

Université de Montréal

L'évaluation du raisonnement clinique en physiothérapie

Par

Jean-Pierre Dumas

Département d'administration et fondements de l'éducation

Faculté des sciences de l'éducation

Thèse présentée à la Faculté des Études Supérieures en vue de l'obtention du grade de
Philosophia Doctor (Ph.D.) en science de l'éducation option Mesure et évaluation

Juin 2016

© Jean-Pierre Dumas 2016

Identification du jury

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée:

L'évaluation du raisonnement clinique en physiothérapie

présentée par :

Jean-Pierre Dumas

a été évaluée par un jury composé des personnes suivantes :

Joseph-Omer Dyer	Président-rapporteur
Jean-Guy Blais	Directeur de recherche
Bernard Charlin	Codirecteur de recherche
Marie-Claude Audétat	Membre du jury
Jean-Sébastien Renaud	Examineur externe
Serge J. Larivée	Représentant du doyen de la FES

Thèse acceptée le 25 novembre 2016

RÉSUMÉ

Dans le contexte de l'enseignement universitaire en sciences de la santé, des changements importants se sont produits au cours des dernières décennies dans les programmes de formation professionnelle. Un de ces changements importants est le fait que ces programmes de formation professionnelle adoptent maintenant, pour la plupart, une approche par compétence. Parallèlement à ce changement de paradigme dans la formation, les professions de la santé considèrent de plus en plus le raisonnement clinique comme étant central à la compétence du clinicien et lui réservent une place plus importante dans la formation. Ces deux changements dans la formation en science de la santé ont entraîné des questionnements quant aux méthodes d'évaluations du raisonnement clinique dans un contexte d'approche par compétence. Une des réponses à ces questionnements a été l'émergence de tests permettant de poser un jugement sur le raisonnement clinique dans le domaine de la médecine. En effet, au cours des 20 dernières années, des outils tels que le test de concordance de script (TCS) et les problèmes de raisonnement clinique (PRC) ont été développés et ont fait l'objet de plusieurs études dans les différentes spécialités de la médecine principalement. Bien que la physiothérapie partage plusieurs caractéristiques avec le domaine de la médecine, elle possède aussi plusieurs différences. En ce sens, il était pertinent de se questionner sur l'applicabilité de ces outils dans le contexte spécifique de la physiothérapie.

Afin de collecter des données, des versions de ces tests ont été développées selon une table de spécification d'un cours du programme de physiothérapie de l'Université de Sherbrooke et ont été administrées à 93 étudiants volontaires des cohortes de deuxième, troisième et quatrième année du programme. Selon la procédure d'attribution des scores de ces tests, un panel d'experts (n=19) dans le domaine a aussi répondu aux items.

Les résultats obtenus nous indiquent que les scores du TCS permettent de discerner des différences significatives entre les étudiants novices (2^e année) et les étudiants finissants de 4^e année du programme de physiothérapie, et ce, avec une taille d'effet importante (Cohen $d = 0,85$). Par contre, il n'existe pas de différence significative entre le score moyen de la cohorte de 3^e année et celui des deux autres cohortes à ce test. Des différences significatives ont été

notées entre les scores du panel d'experts et celui des trois cohortes d'étudiants. Du côté des résultats aux PRC, les scores à ce test ne nous permettent pas de distinguer le niveau de développement de raisonnement clinique des étudiants en physiothérapie entre les différentes années de formation. Par contre, tout comme avec le TCS, des différences significatives ont été notées entre les scores du panel d'experts et celui des trois cohortes d'étudiants. De plus, nous avons analysé et catégorisé les 183 formulaires de stage des étudiants ayant participé à l'étude. Ces formulaires ont été catégorisés selon les observations au niveau du raisonnement clinique effectuées par les superviseurs de stage. Les résultats de cette partie de l'étude nous suggèrent que les scores des PRC sont discriminants en fonction des résultats de stage pour les différentes catégories de résultats. Du côté des scores au TCS, les étudiants ayant eu des difficultés de raisonnement clinique en stage se démarquent aussi par des scores plus faibles au TCS. Conséquemment, il est possible que ces outils puissent également permettre de dépister des étudiants ayant des difficultés de raisonnement clinique et pouvant bénéficier d'une remédiation à ce niveau. Par contre, en tenant compte de ces résultats encourageants, mais aussi du peu d'étudiants ayant eu des difficultés de raisonnement clinique en stage dans cette recherche, il sera nécessaire de confirmer ce rôle de dépistage possible avec une étude de plus grande envergure.

Mots clés : Évaluation, enseignement de la physiothérapie, test de concordance de script, problèmes de raisonnement clinique, stage.

ABSTRACT

Major changes have occurred in the last few decades in teaching the health sciences at the university level in professional training programs. One of these major changes is that most professional training programs have now adopted a competency-based approach. At the same time that this change in training paradigm was occurring, health professions were increasingly considering clinical reasoning as being at the core of clinician competency and were granting it increasing importance in their programs. These two changes in health-science training put into question the methods for assessing clinical reasoning in a competency-based approach. One of the responses to this questioning was the emergence of tests to assess clinical reasoning in the field of medicine. Indeed, over the last 20 years, instruments such as the script concordance test (SCT) and clinical reasoning problems (CRPs) were developed and have been studied primarily in the various medical specialties. While physiotherapy and medicine share a number of characteristics, they also have several differences. As a result, the appropriateness of using these tools in the specific context of physiotherapy is an issue that should be explored.

In an attempt to gather data, versions of these tests were developed using a table of specifications for a course in the physiotherapy program at the Université de Sherbrooke and were administered to 93 volunteer students from the second, third, and fourth year of the program. In accordance with the scoring procedure for these tests, a panel of experts in the field ($n = 19$) also responded to the items.

The results obtained indicate that the SCT revealed significant differences between novice students (second year) and graduating students (fourth year) in the physiotherapy program with a significant effect size (Cohen $d = 0.85$). In contrast, there were no significant differences between the average score for third-year students and the other two cohorts. Significant differences were noted between the scores of the panel of experts and those of the students from the three cohorts. With respect to CRPs, the scores do not allow us to distinguish the level of clinical-reasoning development of physiotherapy students in the three years of training studied. On the other hand, as with the SCT, there were significant differences between the scores of the panel of experts and those of the students from the three cohorts. Moreover, we analyzed and

categorized the 183 rotation forms of the students who participated in the study. These forms were categorized according to level of clinical reasoning based on the observations of rotation supervisors. The results of this part of the study suggest that the CRP scores were discriminating in terms of rotation results for the various categories of results. As for the SCT, the students who encountered difficulties with clinical reasoning during rotations also had the lowest SCT scores. Consequently, these tools may also be able to screen for students who have problems with clinical reasoning and could benefit from remedial assistance at this level. On the other hand, while these results are encouraging, the population sample of students with clinical-reasoning difficulties during rotations in this research was small. Therefore, the role of any possible screening requires confirmation through a broader study.

TABLE DES MATIÈRES

RÉSUMÉ	i
ABSTRACT	iv
TABLE DES MATIÈRES	vi
LISTE DES TABLEAUX.....	xii
LISTES DES FIGURES	xv
LISTE DES SIGLES ET ABRÉVIATIONS	xvi
DÉDICACE	xvii
REMERCIEMENTS.....	xviii
INTRODUCTION	1
CHAPITRE 1 LA PROBLÉMATIQUE.....	3
1.1 Contexte de la physiothérapie au Québec	3
1.2 L'évolution de l'évaluation des apprentissages en milieu universitaire	5
1.2.1 Caractéristiques de l'évaluation des apprentissages dans une approche par objectifs7	
1.2.2 Caractéristiques de l'évaluation des apprentissages dans une approche par compétences.....	9
1.3 L'évaluation du raisonnement clinique.....	17
1.3.1 Questions à choix multiples (QCM).....	18
1.3.2 Questions de type ouvertes (réponses construites)	19
1.3.3 Examen par éléments clés (<i>Key features</i>).....	20
1.3.4 Questions à appariement étendu (<i>Extended matching questions</i>).....	21
1.3.5 Test de concordance des scripts (TCS).....	22
1.3.6 Problèmes de raisonnement clinique (<i>Clinical reasoning problem</i>).....	24
1.3.7 Les examens cliniques objectifs structurés (ECOS).....	26

1.3.8	L'évaluation du raisonnement clinique en contexte de stage	26
1.4	Questions de recherche	28
CHAPITRE 2 CADRE CONCEPTUEL		31
2.1	Le construit du raisonnement clinique	31
2.1.1	Comparaison des différentes définitions et termes utilisés en lien avec le raisonnement clinique	32
2.1.2	Comparaison des différents modèles du raisonnement clinique et leurs évolutions	33
2.1.3	Influence du paradigme interprétatif sur les modèles de raisonnement clinique	38
2.1.4	L'importance du contexte sur les dimensions du raisonnement clinique	40
2.1.5	Le raisonnement clinique en tant que composante de la compétence clinique en sciences de la santé	41
2.2	Cadre d'évaluation du raisonnement clinique et de la compétence clinique :	42
2.2.2	Perspective développementale de la compétence clinique	44
2.3	Validité et processus de validation des tests de raisonnement clinique	48
2.3.1	Évolution du concept de validité.....	48
2.3.2	Sources d'évidence de la qualité du contenu d'un test.....	52
2.3.3	Sources d'évidence de la fidélité d'un test	53
2.3.4	Sources d'évidence de la qualité du processus latent d'un test	53
2.3.5	Sources d'évidence de l'utilité d'un test.....	54
2.3.6	Sources d'évidence de l'impact d'un test	54
2.3.7	Sources d'évidence du réseau nomologique d'un test	54
2.3.8	Synthèse du processus de validation.....	55
CHAPITRE 3 REVUE DE LITTÉRATURE SUR LES OUTILS D'ÉVALUATION DU RAISONNEMENT CLINIQUE RETENUS		57

3.1	Test de concordance de script	57
3.1.1	Sources d'évidence de la qualité du contenu du TCS.....	57
3.1.2	Sources d'évidence de la fidélité du TCS	60
3.1.3	Sources d'évidence de la qualité du processus latent du TCS	62
3.1.4	Sources d'évidence de l'utilité du TCS	62
3.1.5	Sources d'évidence de l'impact du TCS.....	63
3.1.6	Sources d'évidence de la qualité du réseau nomologique du TCS	63
3.2	Problèmes de raisonnement clinique (PRC)	66
CHAPITRE 4 LA MÉTHODOLOGIE.....		71
4.1	Posture de recherche et type d'étude.....	71
4.2	Description des participants et du milieu de collecte de donnée	72
4.2.1	Les étudiants et leur programme de formation	72
4.2.2	Panel d'experts.....	73
4.3	Tests d'évaluation du raisonnement clinique.....	75
4.3.1	Description des tests et du domaine évalué	75
4.3.1.1	Description du TCS.....	75
4.3.1.2	Description des PRC	77
4.3.2	Étapes du développement des tests de raisonnement clinique.....	82
4.3.2.1	Construction des tests.....	82
4.3.2.2	Rédaction des vignettes et des items du TCS.....	83
4.3.2.3	Rédaction des vignettes et des items des PRC	84
4.3.2.4	Évaluations des vignettes et des items des tests de raisonnement clinique.....	85
4.3.3	Collecte de données	86
4.3.4	Méthodes pour établir les scores.....	88
4.3.5	Temps de passation des tests de raisonnement clinique	98

4.3.6	Résultats du questionnaire sur la perception et appréciation des tests de raisonnement clinique	99
4.3.6.1	Première question	99
4.3.6.2	Deuxième question	100
4.3.6.3	Troisième question	103
4.4	Analyse de contenu des formulaires de stage	104
4.4.1	Élaboration d'une grille d'analyse et catégorisation des commentaires des superviseurs.....	106
4.5	Stratégie d'analyse	109
4.5.1	Logiciels utilisés pour l'analyse et la présentation des données.....	111
4.6	Ressources humaines et financières	111
4.7	Considérations éthiques	112
CHAPITRE 5 ANALYSE DE RÉSULTATS		113
5.1	Première partie analyse des scores des tests de raisonnement clinique	114
5.1.1	Résultats au test de concordance de script.....	114
5.1.1.1	Analyse des réponses du panel aux items du test de concordance de script .	114
5.1.1.2	Analyse des réponses des étudiants aux items du test de concordance de script	117
5.1.1.3	Analyse descriptive de la distribution des scores du test de concordance de script	121
5.1.1.4	Analyse comparative de la distribution des scores du test de concordance de script	125
5.1.2	Résultats aux problèmes de raisonnement clinique	127
5.1.2.1	Analyse des réponses du panel aux PRC	127
5.1.2.2	Analyse des réponses des étudiants aux items des problèmes de raisonnement clinique	128

5.1.2.3	Analyse descriptive de la distribution des scores des problèmes de raisonnement clinique	129
5.1.2.4	Analyse comparative de la distribution des scores des problèmes de raisonnement clinique	132
5.1.3	Corrélation entre les scores du test de concordance de script et des problèmes de raisonnement clinique.....	134
5.2	Deuxième partie : mise en relation entre les résultats aux tests de raisonnement clinique et les observations de stage.....	136
5.2.1	Catégorisation des résultats de stage.....	136
5.2.1.1	Évaluation de l'accord sur les catégorisations	139
5.2.2	Résultats des tests de raisonnement clinique par catégorie	140
5.2.3	Triangulation entre les résultats de stage et les résultats aux tests de raisonnement clinique	146
5.2.4	Profil des étudiants ayant des scores qui se sont démarqués positivement ou négativement aux tests de raisonnement clinique.....	148
CHAPITRE 6 DISCUSSION DES RÉSULTATS		150
6.1	Première question de recherche	150
6.1.1	Résultats en lien avec le TCS.....	150
6.1.2	Résultats en lien avec le PRC	155
6.1.3	Comparaison entre les deux outils	159
6.2	Deuxième question de recherche	160
6.3	Limites de l'étude.....	163
6.4	Pistes de recherche	165
CONCLUSION.....		168
RÉFÉRENCES		171
ANNEXES.....		i

ANNEXE 1 CURRICULUM DU PROGRAMME DE PHYSIOTHÉRAPIE DE L'UNIVERSITÉ DE SHERBROOKE	ii
ANNEXE 2 LETTRE D'APPROBATION PAR LE COMITÉ D'ÉTHIQUE ET FORMULAIRE DE CONSENTEMENT	iii
ANNEXE 3 FORMULAIRE D'ÉVALUATION DE STAGE FOCUS	xi
ANNEXE 4 TEST DE CONCORDANCE DE SCRIPT	xxxiv
ANNEXE 5 PROBLÈMES DE RAISONNEMENT CLINIQUE	lxi
ANNEXE 6 TABLEAU DE SPÉCIFICATIONS TCS ET PRC	xc
ANNEXE 7 SONDAGE ÉTUDIANTS.....	xciii
ANNEXE 8 SONDAGE EXPERTS.....	xcv
ANNEXE 9 GRILLE D'ANALYSE DES FORMULAIRES DE STAGE	xcviii
ANNEXE 10 CITATIONS DANS LA LANGUE D'ORIGINE.....	xcix
ANNEXE 11 EXEMPLE DE TYPES DE QUESTIONS DE RAISONNEMENT CLINIQUE APPLIQUÉS À LA PHYSIOTHÉRAPIE	c

