étant donné que leurs scores étaient différents de ceux des autres membres du panel, situés deux écarts-types en dessous de la moyenne par rapport aux 21 autres expertes.

Établissement des scores : méthode des scores combinés

La méthode des scores combinés permet de mieux discriminer les participants en fonction du niveau d'organisation de leurs connaissances (scripts) et de leur expertise^[10]. Elle prend en compte qu'il existe une variabilité dans le processus du raisonnement

À l'un des items, 10 expertes répondent « +2 », 8 répondent « +1 », 2 répondent « 0 » et aucune ne répond « -1 » ou « -2 ». Pour obtenir les scores, il s'agit de repérer la réponse qui a été la plus choisie par les expertes (dans ce cas-ci, la réponse « +2 » a été choisie par 10 expertes). Cette réponse donnera maintenant un point à toutes les personnes évaluées qui la choisiront. Ensuite, il faut diviser toutes les réponses par le nombre d'expertes le plus élevé à avoir choisi la réponse. Dans cet exemple, il faut diviser par 10. Les réponses non choisies par les expertes donnent $0/10 = 0,0$ point et les autres réponses donnent un crédit en fonction du nombre d'expertes qui les ont choisies, soit $8/10 = 0,8$ point pour la réponse « +1 » et $2/10 = 0,2$ pour la réponse « 0 ». Un participant non expert qui choisirait la réponse « +1 » obtiendrait alors $0,2/1$ pour cet item.

	-2	-1	0	1	2	Instructions
Nombre de réponses parmi les expertes	0	0	2	8	10	Repérer la réponse modale (la plus choisie)
Mécanisme de création des scores	0	0	2/10	8/10	10/10	Diviser par le nombre d'expertes ayant donné la réponse la plus choisie
Crédit pour l'item	0	0	0,20	0,80	1	Donner les points aux personnes examinées pour cet item

Fig. 2. Méthode des scores combinés. Exemple de réponses à un item par 20 expertes.

clinique des experts et, par conséquent, qu'il n'y a pas lieu de ne retenir qu'une seule bonne réponse par item. On utilise les réponses des experts recrutés pour construire un barème de scores du TCS, où chaque réponse des experts possède une valeur intrinsèque et reçoit une pondération en fonction du nombre d'experts qui l'ont choisie. Plus précisément, la réponse à un item la plus souvent choisie par les experts (la réponse modale) vaut un point ; les autres réponses des experts reflètent la variabilité dans l'interprétation des données et donnent droit à une fraction d'un point, en fonction du nombre d'experts qui les ont choisies^[10]. La figure 2 montre un exemple de détermination des scores de novices à un item, en fonction des réponses des experts.

Évaluation de l'appréciation du TCS-DC

Après avoir rempli le TCS, les participantes devaient répondre à quelques questions (à l'aide d'une échelle

de type Likert) sur leur appréciation du TCS-DC en tant que méthode d'évaluation du raisonnement clinique des infirmières, dont le contenu traitait de la détérioration clinique sur les unités de soins de médecine ou de chirurgie.

Validité de contenu

Des expertes évaluatrices ont jugé la pertinence et la clarté du TCS-DC global, des scénarios et des items à l'aide des indices de la validité du contenu (IVC). L'IVC est une méthode décrite par Polit et Beck^[21] qui est très utilisée pour vérifier la validité de contenu d'instruments développés en sciences infirmières. L'IVC donne un indice de l'accord qu'il y a entre plusieurs experts concernant la pertinence et la clarté du contenu évalué par l'instrument de mesure. Cette méthode consiste à demander à un groupe d'experts de coter chacun des items de l'instrument en fonction

de leur pertinence et de leur clarté, grâce à deux échelles de Likert à quatre niveaux (1. non pertinent, 2. peu pertinent, 3. assez pertinent, 4. Pertinent ; 1. non clair, 2. peu clair, 3. assez clair, 4. clair).

