

# Le potentiel diagnostique des questions d'un test de concordance de scripts pour évaluer le raisonnement clinique infirmier

*The diagnostic potential of questions in a script concordance test to assess nurses' clinical reasoning*

Katia DUMONT<sup>1</sup>, Nathalie LOYE<sup>1</sup>, Johanne GOUDREAU<sup>2</sup>

<sup>1</sup> Faculté des sciences de l'éducation, Université de Montréal, Canada

<sup>2</sup> Faculté des sciences infirmières, Université de Montréal, Canada

Manuscrit reçu le 20 septembre 2014 ; commentaires éditoriaux formulés aux auteurs le 23 mai et le 6 juin 2015 ; accepté pour publication le 8 juin 2015

## Mots-clés

Raisonnement clinique infirmier ; Évaluation diagnostique ; Test de concordance de script

**Résumé – Contexte :** Pour guider les programmes de formation, il est nécessaire de pouvoir poser une évaluation diagnostique du raisonnement clinique des infirmières. Or, les outils disponibles à cet effet présentent des problèmes majeurs. Un outil d'évaluation novateur, le test de concordance de script, constitue une avenue prometteuse. **But :** Explorer la possibilité d'utiliser les questions d'un test de concordance de script comme outil d'évaluation diagnostique du raisonnement clinique d'étudiantes infirmières. **Données :** Des transcriptions d'entrevues réalisées auprès d'infirmières et d'étudiantes infirmières ( $n = 55$ ) ont été récupérées d'une étude antérieure<sup>[1]</sup>. Ces entrevues consistaient en la verbalisation à voix haute des pensées des participantes alors qu'elles répondaient à des questions d'un test de concordance de script. **Analyse :** Les transcriptions ont été codées afin d'identifier les catégories et les stratégies de pensée utilisées par les participantes pour répondre aux questions du test de concordance de script. Puis, des matrices Q ont été constituées afin de formaliser les relations entre les catégories/stratégies de pensée et les questions du test pour chacun des groupes de participantes. **Résultats :** Chaque groupe de participantes a utilisé différemment les catégories et stratégies de pensée pour répondre aux questions du test. Les questions du test pourraient permettre d'évaluer les forces et les faiblesses des étudiantes en lien avec des processus

cognitifs spécifiques à leur niveau de développement. Le raisonnement des expertes est trop différent de celui des étudiantes pour être utilisé comme point de référence. **Conclusions :** Les étudiantes infirmières utilisent des catégories et stratégies de pensée différentes selon les groupes, et certaines questions d'un TCS présentent un potentiel diagnostique pour évaluer le raisonnement clinique infirmier niveau par niveau (1<sup>ère</sup>, 2<sup>e</sup>, et 3<sup>e</sup> année).

### Keywords

Nurse clinical reasoning; Diagnostic evaluation; Script concordance test

**Abstract – Background:** The development of training programs requires an evaluation of nurses' clinical reasoning. However, there are issues with the evaluation, which uses existing tools. The script concordance test is a promising and innovative evaluation alternative. **Purpose:** To explore the possibility of using questions from a script concordance test as an evaluation tool in assessing nursing students' clinical reasoning abilities. **Data:** Transcripts of interviews with nurses and nursing students ( $n = 55$ ) from a previous study<sup>[1]</sup> were used. During the interviews, participants were asked to share their thoughts aloud while taking a script concordance test. **Analysis:** Transcripts were coded to identify thinking categories and strategies used by participants to answer questions from the script concordance test. Q-matrices were then developed to formalize relationships between thinking categories/strategies and the test questions in each group of participants. **Results:** Each group used different thinking categories/strategies to answer questions in the script concordance test. These questions could help assess nursing students' strengths and weaknesses as they relate to their development level. The reasoning of nursing experts was so unlike that of students that they could not be used as a benchmark for diagnostic evaluation. **Conclusion:** Depending on groups student nurses use different thinking categories and strategies, and some SCT questions show a diagnostic potential in evaluating nurses' clinical reasoning per level (1st, 2nd and 3rd year).

## Introduction

Les programmes de formation des professionnels de la santé s'inscrivent de plus en plus dans une approche axée sur le développement de compétences<sup>[2]</sup>. Pour les infirmières, le raisonnement clinique infirmier (RCI) est vu comme une compétence fondamentale pour l'exercice de leur profession et comme un défi prioritaire de leur formation<sup>[3]</sup>. Développer et évaluer des stratégies éducatives pour aider les infirmières à développer leur RCI pourrait contribuer à rendre les soins plus sûrs pour les patients<sup>[4]</sup>.

Puisque l'évaluation a un impact important sur les stratégies d'apprentissage des étudiants<sup>[5]</sup>, il paraît essentiel de développer des instruments d'évaluation valides et fidèles pour identifier les problèmes de RCI dans une approche diagnostique. De tels instruments pourraient donner des renseignements précis sur le niveau de développement des compétences des étudiantes et sur leurs difficultés, afin de guider les décisions pédagogiques<sup>[6]</sup>.

Or, les instruments qui sont traditionnellement utilisés pour évaluer le RCI présentent des problèmes majeurs<sup>[7]</sup>, notamment un manque de sensibilité et de spécificité<sup>[8]</sup>. La plupart de ces instruments mesurent des comportements observables qui ne témoignent pas complètement de la nature complexe de ce processus cognitif<sup>[9]</sup>. De ce fait, ces méthodes s'avèrent peu cohérentes avec les approches d'enseignement où les apprentissages sont contextualisés afin de contribuer au développement des compétences des professionnels. Évaluer la capacité des étudiants à utiliser un RCI approprié aux situations cliniques est une tâche difficile qui exige des mesures d'évaluation pertinentes<sup>[10]</sup>. Ce projet de recherche visait donc à explorer des stratégies novatrices pour évaluer le RCI de manière diagnostique et contextualisée.

Dans cette perspective, plusieurs études dans différentes disciplines et dans différents contextes indiquent la validité, la fiabilité, la faisabilité et l'applicabilité d'un outil novateur d'évaluation du raisonnement clinique : le test de concordance de scripts (TCS)<sup>[11]</sup>. Le TCS est un examen standardisé,

construit à partir de situations professionnelles authentiques où plusieurs scénarios cliniques sont présentés à l'étudiant. Outre la richesse des connaissances factuelles, le TCS permet d'apprécier la qualité de l'organisation des connaissances, de même que la pertinence de cette organisation pour agir dans des situations cliniques données<sup>[9, 10, 12]</sup>. À ce jour, des TCS ont été évalués et validés pour évaluer le raisonnement clinique des infirmières<sup>[13-16]</sup>. Toutefois, les résultats de recherche actuels ne permettent pas de conclure quant à la pertinence d'un TCS pour l'évaluation diagnostique.