LISTE DES TABLEAUX

Tableau I Études ayant comparé le niveau de raisonnement clinique de groupe à l'aide du TCS	64
Tableau II Années d'expérience du panel d'experts	74
Tableau III Étapes de développement des tests de raisonnement clinique	82
Tableau IV TCS (Item 1) Exemple de grille pour établir le score à un item	89
Tableau V Énoncés des diagnostics des experts sur la vignette 2 (1 ^{re} partie).....	89
Tableau VI Catégorie de diagnostic émis par les experts pour la vignette 2 et pointage attribué pour ce diagnostic	91
Tableau VII Profil des réponses sur les caractéristiques de la bursopathie à la vignette 2 à partir des réponses des experts qui ont choisi ce diagnostic (n=11).....	92
Tableau VIII Pointage attribué pour l'item 1 du profil bursopathie de la vignette 2.....	92
Tableau IX Profil des réponses sur les caractéristiques de la capsulite à la vignette 2 à partir des réponses de sept experts qui ont choisi ce diagnostic n=7	93
Tableau X Pointage attribué pour l'item 1 du profil capsulite de la vignette 2	93
Tableau XI Exemple de calcul de la vignette 2 pour le candidat 4U2.....	94
Tableau XII Tableau d'attribution des scores pour les réponses ouvertes du PRC	95
Tableau XIII Compilation des scores des huit vignettes du PRC pour le candidat 4U2 selon la méthode préconisée par Groves (2002)	97
Tableau XIV Compilation des scores des huit vignettes du PRC pour le candidat 4U2 en attribuant une valeur égale aux items de réponses ouvertes	98
Tableau XV Temps de passation moyen, minimum et maximum des tests par cohorte	99
Tableau XVI Fréquences relatives des réponses à la question 1 du sondage aux étudiants...	100
Tableau XVII Fréquences relatives des réponses à la question 2 du sondage aux étudiants..	102
Tableau XVIII Fréquences relatives des réponses à la question 3 du sondage aux étudiants	103
Tableau XIX Catégorisation de l'évaluation du raisonnement clinique effectuée à partir des commentaires des formulaires FOCUS	107
Tableau XX Catégorisation de l'évaluation du raisonnement clinique effectuée à partir des commentaires des formulaires FOCUS pour les étudiants qui ont eu seulement un stage dans le domaine musculosquelettique.....	109

Tableau XXI Ressources humaines ayant été rémunérées dans le cadre de l'étude.....	111
Tableau XXII Tableau des scores des experts (calculé en excluant leur profil de réponse du calcul) au TCS (82 items)	115
Tableau XXIII Profil de la variabilité des réponses aux items par le panel d'expert	116
Tableau XXIV Items ayant des réponses du panel à plus de 2 catégories du mode.....	117
Tableau XXV Corrélations item-total pour l'ensemble des trois cohortes.....	118
Tableau XXVI Corrélations item-total pour la cohorte de 3 ^e année.....	119
Tableau XXVII Coefficient alpha pour les différentes versions du test et les différentes cohortes	120
Tableau XXVIII Effet de l'optimisation sur les moyennes et écarts types.....	121
Tableau XXIX Statistiques descriptives des scores au TCS optimisé (26 vignettes, 67 items)	122
Tableau XXX Test de Kolmogorov-Smirnov et Shapiro-Wilk sur la normalité des distributions	125
Tableau XXXI ANOVA des scores du TCS Test Post Hoc : Bonferroni	126
Tableau XXXII Taille d'effet de la différence des scores du TCS entre les groupes.....	126
Tableau XXXIII Tableau des scores des experts au PRC (8 vignettes, 218 items).....	128
Tableau XXXIV Coefficient alpha par items sur les scores des PRC (218 items pour les différentes cohortes).....	129
Tableau XXXV Statistiques descriptives des scores au PRC (8 vignettes, 218 items)	130
Tableau XXXVI Test de Kolmogorov-Smirnov et Shapiro-Wilk sur la normalité des distributions.....	132
Tableau XXXVII ANOVA sur les scores des PRC Test post hoc : Bonferroni.....	133
Tableau XXXVIII Taille d'effet de la différence des scores du PRC entre les groupes d'experts et les cohortes d'étudiants.....	133
Tableau XXXIX Corrélation entre les scores du test de concordance de script et des problèmes de raisonnement clinique	136
Tableau XL Nombres de formulaires de stages analysés par cohorte	137
Tableau XLI Nombre et type de difficulté de raisonnement clinique notés lors de l'analyse de contenu des formulaires de stage	139
Tableau XLII Cotation des deux évaluateurs sur les catégories de stage	140

Tableau XLIII Scores aux tests de raisonnement clinique par catégorie pour la cohorte de 2 ^e année	141
Tableau XLIV Scores aux tests de raisonnement clinique par catégorie pour la cohorte de 3 ^e année	142
Tableau XLV Scores aux tests de raisonnement clinique par catégorie pour la cohorte de 4 ^e année	143
Tableau XLVI Scores aux tests de raisonnement clinique par catégorie pour les trois cohortes	144
Tableau XLVII Scores aux tests de raisonnement clinique par catégorie pour les cohortes de 3 ^e et 4 ^e année	145
Tableau XLVIII Triangulation des informations sur les étudiants ayant eu des difficultés en stage au niveau du raisonnement clinique	146
Tableau XLIX Profil des étudiants à moins de deux écarts types de la moyenne de leurs groupes aux tests de raisonnement clinique	148
Tableau L Étudiants ayant eu un score à plus de deux écarts types de la moyenne de leurs groupes à un des tests de raisonnement clinique	148

LISTES DES FIGURES

Figure 1 Cadre conceptuel inspiré des modèles de la littérature sur le raisonnement clinique	47
Figure 2. Taxonomie des procédures d'évaluation des tests selon Lissitz & Samuelsen (2007)	52
Figure 3 Exemple de vignette de test de concordance de script comprenant trois items.....	76
Figure 4 Énoncés proposés pour chaque niveau de l'échelle de type Likert du TCS.....	76
Figure 5 Exemple de vignette de problèmes de raisonnement clinique.....	81
Figure 6 Énoncés proposés pour chaque niveau de l'échelle de type Likert des PRC	82
Figure 8 Histogramme de fréquence des scores au TCS pour chacune des cohortes et du panel	124
Figure 10 Histogramme des scores des PRC pour chacune des cohortes et du groupe d'expert	131

LISTE DES SIGLES ET ABRÉVIATIONS

TCS	Test de concordance des scripts
PRC	Problèmes de raisonnement clinique
OPPQ	Ordre professionnel de la physiothérapie du Québec
ECOS	Examen clinique objectif structuré
<i>FCAMT</i>	<i>Fellow of the Canadian Academy of Manipulative Physiotherapy</i>
CPASS	Centre de pédagogie appliquée aux sciences de la santé
<i>SPSS</i>	<i>Statistical Package for the Social Sciences</i>
<i>MSE</i>	<i>Mean square Error</i>
<i>DTI</i>	<i>Diagnostic Thinking Inventory</i>

DÉDICACE

Je dédie cette thèse à mes filles à qui je souhaite d'avoir la persévérance nécessaire pour réaliser tous les projets importants d'une vie.

REMERCIEMENTS

J'aimerais d'abord remercier chaleureusement mon équipe de direction, Jean-Guy Blais et Bernard Charlin. Votre expertise et votre rigueur scientifique m'ont permis de mener à bien ce projet. Grâce à votre encadrement et à votre regard vigilant, j'ai eu la latitude suffisante pour apprendre de mes erreurs et suffisamment de conseils pour me guider vers des choix judicieux. Je suis reconnaissant d'avoir eu le privilège de m'inspirer de deux modèles d'excellence.

Merci également à toute l'équipe de professeurs du programme de mesure et évaluation de la Faculté d'éducation de l'Université de Montréal que j'ai eu la chance de côtoyer au cours des dernières années et qui ont tous, à leurs manières, contribué à alimenter la réflexion nécessaire pour faire avancer ce projet.

Je remercie aussi sincèrement le support accordé par l'École de réadaptation et par la Faculté de médecine et des sciences de la santé de l'Université de Sherbrooke qui ont rendu possible cette thèse. Investir dans la formation professorale en diminuant la tâche d'enseignement d'un des membres d'une si petite équipe, alors que le programme venait à peine de naître, était un pari ambitieux... Un chaleureux merci à mes collègues professeurs et au personnel de soutien qui ont dû compenser ma plus faible disponibilité pendant plusieurs années. Merci de votre support indéfectible et de votre patience. Un merci particulier à ma collègue la Pre Annick Bourget qui, ayant elle aussi traversé une récente aventure doctorale, ne manquait pas une occasion de s'enquérir de l'évolution de mon projet et de prodiguer des encouragements toujours appréciés.

Un énorme merci à Pre Johanne Desrosiers et à Pr Michel Tousignant, mes patrons au début de l'aventure, qui ont su trouver la bonne dose d'arguments et d'encouragements pour me convaincre de débiter ce projet doctoral. Il me fera un énorme plaisir de réinvestir cette expertise nouvellement acquise au sein de l'École de réadaptation et de la Faculté.

Merci aux étudiants de l'École de réadaptation qui ont accepté de participer en grand nombre à l'étude ainsi qu'aux cliniciens experts en physiothérapie qui ont été impliqués dans les différentes phases de développements et de validations des outils. À cet égard un grand merci

à la coordonnatrice des stages du programme de physiothérapie, Mme Anne-Marie Côté pour son implication active à plusieurs des étapes importantes du projet.

Les commentaires, suggestions et encouragements de toutes ces différentes personnes m'ont confirmé de la pertinence de poursuivre le projet malgré les inévitables difficultés reliées à sa réalisation.

Je me dois également de remercier ma famille et mes amis pour leur incroyable support au cours des dernières années. Merci à David et Daniel, mes partenaires de tennis, ne vous inquiétez pas, je devrais encore me souvenir comment utiliser une raquette! Merci à mère Jacqueline, 87 ans, qui a su à sa manière, m'encourager à trouver des réponses à mes questions quelle que soit la nature des questions. Un énorme merci à Chantal, l'amour de ma vie depuis près de 30 ans qui m'a permis de garder un équilibre, même dans les périodes les plus intenses. Merci à mes filles, Frédérique et Catherine, qui ont dû, à trop d'occasions, accepter de partager leur père pour qu'il puisse travailler sur son projet doctoral. Merci d'avoir été à la fois compréhensives et rassurantes par votre présence pendant ces moments importants !

INTRODUCTION

La présente thèse s'inscrit dans le domaine de la validation d'outils d'évaluation du raisonnement clinique en physiothérapie. Le choix de ce thème est apparu en constatant dans mes années de pratique comme enseignant, que bien que le raisonnement clinique soit considéré comme un élément fondamental dans la compétence des physiothérapeutes, l'évaluation de celui-ci est difficile et complexe.

Tel qu'abordés dans le chapitre de la problématique, plusieurs facteurs peuvent être reliés à ce constat, tout d'abord il est généralement reconnu que la complexité et l'aspect multidimensionnel du raisonnement clinique rendent ardu la tâche de l'évaluer (Charlin, Bordage, & van der Vleuten, 2003). Le raisonnement clinique incorpore ainsi plusieurs composantes inter-reliées tel que les processus cognitifs, les connaissances, la métacognition et l'ensemble de ces composantes sont modulées par le contexte de la situation. Conséquemment, la complexité du raisonnement clinique questionne la capacité d'un seul outil à tenir compte de ces multiples facettes. De plus, pour le domaine de la physiothérapie, peu d'outils permettant spécifiquement de poser un jugement sur le raisonnement clinique ont fait l'objet d'étude de validation, la majorité des outils d'évaluation du raisonnement clinique ayant été développés pour le domaine de la médecine.

Dans un contexte d'émergence des approches curriculaires par compétences qui favorise l'évaluation en situation authentique, il existe aussi un questionnement sur le lien existant entre les résultats de ces outils et la performance de l'étudiant en clinique. La performance en situation clinique, malgré toutes les limitations inhérentes à son évaluation est considérée présentement, comme l'information la plus pertinente dans le jugement de la compétence de l'étudiant en sciences de la santé (Van der Vleuten, Schuwirth, Scheele, Driessen, & Hodges, 2010). La présence d'un lien entre la performance en situation authentique et les tests de raisonnement clinique pourrait possiblement être un facteur favorisant l'acceptation de ces derniers dans les programmes de formation en physiothérapie et dans les autres sciences de la santé.

Ce projet doctoral propose ainsi d'aborder, à plusieurs niveaux, certains éléments de l'évaluation du raisonnement clinique en physiothérapie pour nous permettre de jeter un éclairage sur ce thème relativement peu exploré jusqu'à maintenant.

Cette thèse est divisée en six chapitres. Le premier est consacré à la problématique et présente le contexte général de la recherche ainsi que les problèmes auxquels nous nous intéressons dans le cadre de cette étude. Ce chapitre se termine par la question générale de recherche et les deux questions spécifiques. Le second chapitre est dédié au cadre conceptuel qui présente le point de vue des différents auteurs qui se sont intéressés aux concepts clés pertinents à cette thèse. Les cadres théoriques reliés au raisonnement clinique et à la validation des outils y occupent une place prépondérante. Le troisième chapitre effectue une revue de littérature spécifique aux deux outils retenus pour nous permettre d'évaluer le raisonnement clinique des étudiants en physiothérapie. Le quatrième chapitre présente la méthodologie sur laquelle nous nous sommes appuyés afin de collecter, traiter et analyser les données. Le cinquième chapitre est dédié à la présentation des résultats. Enfin, le sixième chapitre présente les principales interprétations issues du traitement des données de façon à répondre à nos questions spécifiques de recherche. Ce chapitre dégage également les limites de la présente recherche et propose des pistes de recherche future. De son côté, la conclusion présente une réponse à la question générale de recherche.

CHAPITRE 1

LA PROBLÉMATIQUE

Ce chapitre présente les problèmes et le contexte reliés à l'évaluation du raisonnement clinique dans les programmes de formation universitaire en physiothérapie. Dans un premier temps, nous mettrons en parallèle l'évolution rapide de la profession au Québec avec l'importance grandissante qui est accordée au raisonnement clinique dans la pratique. Dans la deuxième section, nous décrirons l'évolution de l'évaluation des apprentissages en milieu universitaire en comparant et en contrastant les perspectives des différents paradigmes en éducation à ce niveau. Enfin dans la troisième section, nous aborderons les caractéristiques et les limites des différents outils et situations développées pour évaluer le raisonnement clinique. Ce chapitre se termine avec la présentation des questions de recherche.

1.1 Contexte de la physiothérapie au Québec

Au Québec et dans le reste du monde, la physiothérapie a connu une évolution rapide au cours des dernières années. Cette profession est passée de profession paramédicale travaillant sous supervision médicale à une profession ayant de plus en plus d'autonomie. (Cleather, 1995). Au niveau législatif cette profession est définie, selon l'Ordre professionnel de la physiothérapie du Québec comme : « une discipline de la santé de première ligne intervenant au niveau de la prévention et promotion de la santé, de l'évaluation, du diagnostic, du traitement et de la réadaptation des déficiences et incapacités touchant les systèmes neurologique, musculosquelettique et cardiorespiratoire de la personne » (Ordre professionnel de la physiothérapie, 2012).