Analyse des données

Des analyses descriptives ont été réalisées pour toutes les données recueillies. Pour répondre aux objectifs de la recherche, les analyses suivantes ont été effectuées. La cohérence interne du TCS-DC a été évaluée par le coefficient alpha de Cronbach. Les scores des participantes, des novices comme des expertes, sont exprimés en moyennes ramenées sur $100 \pm$ écarts-types, suivies de l'étendue, après avoir vérifié la normalité la distribution des scores à l'aide du test de Kolmogorov-Smirnoff. Les scores moyens des deux groupes étaient comparés par un test *t* de Student. Les scores moyens des participantes au temps 1 et au temps 2 ont été comparés par un test *t* de Student pour échantillons appariés. Pour estimer la corrélation entre le temps 1 et le temps 2 (données continues), indiquant la stabilité temporelle du TCS, un coefficient de corrélation intra-classe (CCI) été utilisé. Une relation entre l'expertise clinique et les scores au TCS a été vérifiée à l'aide d'une corrélation de Pearson. Une valeur bilatérale de $p < 0,05$ était considérée comme statistiquement significative. Le logiciel SPSS, version 18 a été employé.

Des IVC ont été calculés, respectivement pour la pertinence et pour la clarté, pour chaque partie du scénario (situation clinique, hypothèse, information complémentaire, item ; IVC-I item) et pour chaque scénario en leur entier (IVC-I scénario), ainsi que pour le test total (IVC-S). Les IVC-I item ont été calculé en dénombrant les scores 3 ou 4 obtenus et en divisant par le nombre d'expertes. Les IVC-I scénarios ont été calculés en faisant la moyenne des IVC-I items. Enfin, les IVC-S ont été calculés en faisant la moyenne de tous les IVC-I scénarios, appelée IVC-S/Average par Polit et Beck^[21]. Les scores des IVC-I item et des IVC-I scénarios obtenus devaient être supérieurs à 0,80 pour être jugés comme acceptables, tandis que les scores obtenus pour les IVC-S devaient être supérieurs à 0,90^[21].

Résultats

Le TCS-DC a été effectivement administré à 49 étudiantes (novices), soit près de 25 % de la population accessible. Deux novices ne satisfaisaient pas les critères d'inclusion et ont été exclues du groupe des novices, mais leurs réponses ont été conservées pour le test-retest. Dix de ces étudiantes ont participé au test-retest pour un total de douze participations. Les novices avaient entre 21 et 47 ans, avec un âge moyen de 25 ans. Elles œuvraient dans différents milieux de soins (trois en santé communautaire, 19 en médecine ou en chirurgie, 14 en soins critiques, quatre dans des milieux autres et trois ne travaillaient pas). Des novices ($n = 19$) présentant les mêmes caractéristiques, mais n'étudiant pas à l'université, ont aussi prétesté le TCS. Sept infirmières et deux CEPI ont vérifié si les instructions pour accéder au TCS et les instructions décrivant comment le remplir étaient claires, si les scénarios étaient similaires à la pratique clinique, si le vocabulaire était facilement compris, puis 10 infirmières ont évalué le temps nécessaire pour répondre au questionnaire.

Description du TCS-DC

La version finale du TCS-DC, après son optimisation, c'est-à-dire soit après le retrait des items ayant une corrélation item-total faible (inférieure à 0,05 ou négative), soit après avoir effectué une analyse des réponses des participantes *a posteriori* permettant d'augmenter sa cohérence interne et sa validité de contenu, possède 20 scénarios et 68 items. Selon les données du prétest, le temps pour répondre à la version préliminaire de 22 scénarios et 88 items variait de 25 à 50 minutes, avec un temps moyen de 40 minutes.

Validité de contenu

Des expertes évaluatrices ont évalué le contenu à l'aide des indices de la validité de contenu (IVC). Les IVC finaux, obtenus après les deux rondes d'évaluation par les expertes évaluatrices (où les scénarios

avaient été reformulés et aucun n'avait été retiré) et après l'optimisation du TCS-DC) sont les suivants. Les IVC-I items variaient de 0,80 à 1,00 pour la pertinence et de 0,75 à 1,00 pour la clarté. Les IVC-I scénarios varient entre 0,95 et 1,00 pour la pertinence et 0,96 à 1,00 pour la clarté. Les IVC-S de la pertinence et de la clarté du TCS-DC ont respectivement été évalués à 0,995 et à 0,993.