Déoulant de ces considérations, la présente étude explorait la possibilité d'utiliser des vignettes et questions du TCS de Deschênes et al.<sup>[13]</sup> comme outil d'évaluation diagnostique du RCI auprès d'étudiantes infirmières en contexte de formation universitaire, afin de cerner leurs forces et leurs difficultés. De ce but, nous formulons les questions spécifiques de recherche suivantes : « Dans quelle mesure chaque question d'un TCS visant à évaluer le RCI peut-elle être reliée à des catégories et à des stratégies de pensée spécifiques ? » et « Comment peut-on décrire le potentiel diagnostique d'un TCS pour évaluer le RCI ? ».

## Contexte théorique

Le raisonnement clinique infirmier

Dans le domaine de la santé, le raisonnement clinique est considéré comme l'ensemble des pensées et des processus de décision associés à la pratique clinique. Il permet aux praticiens de choisir la meilleure action dans un contexte particulier<sup>[17]</sup>. Le raisonnement clinique consiste à acquérir, organiser et interpréter des informations, proposer et tester des hypothèses, et à évaluer des diagnostics alternatifs pour, finalement, engager un traitement<sup>[18]</sup>. Cette activité intellectuelle se fonde sur la synthèse de l'information obtenue à partir d'une situation clinique afin de prendre des décisions de diagnostic et de prise en charge du patient, en intégrant les connaissances et expériences antérieures<sup>[9]</sup>.

Des études menées en sciences infirmières ont permis de décrire les processus cognitifs que les infirmières utilisent spécifiquement afin de comprendre et donner un sens aux données recueillies auprès d'un patient et ainsi, établir un plan de traitement qui réponde à ses besoins<sup>[19]</sup>. Ces processus seraient caractérisés par l'utilisation de connaissances spécifiques à chaque discipline et de stratégies de pensée<sup>[20, 21]</sup>. Ces dernières constituent un élément fondamental du raisonnement clinique et ont été identifiées par Fonteyn<sup>[22]</sup> dans une étude descriptive auprès de 15 infirmières reconnues pour leur expertise dans leur domaine. Cette étude a permis d'identifier 13 stratégies de pensée utilisées par les infirmières dans leur pratique (Tableau I), qui se déclinent en 61 catégories d'application spécifique. D'autres recherches indiquent que la qualité du RCI serait corrélée positivement à l'expertise de l'infirmière, notamment en ce qui a trait à la systématisation de la collecte de données et à l'exactitude diagnostique<sup>[13]</sup>.

L'évaluation du raisonnement clinique

Plusieurs instruments pour évaluer le raisonnement clinique ont été développés dans le domaine des sciences de la santé. Citons, par exemple, le *Patient Management Problem*, la grille d'évaluation globale, le questionnaire à choix multiples, l'examen clinique objectif structuré et le *Clinical Reasoning Problem*<sup>[9]</sup>. La majorité de ces outils a été développée dans le contexte de la médecine, et peu sont disponibles pour évaluer le raisonnement clinique des infirmières. Certains de ces outils s'attardent à des connaissances factuelles, plutôt qu'à la capacité de résolution de problème des étudiants. D'autre se fondent sur un nombre limité d'observations dans un petit échantillon de situations cliniques pour évaluer la compétence des étudiants. Ceci ne témoignerait pas des subtilités et de la contextualité d'un processus cognitif comme le raisonnement clinique. Des difficultés liées à de longues périodes d'observation par un même observateur ont aussi été notées. Lorsqu'une comparaison avec un groupe d'experts sert de point de référence, il demeure difficile de déterminer une démarche experte constante à laquelle comparer celle des étudiants<sup>[9, 23]</sup>.

**Tableau I.** Proportions d'utilisation minimales et maximales et les valeurs seuil pour les 21 catégories retenues.

Stratégies de pensée	Catégories de pensée	Proportion		Valeur seuil
		% min	% max	%
Faire des choix	a. actions	0	64	32
Chercher des informations manquantes	b. approfondir ou préciser une évaluation	7	92	46
	c. histoire du patient	0	46	23
	d. médication	0	50	25
	e. ressources	0	43	21,5
Juger de la valeur	f. conclusions d'évaluation	0	46	23
	g. traitements (ou interventions)	0	71	35,5
Relier des informations	h. histoire du patient et problèmes/ situation du patient	0	71	35,5
	i. information sur le traitement /autres type d'info.	0	57	28,5
	j. conclusions d'évaluations/autres informations	7	71	35,5
Énoncer une règle de pratique	k. politique	0	71	35,5
Faire des hypothèses	l. identifier besoins du patient	0	50	25
	m. spéculer sur le statut du patient	0	64	32
	n. déterminer la cause	0	93	46,5
Énoncer une proposition	o. choisir un plan d'action	0	71	35,5
Fournir des explications	p. sur des actions	0	43	21,5
	q. interventions infirmières	7	54	27
Ordonner des concepts par priorités	r. focus sur plan d'action	0	43	21,5
Conclure	s. condition du patient ou statut	0	43	21,5
	t. une situation	0	36	18
Autres stratégies de pensée	u. faire une pause réflexion	7	86	43

Or, le test de concordance de script (TCS) est standardisé et construit à partir de situations professionnelles authentiques<sup>[10]</sup>. Il consiste à présenter aux étudiants une série de situations cliniques authentiques, puis à leur demander de considérer l'effet d'une nouvelle information sur une hypothèse crédible. La situation décrite pose un problème même pour un expert puisqu'il manque un ou plusieurs éléments, ou que certaines données sont incomplètes, ambiguës ou difficilement interprétables<sup>[12, 24-26]</sup>. Une échelle de type Likert à cinq niveaux est proposée pour répondre à chaque question. Un groupe

d'experts du domaine sert de référence. La comparaison avec leurs réponses permet d'attribuer un score à chaque étudiant. Le crédit donné aux étudiants pour chaque réponse est fonction du nombre d'experts qui ont fourni la même réponse qu'eux<sup>[9, 27]</sup>.

Le TCS est conçu pour apprécier la qualité et la pertinence de l'organisation des connaissances de l'étudiant pour agir efficacement dans un contexte de pratique professionnelle<sup>[13]</sup>. Contrairement au questionnaire à choix multiples, le TCS évalue les compétences des candidats et comment ils utilisent l'information pour confirmer ou éliminer des

hypothèses<sup>[27, 28]</sup>. Dans plusieurs domaines de la médecine, diverses études visent à développer des TCS comme modalité d'évaluation formative notamment en neurologie<sup>[29]</sup>, en radio-oncologie, en chirurgie et en sciences infirmières<sup>[30]</sup>.