En effet, au Québec depuis 1990, la physiothérapie est considérée comme une profession de première ligne dans le système de santé, ce qui implique que les patients peuvent consulter ces professionnels sans consultation médicale préalable. Ce changement a amené des modifications au niveau des curriculums de formation universitaire en physiothérapie. Ainsi, un plus grand accent de la formation a été placée sur le diagnostic différentiel, de façon à permettre aux physiothérapeutes d'identifier les patients qui ne seraient pas appropriés pour

leurs traitements (condition systémique, cancer, etc.) et de déterminer l'intervention la plus pertinente pour ceux qui le sont. (Jette, Ardleigh, Chandler, & McShea, 2006).

Un autre changement législatif qui a eu des implications sur la formation des physiothérapeutes est la redéfinition des actes qui sont réservés aux physiothérapeutes et aux thérapeutes en réadaptation physique (formation collégiale) par le règlement 94m du Code des professions adopté en 2010. Par ce règlement, le physiothérapeute se voit attribuer la responsabilité de l'évaluation neuromusculosquelettique de même qu'un certain encadrement de la pratique des thérapeutes en réadaptation (Code des professions, 2010b). Tout comme l'accès direct, cette réorganisation du travail met l'accent sur le raisonnement clinique du physiothérapeute durant l'évaluation du problème de consultation.

De plus, plusieurs autres changements sont à prévoir dans les prochaines années au niveau de l'ajout d'autres responsabilités professionnelles comme la référence pour des tests d'imagerie et de prescription de médication dans le champ de compétences des physiothérapeutes (Desmeules, 2011). Ces changements connus sous le terme « pratique avancée » sont déjà présents, à différents niveaux, dans d'autres juridictions (Ontario, Angleterre, Australie, etc.) et pourraient également survenir au Québec dans un proche avenir (Aiken, 2012).

En lien avec ces changements présents et futurs et en accord avec la situation canadienne, le niveau de formation requis pour les physiothérapeutes du Québec est devenu la maîtrise professionnelle depuis 2010 (Code des professions, 2010a). Ce rehaussement du diplôme d'entrée à la profession s'est avéré nécessaire compte tenu des nouvelles exigences du marché du travail et du niveau de responsabilités attendues des physiothérapeutes. Les universités québécoises ont donc réajusté la formation à plusieurs niveaux et en particulier au chapitre de la place prépondérante du raisonnement clinique dans les curriculums de formation. Par exemple le programme de l'Université de Sherbrooke utilise à partir de la 2^e année de formation, l'apprentissage du raisonnement clinique comme méthode d'apprentissage principale. L'apprentissage du raisonnement clinique permet à l'étudiant de mobiliser ses ressources et d'exercer son jugement dans différentes histoires de cas réalistes (Chamberland, 1998).

Cette place prépondérante du raisonnement clinique dans les curriculums de formation en physiothérapie fait écho à l'importance accordée à la recherche dans ce domaine dans les dernières années et rejoint en ce sens la place accordée au raisonnement clinique en médecine (Christensen, 2009; Cruz, Moore, & Cross, 2012; Edwards, Jones, Carr, Braunack-Mayer, & Jensen, 2004; Hendrick, Bond, Duncan, & Hale, 2009; Josephson, Bulow, & Hedberg, 2011; May, Greasley, Reeve, & Withers, 2008).

Même s'il est de plus en plus reconnu que le raisonnement clinique est crucial à une pratique de qualité et malgré l'importance accordée à l'enseignement du raisonnement clinique dans les programmes universitaires en physiothérapie, il n'existe, à notre connaissance, que très peu d'études sur les outils pouvant évaluer le raisonnement clinique dans ce domaine (Cohen, Fitzgerald, Lane, & Boninger, 2005). Dans ce contexte, il est pertinent de se demander si les méthodes d'évaluation des apprentissages du raisonnement clinique utilisées dans les programmes universitaires en physiothérapies sont encore appropriées compte tenu des nouvelles compétences attendues des diplômés universitaires en physiothérapie (Verheyden, Handgraaf, Demirci, & Grüneberg, 2011). La prochaine section de ce chapitre traitera des différentes perspectives qui ont influencé l'évaluation des apprentissages au cours des dernières années dans le milieu universitaire.

1.2 L'évolution de l'évaluation des apprentissages en milieu universitaire

L'évaluation des apprentissages en milieu universitaire a également connu plusieurs changements au cours des dernières années. Plusieurs facteurs sont à l'origine ou ont fortement influencé ces changements. Un de ces facteurs est la transition progressive du paradigme d'enseignement vers un paradigme d'apprentissage dans les maisons d'enseignement supérieur (Prégent, Bernard, & Kozanitis, 2009).

La perspective centrée sur l'enseignement était le paradigme dominant en éducation durant les dernières décennies du 20^e siècle (Huba & Freed, 2000). Ce paradigme sous-entend que la qualité de l'enseignement est le critère déterminant de l'apprentissage (Legendre, 2004). Conséquemment et dans cette perspective, l'accent est placé davantage sur les procédés d'enseignement plutôt que sur les processus d'apprentissages. Le procédé d'enseignement le

plus souvent associé à ce courant est le cours magistral, considéré comme le mode privilégié de la transmission du savoir (Vienneau, 2011). Cette conception, qui puise sa source dans les théories béhavioristes de l'apprentissage, suppose que le professeur est responsable de la mise en place des choix pédagogiques et des renforcements nécessaires, afin d'obtenir le comportement attendu de l'étudiant, celui-ci ayant le rôle « passif » de réagir aux stimuli de l'environnement que lui offre l'enseignant (Huba & Freed, 2000).

Du courant béhavioriste découle aussi l'approche par objectifs pour l'évaluation des apprentissages (Desbiens, 2005). Généralement dans cette approche, on détermine *à priori* l'ensemble des connaissances, habiletés et attitudes qui seront enseignées et évaluées. Des taxonomies ont ainsi été proposées pour organiser et découper les objectifs spécifiques et généraux, de manière à rendre explicite ce qui est attendu des étudiants. Une des plus connues de ces taxonomies est celle de Bloom qui propose un découpage en trois domaines : cognitif, psychomoteur et affectif. Les domaines sont par la suite divisés en niveaux hiérarchisés. Par exemple, les objectifs pédagogiques du domaine cognitif incluent la connaissance, la compréhension, l'application, l'analyse, la synthèse et l'évaluation (Bloom, 1973). De ces objectifs découleront les moyens pédagogiques utilisés par l'enseignant. Les objectifs représentent, en quelque sorte, les termes du contrat pédagogique qui unit l'enseignant et l'étudiant (Legendre, 2004).

En convergence avec le courant behavioriste, le paradigme d'enseignement et à l'approche par objectifs, s'est développée la perspective docimologique de l'évaluation (Nguyen & Blais, 2007). Cette perspective, fortement présente dans les facultés de médecine et des sciences de la santé, privilégie les mesures faites à l'aide de tests ou de tâches standardisés pour recueillir des informations sur l'apprentissage des étudiants (Jouquan, 2002). On considère qu'une tâche est standardisée quand : « a) tous les étudiants répondent aux mêmes questions; b) tous les étudiants reçoivent les mêmes instructions claires et précises; c) aucun étudiant ne profite d'avantages externes à l'objet de mesure par rapport aux autres; d) pour tous les étudiants, un système prévu de correction s'applique uniformément. » (Morissette, 1984, 93). Cette approche met l'accent sur les qualités psychométriques (validité et fidélité) des données recueillies à l'aide des outils utilisés (Desbiens, 2005). À partir de ces données, une interprétation de nature normative ou

critériée pourra être effectuée pour en arriver à prendre une décision sur un candidat. Une interprétation normative est effectuée lorsqu'on regarde la position relative d'un individu par rapport à un groupe pour porter un jugement sur le niveau de cet individu, et décider par la suite si ce niveau est acceptable ou non. Dans le cas de l'interprétation critériée, elle s'effectue en lien avec des facteurs de référence établis par le programme et implique un seuil de réussite sans égard à la performance du groupe (Morissette, 1984).

Ces différents paradigmes et approches se sont influencés mutuellement à un tel point sur le plan de l'évaluation des apprentissages, qu'il devient difficile de déterminer les influences qui relèvent de l'un ou l'autre de ces paradigmes et approches (Nguyen & Blais, 2007). Nous regarderons dans la prochaine partie de cette section les particularités de l'évaluation des apprentissages dans une approche par objectifs.

1.2.1 Caractéristiques de l'évaluation des apprentissages dans une approche par objectifs

Dans l'approche par objectifs, l'évaluation des apprentissages portera de façon directe sur les objectifs spécifiques déterminés par le professeur ou le programme d'étude. En conformité avec le courant behavioriste, l'évaluation des apprentissages se concentrera sur les comportements observables qui découlent des actions pédagogiques (Vienneau, 2011). Ainsi, à partir de l'objectif spécifique, un certain nombre de questions (items) seront développées.

« Chaque item est en fait un stimulus qui doit provoquer chez tous les élèves une réaction précise qui constituera la preuve du comportement souhaité. Ce comportement constitue par ailleurs un indice de la présence de la connaissance, de l'habileté ou de l'attitude définie dans le programme » (Morissette, 1984, p.87).

Ce découpage analytique en petites portions de connaissances, d'habiletés, ou d'attitudes est une des caractéristiques des examens ou tests¹ utilisés dans cette approche (Legendre, 2004). Ce découpage découle d'une organisation par objectifs en domaines cognitif, psychomoteur et affectif, ce qui est aussi une des principales critiques adressées à l'évaluation des apprentissages

¹ « Lorsque les procédures de conception, d'administration et d'analyse seront plus rigoureuses, plus systématiques, on parlera de test plutôt que d'examen » (Morissette, 1984, p. 86).

dans cette approche (Tardif, 2006). Ces différents domaines étant imbriqués la majorité du temps, leur division risque de faire perdre de vue certains aspects essentiels (Legendre, 2005).

Plusieurs critiques de l'évaluation des apprentissages selon l'approche par objectifs sont reliées à des limitations imposées par les utilisateurs plutôt que des limitations de l'approche elle-même (Nguyen & Blais, 2007). Par exemple, il est fréquemment mentionné que l'approche par objectifs a tendance à collecter des données sur les objectifs de faible niveau taxonomique et à privilégier le résultat plutôt que le processus (Scallon, 2004; Tardif, 2006). Ces critiques sont probablement plus en lien avec la difficulté de concevoir des questions qui vont permettre d'évaluer des niveaux taxonomiques élevés et des processus, qu'avec les prémisses de cette approche (Nguyen & Blais, 2007). En effet, peu importe les pratiques pédagogiques utilisées, il est difficile de développer des questions pouvant nous permettre d'évaluer des habiletés cognitives complexes ou des processus. Cette difficulté peut faire qu'en pratique une grande partie des outils d'évaluations des apprentissages se situe au niveau de la mémorisation ou de l'apprentissage de surface tel que l'étude de Palmer *et al.* (2010) le suggère dans le contexte de la médecine.

Probablement aussi en lien avec son affinité pour l'approche docimologique, dans l'approche par objectifs, la recherche d'une seule bonne réponse est privilégiée au détriment du processus de résolution de problèmes (Tardif, 1993). Or, une méthode d'évaluation qui cible le processus n'est pas incompatible avec l'approche par objectifs (Nguyen & Blais, 2007), mais en pratique, elle fait peu partie de la culture.

Dans une critique des tests standardisés, Resnick et Resnick (1992) questionnent deux postulats associés à cette perspective; premièrement celui de la « décomposabilité », qui stipule que chaque apprentissage peut être divisé et que l'évaluation isolée de chacune des composantes peut nous informer de la compétence globale. De plus, ces auteurs s'interrogent sur le postulat de la décontextualisation qui stipule que chaque composante de compétences est fixe, peu importe le contexte où elle est présente. Comme plusieurs auteurs (Perrenoud, 2002; Prigent *et al.*, 2009; Tardif, 1993) l'ont noté, même si les étudiants semblent avoir acquis des connaissances dans des situations académiques, le transfert dans des situations professionnelles n'est pas facile et est au cœur de la remise en question du découpage taxonomique.

Une autre des raisons qui a conduit à une remise en question des perspectives du paradigme centré sur l'enseignement, de l'approche par objectifs et de l'approche docimologique, est l'intérêt grandissant pour le courant constructiviste en éducation par rapport aux conceptions béhavioristes. Selon la conception constructiviste, tout apprentissage ne survient pas forcément à la suite d'un enseignement et tout enseignement ne conduit pas nécessairement à un apprentissage (Legendre, 2004). La prochaine partie traitera des caractéristiques du paradigme centré sur l'apprenant et de l'approche par compétences qui a été choisie par un grand nombre de programmes universitaires en sciences de la santé incluant la physiothérapie (Vissers, van Daele, & Denekens, 2012). Cette section décrira aussi les caractéristiques essentielles de l'évaluation des apprentissages dans ce courant pédagogique.

1.2.2 Caractéristiques de l'évaluation des apprentissages dans une approche par compétences

Le changement du paradigme centré sur l'enseignement à celui centré sur l'apprenant a débuté durant les années 80 et se poursuit aujourd'hui dans les différents milieux d'enseignement incluant l'université (Huba & Freed, 2000). Ce courant, de même que sa « proche parente » l'approche par compétences, est inspiré des courants constructivistes et socioconstructivistes en éducation (Legendre, 2004). Dans le courant constructiviste, l'apprentissage est défini comme un processus de construction personnelle de la réalité qui implique activement l'individu (Vienneau, 2011). La perspective du socioconstructiviste ajoute la notion de l'importance des interactions sociales dans la construction des connaissances (Legendre, 2008). Un engagement profond de l'étudiant dans la démarche d'apprentissage est attendu dans ces deux conceptions. Ces courants ayant influencé l'évaluation des apprentissages dans le paradigme centré sur l'apprenant et dans l'approche par compétences de façon similaire, elles seront donc traitées ensemble dans cette section.

Dans une perspective constructiviste, l'étudiant doit participer activement à son apprentissage par des activités qui doivent susciter ses questionnements, tout en étant en interaction avec ses pairs. Ces activités doivent favoriser la réactivation des connaissances antérieures de l'étudiant et doivent promouvoir des situations d'autoapprentissage, des projets et de la résolution de problèmes (Vienneau, 2011). Dans ce paradigme, le choix des situations pour contextualiser les

apprentissages dans des situations authentiques est important, et ce, en favorisant le recours à de multiples ressources (Legendre, 2004).

Dans une perspective globale, l'approche par compétences, inspirée du constructivisme et socioconstructivisme, essaie de décloisonner les disciplines pour se rapprocher de l'utilisation des savoirs qui est faite en situation réelle (Tardif, 2006). Les différentes ressources et les différents contextes du programme de formation et d'évaluation des apprentissages doivent être agencés et planifiés de façon optimale pour favoriser les intégrations.

Dans un premier temps, nous définirons le concept de compétence, celui-ci étant évidemment central à l'approche et à l'objet d'évaluation. Bien qu'il existe un grand nombre de définitions de la compétence (Coulet, 2011), la plupart de celles-ci impliquent la notion d'action ou d'agir par opposition à un objectif d'apprentissage qui, dans bien des cas, peut être de nature théorique ou lié à un savoir spécifique (déclaratif, procédural ou conditionnel).