Fidélité

Le coefficient alpha de Cronbach du TCS-DC s'élève à 0,76 indiquant une bonne cohérence interne. L'étude de la stabilité temporelle a révélé que les scores moyens au temps 1 (T1) ($62,9 \pm 7,2$) et au temps 2 (T2) ($65,4 \pm 4,7$) n'étaient pas statistiquement différents l'un de l'autre ($p > 0,05$). En vérifiant la présence d'une corrélation entre les deux temps, un CCI modéré de 0,554 (intervalle de confiance à 95 % = 0,002, 0,847 ; $p < 0,05$) a été obtenu. Cela signifie que les scores étaient modérément comparables entre le T1 et au T2, puisque le CCI se situe entre 0,400 et 0,600^[22].

Validité de construit discriminante

Les scores moyens ramenés sur 100 obtenus par le groupe de novices étaient de $64,2 \pm 8,5$; ceux des expertes étaient de $80,2 \pm 5,6$. Il y avait une différence statistiquement significative entre les deux groupes ($p < 0,05$) ainsi qu'une augmentation des scores selon l'expertise clinique. Il existait une relation positive statistiquement significative ($r = 0,525$; $p < 0,05$) entre les scores moyens obtenus au TCS-DC et le nombre d'années d'expérience clinique.

Appréciation du TCS-DC

L'appréciation du TCS-DC par les novices et les expertes est présentée dans le Tableau II. Les éléments ayant été les plus fortement appréciés par plus de 70 % des participantes sont la ressemblance des scénarios du TCS-DC avec les situations cliniques

rencontrées dans la pratique clinique et la possibilité d'utiliser le TCS-DC pour favoriser des apprentissages. Les novices ont également estimé que le TCS-DC permettait d'évaluer leur capacité à évaluer, à surveiller, à interpréter des signes de détérioration clinique ou à intervenir en contexte de détérioration clinique tandis que les expertes avaient une appréciation modérée de cette caractéristique. Le TCS-DC a aussi reçu une évaluation modérée par 40 à 69 % des novices comme des expertes en ce qui a trait à son utilité lors d'entrevue ou d'évaluation sommative, ou comme une méthode d'évaluation efficace du raisonnement clinique. Finalement, les novices ont trouvé que le TCS-DC était modérément agréable à remplir tandis que les expertes l'ont trouvé agréable à remplir.

Discussion

Deux études précédentes en sciences infirmières, qui s'étaient intéressées à la construction et la validation d'un TCS^[15, 16], avaient démontré que le TCS pouvait être utilisé en sciences infirmières. Les résultats de notre étude vont dans le même sens. Le TCS-DC comporte, après optimisation, 20 scénarios et 68 items évaluant le raisonnement clinique infirmier en contexte de détérioration clinique en unités de soins de médecine ou de chirurgie. Sa cohérence interne est satisfaisante. Ceci peut sans doute être mis au crédit du nombre élevé d'items du questionnaire et du recrutement d'un échantillon d'experts assez important. Sa stabilité temporelle est modérée et devrait être réévaluée avec un plus grand échantillon, afin d'avoir une meilleure idée de celle-ci. En effet, l'intervalle de confiance très large du coefficient de corrélation calculé démontre l'imprécision de la mesure qui est liée à la petite taille des échantillons. Outre la petite taille d'échantillon, cette variation et ce CCI modéré pourraient s'expliquer par la hausse moyenne de 2,5 % des scores au T2 qui n'a pas été jugée statistiquement significative au test *t* pour échantillons appariés, mais qui indique que les réponses des étudiantes s'étaient modifiées d'un temps à l'autre.

Tableau II. Appréciation du test de concordance de script.