Plusieurs recherches ont été conduites sur des TCS afin d'en évaluer les qualités psychométriques. Ces études ont abordé la validité, la fidélité, la faisabilité et l'applicabilité dans différentes disciplines et divers contextes, et soutiennent positivement le TCS<sup>[11]</sup>. Il a montré sa capacité à détecter les candidats plus expérimentés cliniquement et à distinguer les différents groupes de participants en fonction de leur niveau d'expertise<sup>[12, 31, 32]</sup>. Dans le domaine des sciences infirmières, Deschênes<sup>[13]</sup> a développé un TCS afin d'évaluer le raisonnement clinique empreint d'une philosophie de soins humaniste dans des situations de soins courantes de la pratique infirmière. Ce TCS a montré des propriétés psychométriques adéquates (alpha de Cronbach = 0,86, discrimination entre experts et étudiants) et offre ainsi une méthode standardisée d'évaluation du RCI en sciences infirmières<sup>[14]</sup>. Le choix de ce TCS semblait donc pertinent pour cette étude qui consistait à explorer une avenue novatrice pour évaluer le RCI.

Vers une utilisation diagnostique  
du test de concordance de script

Étudier le potentiel diagnostique du TCS revient à s'interroger sur sa capacité à distinguer les particularités individuelles des étudiants, et à déterminer leurs forces et leurs faiblesses en « s'intéress[ant] à la démarche qui [leur] permet d'aboutir à la réponse, plutôt qu'à la réponse elle-même »<sup>[33, p.52]</sup>. Il faut donc chercher quelles stratégies de pensée spécifiques sont sollicitées par chaque question du TCS et vérifier si elles varient d'une question à une autre.

Une manière de formaliser les relations entre les processus cognitifs, les stratégies de pensée par exemple, et les questions d'un test a été proposée par Tatsuoka<sup>[34]</sup> sous la forme d'une matrice nommée Q. Elle prend la forme d'un tableau dans lequel les questions constituent les colonnes et les processus cognitifs, les lignes. Une valeur 1 à l'intersection d'une ligne et d'une colonne indique que le processus est

requis par la question. La matrice Q fournit la structure diagnostique d'un test. Une fois cette structure connue, il est possible de décrire les forces et les faiblesses d'un étudiant en considérant les processus cognitifs que partagent les questions réussies et celles non réussies, ce qui constitue une évaluation diagnostique.

Ainsi, structurer un TCS de RCI comme celui de Deschênes<sup>[13]</sup> sous la forme d'une matrice Q pourrait être la première étape pour utiliser un tel test à des fins diagnostiques. Pour élaborer cette matrice, il est possible d'analyser les liens entre les vignettes et questions d'un TCS et les stratégies (et les catégories) de pensée de Fonteyn<sup>[22]</sup>. Ainsi, il devrait être possible de formaliser les relations entre chaque question et les stratégies nécessaires pour répondre à une question.

## Méthodes

### Participant·es et collecte des données

Cette étude s'appuie sur l'analyse secondaire des données d'une étude antérieure<sup>[1]</sup>. L'étude antérieure visait à modéliser l'apprentissage du RCI, du début de l'apprentissage jusqu'à l'expertise. Les participant·es ( $n = 66$ ) étaient des infirmières nouvellement diplômées ( $n = 11$ , 12-18 mois dans le milieu clinique) ou expertes ( $n = 14$ , plus de 5 ans d'expérience) et des étudiantes infirmières de première ( $n = 14$ ), deuxième ( $n = 14$ ) et troisième ( $n = 13$ ) années de baccalauréat en sciences infirmières. Les participant·es ont été recrutées afin d'obtenir une variation maximale de leurs stades de développement de RCI. Dans la présente étude, les données des infirmières nouvellement diplômées n'ont pas été utilisées, en raison de la portée de l'étude, qui s'intéressait au contexte de la formation initiale.

Les données de l'étude antérieure<sup>[1]</sup> ont été recueillies grâce à des extraits du TCS de Deschênes<sup>[13]</sup> : 13 questions reliées à cinq situations cliniques (vignettes). Il est à noter que ce sont uniquement les vignettes et les questions, et non la procédure complète de réponse aux questions du TCS que ces chercheurs ont employé. Un exemple de

Vignette 4 (V4) : Vous visitez à domicile M et Mme Lafrance. Mme Lafrance souffre d'un cancer des os en phase terminale, elle désire mourir chez elle entourée de sa famille. Vous devez lui administrer ses injections d'analgésique à des heures régulières. À votre arrivée, M Lafrance a le faciès tendu et les larmes aux yeux, il se dit dépassé par la situation et estime que sa femme est très souffrante.

	Si vous pensez que :	Et qu'alors vous trouvez que :	
V4 Q1	Si vous pensez à demander à M. Lafrance la meilleure façon de les aider lui et sa femme pour le moment	M. Lafrance vous dit qu'il trouve que sa femme reçoit trop d'injections d'analgésiques et qu'il a peur que le médicament accélère son décès	Comment cette information influence-t-elle l'idée initiale?
V4 Q2	Si vous pensez à faire verbaliser M. Lafrance sur sa propre perception de la mort	M. Lafrance vous dit que seul Dieu peut décider de l'heure de la mort d'une personne	Comment cette information influence-t-elle l'idée initiale?

Fig. 1. Exemple de vignette et questions reliées.

vignette et de questions liées est présenté dans la figure 1.

Lorsqu'elles répondaient aux questions du TCS, les participantes étaient invitées à verbaliser leurs pensées à voix haute selon la méthode de *think aloud*<sup>[35]</sup>. Cette méthode est reconnue pour son utilité à collecter des données permettant d'élucider les processus cognitifs sous-jacents au RCI<sup>[20]</sup> et a été utilisé dans l'étude de Fonteyn<sup>[22]</sup> décrite précédemment. Ces données ont été enregistrées sur bande audio, puis transcrites. Ensuite, elles ont fait l'objet d'une analyse de protocole, méthode qui consiste à identifier les concepts utilisés par les participants lorsqu'ils raisonnent afin d'identifier les liens qu'ils font entre ces concepts et les processus de raisonnement qu'ils emploient<sup>[35]</sup>. Cette analyse s'est basée sur le modèle des stratégies de pensées de Fonteyn<sup>[22]</sup>. Les 61 catégories de stratégies de pensée ont fait office de codes pour l'analyse qualitative des données. L'analyse de protocole a été réalisée par une assistante de recherche et deux chercheurs ont analysés 15 % du matériel afin de valider la codification de l'assistante de recherche. Au final, 55 transcriptions (1/participante) et le cahier de codification ont été utilisés dans la présente étude.

#### Analyse des données

D'emblée, il est à noter que pour toute l'analyse, nous n'avons pas tenu compte des bonnes ou des mauvaises réponses des participantes, mais uniquement de la

manière dont elles s'y sont prises pour répondre aux questions. Pour commencer, nous avons divisé les transcriptions de chaque participante selon les questions du TCS. Ensuite, nous avons isolé les codes des catégories de pensée<sup>[22]</sup> dans les transcriptions pour chacune des questions du TCS. Un total de 727 codes combinés (question-catégorie) a été obtenu dans les 55 transcriptions.