Pour aborder l'évaluation des compétences, nous nous appuyerons sur la définition de compétence de Tardif qui s'inscrit dans cette vision de la littérature francophone : « Savoir-agir complexe prenant appui sur la mobilisation et la combinaison efficaces d'une variété de ressources internes et externes à l'intérieur d'une famille de situations. » (Tardif, 2006, p. 22). Cette définition semble, pour le moment, représenter une synthèse du construit théorique de la compétence en sciences de l'éducation dans le monde francophone (Nguyen & Blais, 2007; Prigent *et al.*, 2009). Le concept de compétence semble mieux accepté au niveau de la formation professionnelle d'où il tient ses origines. Il est, par contre, plus controversé dans la formation générale (Allal, 1999). Ses détracteurs y voyant une intrusion du monde du travail dans la formation scolaire et craignent de perdre une formation générale faite de savoirs qui ne sont pas nécessairement orientés vers un aspect utilitaire (Hirtt, 2009).

En considérant les importants aspects qui différencient l'approche par objectifs de l'approche par compétences, il apparaît évident que l'évaluation des apprentissages doit aussi refléter ce changement de vision (Tardif, 2006). D'une approche à l'autre, l'objet de l'évaluation passe donc du jugement de façon isolée sur une connaissance, une habileté ou une attitude, au jugement de la mobilisation de l'ensemble de ces éléments dans une situation donnée.

La conception de l'apprentissage qui soutient l'approche par compétences questionne les façons de faire qui sont utilisées afin d'évaluer les apprentissages dans une approche par objectifs. À la lumière des travaux de différents auteurs francophones (Le Boterf, 2007; Perrenoud, 2004; Scallon, 2004; Tardif, 2006), certaines caractéristiques de l'évaluation en approche par compétences semblent faire consensus. Une de ces conceptions est d'accorder la priorité à l'évaluation des compétences en y intégrant les ressources et le processus de résolution de problème.

Cette conception de l'évaluation en approche par compétences est probablement une des plus controversées. Elle implique d'évaluer comment l'étudiant mobilise et combine l'ensemble des ressources dans un contexte précis, en évitant d'évaluer les ressources comme entités séparées (Tardif, 2006). Selon LeBoterf (2007), la compétence ne réside pas dans les ressources, mais dans la mobilisation de celles-ci. La compétence étant définie par la situation, évaluer des ressources sans contexte perd son sens, mentionne-t-il. Cette conception contraste donc de façon importante avec les pratiques fragmentées et analytiques de l'approche par objectifs, qui propose de découper chaque connaissance et habileté en tranche fine.

Ces différences ont alimenté de nombreux débats sur la question (Baillargeon, 2009; Hirtt, 2009). Pour plusieurs, il est difficile de concevoir qu'il n'est pas indispensable de s'assurer de la maîtrise du langage avant d'évaluer la composition de texte par exemple ou qu'il n'est pas important d'évaluer les connaissances d'anatomie humaine d'un chirurgien, préalablement à l'évaluation de ses compétences chirurgicales (Hirtt, 2009).

Bien qu'il existe un consensus qui préconise que l'évaluation des compétences se fasse en contexte de situation complexe (Joughin, 2009; Scallon, 2004; Tardif, 2006), il demeure que cette caractéristique pose un défi technique pour l'évaluateur qui essaie de donner une rétroaction la plus précise possible sur une tâche, en essayant de cibler la ressource ou la capacité à mobiliser cette ressource, le processus ou la particularité de la situation qui pose problème pour l'étudiant (Scallon, 2004). Le changement de culture impliqué est majeur et exige de bien cerner l'information que chaque situation d'évaluation peut nous apporter (Tardif, 2006). Cette complexité de l'évaluation demande une planification importante des situations qui seront

utilisées dans le programme de formation, afin d'en retirer les informations pertinentes sur la progression et ultimement sur la compétence de l'étudiant (Gerard & Roegiers, 2010).

Même si l'approche par compétences ne peut revendiquer l'exclusivité de l'évaluation des processus ou des cheminements, plusieurs auteurs associés à ce courant en ont fait une caractéristique importante (Scallon, 2004; Tardif, 2006). Le fait de ne considérer que le résultat d'une tâche ou la réponse à une question fait perdre beaucoup d'informations pertinentes à l'apprentissage. En effet, une mauvaise réponse ou tâche incomplète ne permet pas toujours de savoir où se situe la difficulté de l'étudiant. Quoique, souvent difficile dans un contexte sommatif, le fait de s'intéresser au cheminement de l'étudiant et de donner une rétroaction sur celui-ci peut permettre la progression des apprentissages de l'étudiant peu importe l'approche pédagogique (Scallon, 2004).

Une autre notion qui semble faire consensus chez la plupart des auteurs associés à l'évaluation des compétences est l'importance des situations d'évaluation authentiques (Wiggins, 1989).

Selon Wiggins (1989), l'évaluation authentique doit :

- être réaliste;
- favoriser le jugement et l'innovation;
- demander aux étudiants d'accomplir une réalisation plutôt que simplement dire, répéter ou reproduire;
- simuler le contexte d'une situation de travail;
- proposer des tâches complexes;
- favoriser la consultation, la rétroaction et l'amélioration.

Cette façon de faire est un bel exemple d'intégration de l'apprentissage et de l'évaluation. L'utilisation de situations complexes de la vie quotidienne ou professionnelle permet à l'étudiant de faire des apprentissages et à l'enseignant de juger de la mobilisation des ressources à l'intérieur de cette situation. Ces situations seraient appropriées pour évaluer les problèmes complexes, pour lesquels il n'existe pas qu'une seule bonne réponse (Huba & Freed, 2000). L'aspect réel du problème à résoudre permettrait aussi d'assurer des conditions optimales de

motivation chez l'étudiant en donnant du sens à son apprentissage (Huba & Freed, 2000) particulièrement en formation professionnelle.

L'évaluation en situation authentique implique de recourir à des critères multiples dans chaque cas (Tardif, 2006). Pour bâtir une situation d'évaluation compatible avec ce principe, il faut aller au-delà des frontières disciplinaires; il faut que la tâche permette une combinaison et une intégration des ressources qui n'appartiennent pas à la même discipline. Ceci sous-entend un accord des enseignants d'un même programme sur les situations et les critères d'évaluation d'une même tâche. La mise en commun des différentes expertises des enseignants et des données recueillies est indispensable pour assurer le bon fonctionnement de ce type de situation d'évaluation (Gerard & Roegiers, 2010; Prigent *et al.*, 2009).

Bien que ces situations authentiques semblent présenter plusieurs avantages au niveau de la validité apparente pour évaluer des savoirs agir complexes (Tardif, 2006), celles-ci demandent beaucoup de temps à l'étudiant (et au professeur), et permettent d'évaluer somme toute peu de ressources dans une seule situation (Jouquan, 2002). La performance étant spécifique à un contexte, son évaluation à l'aide d'un nombre restreint de tâches nuit à la généralisation qui peut être faite (Shavelson, Baxter, & Gao, 1993).

Plusieurs facteurs, comme le fait que l'étudiant ait rencontré une situation trop complexe ou que le résultat de l'activité soit peu concluant, peuvent rendre plus difficile la tâche de déterminer quelles ressources ne sont pas bien maîtrisées par l'étudiant et influencer le jugement de l'évaluateur. S'y ajoute la difficulté d'établir *a priori* les critères d'évaluation pour des tâches variées et pour lesquelles il va être difficile d'estimer le temps nécessaire pour les réaliser (Scallon, 2004). Donc, plusieurs éléments liés à la non-standardisation des situations d'évaluation peuvent être vus comme des problèmes importants selon les critères de l'approche docimologique et potentiellement des obstacles à l'utilisation des situations authentiques en évaluation certificative (Jouquan, 2002). L'ensemble de ces arguments nous incite à bien choisir les moments et la nature des épreuves d'évaluation au travers desquels des situations authentiques pourront être utilisées pour juger de la compétence des étudiants.

Le jugement est une autre des caractéristiques qui fait partie intégrante de l'évaluation dans une approche par compétences (Gerard, 2002). L'utilisation de situations authentiques ou de certains autres outils, comme les cartes conceptuelles ou le portfolio, implique souvent une absence de données quantitatives souvent associées à des mesures (Blais, 2012) et exige que le jugement de l'évaluateur soit prépondérant (Scallon, 2004).

Lorsqu'évalué avec les critères de l'approche docimologique, cet aspect peut être considéré comme un point faible de l'évaluation de compétence ou de performance (Khattri & Sweet, 1996). En effet, il est en général admis que le jugement d'un évaluateur peut être biaisé par différents facteurs internes ou externes compromettant ainsi la validité de l'évaluation. Par exemple, la préférence à l'égard de certaines caractéristiques d'une production (effet de halo) ou la position d'un travail dans le lot à évaluer (effet de l'ordre de correction) peuvent influencer de façon importante l'évaluation de ce travail (Scallon, 2004).

Selon Scallon (2004), le recours au jugement professionnel demeure incontournable dans une approche par compétences, même s'il en reconnaît certaines limitations. Dans le but de minimiser ces limitations, la formation des évaluateurs a été proposée dans le but d'améliorer la fidélité inter-évaluateurs (Kogan, Holmboe, & Hauer, 2009; Van der Vleuten, Schuwirth, Scheele, Driessen, & Hodges, 2010). Bien que peu d'études le démontrent clairement pour l'instant, ce besoin semble particulièrement important dans les cas où des évaluateurs externes (par exemple, les superviseurs de stage) sont interpellés pour poser des jugements sur la performance d'étudiants, sans avoir de formation professionnelle en pédagogie (Holmboe, Hawkins, & Huot, 2004).

L'adoption de l'approche par compétences comme pratique pédagogique s'est imposée au Québec dans les dernières années, au niveau des programmes de formation générale et professionnelle (Prégent *et al.*, 2009). Cependant, selon Perrenoud, l'implantation des caractéristiques de l'évaluation des apprentissages de l'approche par compétences s'est faite avec un certain retard par rapport aux pratiques d'enseignement. « On souffre de la relative légèreté des systèmes éducatifs qui mettent en place des curricula par compétences avant de savoir comment les évaluer. » (Perrenoud, 2005; p. 5). Cette façon de faire peut ainsi conduire à certaines incohérences entre les pratiques d'enseignement et les pratiques évaluatives. Des

changements majeurs ne peuvent être mis en place s'il n'y a pas au préalable de plan au niveau de l'évaluation des apprentissages. « L'évaluation est indissociable du processus pédagogique, à tel point que certains qualifient celui-ci de processus d'enseignement-apprentissage-évaluation » (Gerard & Roegiers, 2010, p.1).

Malgré des efforts en ce sens, l'approche par compétences n'a pas encore développé de maturité par rapport à ses pratiques évaluatives (Jonnaert, 2009; Perrenoud, 2005). Plusieurs questions restent encore en suspens : par exemple, devrait-on évaluer les compétences dans des situations connues ou nouvelles? Si une nouvelle situation est retenue, sur quelle base de correspondance peut-on établir la validité de la situation d'évaluation par rapport aux situations d'apprentissage et à la famille de situations (Scallon, 2004)? Selon Scallon, même si les termes reliés à la validité restent les mêmes qu'avec l'approche docimologique, « Il faut cependant resituer ces termes dans une approche par compétences, ce qui change la perspective dans plusieurs cas » (Scallon, 2004, p. 261). Ce travail qui consiste à redéfinir les facettes de la validité reste à faire en grande partie, dans le domaine de l'évaluation des compétences, selon Jonnaert (2009).

De plus, selon Tardif (2006), l'évaluation des compétences ne pourra être complètement satisfaisante que lorsque des modèles cognitifs² de l'apprentissage permettront de baliser leurs développements. Or, malheureusement, ces modèles font encore largement défaut.

« À l'heure actuelle, dans la plupart des programmes de formation axés sur le développement de compétences et pour la majorité des compétences ciblées, il n'existe aucun modèle cognitif de l'apprentissage. Dans une large mesure, cet état de fait fragilise l'évaluation des apprentissages dans ces programmes... » (Tardif, 2006, p. 122).

Donc, dans l'optique de Tardif (2006), tant que les modèles et les trajectoires de développement de compétences qui en découleront n'auront pas été développés, validés et que les programmes de formation ne s'y seront pas conformés, une évaluation des compétences complètement cohérente ne sera pas possible.

² Selon le National Research Council, ces modèles issus des données probantes font référence à la manière avec laquelle les étudiants construisent leurs compétences dans un domaine. (Tardif, 2006)

Un changement de paradigme ne s'effectue pas sans difficulté, plusieurs défis restent présents. Dans les prochaines années, l'approche par compétences adoptée par de nombreux programmes professionnels universitaires devra démontrer sa capacité à s'implanter de façon intégrale à la fois comme approche d'enseignement et également comme approche d'évaluation.

Comme mentionné précédemment, il existe une forte convergence des concepts liés au courant béhavioriste, à l'approche docimologique, à l'approche par objectifs et au paradigme centré sur l'enseignement. Il existe aussi une « parenté conceptuelle » entre le courant socioconstructiviste, l'approche par compétences et le paradigme centrée sur l'apprenant (Nguyen & Blais, 2007). Bien qu'il serait réducteur d'opposer systématiquement ces deux grandes perspectives, il demeure que celles-ci ont des visions différentes par rapport à l'évaluation des apprentissages. Cependant, est-ce que certains aspects de ces perspectives peuvent être compatibles ou même complémentaires?

Plusieurs tenants de l'approche par compétences préconisent une coupure radicale par rapport aux pratiques associées à l'évaluation des apprentissages de l'approche par objectifs (Jonnaert, 2009; Tardif, 2006). Cependant on peut s'interroger sur la nécessité de remettre en question l'ensemble des pratiques évaluatives dans une approche par compétences. Pour éviter de « jeter le bébé avec l'eau du bain », il peut être pertinent de porter un regard attentif sur les pratiques d'évaluations qui sont compatibles et celles qui sont à proscrire dans une approche par compétences. Par exemple dans le but de favoriser le développement des compétences dans un programme de formation professionnelle, doit-on éviter d'évaluer les connaissances séparément des compétences associées?

« Selon certains didacticiens des disciplines, comme Jean-Pierre Astolfi, on pourrait se passer complètement du concept de compétence, à condition de définir les connaissances comme des ressources dont on peut se servir non seulement pour répondre à des questions d'examen hors de tout contexte d'action, mais pour résoudre des problèmes, prendre des décisions, conduire des projets, etc. » (Perrenoud, 2005, p. 2).

Cette affirmation remet en perspective l'importance de la connaissance contextualisée dans l'apprentissage, sans que celle-ci soit nécessairement assujettie au concept de compétence. En ce sens, il apparaît que l'évaluation des ressources contextualisées peut représenter une des données pertinentes qui permet d'effectuer un jugement évaluatif, surtout en début de

programme de formation professionnelle. Il semble, de plus, possible de contextualiser les savoirs sans toutefois les évaluer avec toutes les caractéristiques préconisées par l'approche par compétences, car, dans beaucoup de circonstances, une telle pratique n'est pas réalisable. Par exemple, il est probablement utopique et non souhaitable d'évaluer les apprentissages en situation authentique d'un étudiant de première année en sciences de la santé, la charge cognitive associée à ce scénario n'étant pas optimale pour un novice à ce stade (Schuwirth & van der Vleuten, 2011). Par contre, à mesure que l'étudiant développe un bagage suffisant de ressources pour pouvoir les mobiliser dans différentes situations de la vie professionnelles, il devient alors plus réaliste d'évaluer ses compétences dans des situations professionnelles crédibles.