	Novices		Expertes	
	Nombre	%	Nombre	%
Les scénarios de ce questionnaire décrivaient des situations de détérioration clinique qui pourraient être rencontrées dans ma pratique	33	70,2	18	85,7
J'ai trouvé que le questionnaire permettait d'évaluer mes capacités à évaluer, à surveiller, à interpréter des signes de détérioration ou à intervenir en contexte de détérioration clinique	33	70,2	13	61,9
Je crois que ce questionnaire pourrait être utilisé à des fins d'évaluation lors d'un cours ou d'une entrevue	27	57,4	9	42,9
Je crois que le test de concordance de script représente une méthode efficace pour évaluer mon raisonnement clinique/ jugement clinique	22	46,8	12	57,1
Je crois que cette méthode d'évaluation pourrait favoriser des apprentissages	34	72,3	15	71,4
J'ai trouvé le questionnaire agréable à remplir	20	42,5	15	71,4

Cette valeur non parfaite de stabilité temporelle peut aussi s'expliquer par la présence d'un biais de mémoire ou de maturation. La validité du contenu, évaluée à l'aide de l'IVC, a obtenu de très bons résultats. Le contenu du TCS a été jugé comme pertinent et clair étant donné que les IVC-I étaient majoritairement supérieurs à 0,80 et que les IVC-S étaient supérieurs à 0,90^[21]. L'examen de la validité de construit discriminante a révélé que les scores des novices étaient significativement moins élevés que ceux des expertes : une relation linéaire entre les scores et le niveau d'expertise clinique a été observée. Enfin, de manière générale, le TCS-DC a été assez bien apprécié. Ainsi, les résultats de cette étude démontrent que ce TCS-DC est une méthode d'évaluation appropriée pour la mesure du raisonnement clinique infirmier dans un contexte de détérioration clinique parce qu'il est un outil fidèle et valide.

À notre connaissance, cette étude est la première à avoir élaboré un TCS selon les derniers critères recensés dans les écrits scientifiques. Nous avons fait une évaluation complète de la fidélité (cohérence interne et stabilité temporelle) et de la validité (validité de contenu et validité de construit discriminante), au-delà de ce qui était recommandé dans les articles scientifiques. En effet, l'étude de la fidélité des TCS avait été rapportée comme étant insatisfaisante^[13]

puisque seule la cohérence interne des TCS était évaluée, ce qui ne rendait pas compte de la constance de ses résultats. Par ailleurs, afin d'optimiser la qualité des situations cliniques possibles à inclure dans le TCS-DC, six personnes (trois infirmières collaboratrices et les trois chercheurs) au lieu de deux^[10] ont collaboré à cette étape. Ensuite, pour la première fois, la validité du contenu a été évaluée : auparavant, la validité apparente des TCS était évaluée à l'aide de l'avis de deux experts, tandis que dans cette étude, cinq expertes évaluatrices ont vérifié la pertinence et la clarté de tous les éléments constituant le TCS. Enfin, un pré-test a été effectué auprès d'infirmières et de CEPI appartenant à la population cible.

Notre étude comporte néanmoins certaines limites. La méthode d'échantillonnage pour recruter les participants était non probabiliste et de convenance. Par ailleurs, le TCS-DC était un questionnaire administré en ligne sur internet, ce qui a rendu difficile le contrôle de la participation : les étudiants comme les experts ont pu utiliser des aides externes pour remplir le questionnaire, ce qui peut avoir biaisé leurs scores. L'impact d'un biais de mémoire et d'un biais de maturation peut aussi être discuté puisque nous avons évalué la stabilité temporelle à l'aide d'un test retest et que nous n'avons pu contrôler le moment de réponse au re-test.