Puis, nous avons produit un tableau de fréquence d'utilisation des codes combinés. Ce tableau indiquait le nombre de participantes ayant utilisé une catégorie de pensée pour répondre à une question. Ceci a permis de calculer les proportions d'utilisation des catégories de pensée pour chaque question (nombre de participantes ayant utilisé une catégorie de pensée/nombre total de participantes, exprimé en %). Étant donné que la présente étude s'intéressait au potentiel diagnostique des questions du TCS, nous avons fait le choix d'éliminer les catégories de pensée sous-utilisées et sur-utilisées. La sous-utilisation était définie comme une proportion d'utilisation inférieure à 25 % pour une catégorie de pensée et ce, pour toutes les questions. Le critère pour déterminer une sur-utilisation était que pour toutes les questions, la proportion d'utilisation d'une catégorie de pensée soit supérieure à 50 %. En procédant ainsi, nous avons retenu 21 des 61 catégories de pensée. Les autres ont été rejetées en raison d'une utilisation trop faible ou trop forte. Par ailleurs, deux stratégies de pensée ont été éliminées car toutes les catégories de pensée associées avaient été rejetées.



Par la suite, nous avons construit des graphiques pour chacune des 21 catégories de pensée. Ces graphiques présentaient les proportions d'utilisation des catégories de pensée pour chaque question, en tenant compte de l'appartenance aux groupes (1<sup>ère</sup>, 2<sup>e</sup> et 3<sup>e</sup> année de baccalauréat, expertes). Étant donné que nous souhaitions identifier les catégories de pensée utilisées pour chaque question, nous avons décidé d'établir un lien entre une question et une catégorie lorsqu'elle est utilisée par plus de 50% de l'utilisation maximum : la valeur seuil fait l'objet de la dernière colonne du tableau I. Par exemple, pour la catégorie « faire des choix d'action », le pourcentage le plus élevé est de 64 % ; nous avons donc seulement sélectionné les questions pour lesquelles le pourcentage est supérieur à 32 %.

À partir de ces résultats, nous avons construit une matrice Q<sup>34</sup>, où chaque ligne correspondait à une catégorie de pensée et chaque colonne à une question du TCS. Dans les cases, nous avons inscrit les groupes de participantes qui ont utilisé de manière significative une stratégie de pensée pour répondre à une question, en tenant compte des valeurs seuils déterminées précédemment. Nous avons construit une autre matrice Q pour les stratégies de pensée.

Par la suite, nous avons calculé les degrés d'accord entre les groupes pour les catégories de pensée et pour les stratégies de pensée. Le but de cette opération était de quantifier les différences entre les groupes. Pour ce faire, nous avons calculé le nombre de fois où des groupes se retrouvaient ensemble dans une case de la matrice. Ces nombres ont été divisés par le nombre de croisements effectifs entre les catégories et les questions ( $n = 124$ ). Nous avons procédé de la même manière pour les stratégies de pensée, où le nombre de croisements effectifs était inférieur ( $n = 96$ ).

Finalement, nous avons calculé les degrés d'accord entre les questions considérées deux par deux, dans le but d'identifier les ressemblances éventuelles des combinaisons de catégories de pensée et de stratégies de pensée, entre deux questions. À cette fin, nous avons calculé le nombre de fois où deux questions nécessitaient, ou non, les mêmes catégories de pensée et avons divisé par 21 (nombre de catégories de pensée). Nous avons procédé de la même

manière pour les stratégies de pensée, en divisant par 11 (nombre de stratégies de pensée).

### Considération éthiques

L'étude antérieure avait été approuvée par les comités d'éthique des deux milieux cliniques participants ainsi que par celui de l'université des chercheurs. Le consentement libre et éclairé de chaque participante a été sollicité par un agent de recherche. Les données ont été dénominalisées afin d'assurer la confidentialité et les participantes pouvaient se retirer de l'étude à tout moment.

La présente étude a été approuvée par le comité d'éthique de l'université de la chercheuse principale. Les renseignements identificatoires des participantes n'étant pas nécessaire, la confidentialité a été respectée. Lors de l'étude antérieure, les participantes avaient consenti à une étude sur le développement du RCI et la présente étude demeurait dans cette thématique. Ainsi, les participantes n'ont pas eu à signer un nouveau consentement pour la présente étude.

## Résultats

Le tableau I présente les proportions d'utilisation minimales et maximales et les valeurs seuil pour les 21 catégories retenues. Il met en évidence les grandes variations d'utilisation de chaque catégorie de pensée, avec des proportions d'utilisation maximales variant de 36 à 93 % et des proportions d'utilisation minimales entre 0 et 7 %.

Le tableau II contient la matrice Q pour les catégories de pensée et le tableau III, celle pour les stratégies de pensée. Ces tableaux mettent en évidence les catégories et les stratégies de pensée utilisées pour chaque question, par chaque groupe et montrent d'importantes différences entre les groupes. Le tableau IV, quant à lui, présente les pourcentages d'accord entre les groupes pour les catégories et les stratégies de pensée. Il indique des disparités importantes entre les groupes quant aux catégories utilisées. De manière assez logique, les pourcentages d'accord entre les groupes sont toujours plus élevés pour les stratégies de pensée qui regroupent plusieurs catégories. Étant donné le faible pourcentage d'accord entre

Tableau II. Matrice Q pour les catégories de pensée.

Stratégies	Catégories	V1q2	V1q3	V1q4	V2q2	V2q3	V2q4	V3q2	V3q3	V4q2	V4q3	V5q3	V5q4	V5q5	% des questions
1	a			13E				12E	2E						23 %
2	b				123	123	13			1			123	12	46 %
3	c		3	13		1		123E				3E	3		46 %
4	d	13E					123		123	1E					31 %
5	e			1				12			E				23 %
6	f			13	3						2				23 %
7	g	E							3	12E					23 %
8	h		12	23				23E				2E			31 %
9	i						123		1	E					23 %
10	j	123	13	2	3	23	2	2		2	12	3	23	1	92 %
11	k												2E		8 %
12	l	3						2E	2	E			1		38 %
13	m	123E	23E		123E	123	12E							12E	46 %
14	n		E	2E	2	12E		23E			2E	12	3		62 %
15	o	12				3		3	1				1	12E	46 %
16	p			1E				123E	123E	3					23 %
17	q	13	3	3	3	3	3	13	3		1			E	77 %
18	r	3		3		3		13E	13			23	3	12	62 %
19	s			123E	3E	3		3E	3	E	1	13	1		69 %
20	t	3	3E	23		23	2	2	3	E	123E		23E	2	85 %
21	u		12	1	1	23	23	12	12	1	12		1		69 %
Catégories sollicitées		43 %	38 %	62 %	38 %	52 %	38 %	67 %	52 %	48 %	38 %	29 %	52 %	33 %	

1 = 1<sup>ère</sup> année, 2 = 2<sup>e</sup> année, 3 = 3<sup>e</sup> année, E = expertes. La signification des chiffres et des lettres des deux premières colonnes est indiquée dans le tableau 1



Tableau III. Matrice Q pour les stratégies de pensée.