Dans la prochaine section nous discuterons des différents outils d'évaluation qui ont été développés pour évaluer le raisonnement clinique et comment ceux-ci s'intègrent dans les différents paradigmes qui influencent l'évaluation des apprentissages.

1.3 L'évaluation du raisonnement clinique

L'évaluation du raisonnement clinique est généralement considérée comme une tâche complexe (Charlin, Bordage, & van der Vleuten, 2003; Gruppen & Frohna, 2002; van der Vleuten, Norman, & Schuwirth, 2008). Cette complexité tient en partie au fait que le raisonnement clinique est imbriqué aux connaissances et aux habiletés en tant que composante de la compétence clinique (Charlin *et al.*, 2003). De plus, il existe en général plusieurs bonnes réponses à un problème clinique et différents processus pour le résoudre. La difficulté d'évaluer le raisonnement clinique, tout comme dans le cas de questions de niveaux taxonomiques élevés, peut inciter les concepteurs de question d'examen à privilégier l'évaluation de connaissance pour lesquels les réponses sont claires et ne peuvent être remises en question (Palmer & Devitt, 2007).

Principalement développés dans un contexte médical, plusieurs types d'outils sont actuellement utilisés dans ce domaine afin d'évaluer le raisonnement clinique. Nous traiterons des caractéristiques des outils ayant fait l'objet du plus grand nombre d'études et verrons comment certains de ces outils pourraient être utilisés pour évaluer le raisonnement clinique en

physiothérapie. Dans le cas des outils papier crayons³, certains de ceux-ci reposent sur un type de question spécifique, d'autres sont plutôt un assemblage de plusieurs types de questions. Nous décrirons de plus, des situations et contextes dans lesquelles le raisonnement clinique peut être évalué.

1.3.1 Questions à choix multiples (QCM)

Bien que la perception de l'utilité ou de la pertinence des questions à choix multiples (QCM) varie beaucoup d'un auteur à l'autre, certains auteurs tels que Prégent *et al.* (2009) ou Joughin, (2009) iront jusqu'à remettre en question complètement leur utilisation à l'université particulièrement dans une approche par compétences. Les critiques portent, entre autres, sur l'incitation de ce type de question vers un apprentissage de surface (Bennett, 1993; Crooks, 1988) et sur la difficulté de concevoir des questions qui ne comportent pas d'indices aidant à répondre sans maîtriser la matière évaluée (Prégent *et al.*, 2009). Plusieurs études (Martinez, 1991; Ward, 1982) ont démontré que, pour la même matière évaluée, les réponses construites auraient en général un score inférieur aux questions à choix de réponses. Ce phénomène serait probablement partiellement attribuable à l'effet d'indice qui est moins élevé dans les réponses construites (Schuwirth & van der Vleuten, 2004).

Il demeure cependant que cet outil est encore très utilisé dans les sciences de la santé (McCoubrie, 2004). Mais dans le contexte du raisonnement clinique, qui se situe clairement à un niveau de complexité au-dessus des simples connaissances factuelles, nous pouvons nous questionner à savoir si ce type d'outil est approprié? Le résultat d'une étude effectuée par Schuwirth en 2001 suggère que l'utilisation d'une vignette clinique qui comporte un contexte implique un processus de résolution de problème qui diffère de la simple vérification de la véracité d'un énoncé, comme dans le cas d'une question à choix multiples qui ne comporterait pas de donnée clinique. Le fait d'incorporer un contexte clinique à une question ajoute une dimension de décision contextuelle qui implique l'utilisation de stratégies cognitives différentes de l'expert par rapport au novice (Schuwirth, Verheggen, van der Vleuten, Boshuizen, &

³ Les outils papier crayon regroupent de façon générale les tests dont les questions sont lues par les candidats et auxquelles ils répondent par écrit et sont traditionnellement effectué en salle de classe.

Dinant, 2001), ce qui ne semble pas être le cas dans les questions portant sur les connaissances factuelles.

Les avantages fréquemment mentionnés de cet outil sont de permettre un large échantillonnage de questions ayant plusieurs contextes différents, ce qui est important compte tenu de la spécificité de contenu. En effet, il a été démontré de façon empirique que la corrélation entre la performance sur différents problèmes ayant différents contextes est faible (Elstein, Schulman, & Sprafka, 1978). Cette constatation est à l'origine de la tendance actuelle en évaluation sommative dans les sciences de la santé, de favoriser de multiples questions courtes ayant différents contextes, en opposition à des questions longues dans un nombre de contextes limités (Charlin *et al.*, 2003). Un autre avantage important des questions à choix multiples, compte tenu de ressources limitées, est le temps de correction qui s'en trouve minimisé, cette correction pouvant être effectuée de façon automatisée.

D'autre part, il existe plusieurs désavantages à ce type de question (même avec une vignette clinique), dans le contexte de l'évaluation de raisonnement clinique. Premièrement, une seule bonne réponse est généralement demandée, ce qui n'est souvent pas le cas dans les situations professionnelles complexes (Schön, 1983). Cela peut ainsi occasionner des raisonnements réducteurs et peu nuancés chez l'étudiant et pourrait provoquer une adaptation de l'enseignement à cette forme de question. En effet, il est possible que certaines méthodes d'évaluation puissent encourager certaines formes d'enseignement destinées à préparer spécifiquement l'étudiant pour le type de questions (*teaching for the test*), plutôt que de favoriser son apprentissage au sens plus large (Bennett, 1993). Un autre désavantage majeur des QCM est de s'attarder uniquement au résultat du raisonnement, sans permettre l'accès au processus et potentiellement aux erreurs d'interprétation de l'étudiant (Scallon, 2004). Sans cette information, il devient très difficile de pouvoir donner une rétroaction utile à l'apprentissage.

1.3.2 Questions de type ouvertes (réponses construites)

Plusieurs types de questions ouvertes existent dans le contexte de l'évaluation du raisonnement clinique. Dans ce contexte, le terme *Modified essay question* est souvent utilisé dans la littérature anglophone en présence d'une situation clinique. Ce type de questions est souvent perçu comme

plus authentique et moins influencé par l'effet d'indice que les QCM (Stratford & Pierce-Fenn, 1985). Par contre, les résultats de ces deux types de questions sont généralement hautement corrélés lorsque celles-ci portent sur le même thème (Schuwirth & van der Vleuten, 2004). Le thème pouvant porter sur une connaissance spécifique ou une décision à prendre dans un certain contexte.

La conception de ce type de question est habituellement plus simple, mais la production d'une échelle de correction basée sur un consensus est complexe (Charlin *et al.*, 2003). Cette complexité du processus de correction pourrait inciter les concepteurs à créer des questions qui vont évaluer les connaissances factuelles plutôt que le raisonnement clinique (Palmer & Devitt, 2007; Palmer, Duggan, Devitt, & Russell, 2010). De plus, parce que ces questions sont plus longues, la fidélité de ce type d'examen (qui comporte moins de questions par heure de passation) est plus faible (Schuwirth & van der Vleuten, 2004). Cependant, ce format peut être le plus approprié dans certains cas, pour évaluer quelques aspects du raisonnement clinique. Mentionnons, à titre d'exemple, la génération d'un diagnostic (sans suggérer de réponses) ou pour faire expliciter l'étudiant sur ces choix thérapeutiques. Considérant ces facteurs, la combinaison des formats, telle que préconisée dans les examens par éléments clés, peut comporter des avantages pour évaluer le raisonnement clinique.

1.3.3 Examen par éléments clés (*Key features*)

Dans les examens par éléments clés, une série de questions est posée sur les éléments les plus importants d'une vignette clinique réaliste. L'élément clé représente une étape critique de la résolution d'un problème ou d'une décision qui représente un défi où l'étudiant est plus susceptible de commettre une erreur. Plusieurs types de décisions peuvent être évalués, que ce soit d'un point de vue diagnostique, pronostique ou thérapeutique. Plusieurs formats de question peuvent être utilisés : à choix multiples, à longue liste ou à courtes réponses, entre autres dans le but de minimiser les effets d'indices (Farmer & Page, 2005). Dans ce contexte, les questions à réponses courtes (construites) exigent plus de ressources, car celles-ci doivent être corrigées individuellement. Toutefois, cela permettrait une meilleure discrimination du niveau des candidats (Farmer & Page, 2005). Ce type de questions par éléments clés permet d'utiliser un grand nombre de contextes dans un assez court intervalle de temps, répondant ainsi aux

exigences créées par la spécificité de contenu. Selon le contexte, plusieurs bonnes réponses peuvent être acceptées et le nombre de réponses attendu est spécifié. Ce type de questions a graduellement remplacé, au cours des 15 dernières années, les PMP (*Patient Management Problem*) dans certains tests à enjeux élevés des organisations médicales de différents pays (van der Vleuten *et al.*, 2008). Même si ce type d'examen corrige en bonne partie les défauts des PMP (faible fidélité et corrélation élevée avec les connaissances), il nous apparaît que relativement peu d'études se sont penchées sur ses qualités psychométriques et sa capacité à nous permettre d'évaluer le raisonnement clinique. Pour une revue de littérature sur les éléments clés se référer à l'article de Hrynchak *et al.*, 2014. Pour un exemple de ce type de question voir annexe 11.

1.3.4 Questions à appariement étendu (*Extended matching questions*)

Les questions à appariement étendu représentent une variation des questions à choix multiples. Ce type de question présente d'abord les choix et demande d'associer la bonne réponse avec les énoncés des vignettes cliniques portant sur un thème. Par rapport aux choix de réponses qui sont typiquement de quatre ou cinq, ce type de questions peut en comporter jusqu'à 23 (Case, Swanson, & Ripkey, 1994), incluant quelques vignettes cliniques par thème. Les arguments en faveur de ce type de questions soulignent principalement la diminution de la probabilité de trouver la bonne réponse en se fiant à la chance, ce qui améliore la discrimination par rapport aux questions à choix multiples conventionnelles (Case *et al.*, 1994). Ce style de question a été originellement conçu pour évaluer la reconnaissance de modèles types (*pattern recognition*), en assumant que l'étudiant se ferait une idée de la réponse en regardant la vignette avant de consulter les choix possibles. Certaines études démontrent par contre une augmentation du temps utilisé par l'étudiant en fonction du nombre de choix de réponses offerts (Swanson, Holtzman, & Allbee, 2008; Swanson, Holtzman, Clauser, & Sawhill, 2005), suggérant que l'étudiant considère chacune des différentes réponses. Pour cette raison, il est conseillé de diminuer le nombre de choix de réponses afin d'optimiser le temps de passation et le nombre de questions et contextes possibles (Swanson *et al.*, 2008).

Dans le but de déterminer si le format des questions influence le processus de raisonnement utilisé, une étude (Heemskerk *et al.*, 2008) a comparé le raisonnement clinique utilisé dans les

questions à appariement étendu et les questions à courtes réponses (construites). Cette étude a démontré que le type de question à appariement étendu entraînerait plus souvent un raisonnement hypothético-déductif qu'une question ouverte pour le même contenu. Donc, contrairement à ce que les concepteurs de ce type de questions avaient prévu au départ, il semble que les étudiants vont aller tester, dans un mode hypothético-déductif, chacun des choix offerts dans une question à appariement étendu. Dans leur étude, Coderre *et al.* (2004) ont démontré qu'il n'y a pas de différence entre les processus de raisonnement clinique utilisés pour résoudre des questions à appariement étendu et les choix multiples classiques. En utilisant un protocole d'explicitation à voix haute, cette étude a constaté l'utilisation de processus similaires (reconnaissance de modèles types, hypothético-déductif) dans les deux types de questions.

Bien qu'il existe probablement des différences entre les questions à choix multiples et les réponses construites (Bennett, 1993), les résultats des études de Coderre *et al.* (2004) et Heemskerk *et al.* (2008), de même que celles de Beullens *et al.* (2005, 2006) suggèrent que des QCM, des questions à appariement étendu ou des questions à réponses courtes peuvent, si elles sont bien construites, évaluer le raisonnement clinique plutôt que le simple rappel de connaissance (Beullens, Struyf, & Van Damme, 2005, 2006). Pour un exemple de ce type de question voir annexe 11.

1.3.5 Test de concordance des scripts (TCS)

Le test de concordance des scripts (TCS) est un outil d'évaluation du raisonnement clinique qui comporte un seul format de question utilisant un support de réponse de type Likert. Le TCS est un outil qui semble de plus en plus utilisé dans les différents domaines de la médecine (Lubarsky, Charlin, Cook, Chalk, & van der Vleuten, 2011), mais également de façon émergente dans les autres sciences de la santé (Cohen *et al.*, 2009; Deschênes, Charlin, Gagnon, & Goudreau, 2011). Basé sur la théorie des scripts, ce test nous permet d'évaluer l'interprétation des données et la nature des associations entre les éléments organisés en scripts.

Le TCS nous permet de comparer le niveau de concordance entre les scripts des personnes évaluées et ceux d'un panel d'experts de référence dans des problèmes de structures complexes

ou « mal définis⁴ ». Une des particularités de ce test, par rapport aux questions à choix multiples, est que plus d'une seule bonne réponse à la question est acceptée (Fournier, Demeester, & Charlin, 2008). Ainsi, pour être utilisées en situation d'évaluation, les réponses des experts devront avoir une certaine variabilité et, en fonction du nombre de réponses des experts du panel, un certain nombre de points sera attribué à la réponse. La réponse choisie par la majorité des panélistes se verra attribuer le score de 1, tandis que les autres choix se verront attribuer une proportion de la réponse modale, en fonction du nombre d'experts qui ont choisi cette réponse. Une question qui n'aurait aucune variabilité dans les réponses serait rejetée, puisqu'elle ne pourrait être qualifiée de question représentant un problème à structure complexe. Il existe dans la littérature, plusieurs variations dans la façon de calculer les scores au TCS, nous y reviendrons dans le chapitre sur la méthodologie.

Les vignettes utilisées dans le TCS sont bâties de façon à nous permettre d'estimer la capacité à interpréter une information clinique dans un contexte authentique (ou une facette de celui-ci). Bien qu'elles ne puissent prétendre à reproduire la complexité de la pratique en situation réelle, elles devraient représenter une situation réaliste de pratique clinique. Chaque question associée à une vignette ne vise pas à évaluer le résultat du raisonnement clinique, mais chacune des microdécisions qui vont permettre d'en arriver au résultat (Charlin, Desaulniers, Gagnon, Blouin, & van der Vleuten, 2002). De plus, celles-ci sont bâties de façon à ne pas pouvoir se répondre par simple rappel de connaissances (Lubarsky *et al.*, 2011).