Par ailleurs, notre étude amène à réfléchir à la définition de l'expertise en sciences infirmières et au choix des experts pour constituer le panel de référence permettant le calcul des scores. En effet, lors du processus de validation du TCS-DC, deux expertes ont dû être retirées du panel d'expertes selon la méthode proposée par Gagnon et al.^[20], puisque leurs réponses étaient discordantes par rapport à celles des autres membres du panel et que leurs scores se situaient à plus de deux écarts-types en dessous de la moyenne en comparaison avec les scores des autres expertes. Deschênes^[15] avait fait une constatation semblable et avait retiré trois de ses experts. Elle expliquait ces résultats par le fait que ces trois experts avaient probablement un processus de raisonnement clinique différent de la plupart des autres experts et que les critères pour choisir les experts en *Human caring* devaient manquer de spécificité. Nous pensons que, comme les expertes retirées de notre étude avaient un grand nombre d'années d'expérience en soins critiques, leur processus de raisonnement clinique pouvait aussi être différent. Les données socio-démographiques collectées et le degré d'appréciation ne montraient pas de différence entre ces deux expertes et les autres membres du panel, excepté pour une des deux expertes exclues qui était sans emploi au moment de sa participation. Ainsi, nous ne pouvons expliquer ces différences de manière objective et vérifiable. Une hypothèse d'explication que nous pouvons aussi émettre est que le critère d'exclusion de ne pas avoir travaillé activement auprès des patients était trop large et pourrait être réduit. Une autre hypothèse d'explication est que les expériences vécues par ces deux expertes au cours de leur carrière ont mené à la création ou à l'organisation de scripts différents de ceux de la plupart des autres expertes recrutées dans cette étude. Enfin, une dernière hypothèse que nous pouvons émettre est que ces experts pouvaient ne pas avoir suivi de formation continue, de sorte que leurs connaissances n'étaient pas totalement à jour. En somme, nous pensons que pour améliorer la sélection des experts, des critères d'inclusion plus spécifiques devraient être utilisés. Étant donné que l'expertise est influencée par l'expérience clinique et les connaissances, il serait intéressant qu'en plus de prendre en compte le nombre

d'années d'expérience et le niveau de scolarité, on tienne compte de la mise à jour continue des connaissances de l'expert à l'aide de la formation continue.

De plus, notre étude répond partiellement à la question de l'acceptabilité d'un TCS-DC de la part des étudiantes en sciences infirmières. Les étudiantes, comme les expertes, ont jugé que les scénarios du TCS-DC ressemblaient à leur pratique et croient que cette méthode pourrait faciliter des apprentissages. Elles sont toutefois moins favorables à son utilisation pour l'évaluation sommative ou en situation d'entrevue. Les novices l'ont aussi trouvé moyennement agréable à remplir. Par ailleurs, seul le quart de la population accessible l'a fait. Pour sa part, Latreille^[16] avait utilisé un questionnaire d'opinion qui avait démontré que le TCS était jugé comme assez difficile (niveaux de 3 à 5, sur une échelle de difficulté maximale à 5) par 80 % des étudiantes de troisième année en sciences infirmières. Elle expliquait cette réaction par le format inhabituel du TCS et par l'ambiguïté des items, mentionnés dans les commentaires des étudiantes. Ces résultats démontrent que, comme le TCS en sciences infirmières est très nouveau, il serait nécessaire d'en expliquer davantage les principes aux participantes pour une meilleure acceptabilité car, bien qu'elles aient reçu les explications nécessaires et qu'elles aient fait l'expérience d'un exemple avec l'étudiante-chercheuse, le format inhabituel pour les infirmières crée un inconfort. Toutefois, la longueur du TCS-DC (40 minutes en moyenne avant son optimisation) constituera toujours un frein à son utilisation volontaire.

En conclusion, cette étude est originale et novatrice : elle est la troisième en sciences infirmières et la première à avoir évalué de façon exhaustive la fidélité et la validité. Les résultats principaux de cette étude se résument : 1) à la construction d'un TCS évaluant le raisonnement clinique infirmier en contexte de détérioration clinique en unités de médecine et de chirurgie ; 2) à la démonstration d'une bonne cohérence interne et d'une stabilité temporelle modérément satisfaisante du TCS-DC ; 3) à la démonstration de la capacité du TCS à distinguer les infirmières novices des infirmières expertes en fonction de leur niveau d'expérience avec les situations