Stratégies	V1q2	V1q3	V1q4	V2q2	V2q3	V2q4	V3q2	V3q3	V4q2	V4q3	V5q3	V5q4	V5q5	% des questions
1. Faire des choix			13E				12E	2E						23 %
2. Chercher infos manquantes	13E	3	13	123	123	123	123E	123	1E	E	3E	123	12	100 %
3. Juger de la valeur	E		13	3				3	12E	2				46 %
4. Relier des infos	123	123	23	3	23	123	23E	1	2E	12	23E	23	1	100 %
5. Énoncer règle de pratique												2E		8 %
6. Faire des hypothèses	123E	23E	2E	123E	123E	12E	23E	2	E	2E	12	13	12E	100 %
7. Énoncer une proposition	12				3		3	1				1	12E	46 %
8. Fournir des explications	13	3	13E	3	3	3	123E	123E	3	1			E	85 %
9. Ordonner concepts/priorité	3		3		3		13E	13			23	3	12	62 %
10. Conclure	3	3E	123E	3E	23	2	23E	3	E	123E	13	123E	2	100 %
11. Autres stratégies		12	1	1	23	23	12		1	12		1		69 %
% des stratégies	73 %	55 %	82 %	64 %	73 %	55 %	82 %	82 %	64 %	55 %	45 %	73 %	64 %	
1 = 1 <sup>ère</sup> année, 2 = 2 <sup>e</sup> année, 3 = 3 <sup>e</sup> année, E = expertes.														

**Tableau IV.** Pourcentages d'accords entre les groupes pour les catégories et les stratégies de pensée.

Groupes	Catégories		Stratégies	
	Accords	Degrés d'accord	Accords	Degrés d'accord
	n	%	n	%
1 et 2	29	23	30	31
1 et 3	23	19	27	28
1 et E	16	13	19	20
2 et 3	27	22	29	30
2 et E	22	18	24	25
3 et E	18	15	21	22
1, 2, 3	15	12	17	18
1, 2, E	10	8	14	15
1, 3, E	10	8	13	14
2, 3, E	11	9	13	14
1, 2, 3, E	7	6	9	9

1 = 1<sup>ère</sup> année, 2 = 2<sup>e</sup> année, 3 = 3<sup>e</sup> année, E = expertes.

les quatre groupes (6 % pour les catégories et 9 % pour les stratégies), nous pouvons conclure que, de manière générale, les quatre groupes n'utilisent pas les mêmes catégories et stratégies de pensée pour répondre aux questions.

Les tableaux V et VI présentent les degrés d'accord entre les questions deux par deux pour les catégories de pensée (tableau V) et les stratégies de pensée (tableau VI). Lorsqu'on s'intéresse à l'accord calculé sur les catégories de pensées réellement partagées (nombres entre parenthèses), les valeurs varient entre 2 et 12 sur 21. Par exemple, V1Q4-V3Q2 ont 57 % de présences de catégories de pensée communes (12/21). Ainsi, les nombres entre parenthèses montrent que les accords qui semblent être élevés, sont davantage constitués des absences communes de catégories/stratégies de pensée que des présences communes. Il est à noter que les questions partagent toujours au moins deux catégories de pensée et au plus neuf.

Les constats suivants peuvent être dégagés de ces résultats :

1) Les deux tiers des catégories de pensée de Fonteyn<sup>[22]</sup> sont peu utilisées ou utilisées de manière quasi-systématique par les participantes pour les 13 questions de notre étude (voir tableau I).

- 2) Pour tous les groupes confondus, les questions retenues sollicitent entre 29 % et 67 % des catégories de pensées. Cependant, certaines catégories sont utilisées dans presque toutes les questions (92 %) alors que d'autres ne sont utilisées que dans très peu de questions (8 %) (voir tableau II).
- 3) De manière logique, les stratégies de pensée sont utilisées pour plus de questions que les catégories de pensée puisque l'on regroupe plusieurs catégories en une seule stratégie. Par exemple, on observe que quatre stratégies de pensée sont systématiquement utilisées pour répondre à toutes les questions.
- 4) Des disparités importantes existent entre les groupes de participants, c'est-à-dire que les quatre groupes n'utilisent pas les mêmes catégories de pensée pour répondre à chacune des questions et les utilisent différemment. Il en est de même pour les stratégies de pensée (voir tableau IV).
- 5) L'analyse détaillée des catégories et stratégies partagées par les questions deux par deux pour chaque groupe montre qu'il n'y a pas vraiment de combinaisons type. Nous en déduisons que chaque question est différente car elle requiert une nouvelle combinaison de catégories/stratégies de pensée pour y répondre, et que chaque groupe se distingue dans son raisonnement.

**Tableau V.** Degrés d'accord entre les questions deux par deux, pour les catégories de pensées.

Questions	V1q2	V1q3	V1q4	V2q2	V2q3	V2q4	V3q2	V3q3	V4q2	V4q3	V5q3	V5q4	V5q5
V1q2													
V1q3	12/21 (4)												
V1q4	7/21 (4)	14/21 (7)											
V2q2	10/21 (3)	15/21 (5)	12/21 (6)										
V2q3	13/21 (6)	16/21 (7)	12/21 (8)	16/21 (7)									
V2q4	14/21 (5)	15/21 (5)	8/21 (4)	15/21 (5)	14/21 (6)								
V3q2	10/21 (6)	13/21 (7)	18/21 (12)	9/21 (5)	14/21 (9)	6/21 (4)							
V3q3	15/21 (7)	6/21 (2)	9/21 (6)	6/21 (2)	9/21 (5)	10/21 (4)	12/21 (8)						
V4q2	12/21 (5)	8/21 (3)	7/21 (5)	11/21 (4)	10/21 (5)	14/21 (6)	9/21 (6)	14/21 (7)					
V4q3	10/21 (3)	15/21 (5)	16/21 (9)	17/21 (6)	14/21 (6)	13/21 (4)	13/21 (7)	8/21 (3)	11/21 (4)				
V5q3	10/21 (2)	15/21 (4)	14/21 (6)	13/21 (3)	14/21 (5)	9/21 (1)	13/21 (6)	8/21 (2)	9/21 (2)	13/21 (3)			
V5q4	11/21 (4)	11/21 (5)	11/21 (7)	11/21 (5)	17/21 (9)	9/21 (4)	14/21 (9)	9/21 (5)	12/21 (6)	12/21 (5)	14/21 (5)		
V5q5	17/21 (6)	14/21 (4)	9/21 (5)	14/21 (4)	17/21 (7)	16/21 (5)	10/21 (5)	11/21 (4)	10/21 (3)	12/21 (3)	11/21 (2)	13/21 (5)	

## Discussion

Cette étude, basée sur l'analyse secondaire de données d'une étude antérieure<sup>[1]</sup>, explorait la possibilité d'utiliser des vignettes et des questions du TCS de Deschênes<sup>[13]</sup> comme outil d'évaluation diagnostique du RCI. Pour ce faire, des matrices Q<sup>[34]</sup> ont été constituées pour relier les 13 questions du TCS à des catégories et des stratégies de pensée spécifiques au RCI<sup>[22]</sup> et ont été analysées pour identifier les distinctions et les similitudes entre les groupes à l'étude.