Même si le TCS a été utilisé principalement pour évaluer les aspects diagnostiques, d'investigation et thérapeutique, certaines études l'ont utilisé au niveau d'aspects tels que l'éthique (Tsai, Chen, & Lei, 2012) et le *caring* (Deschênes *et al.*, 2011). De plus, cette forme de test semble appropriée pour évaluer le raisonnement clinique à différents moments de la formation, du préclinique (Humbert *et al.*, 2011) à la résidence (Lubarsky, Chalk, Kazitani, Gagnon, & Charlin, 2009), avec de possibles applications en évaluations des cliniciens en exercice (Goulet, Jacques, Gagnon, Charlin, & Shabah, 2010) et en formation continue, en

⁴ Les problèmes « mal définis » nécessiteraient des connaissances et habiletés différentes des problèmes « bien définis » et seraient plus représentatifs de la pratique professionnelle selon Schön (1983).

ajustant bien sûr le niveau des questions. Par contre, le test a été peu utilisé jusqu'à maintenant dans le contexte de la physiothérapie, seulement une étude a été répertoriée dans ce domaine de recherche (Cohen *et al.*, 2009).

Sur le plan des avantages, tout comme les questions QCM, le TCS permet de diminuer le problème de la spécificité de contenu en évaluation, en permettant d'inclure un nombre important de questions ayant des contextes différents. Il requiert relativement peu de ressources pour être bâti et corrigé. Il peut être utilisé également pour donner une rétroaction formative (Humbert *et al.*, 2011). Par contre, il est nécessaire de faire appel à un panel de référence pour établir les scores, ce qui rend le processus plus complexe, mais lui confère plus de crédibilité.

1.3.6 Problèmes de raisonnement clinique (*Clinical reasoning problem*)

Les problèmes de raisonnement clinique (PRC) sont une modalité d'évaluation du raisonnement clinique apparue dans la littérature au début des années 2000. Ce type de test propose des vignettes cliniques avec un scénario réaliste qui comportent des informations sur l'histoire et sur les résultats de l'examen physique d'un patient (Groves, Scott, & Alexander, 2002). À partir de ce scénario, il est demandé d'émettre des hypothèses diagnostiques et de relever les éléments qui appuient ou questionnent ces hypothèses, tout en donnant un poids relatif à ces éléments dans le diagnostic. Les PRC⁵, tout comme le TCS, incluent des vignettes qui sont développées par un groupe d'experts et font appel à un autre groupe d'expert pour bâtir l'échelle de correction. Par contre, à la différence du TCS, les PRC utilisent des réponses construites plutôt que des choix de réponse et offrent une grande quantité d'information à analyser avant qu'il ne soit demandé au candidat de poser un jugement. Le TCS, quant à lui, évalue l'effet d'une courte information sur une décision. Malgré ces différentes particularités, les deux tests aspirent à nous permettre d'évaluer le processus plutôt que le seul résultat du raisonnement clinique (Groves *et al.*, 2002). De plus, les deux tests associent les informations données à un vecteur comportant un poids relatif pour appuyer ou infirmer une hypothèse (Schuwirth *et al.*, 2001).

⁵ Dans le texte le pluriel sera utilisé pour les PRC et le singulier pour le TCS.

De par sa conception, le score des PRC sera influencé par la quantité d'information pertinente identifiée et correctement interprétée, ce qui a tendance à récompenser l'exhaustivité de la démarche. Or, il est généralement reconnu que celle-ci n'est pas toujours un synonyme de meilleur résultat en raisonnement clinique (Charlin et al., 2003; Page & Bordage, 1995; van der Vleuten *et al.*, 2008), les experts étant en mesure de reconnaître des modèles types (*pattern recognition*) efficacement avec peu d'éléments cliniques identifiés de façon explicite. Il peut toutefois être argumenté que la non-reconnaissance ou la mauvaise interprétation des données cliniques se révèle être une source importante d'erreurs diagnostiques (Groves, O'Rourke, & Alexander, 2003b; Norman & Eva, 2010), d'où l'importance d'évaluer la reconnaissance et l'interprétation des données. En ce qui a trait à la rédaction des questions PRC et de leurs validations, le processus⁶ semble comparable aux TCS. Cependant, en lien avec l'utilisation de réponses construites, le temps de correction des PRC tel que décrit dans Groves *et al.* (2002) serait plus long que le TCS.

Jusqu'à maintenant les PRC ont été utilisés au niveau du diagnostic seulement, mais il serait probablement possible de les utiliser pour d'autres aspects du raisonnement clinique comme le pronostic par exemple. Globalement, cet outil semble présenter un potentiel intéressant pour la physiothérapie en incorporant un type de question à réponses ouvertes qui dans certains cas est plus approprié que la réponse à choix multiples pour évaluer le raisonnement clinique. D'ailleurs, une forme similaire de cet outil d'évaluation est utilisée par plusieurs programmes de physiothérapie dans le monde, au niveau formatif (Jones, 2008).

La prochaine partie de cette section présentera l'utilisation de situation et de contexte où le raisonnement clinique peut être évalué dans le cadre de tâches professionnelles. Ces situations demandent à l'étudiant d'intégrer les différentes ressources nécessaires à la prise en charge partielle ou complète d'un patient réel ou simulé.

⁶ Sera discuté dans le chapitre de méthodologie.

1.3.7 Les examens cliniques objectifs structurés (ECOS)

Les examens cliniques objectifs structurés (ECOS) sont des situations fréquemment employées pour évaluer les habiletés cliniques en situation simulée et en contexte académique. Ce type d'examen comprend en général plusieurs stations (en général plus de 10), de courte durée (10 à 20 minutes), et qui incluent un scénario clinique impliquant un patient simulé. Tous les étudiants sont exposés aux mêmes situations et évalués à l'aide d'une grille d'observation spécifique à la station. Bien que très coûteux en ressources, ce type d'évaluation est courant dans les sciences de la santé et présente, en général, une bonne fidélité et validité (Pell, Fuller, Homer, & Roberts, 2010). Malgré les ressources investies dans l'ECOS, celui-ci est peu utilisé pour évaluer le raisonnement clinique des candidats (Charlin *et al.*, 2003; Page & Bordage, 1995; van der Vleuten *et al.*, 2008). Les réticences exprimées proviennent du fait que l'ECOS ne permet d'évaluer que des comportements observables, ce qui n'est souvent pas le cas avec le raisonnement clinique dans une situation simulée où seul le résultat du processus reste accessible.

1.3.8 L'évaluation du raisonnement clinique en contexte de stage

À travers le monde, les situations de stage font partie intégrante de la formation du physiothérapeute (Lekkas *et al.*, 2007). Les stages constituent en quelque sorte une réminiscence de la formation de l'apprenti qui a été la base de la formation en sciences de la santé pendant des décennies (Cox *et al.*, 2006). Durant cette période de formation, le stagiaire apprend et est évalué dans des situations authentiques caractéristiques de sa future profession. Le fait que l'évaluation et l'apprentissage en stage sont intimement imbriqués comporte certains avantages et désavantages en lien avec la polarité des concepts d'évaluation formative et sommative (Tanggaard & Elmholdt, 2008). L'étudiant étant en général plusieurs semaines avec le même superviseur, cette longue période d'observation permet au superviseur de voir la progression de l'étudiant et non seulement sa performance instantanée. Par contre, le fait d'être évalué pendant son apprentissage peut rendre l'étudiant craintif de faire des erreurs, au point d'être inhibé dans son apprentissage et de ne pas tirer le maximum de la situation d'apprentissage. Ce sentiment d'être constamment évalué par le superviseur peut être un frein à poser des questions et ainsi exhiber son manque de connaissance (Scallan, 2000).

Le danger potentiel de combiner évaluation formative et sommative par le même évaluateur a été souligné par plusieurs auteurs (Black, 1995; Crooks, 1988; Tan, 1992), mais il demeure que ce modèle est le plus commun en sciences de la santé présentement. Dans l'ensemble des programmes de formation en sciences de la santé, il est nécessaire de réussir ces stages pour graduer et avoir accès à la pratique professionnelle. Par définition les stages constituent un moment important pour évaluer les compétences des étudiants en physiothérapie ainsi que le raisonnement clinique qui en est une composante majeure (Dalton, Davidson, & Keating, 2011).

Les outils utilisés en contexte de stage encadrent le jugement global d'un expert lors d'une évaluation formative ou sommative de la performance d'un étudiant en situation clinique (Kogan *et al.*, 2009). Des outils tels que le mini CEX en médecine ont été développés pour permettre l'évaluation de compétences variées tels la communication ou le professionnalisme (Pelgrim *et al.*, 2011) et donner une rétroaction à l'étudiant.

En physiothérapie, les outils d'évaluation en contexte de stage sont en général bâtis à partir du référentiel de compétence du programme universitaire de formation ou des organismes de régulation de la profession (Dalton *et al.*, 2011; Roach *et al.*, 2012). Comme dans le cas du mini CEX, ces outils incorporent des aspects quantitatifs et qualitatifs de la performance de l'étudiant évaluée par le jugement du superviseur de stage (Roach *et al.*, 2012). Par exemple dans plusieurs de ces outils, le superviseur doit établir le niveau de l'étudiant sur une échelle ordinale et soutenir son jugement par des observations sur les performances de l'étudiant évalué.

Ces outils d'observation en situation clinique authentique incorporent le raisonnement clinique comme une des composantes de la performance, mais sans toutefois essayer de décortiquer les processus impliqués (Charlin *et al.*, 2003). Bien qu'elle comporte des avantages, cette approche globale peut faire en sorte de laisser passer des difficultés importantes sur le plan du raisonnement clinique du stagiaire, sans pouvoir clairement identifier où se situe le problème (Audétat, Laurin, & Sanche, 2011). Même si les superviseurs ont en général une idée intuitive assez rapide des difficultés de raisonnement clinique de leurs stagiaires, ceux-ci éprouvent de la difficulté à poser un diagnostic pédagogique précis à ce niveau et à expliciter la façon dont ils effectuent cette tâche (Audétat, Faguy, Jacques, Blais, & Charlin, 2011; Jette *et al.*, 2007).

1.4 Questions de recherche

Bien que la physiothérapie partage plusieurs caractéristiques communes avec la médecine, elle se distingue aussi par plusieurs différences au niveau du modèle de pratique et de la relation avec le patient (Edwards *et al.*, 2004; Eisenberg, 2012). Les outils d'évaluation du raisonnement clinique décrits dans la littérature ont été utilisés en grande majorité dans un contexte médical. En ce sens, il est pertinent de se questionner sur leur applicabilité dans le contexte spécifique de la physiothérapie.

La plupart de ces outils d'évaluation du raisonnement clinique développés en médecine sont de type papier crayon, ceux-ci possèdent plusieurs avantages, entre autres, la possibilité d'évaluer en peu de temps plusieurs situations de pratique (Charlin *et al.*, 2003). Par définition, ces situations de pratique sont contextualisées et pourraient possiblement nous permettre d'inférer plus judicieusement sur le niveau de compétence des étudiants en physiothérapie que des tests qui évaluent seulement des connaissances factuelles décontextualisées. Il est de plus possible que certains de ces tests puissent être complémentaires et puissent nous permettre d'inférer sur des dimensions ou aspects différents du raisonnement clinique (Dory, Gagnon, Vanpee, & Charlin, 2012) d'où l'intérêt d'utiliser plus d'un outil d'évaluation. Ces outils pourraient ainsi nous permettre d'évaluer les ressources et le jugement de façon intégrée dans les différentes familles de situations tel que préconisée par l'approche par compétence (Scallon, 2004) tout en étant complémentaire à l'évaluation plus globale des situations professionnelles authentiques en contexte de stage.

De tous les instruments disponibles dans la littérature médicale pouvant permettre de poser un jugement sur le raisonnement clinique, le TCS et le PRC présentent, selon notre appréciation de la littérature sur le sujet, un bon potentiel pour étudier ces questions dans le contexte de la physiothérapie. Nous croyons, de plus, que ces deux instruments présentent une compatibilité conceptuelle par l'utilisation d'un panel d'experts pour attribuer les scores et une certaine complémentarité dans les processus de raisonnement clinique utilisés. En effet, l'utilisation du PRC implique d'émettre une hypothèse, une étape importante du raisonnement clinique utilisée par le clinicien en situation authentique. Dans le cas du TCS où l'hypothèse est fournie, un des

avantages réside dans le format qui permet d'avoir recours à un grand nombre de contextes différents. Nous discuterons, plus en détails des qualités de ces instruments dans le chapitre sur la revue de littérature des instruments d'évaluation du raisonnement clinique choisis (chapitre 3).

Jusqu'à maintenant peu d'études ont relié les résultats des tests de raisonnement clinique aux performances des étudiants en situation clinique authentique. Étant donné qu'il semble y avoir un consensus sur le fait que le raisonnement clinique est central à la compétence clinique du physiothérapeute, si les outils d'évaluation papier crayon évaluent bel et bien ce construit, ils devraient être reliés avec la performance clinique observée en milieu authentique. De plus, si cette association se précise, elle pourrait possiblement nous permettre d'identifier précocement les étudiants étant à risque d'éprouver des difficultés de raisonnement clinique en milieu de stage pour pouvoir leur offrir une remédiation appropriée.

Nous croyons donc qu'il est pertinent scientifiquement et socialement d'étudier les outils d'évaluation du raisonnement clinique tel que le TCS et les PRC chez les étudiants en physiothérapie pour en déterminer le potentiel d'utilisation dans ce domaine. Pour ces raisons, dans le cadre de cette recherche nous posons les questions de recherches suivantes :

Question générale de recherche

Est-ce que le test de concordance des scripts (TCS) et les problèmes de raisonnement cliniques (PRC) développés en contexte médical permettent d'évaluer le raisonnement clinique en physiothérapie?

Questions spécifiques de recherche

- 1 Est-ce que le test de concordance des scripts (TCS) et les problèmes de raisonnement cliniques (PRC) peuvent nous permettre de distinguer le niveau de développement du raisonnement clinique des étudiants universitaires en physiothérapie?
- 2 Est-ce qu'il y a une relation entre les résultats aux tests de raisonnement clinique (TCS, PRC) et les résultats de stage?

Le prochain chapitre abordera le cadre conceptuel qui soutiendra l'ensemble du projet de recherche et nous permettra d'en analyser les résultats.

CHAPITRE 2

CADRE CONCEPTUEL

Le sujet d'étude de cette recherche doctorale est l'évaluation du raisonnement clinique en physiothérapie. Dans le but d'expliquer les choix conceptuels qui permettront de comprendre les enjeux du projet, le présent chapitre comportera trois parties. La première introduira le construit du raisonnement clinique en comparant les termes rattachés conceptuellement au raisonnement clinique, les différents modèles, leur évolution historique et les différentes dimensions du cadre conceptuel retenues pour l'évaluation du raisonnement clinique. La deuxième partie pour sa part, traitera de l'évaluation du raisonnement clinique dans le contexte plus global du développement des compétences cliniques en sciences de la santé. La troisième partie traitera quant à elle, de validité et de processus de validation des tests et de son application au contexte du raisonnement clinique. Pour chacune des parties, nous tenterons d'expliquer les raisons qui ont mené à certains choix et leurs implications.