de détérioration clinique et d'une association positive entre les scores et le nombre d'années d'expérience ; 4) à la documentation d'excellents IVC pour le TCS bien que les novices comme les expertes aient trouvé moins de ressemblance avec la pratique clinique que les experts évaluateurs ; 5) à la documentation d'une acceptabilité satisfaisante par les participantes, comparable à celle mise en évidence dans une autre étude en sciences infirmières. Des études permettant de perfectionner le TCS pourront être entreprises. Entre autres, une étude devrait reprendre l'étude de la stabilité temporelle du TCS-DC afin de vérifier si le TCS possède une meilleure stabilité temporelle lorsqu'un plus grand échantillon est utilisé. Il conviendrait également de vérifier si le TCS est en lui-même un outil d'apprentissage, ce qui permettrait de vérifier sa sensibilité au changement, compte tenu du constat d'une hausse moyenne du score de 2,5 points lors de la deuxième administration et puisque les participants ont trouvé que le TCS-DC pourrait être utilisé pour favoriser des apprentissages. Il faudrait aussi vérifier sur un plus grand nombre de participants si les scores aux TCS sont influencés par l'expérience en soins critiques ou par un cours en soins critiques. D'autres études pourraient également vérifier la validité de construit convergente ou faire une analyse factorielle, entre autres analyses, afin de renforcer la démonstration des qualités psychométriques du TCS-DC. Finalement, puisque les études en sciences infirmières portant sur le TCS ont su démontrer le potentiel du TCS dans l'évaluation du raisonnement clinique infirmier, d'autres études auraient intérêt à être entreprises pour évaluer le raisonnement clinique infirmier dans différents contextes et milieux de soins.

Contributions

Le projet de recherche présenté dans cet article a été élaboré dans le cadre d'un travail académique (maîtrise). L'étudiante-chercheuse a reçu deux bourses pour effectuer ce projet de recherche : une bourse d'excellence institutionnelle de l'Université de Sherbrooke et une bourse de la Fondation de recherche en

sciences infirmières du Québec (FRESIQ), en partenariat avec le Ministère de l'éducation, du loisir et du sport du Québec (MELS). Accompagnée par ses directrices de maîtrise (Andréanne Tanguay et Cécile Michaud), l'étudiante-chercheuse (Sarah Côté) a effectué la recension des écrits, planifié l'étude, élaboré le protocole de recherche, élaboré le TCS, l'a validé auprès d'expertes et l'a pré-testé, obtenu les autorisations éthique et administrative, recruté les participantes, recueilli et analysé les données, interprété les résultats, rédigé l'article et fait les corrections jusqu'à satisfaction de l'ensemble des auteurs. Bernard Charlin et Robert Gagnon ont agi à titre d'experts pour faire l'analyse et l'interprétation des données ainsi que pour rédiger l'article. Leurs recherches antérieures ont grandement servi pour l'élaboration et la validation du TCS. Ils ont finalement corrigé l'article.

Approbation éthique

Le projet d'étude a été accepté par le Comité d'éthique de la recherche en santé chez l'humain du Centre de recherche clinique Étienne-Le Bel du Centre Hospitalier Universitaire de Sherbrooke (CHUS). La participation à cette étude était faite sur une base volontaire. En acceptant de remplir le TCS-DC, les participantes donnaient ainsi leur consentement de manière implicite. L'autorisation de recruter dans l'École des sciences infirmières de l'université concernée a été obtenue du directeur avant de commencer l'étude.

Déclaration d'intérêts

SC déclare n'avoir aucun conflit d'intérêts, AT déclare n'avoir aucun conflit d'intérêts, R.G déclare n'avoir aucun conflit d'intérêts, BC déclare n'avoir aucun conflit d'intérêts, CM déclare n'avoir aucun conflit d'intérêts.