Nos résultats mettent en évidence un potentiel diagnostique des questions du TCS pour évaluer le RCI niveau par niveau (première, deuxième, troisième années de formation). L'analyse des matrices Q permet de constater que le nombre de catégories ou

de stratégies de pensée sollicitées présente une importante variation entre les questions, mais également d'un groupe à l'autre. De plus, les matrices Q nous permettent de voir qu'au-delà des variations quantitatives, les quatre groupes n'utilisent pas les mêmes catégories de pensée et les combinent de manière différente pour chaque question. Cette situation se répète pour les stratégies de pensée, mais les distinctions sont moins prononcées.

Ceci est notamment illustré par les pourcentages d'accords faibles entre les groupes. Par exemple, les étudiantes de première et deuxième année se ressemblent en ce qu'elles utilisent plus souvent les pauses de réflexion (catégorie u) et la recherche d'information manquante (stratégie 2). Les étudiantes de troisième année, quant à elles, font davantage de

**Tableau VI.** Degrés d'accord entre les questions deux par deux, pour les stratégies de pensées.

Questions	V1q2	V1q3	V1q4	V2q2	V2q3	V2q4	V3q2	V3q3	V4q2	V4q3	V5q3	V5q4	V5q5
V1q2													
V1q3	7/11 (5)												
V1q4	8/11 (7)	8/11 (6)											
V2q2	7/11 (6)	9/11 (6)	9/11 (7)										
V2q3	9/11 (7)	9/11 (6)	8/11 (7)	8/11 (6)									
V2q4	7/11 (5)	11/11 (6)	8/11 (6)	9/11 (6)	9/11 (6)								
V3q2	8/11 (7)	8/11 (6)	9/11 (8)	7/11 (6)	10/11 (8)	8/11 (6)							
V3q3	10/11 (8)	6/11 (5)	9/11 (8)	7/11 (6)	8/11 (7)	6/11 (5)	9/11 (8)						
V4q2	9/11 (6)	10/11 (6)	9/11 (7)	11/11 (7)	8/11 (6)	10/11 (6)	7/11 (6)	7/11 (6)					
V4q3	8/11 (6)	10/11 (6)	9/11 (7)	11/11 (7)	8/11 (6)	10/11 (6)	7/11 (6)	7/11 (6)	11/11 (7)				
V5q3	8/11 (5)	8/11 (4)	7/11 (5)	7/11 (4)	8/11 (5)	8/11 (4)	7/11 (5)	7/11 (5)	7/11 (4)	7/11 (4)			
V5q4	7/11 (6)	7/11 (5)	6/11 (6)	6/11 (5)	9/11 (7)	7/11 (5)	8/11 (7)	6/11 (6)	6/11 (5)	6/11 (5)	8/11 (5)		
V5q5	10/11 (6)	8/11 (5)	7/11 (6)	7/11 (5)	10/11 (7)	8/11 (5)	9/11 (7)	9/11 (7)	7/11 (5)	7/11 (5)	9/11 (5)	8/11 (6)	

conclusions (stratégie 10) que leurs collègues. Elles sont aussi en mesure de fournir des explications (stratégie 8) et de relier des informations (stratégie 4), ce que les participantes des groupes de première et deuxième année font moins fréquemment. D'un autre côté, certaines catégories/stratégies de pensée sont utilisées en première et deuxième année, mais ne le sont plus en troisième année (ex. : catégorie e et stratégie 5). Comme le rapportent Humbert, Besinger et Miech<sup>[36]</sup>, plus les étudiantes acquièrent de l'expérience et des connaissances, plus leur RCI évolue et se transforme. Nos résultats appuient cette idée et coïncident avec ceux de l'étude antérieure<sup>[1]</sup>. D'après Fero et al.<sup>[4]</sup>, la plupart des études montrent une progression dans le développement des capacités de pensée critique des étudiantes infirmières selon leur

niveau de formation. De plus, selon Smith<sup>[37]</sup>, la transition des étudiants se fait à travers une série de stades de développement du RCI, de novice à experte. Ces différentes étapes nécessitent différentes méthodes d'évaluation des compétences. À cet effet, nous croyons que le TCS présente un potentiel car il est sensible au changement entre les stades de développement.

Ainsi, l'évaluation diagnostique au moyen du TCS devrait tenir compte de chaque stade de développement de cette compétence<sup>[1]</sup>. Ceci vient appuyer l'idée qu'une matrice unique ne pourrait être utilisée pour les trois années de formation universitaire initiale étant donné le manque de stabilité observé dans les tableaux II et III. Ce faisant, les évaluateurs devraient choisir les questions pertinentes compte

tenu du niveau de formation des étudiantes évaluées, en considérant les catégories/stratégies de pensée qu'ils cherchent à évaluer.

Dans cette étude, nous avons aussi examiné les réponses d'expertes afin d'élaborer les matrices Q. Nos résultats montrent clairement que ce groupe ne peut servir de référence pour identifier les catégories ou stratégies de pensées qui posent problème aux étudiantes, car leur manière de raisonner est fort différente. Ces résultats correspondent à ceux de plusieurs études antérieures<sup>[14, 38-40]</sup>.

Cependant, certains spécialistes de l'évaluation diagnostique<sup>[33, 41-43]</sup> rapportent que dans un but diagnostique, un « état de connaissance » comme l'expertise sert de modèle de réponses idéales dans la performance à un test. En d'autres termes, en évaluation diagnostique, il est important d'avoir un modèle d'expert afin de pouvoir inférer les processus cognitifs maîtrisés ou non par un étudiant. Pour déterminer ce que constituerait une performance idéale au TCS pour chaque niveau, il faudrait, entre autre, définir l'expertise dans chaque groupe. Cette démarche permettrait d'identifier des points de référence.