2.1 Le construit du raisonnement clinique

Cette première partie du cadre conceptuel tentera de clarifier les différents termes reliés au raisonnement clinique en abordant certaines différences pouvant exister au sein de cette littérature dans les sciences de la santé. Nous présenterons également les modèles de raisonnement clinique les plus connus et leurs évolutions, bien qu'aucun de ces modèles ne fasse l'unanimité présentement dans la littérature. Les particularités de chacun de ces modèles seront exposées pour mieux comprendre leur apport à notre compréhension du raisonnement clinique. De plus, nous aborderons l'influence du paradigme interprétatif dans la conception du raisonnement clinique de certaines professions de la santé incluant la physiothérapie. Le contexte en tant que partie indissociable du processus de raisonnement clinique sera également discuté particulièrement en considérant son importance dans l'évaluation de celui-ci. Finalement, dans la dernière section de cette partie nous essayerons de définir le lien entre le raisonnement clinique et la compétence clinique.

2.1.1 Comparaison des différentes définitions et termes utilisés en lien avec le raisonnement clinique

Plusieurs termes ont été utilisés historiquement pour décrire les processus cognitifs utilisés par les cliniciens dans leurs démarches avec les patients. Un de ces termes est la prise de décision clinique qui implique la considération de norme ou de modèle statistique dans le choix des actions (Elstein & Schwartz, 2002). Ce terme est aussi associé à une approche de recherche basée sur les notions du théorème de Bayes. Ce modèle analytique décrit que, dans une situation typique, un clinicien aura une estimation *a priori* de la probabilité d'un diagnostic ou d'un pronostic et que par la suite, cette probabilité variera en fonction des informations recueillies auprès du patient. Cette approche préconisée par les tenants de l'approche de la pratique basée sur les données probantes (Sackett, Strauss, & Richardson, 2000) n'est généralement pas utilisée de façon spontanée par les cliniciens. Il appert également que les experts ne se distinguent pas par un meilleur raisonnement probabiliste que les novices (Tversky & Kahnemann, 1982). Les informations provenant, par exemple, des études sur les qualités psychométriques des tests diagnostiques ou des essais cliniques thérapeutiques figurent au nombre des connaissances qu'un clinicien va utiliser, mais celles-ci ne sont probablement pas combinées de façon mathématique afin de prendre une décision (van der Vleuten *et al.*, 2010).

Le terme raisonnement diagnostique est un autre terme fréquemment employé, mais bien qu'un grand nombre d'études aient abordé le raisonnement sous l'angle du diagnostic, ce terme n'implique qu'une des actions du clinicien qui malgré son importance ne reflète pas l'ensemble des processus utilisés (Norman, 2005).

Le terme jugement clinique est également rencontré dans la littérature et une tentative de le différencier avec le raisonnement clinique a été faite en sciences infirmières (Victor-Chmil, 2013) en y soulignant les aspects psychomoteurs et affectifs de la tâche. Cette définition se base sur les travaux de recherche de Tanner (2006), mais ne semble pas faire consensus, puisque le terme jugement clinique est peu utilisé dans la littérature des sciences de la santé autres que les sciences infirmières.

Il nous apparaît clair que le raisonnement clinique est le terme qui apparaît le plus fréquemment dans les écrits scientifiques pour décrire l'ensemble des processus utilisés par les cliniciens pour aborder les problèmes de santé de leurs patients (Fleming, 1991; Gruppen & Frohna, 2002; Higgs, Jones, Loftus, & Christensen, 2008; Holdar, Wallin, & Heiwe, 2013; Simmons, 2010).

Dans le cadre de ce projet, nous retenons donc le terme de raisonnement clinique et la définition qui en est faite de Higgs et Jones (2008). Celle-ci nous apparaît comme assez générique pour englober la plupart des éléments importants reliés au raisonnement clinique dans la littérature et assez précise pour bien situer le concept. « Celui-ci peut être défini comme l'ensemble des processus de pensée et de prises de décisions associés à la pratique clinique, qui permettent au clinicien de réaliser la meilleure action possible dans un contexte donné. » (traduction libre de Higgs & Jones, 2008, p.4)⁷. Nous décrirons dans la prochaine partie de ce chapitre les différents modèles de raisonnement clinique et leurs évolutions historiques de même que les dimensions retenues pour notre cadre conceptuel qui cible l'évaluation.

2.1.2 Comparaison des différents modèles du raisonnement clinique et leurs évolutions

Un des premiers modèles de raisonnement clinique qui est apparu dans la littérature est celui de Elstein (Elstein *et al.*, 1978) et il est caractérisé par l'émergence des processus hypothético-déductifs. Selon ce modèle, les hypothèses apparaissent très tôt à l'esprit du clinicien et un nombre limité d'hypothèses émises sert de cadre pour recueillir des données supplémentaires. Les données cliniques ainsi recueillies sont ensuite interprétées pour déterminer leurs compatibilités avec l'hypothèse testée. Bien que la qualité des hypothèses soit supérieure au niveau des experts, ce processus serait à la fois utilisé par les novices et les experts, de façon semblable (Nendaz, Charlin, Leblanc, & Bordage, 2005). Ce constat remettait en doute la volonté initiale de comprendre les processus cognitifs de l'expert pour ensuite pouvoir les enseigner aux novices. De plus, la spécificité de contenu, c'est-à-dire la faible corrélation entre la capacité de résoudre un problème d'un cas à l'autre, démontre que le raisonnement clinique n'est pas une habileté générique (Norman, 2005).

⁷ Voir annexe 10 pour texte original

Par la suite les modèles de raisonnement clinique ont graduellement incorporé le processus de reconnaissance de patron (*pattern recognition*) (Patel & Groen, 1986) inspiré des recherches en psychologie cognitive. Ce processus est habituellement très rapide et est souvent associé à l'intuition, bien qu'il fasse appel à une construction de connaissances acquises.

Cette prise de décision à partir de patrons se fait principalement dans les cas non complexes (en fonction du niveau d'expertise) associés à des situations cliniques qui sont présentes dans la mémoire du praticien. Ce processus serait plus facilement influencé, consciemment ou inconsciemment, par le contexte et l'état émotif du clinicien (Pelaccia, Tardif, Triby, & Charlin, 2011). Même si la reconnaissance de patron est aussi utilisée par les novices, il est en général plus utilisé chez les experts, car relié à une organisation plus riche des différents types de connaissances (Nendaz *et al.*, 2005). Ce processus qui s'effectue avec beaucoup d'efficacité chez l'expert devient rapidement automatisé au point qu'il est souvent difficile pour celui-ci de préciser le parcours de sa pensée (Charlin, 2011). Malgré le fait que certains auteurs craignent que l'utilisation de ce processus conduise à négliger les hypothèses alternatives (fermeture prématurée) à la suite de la rapidité des décisions prises (Higgs, Jones, Loftus, *et al.*, 2008) il n'y a cependant pas d'évidence empirique claire que l'utilisation de ce système provoque fréquemment des erreurs (Norman & Eva, 2010).

Bien que les deux processus décrits précédemment soient considérés distincts, la littérature considère qu'ils coexistent de façons intégrées dans le raisonnement clinique courant (Eva, 2005; Pelaccia *et al.*, 2011). La première étape étant la reconnaissance non analytique d'un patron, suivi d'une confirmation hypothético-déductive, lorsque nécessaire (Eva, 2005). Considérant cette évolution de la littérature, les modèles de raisonnement clinique ont graduellement incorporé cette interaction entre les deux processus qu'implique le *dual process theory* (Marcum, 2012; Nendaz *et al.*, 2005) avec des processus d'activation automatique d'hypothèses (appelé système 1) et des processus de vérification délibérés (système 2). Ces deux stratégies cognitives seraient complémentaires et interactives pour les novices et les experts (Eva, 2005). L'interaction et le recours optimal à ces deux processus dépendront de plusieurs facteurs, dont la complexité, le temps dont dispose le clinicien, le caractère incertain et les enjeux reliés au cas (Pelaccia *et al.*, 2011). Pour que ces deux processus fonctionnent de

façon optimale, le clinicien doit être en mesure de savoir quand passer de l'un à l'autre (Mamede, Schmidt, Rikers, Penaforte, & Coelho-Filho, 2007). Ce que Schön identifie comme la réflexivité en action doit permettre au clinicien de déterminer quand l'action automatisée du processus non analytique n'est pas suffisante parce que le problème est non typique ou mal défini (Schön, 1983).

Selon Moulton (2007), la possibilité de recourir à ces deux systèmes en temps opportun serait une caractéristique importante de l'expertise de haut niveau. Cette capacité de « ralentir » les processus cognitifs au besoin serait associée à un système de contrôle métacognitif qui permet de réajuster lorsque le patron clinique reconnu comporte des éléments discordants (Moulton, Regehr, Mylopoulos, & MacRae, 2007). La métacognition est d'ailleurs une des dimensions qui apparaît dans plusieurs modèles de raisonnement clinique (Chamberland, 2006; Charlin *et al.*, 2012; Higgs, Jones, Loftus, *et al.*, 2008). Celle-ci est définie comme étant l'action de poser un regard critique sur son raisonnement clinique et de tirer profit de ses expériences pour améliorer ses performances (Mamede et Schmidt, 2004).

Tout comme le processus de résolution de problèmes, la représentation du problème est associée à la dimension cognitive du raisonnement clinique. Par exemple, les modèles de Chamberland (2006) et Charlin (2012) incluent le processus de résolution et de représentation du problème, comme des portions distinctes, mais en interrelation étroite dans leurs modèles de raisonnement clinique. Dans notre contexte, la représentation du problème est définie comme étant une structure cognitive temporaire résultant de la combinaison des connaissances antérieures des étudiants et des informations relatives aux problèmes de santé d'un patient. (Gruppen et Frohna, 2002). La représentation du problème est considérée comme étant cruciale au raisonnement clinique, car elle en constitue le point de départ qui guide la suite des actions. Elle serait une nécessité cognitive, qui permet au clinicien de pouvoir réduire le problème à un niveau où la quantité d'information peut être gérée (Nendaz *et al.*, 2005). À partir de cette représentation du problème peut débuter la résolution de celui-ci.

Les connaissances représentent également une dimension présente dans la plupart des modèles de raisonnement clinique décrits dans la littérature (Chamberland, 2006; Charlin *et al.*, 2012; Higgs, Jones, Loftus, *et al.*, 2008). Tout comme la cognition et la métacognition, les

connaissances sont en interactivité dans les processus de raisonnement clinique. Certains modèles proposent un découpage des types de connaissance qui peuvent être utiles dans un cadre d'analyse. Comme dans le modèle proposé par Charlin (2012), nous incluons trois types de connaissances dans notre cadre conceptuel, les connaissances biomédicale, clinique et expérientielle.

Dans le domaine des sciences de la santé le terme connaissance biomédicale (par exemple, l'anatomie et la physiologie humaine) fait référence à ces ressources qui sont apprises plus souvent en début de curriculum et qui servent de base à la compréhension des phénomènes observés en clinique (Woods, 2007). L'importance de ces connaissances dans le raisonnement clinique a fait l'objet de plusieurs débats dans la littérature, certaines études ayant démontré que les experts utilisaient moins de ce type de connaissances pour résoudre des problèmes cliniques que des novices (Patel & Groen, 1986; Patel, Groen, & Arocha, 1990). Par contre certains auteurs ont démontré que l'apprentissage des connaissances biomédicales est important à plusieurs niveaux, par exemple au niveau de la précision diagnostique (Woods *et al.*, 2006) particulièrement lorsque les cas cliniques sont complexes (Woods, Brooks, & Norman, 2007).

Le concept d'encapsulation des connaissances biomédicales (Boshuizen & Schmidt, 1992) permet de comprendre l'importance relative de celles-ci du niveau novice au niveau expert. Au début de sa formation, un étudiant en science de la santé a des connaissances organisées principalement selon des notions théoriques biomédicales, qui graduellement deviendront organisées selon des réseaux ou scripts cliniques en fonction de leur exposition clinique à des patients réels. Les informations biomédicales, quant à elles, seront graduellement encapsulées, c'est-à-dire organisées en un nombre plus restreint de concepts englobants.

Les connaissances cliniques peuvent être définies comme les connaissances relatives aux signes et aux symptômes de même que leurs relations avec les maladies (Gruppen et Frohna, 2002). Ces connaissances souvent de nature sémiologique contrastent avec les connaissances biomédicales qui mettent l'accent sur les mécanismes causaux (Patel & Kaufman, 2000). Il est généralement admis que les connaissances cliniques doivent être en interaction avec les connaissances biomédicales dans le raisonnement clinique (Gruppen & Frohna, 2002; Norman,

2005) et que celles-ci ne représentent pas deux entités qui fonctionnent en parallèle (Rikers, Schmidt, & Moulaert, 2005).

Les connaissances expérientielles peuvent être définies, quant à elles, comme les connaissances accumulées au fil des années de formation ou de pratique (Norman, 2005). Ce type de connaissances permet de créer un répertoire de cas réels qui va être à la base de la création de scripts. Ainsi, il a été démontré que de pouvoir reconnaître des attributs observés précédemment chez des patients peut améliorer la précision du diagnostic (Brooks, LeBlanc, & Norman, 2000; Norman, Brooks, Regehr, Marriott, & Shali, 1996). Les patients rencontrés au cours de la pratique forment des exemples (cas concrets) à l'intérieur d'un type de maladie.

Pour pouvoir être utiles au raisonnement clinique, les différents types de connaissances doivent être organisés et facilement accessibles. Il est généralement reconnu que les experts auraient une organisation des connaissances plus élaborées qui s'est bâtie en fonction des connaissances acquises et des expériences cliniques. Les données cliniques d'un patient permettent ainsi une activation efficace de ces connaissances, de façon à s'adapter à la situation (Schmidt, Norman, & Boshuizen, 1990). Plusieurs modèles différents d'organisation des connaissances décrits dans la littérature tentent d'expliquer comment le clinicien peut gérer une quantité importante d'information avec des capacités cognitives somme toute limitées. Les scripts sont un de ces modèles d'organisation des connaissances en fonction de l'action qui provient de la psychologie cognitive et qui décrit les architectures de connaissances spécifiques associées à certaines situations (Charlin, Boshuizen, Custers, & Feltovich, 2007). Ces représentations cognitives sont apprises et gardées en mémoire en fonction de contextes spécifiques. Par exemple, en situation de repas au restaurant, un serveur qui amène un menu est une situation qui va activer un script particulier, et pour lequel la plupart des Occidentaux sont en mesure de prévoir la séquence des événements qui vont suivre. Cette organisation des connaissances s'est bâtie graduellement au fil d'expériences similaires et comprend des automatismes qui nécessitent un minimum d'effort pour accomplir la tâche. Le script étant contextualisé, cette même séquence dans un restaurant d'une culture étrangère pourrait activer un script inapproprié pour la circonstance.

Transposé au domaine des sciences de la santé, ce modèle soutient que ce réseau de connaissances organisées, le script, permet au clinicien de mieux comprendre la situation du

patient qui est devant lui et pour lequel des décisions devront être prises (Schmidt *et al.*, 1990). En fonction du type de connaissances organisées, on peut parler de script diagnostique, d'investigation ou thérapeutique (Charlin, Boshuizen, *et al.*, 2007).