Références

1. Jean S. Direction générale du personnel, réseau ministériel. Ministère de la Santé et des Services sociaux [MSSS] Projection de la main-d'œuvre

- infirmière, de 2004-2005 à 2019-2020. 2005 [On-line] Disponible sur : http://www.fiqsante.qc.ca/documents_publications/documents/pmo_inf_rapport2005. Pdf.
2. Liaw S, Scherpbier A, Klainin-Yobas P, Rethans J. A review of educational strategies to improve nurses' roles in recognizing and responding to deteriorating patients. *Int Nurs Rev* 2011;58:296-303.
 3. Ordre des infirmières et infirmiers du Québec [OIIQ]. Champ d'exercice et activités réservées, 2012 [On line]. Disponible sur : <http://www.oiiq.org/pratique-infirmiere/champ-dexercice-et-activites-reservees>.
 4. Albert B, Huesman L. Development of a modified early warning score using the electronic medical record. *Dimens Crit Care Nurs* 2011;30:283-92.
 5. Kyriacos U, Jelsma J, Jordan S. Monitoring vital signs using early warning scoring systems: a review of the literature. *J Nurs Manag* 2011;19:311-30.
 6. Côté S, St-Cyr Tribble D. Le raisonnement clinique des infirmières, analyse de concept. *Recherche en soins infirmiers* 2012;111:13-21.
 7. Hoffman K, Aitken L, Duffield C. A comparison of novice and expert nurses' cue collection during clinical decision-making: verbal protocol analysis. *Int J Nurs Stud* 2009;46(10):1335-44.
 8. Schmidt HG, Norman GR, Boshuizen HP. A Cognitive Perspective on Medical Expertise: Theory and Implications. *Acad Med* 1990;65:611-21.
 9. Whittington J, White R, Haig K, Slock M. Using an automated risk assessment report to identify patients at risk for clinical deterioration. *Jt Comm J Qual Patient Saf* 2007;33:569-74.
 10. 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.
 11. Lubarsky S, Charlin B, Cook D, Chalk C, van der Vleuten C. Script concordance testing: a review of published validity evidence. *Med Educ* 2011;45:329-38.
 12. 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.
 13. Fournier J, Demeester A, Charlin B. Script concordance tests: guidelines for construction. *BMC Med Inform Decis Mak* 2008;8:818.
 14. Lineberry M, Kreiter C, Bordage G. Threats to validity in the use and interpretation of script concordance test scores. *Med Educ* 2013;47:1175-83.
 15. Deschênes M, Charlin B, Gagnon R, Goudreau J. Use of a script concordance test to assess development of clinical reasoning in nursing students. *J Nurs Educ* 2011;50:381-7.
 16. Latreille, M-E. Évaluation du raisonnement clinique d'étudiantes et d'infirmières dans le domaine de la pédiatrie, à l'aide d'un test de concordance de script. Ottawa : Université d'Ottawa, 2012.
 17. Petrucci A, Nouh T, Boutros M, Gagnon R, Meterissian S. Assessing clinical judgment using the Script Concordance test: the importance of using specialty-specific experts to develop the scoring key. *Am J Surg* 2013;205:137-40.
 18. Centre de pédagogie appliquée aux sciences de la santé (CPASS). Test de concordance de script. 2012 [On-line] Disponible sur : <http://www.cpass.umontreal.ca/tcs.html>.
 19. Gagnon R, Charlin B, Lambert C, Carrière B, Van der Vleuten C. Script concordance testing: more cases or more questions? *Adv Health Sci Educ Theory Pract* 2009;14:367-75.
 20. Gagnon R, Lubarsky S, Lambert C, Charlin B. Optimization of answer keys for script concordance testing: should we exclude deviant panelists, deviant responses, or neither? *Adv Health Sci Educ Theory Pract* 2011;16:601-8.
 21. Polit D, Beck C. *Nursing research: generating and assessing evidence for nursing practice*. (8th ed.) Philadelphie (PA) : Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008.
 22. Zegers M, de Bruijne MC, Wagner C, Groenewegen PP, van der Wal G, de Vet HC. The inter-rater agreement of retrospective assessments of adverse events does not improve with two reviewers per patient record. *J Clin Epidemiol* 2010;63:94-102.

Correspondance et offprints : Sarah Côté, École des sciences infirmières, Université de Sherbrooke, 3001, 12^e avenue Nord, Sherbrooke (Québec), J1H 5N4.
Mailto : sarah.cote@usherbrooke.ca