L'adaptation des questions du TCS devrait aussi tenir compte des 40 catégories de pensée que nous avons éliminées. Leur sous-utilisation et leur sur-utilisation pourraient être attribuables à la méthodologie des travaux sur lesquels nous sommes basés, Deschênes<sup>[14]</sup> ayant validé son TCS avec des étudiantes de première année du baccalauréat en sciences infirmières, alors que Fonteyn<sup>[22]</sup> a interviewé des expertes. Notre étude porte donc sur des questions adaptées à des étudiantes de première année de formation avec des catégories/stratégies de pensées attribuées à des expertes. Nos résultats montrent que les catégories/stratégies de pensées utilisées par les expertes sont très différentes des étudiantes. Des études ultérieures pourraient permettre d'identifier d'autres catégories/stratégies de pensée utilisées par les étudiantes, selon les niveaux de formation.

Notre étude est limitée en raison de l'utilisation d'un nombre restreint de vignettes et de questions. En effet, il aurait été pertinent pour notre étude d'avoir des données pour toutes les vignettes et questions du TCS original, ce qui n'était pas possible vu l'utilisation des données d'une étude antérieure déjà

complétée. Ceci aurait pu permettre, par exemple, de solliciter les catégories/stratégies de pensée que nous avons éliminées parce que sous-utilisées dans notre échantillon. En outre, les vignettes utilisées par Fonteyn<sup>[19]</sup> diffèrent de celles employées par Deschênes<sup>[13]</sup>. Une autre piste de recherche pourrait consister à utiliser les vignettes de la première pour élaborer le TCS. Il serait également possible d'utiliser les vignettes de la seconde pour explorer les catégories/stratégies de pensée des étudiantes ou des expertes.

Une autre limite vient du fait que les participantes ont été sélectionnées dans une seule université, ce qui biaise la représentativité de la population étudiante en sciences infirmières. Ainsi, il importe de mentionner que nos résultats ne sont pas généralisables. Il serait intéressant dans une étude ultérieure de comparer des étudiants d'autres institutions d'enseignement et d'autres niveaux de formation (collégial et universitaire). La méthode que nous avons développée pourrait toutefois être réutilisée dans divers contextes.

Par ailleurs, certains de nos choix méthodologiques sont discutables. Par exemple, les critères de sous-utilisation et de sur-utilisation des catégories/stratégies de pensée ont été sélectionnés pragmatiquement. Par ailleurs, nous n'avons pas tenu compte de la qualité des réponses fournies par les participantes, étant donnée la complexité du processus à mettre en place pour différencier les bonnes réponses des moins bonnes et que les conditions nécessaires débordaient les limites de cette étude. Notre démarche repose donc sur l'ensemble des réponses aux questions sans nuance quant à leur qualité. Il serait néanmoins pertinent de s'attarder à ces nuances dans une prochaine étude, ce qui permettrait de formaliser les relations entre chaque question et les stratégies nécessaires à une réponse acceptable.

## Conclusion

Cette étude montre que le TCS a un potentiel diagnostique mais niveau par niveau (1<sup>ère</sup>, 2<sup>ème</sup>, 3<sup>ème</sup> année et expertes). Les résultats de cette recherche mettent en évidence les catégories et stratégies de pensées utilisées pour répondre à certaines questions du TCS

selon les groupes. Nos résultats ont aussi permis de découvrir des disparités importantes entre les groupes : ils montrent notamment que le RCI du groupe d'expertes est si différent de celui des étudiantes qu'il ne peut servir de référence.

Étant donné le manque de procédures systématiques pour identifier les difficultés de raisonnement clinique des étudiantes<sup>[44]</sup>, le TCS pourrait s'avérer utile pour soutenir les programmes d'enseignement en sciences infirmières. Ainsi, il pourrait permettre d'offrir une aide adaptée pour stimuler le développement des stratégies de pensée nécessaires à différents stades cognitifs d'apprentissage.

## Remerciements

Les auteurs tiennent à remercier Patrick Lavoie pour sa contribution.

## Valorisation scientifique

Cet article a été élaboré à partir du travail de mémoire de maîtrise de l'auteur principal.

Le projet de cette a étude a été présenté dans le cadre du 79<sup>ème</sup> Congrès de l'Association francophone pour le savoir (ACFAS), Sherbrooke, mai 2011.

Dumont K., Loye N. & Goudreau J. (2011). Validation d'un test de concordance de script (TCS) à l'aide d'une étude sur le raisonnement clinique infirmier (RCI).

## Contributions

Katia Dumont a conçu le protocole de recherche, effectué le recueil des données, interprété les résultats, effectuée l'analyse statistique et rédigé la version initiale manuscrit. Nathalie Loye a supervisé les travaux de recherche, participé à l'analyse statistique et à l'écriture du manuscrit. Johanne Goudreau a supervisé les travaux de recherche et participé à la révision du manuscrit

## Approbation éthique

La présente étude a été approuvée par Comité plurifacultaire d'éthique de la recherche (CPÉR), Faculté de l'aménagement, de droit, de musique, des sciences de l'éducation et de théologie et de sciences des religions, Université de Montréal 15 décembre 2011

## Déclaration d'intérêt

Aucun auteur ne déclare de conflit d'intérêts en lien avec le contenu de cet article.

## Références

1. Goudreau J, Boyer L, Létourneau D. Clinical nursing reasoning in nursing practice: A cognitive learning model based on a think aloud methodology. *Quality Advanced in Nursing Education - Avancées en formation infirmière* 2014;1:4
2. Frenk J CL, Bhutta ZA, Cohen J, Crips N, Evans T, et al. Health professionals for a new century: Transforming education to strengthen health systems in an interdependent world. *Lancet* 2010;376:1923-58
3. Lindeman CA. The future of nursing education. *J Nurs Educ* 2000;39:5-12
4. Fero LJ, Witsberger CM, Wesmiller SW, Zullo TG, Hoffman LA. Critical thinking ability of new graduate and experienced nurses. *J Adv Nurs* 2009;65:139-48
5. Durak HI, Caliskan SA, Bor S, van der Vleuten C. Use of case-based exams as an instructional teaching tool to teach clinical reasoning. *Med Teach* 2007;29:e170-74
6. Gierl MJ. Making diagnostic inferences about cognitive attributes using the rulespace model and attribute hierarchy method. *J Educ Meas* 2007;44:325-40
7. Kalamideres M, Caire F, Dager F, Brassier G, Moreau JJ. Création d'outils d'autoévaluation pour une formation médicale continue en ligne : Modélisation à propos d'une formation. *Neurochirurgie* 2008;54:21-7



8. Boychuk Duchscher JE. Critical thinking: Perceptions of newly graduated female baccalaureate nurses. *J Nurs Educ* 2003;42:14-27
9. Charlin B, Borgage G, van der Vleuten C. L'évaluation du raisonnement clinique. *Pédagogie Médicale* 2003;4:42-52
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. Carrière B, Gagnon R, Charlin R, Downing S, Bordage G. Assessing clinical reasoning in pediatric emergency medicine: Validity evidence for a script concordance test. *Ann Emerg Med* 2009;53:647-52
12. Charlin B, van der Vleuten C. Standardized assessment of reasoning in contexts of uncertainty: The script concordance approach. *Eval Health Prof* 2004;27:304-19
13. Deschênes MF. Élaboration et évaluation des propriétés psychométriques d'un instrument d'évaluation du raisonnement clinique empreint de human caring (mémoire). Montréal : Université de Montréal, 2006
14. Deschênes MF, 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;50:381-87
15. Côté S, Tanguay A, Gagnon R, Charlin B, Michaud C. Élaboration et validation d'un test de concordance de script pour évaluer le raisonnement clinique des infirmières en contexte de détérioration clinique. *Pédagogie Médicale* 2014;15:7-20
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. Audétat MC, Laurin S. Supervision of clinical reasoning: Methods and a tool to support and promote clinical reasoning. *Can Fam Physician* 2010;56:e127-129
18. Elstein AS, Swchartz A. Clinical problem solving and diagnostic decision making: Selective review of the cognitive literature. *BMJ* 2002;324:729-32
19. Kautz DD, Kuiper R, Pesut DJ, Knight-Brown P, Daneker D. Promoting clinical reasoning in undergraduate nursing students: Application and evaluation of the outcome present state test (opt) model of clinical reasoning. *Int J Nurs Educ Scholarsh* 2005;2:4-19
20. Funkesson KH, Anbäcken EM, Ek AC. Nurses' reasoning process during care planning taking pressure ulcer prevention as an example. A think-aloud study. *Int J Nurs Stud* 2007;44:1109-19
21. Simmons B. Clinical reasoning: Concept analysis. *J Adv Nurs* 2010;66:1151-58
22. Fonteyn M. Thinking strategies for nursing practice. Philadelphie (PA); 1998.
23. Higgs J, Jones M, Loftus S, Christensen N. Clinical reasoning in the health professions. Amsterdam: Butterworth Heinemann; 2008.
24. Caire F, Sol JC, Moreau JJ, Isidori P, Charlin B. Auto-évaluation des internes en neurochirurgie par tests de concordance de script (tcs) : Le processus d'élaboration des tests. *Neurochirurgie* 2004;50:66-72
25. Gagnon R, Charlin B, Lambert C, Carrière B, van der Vleuten C. Script concordance testing: More cases or more questions? *Adv Health Scien Educ* 2009;14:367-75
26. Ruiz JG, Tunuguntla R, Charlin B, Ouslander JG, Symes SN, Gagnon R, et al. The script concordance test as a measure of clinical reasoning skills in geriatric urinary incontinence. *J Am Geriatr Soc* 2010;58:2178-84
27. Marie I, Sibert L, Roussel F, Hellot MF, Lechevallier J, Weber J. Le test de concordance de scripts : Un nouvel outil d'évaluation du raisonnement et de la compétence clinique en médecine interne? *Rev Med Interne* 2005;26:501-7
28. Meterissian S, Zalobotny B, Gagnon R, Charlin B. Is the script concordance test a valid instrument for assessment of intraoperative decision-making skills? *Am J Surg* 2007;193:248-51
29. Lubarsky S, Chalk C, Kazitani D, Gagnon R, Charlin B. The script concordance test: A new tool assessing clinical judgement in neurology. *Can J Neurol Sci* 2009;36:326-31
30. Charlin B, Gagnon R, Lubarsky S, Lambert C, Meterissian S, Chalk C, et al. Assessment in the context of uncertainty using the script concordance test: More meaning for scores. *Teach Learn Med* 2010;22:180-6
31. Park AJ, Barber M, Bent AE, Dooley YT, Dancz C, Sutkin G, et al. Assessment of intraoperative judgment during gynecologic surgery using the script concordance test. *Am J Obstet Gynecol* 2010;203:e1-6

32. Sibert L, Charlin B, Gagnon R, Corcos J, Khalaf A, Grise P. Evaluation of clinical reasoning in urology: Contribution of the script concordance test. *Prog Urol* 2001;11:1213-19
33. Loye N. Quelques nouveaux modèles de mesure. *Mesure et Évaluation* 2005;28:51-68
34. Tatsuoka KK. Rule-space: An approach for dealing with misconceptions based on item response theory. *J Educ Meas* 1983;20:345-354
35. Fonteyn ME, Kuipers B, Grobe SJ. A description of think aloud method and protocol analysis. *Qual Health Res* 1993;3:430-41
36. Humbert AJ, Besinger B, Miech EJ. Assessing clinical reasoning skills in scenarios of uncertainty: Convergent validity for a script concordance test in an emergency medicine clerkship and residency. *Acad Emerg Med* 2011;18:627-34
37. Smith CS. A developmental approach to evaluating competence in clinical reasoning. *J Vet Med Educ* 2008;35:375-81
38. Lucchiari C, Pravettoni G. Cognitive balanced model: A conceptual scheme of diagnostic decision making. *J Eval Clin Pract* 2012;18:82-8
39. Norman G. Research in clinical reasoning: Past history and current trends. *Med Educ* 2005;39:418-27
40. Nouh T, Boutros M, Gagnon R, Reid S, Leslie K, Pace D, et al. The script concordance test as a measure of clinical reasoning: A national validation study. *Am J Surg* 2012;203:530-4
41. Gierl MJ. Making diagnostic inferences about cognitive attributes using the rule-space model and attribute hierarchy method. *J Educ Meas* 2007;44:325-40
42. Loye N. Conditions d'élaboration de la matrice Q des modèles cognitifs et impact sur sa validité et sa fidélité. Thèse de doctorat non-publiée, Ottawa : Université d'Ottawa, 2008
43. Loye N, Caron F, Pineault J, Tessier-Baillargeon M, Burney-Vincent C, et Gagnon M. La validité du diagnostic issu d'un mariage entre didactique et mesure sur un test existant. Dans Raïche G P-CK, Magis D (Dir.), ed. La validité du diagnostic issu d'un mariage entre didactique et mesure sur un test existant. Ste-Foy, Québec: Presses de l'Université du Québec; 2011:11-30.
44. Audétat MC, Dory V, Nendaz M, Vanpee D, Pestiaux D, Junod Perron N, et al. What is so difficult about managing clinical reasoning difficulties? *Med Educ* 2012;46:216-27

---

Correspondance et offprints : Katia Dumont. Centre de médecine comportementale de Montréal (CMCM), Hôpital du Sacré-Cœur de Montréal, 5400 Boulevard Gouin Ouest, Montréal, QC H4J 1C5, Canada. Téléphone + 1 514-338-2222 poste 7820. Mail to : [katia.dumont@crhsc.rtss.qc.ca](mailto:katia.dumont@crhsc.rtss.qc.ca)