2.1.3 Influence du paradigme interprétatif sur les modèles de raisonnement clinique

La majorité des écrits sur le raisonnement clinique en sciences de la santé définit celui-ci dans le cadre d'un modèle cognitif centré sur le clinicien. Ce modèle s'inscrit dans l'approche biomédicale traditionnelle selon laquelle le clinicien détermine ce qui doit être fait pour améliorer la condition de santé du patient. Par contre, on peut questionner la pertinence de ce modèle dans un contexte de changement dans les systèmes de santé où le patient occupe une large place et où la prise de décision thérapeutique est réalisée conjointement avec celui-ci, où de nombreuses occasions de travail interdisciplinaire sont nécessaires et où les ressources du milieu vont fortement influencer les décisions. Dans ce contexte, plusieurs auteurs dans les domaines des sciences autres que médecine (Edwards *et al.*, 2004; Loftus & Smith, 2008), jugent pertinent de redéfinir le raisonnement clinique dans un paradigme interprétatif pour permettre de le concevoir comme une activité socioculturelle construite avec de multiples facettes et interprétations possibles. Selon ce paradigme, le raisonnement clinique est perçu comme une activité humaine qui n'est jamais complète et où, par définition, la recherche d'une seule vérité est incompatible. Ce paradigme permet au raisonnement clinique de s'inscrire dans une approche centrée sur le patient en donnant une plus grande place à la perspective des patients et des autres intervenants impliqués dans les actions prises. Selon Higgs *et al.* (2008), la complexité des interrelations entre les différents aspects du contexte inscrits dans le raisonnement clinique rend difficile son découpage en petites portions et peut induire des conclusions artificielles et réductrices (Higgs, Jones, & Titchen, 2008). Une forme de raisonnement souvent associée à une vision interprétative est le raisonnement narratif qui cherche à comprendre l'expérience du problème de santé du patient à travers sa conception personnelle. Selon Edwards *et al.* (2004), cette exploration de la perspective du patient n'est pas compatible avec une vision empirico-analytique, puisque l'expérience du patient ne peut être validée autrement qu'en établissant avec celui-ci un consensus sur le sens du problème de santé. La recherche de lien de causalité ou la formulation d'hypothèses n'est pas l'objet de ce type de

raisonnement; il s'agira davantage d'établir une compréhension commune de la situation vécue par le patient.

Bien que plusieurs recherches sur le raisonnement clinique en sciences de la santé (Edwards *et al.*, 2004; Fleming, 1991; Jensen, Gwyer, & Shepard, 2000) aient été réalisées dans une optique interprétative, selon Loftus et Smith (2008) cette vision viserait à compléter plutôt qu'à remplacer le paradigme empirico-analytique. Dans cette optique de complémentarité, Edwards *et al.* (2004), dans une étude de cas multiples impliquant des physiothérapeutes en orthopédie, neurologie et soins à domicile, a mis en évidence des processus associés aux deux paradigmes dans le raisonnement clinique de ces professionnels.

Ces auteurs mentionnent que la vision différente et quelquefois contradictoire que sous-tendent ces deux processus semble se concilier dans le raisonnement clinique des physiothérapeutes dans les différents milieux de pratiques (Edwards *et al.*, 2004). En s'appuyant sur le résultat de cette étude, Edwards *et al.* (2004) proposent que le mode de raisonnement des physiothérapeutes soit dialectique, celui-ci alternant entre des processus diagnostiques (empirico-analytique) et narratifs (interprétatif). En accord avec cette vision Jones et Edwards (2008) ont proposé un modèle où le clinicien et le patient occupe tous deux une place aussi importante tout en demeurant en grande interactivité. Nous retenons donc cet aspect dans le cadre conceptuel de notre projet comme une caractéristique du raisonnement clinique en physiothérapie. Cette particularité de la physiothérapie teinte le regard de la profession de façon différente par rapport à la vision biomédicale traditionnellement rencontrée en médecine.

Les dimensions principales du raisonnement clinique retenues pour notre cadre conceptuel d'évaluation du raisonnement clinique sont donc la cognition, la métacognition et les connaissances. Ces dimensions sont basées sur le modèle de Jones et Edwards (2008) proposé pour la physiothérapie. Ces différentes dimensions vont nous permettre d'encadrer le processus d'écriture des vignettes et des items des tests de raisonnement clinique utilisé dans le cadre de ce projet doctoral. Dans un contexte de stage, la représentation du raisonnement clinique des superviseurs cliniques en physiothérapie tient également compte de ces mêmes dimensions (Thomas, 2014) et influence l'évaluation de celui-ci. Le modèle de Jones et Edwards retenu est par contre peu explicite quant aux composantes de ces dimensions. Nous avons ainsi inclus et

décrit certaines composantes qui seront pertinentes dans notre cadre conceptuel de l'évaluation du raisonnement clinique des physiothérapeutes.

2.1.4 L'importance du contexte sur les dimensions du raisonnement clinique

Un autre des aspects du raisonnement clinique qui a été peu explicité jusqu'à tout récemment est le contexte. Dans les études sur le raisonnement clinique, très souvent le contexte est considéré comme une forme de bruit nuisible à la compréhension des phénomènes (Durning, Artino, Schuwirth, & van der Vleuten, 2013). Durning *et al.* (2013) pour leur part, voient le contexte comme un aspect primordial du raisonnement clinique qu'il faut essayer de mieux comprendre plutôt que de chercher à en faire abstraction, parce que difficilement contrôlable. Inspirée de cette vision, nous définirons le contexte dans notre cadre conceptuel comme étant une partie intégrante de la situation à gérer par le clinicien, il implique des acteurs (patient, clinicien, milieu de soins) à un moment particulier, dans un cadre particulier à propos d'une certaine situation. Il inclut le contexte du clinicien (ex. : son cadre de référence, sa charge de travail, ses expériences, etc.), le contexte du patient (ex. : sa situation socio-économique, ses valeurs, sa compréhension du problème de santé, etc.) de même que le milieu de soins (ex. : le milieu de pratique, les ressources financières et matérielles disponibles, la composition de l'équipe multidisciplinaire, etc.). C'est l'interaction de ces différents aspects qui va déterminer la qualité et le succès de l'intervention du professionnel de la santé et le contexte qui rend chaque situation de résolution de problème unique (Higgs & Jones, 2000). Cet aspect unique de chaque situation rend difficile la production de guide de pratique ou d'algorithme essayant de guider les décisions des cliniciens (Durning, Artino, Pangaro, van der Vleuten, & Schuwirth, 2010) et est probablement relié au peu d'adhérence des cliniciens lorsque ceux-ci existent (Lugtenberg, Zegers-van Schaick, Westert, & Burgers, 2009).

De plus, la difficulté de reconnaître les aspects spécifiques du contexte d'un problème est fréquemment nommée comme une cause d'erreur de raisonnement clinique chez les cliniciens (Norman & Eva, 2010). Comme mentionné précédemment cette spécificité de cas ou de contexte rend la tâche d'évaluation du raisonnement clinique beaucoup plus complexe et doit être prise en considération dans l'évaluation du raisonnement clinique que ce soit avec des outils papier crayon ou en situation de stage.

2.1.5 Le raisonnement clinique en tant que composante de la compétence clinique en sciences de la santé

Le raisonnement clinique est en général considéré comme une des composantes majeures de la compétence du clinicien (Higgs, Jones, Loftus, *et al.*, 2008). La notion de compétence est centrale dans les programmes de formations en sciences de la santé (Frank & Danoff, 2007; Gruppen, Mangrulkar, & Kolars, 2012; Morcke, Dornan, & Eika, 2012), mais celle-ci peut être définie de multiple façon. Dans une revue de littérature sur les définitions de compétence en sciences de la santé, Fernandez (2012) constate que la majorité des définitions de la compétence inclut des composantes d'habiletés et de connaissances et que certaines définitions incluent également les attitudes et les valeurs (professionnalisme, éthique, etc.) bien que cette dernière composante demeure plus difficile à circonscrire à travers les définitions retrouvées dans la littérature (Fernandez *et al.*, 2012). En constatant que les connaissances sont également une dimension du raisonnement clinique, dans tous les modèles de raisonnement clinique répertoriés (Chamberland, 2006; Charlin *et al.*, 2012; Gruppen & Frohna, 2002; Higgs, Jones, Loftus, *et al.*, 2008), nous arbitrons le choix de garder comme une partie intégrante de celui-ci, les connaissances étant à notre avis une partie indissociable du raisonnement particulièrement dans un cadre d'analyse à visée évaluative.

Bien qu'il puisse être tentant d'effectuer un rapprochement entre le concept de compétence et celui de raisonnement clinique compte tenu de la complexité et l'importance de celui-ci dans la pratique en sciences de la santé, à notre connaissance, aucun auteur n'a clairement décrit le raisonnement clinique comme une compétence, plusieurs auteurs le décrivent plutôt comme une composante de la compétence (Charlin, 2003) ou bien une habileté (Fernandez, 2012). Comme décrit par plusieurs auteurs (Le Boterf, 2007; Tardif, 2006) associés à la littérature sur les compétences, celle-ci serait un savoir agir complexe qui nécessite la mobilisation de ressources dans une famille de situations. Dans cette optique, nous croyons que le raisonnement clinique est une partie importante, mais celle-ci ne représente pas la totalité des ressources nécessaires pour actualiser une compétence. Le professionnel compétent doit être en mesure de combiner différentes autres ressources de nature psychomotrice ou interpersonnelle pour bien exécuter les tâches attendues de lui par la société (Fernandez, 2012).

Dans le but d'opérationnaliser notre cadre et de bien différencier le raisonnement clinique des autres habiletés, nous définirons celles-ci comme étant tous les savoir-faire (communication, psychomoteur, etc.) autres que cognitifs, cet aspect étant associé au raisonnement clinique dans notre cadre conceptuel. Cette distinction permet, à notre avis, de refléter la place centrale qu'occupe le raisonnement clinique dans la compétence en sciences de la santé sans négliger les autres aspects.

2.2 Cadre d'évaluation du raisonnement clinique et de la compétence clinique :

Cette deuxième partie du cadre conceptuel présentera un cadre de référence d'évaluation relativement bien connu dans les sciences de la santé, ce cadre pourra nous permettre de bien cerner les cibles pertinentes d'évaluation et les liens entre elles au cours de ce travail doctoral. Nous présenterons également dans cette partie, un cadre de référence qui nous permettra de comprendre l'évolution de la compétence clinique à travers les étapes de son développement.

La pyramide de Miller

Le projet de recherche s'intéressant à l'évaluation du raisonnement clinique en physiothérapie, il convient d'établir un cadre qui nous permettra de le situer dans l'ensemble du processus d'évaluation des apprentissages dans ce domaine. La pyramide de Miller (Miller, 1990) offre ce cadre de référence qui nous permet de mieux comprendre le champ d'application des outils d'évaluation qui nous permettront d'inférer sur la compétence de l'étudiant. Dans ce modèle, le niveau de base comprend les connaissances factuelles (biomédicales et cliniques), alors que le deuxième niveau ajoute le savoir comment, auquel on pourrait ajouter le savoir quand, ce qui est essentiellement la contextualisation des connaissances du premier niveau. Ces deux premiers niveaux demeurent théoriques et sont en général associés à des outils crayons papier. Le troisième niveau de cette pyramide est associé à une démonstration pratique de la performance de l'étudiant. Celle-ci est en général faite en milieu académique sur des patients standardisés (en examen clinique objectif structuré) et implique la maîtrise d'éléments des premiers niveaux pour bien réussir la tâche. Le quatrième et dernier niveau s'appliquera à la démonstration de l'agir en milieu réel sur des patients. Le dernier niveau de la pyramide sous-entend évidemment la maîtrise d'éléments des premiers niveaux dans le but de bien réussir les tâches (van der

Vleuten *et al.*, 2010). Les outils d'évaluation du raisonnement clinique se retrouvent en général aux niveaux 2, 3 et 4 de cette pyramide, le niveau 1 évaluant les connaissances factuelles de façon décontextualisée.

Les outils papier crayon du niveau deux de la pyramide de Miller ont généralement comme limitation de ne donner accès qu'au résultat du processus de raisonnement clinique sans nécessairement permettre de déterminer les processus utilisés par les étudiants (Norman, 2005). Cette limitation est bien sûr aussi présente dans beaucoup d'outils sommatifs reliés à l'évaluation des apprentissages (Scallon, 2004).

La pyramide de Miller est un cadre de référence fréquemment utilisé en évaluation des apprentissages dans les sciences de la santé. Par contre, un des aspects de sa description originale et les modifications (Boursicot *et al.*, 2011) qui en ont été faites est controversée. Au départ Miller décrivait le premier niveau de la pyramide comme étant les connaissances, le 2^e niveau étant défini comme la compétence sans élaborer sur ce qu'était la compétence et pourquoi elle était retrouvée à ce niveau. Le troisième niveau est la performance et le 4^e niveau est défini comme l'action (*does*). La relecture qui en est faite par Boursicot *et al.* (2011) définit que le 3^e niveau (*shows how*) de la pyramide évalue la compétence alors que le quatrième niveau (*does*) évalue la performance. Cette distinction entre performance et compétence contraste de façon importante avec la vision de certains auteurs qui voient la performance comme une façon d'inférer sur la compétence, celle-ci étant inobservable (Legendre, 2008). L'article de Boursicot *et al.* (2011) qui semble être un consensus d'un groupe d'experts ne s'appuie sur aucun cadre théorique pour définir ces concepts et nommer de façon différente les comportements observés en situation réelle (4^e niveau) et simulée (3^e niveau). Khan et Ramachandran (2012) questionnent, à juste titre, la nécessité de distinguer ces deux niveaux par des appellations différentes et suggèrent plus simplement l'observation de la performance en milieu simulée (au 3^e niveau) et l'observation de la performance en milieu réel (au 4^e niveau). Ces termes seraient moins susceptibles de créer de la confusion, quel que soit le cadre théorique sous-jacent.

Les différentes dimensions du raisonnement clinique étant toutes interreliées et interdépendantes, la possibilité de les évaluer séparément est difficile et questionnable dans une optique d'évaluation des compétences (Tardif, 2006). Par contre, les résultats des tests papier

crayon de raisonnement clinique que l'on retrouve au 2^e niveau de la pyramide peuvent être vus comme un des multiples indicateurs possibles de la compétence globale de l'étudiant. Dans un programme d'évaluation où l'on veut recourir à plusieurs outils qui vont nous permettre d'évaluer des aspects différents de la compétence de l'étudiant de façon complémentaire, les outils papier crayon ont certainement un rôle à jouer (van der Vleuten *et al.*, 2010). Mais selon les pratiques courantes reliées à ce type de question, celles-ci doivent faire appel à plusieurs éléments de contexte différent (spécificité de contenu) et idéalement être représentatives de la performance de l'étudiant en milieu authentique. Dans cette optique, plusieurs études sur les tests de jugement situationnel (Weekley & Ployhart, 2006), un outil papier crayon originellement développé pour la sélection de candidats en milieu de travail qui présente quelques caractéristiques communes avec le TCS, ont démontré une bonne relation de ces tests avec la performance en milieu authentique. Par exemple, une étude a récemment démontré, dans un contexte de pratique médicale au Royaume-Uni, une bonne capacité de ces tests à prédire la performance en situation réelle de 196 candidats en fin de formation en médecine (Lievens & Patterson, 2011). De surcroît, ces tests de jugement démontraient une valeur additionnelle par rapport aux tests de connaissance.

Bien que le raisonnement clinique ne soit pas nécessairement évalué à tous les niveaux de la pyramide de Miller, ce modèle nous permet de bien cibler le contexte d'évaluation et nous offrir la possibilité d'explorer des liens entre ses différents niveaux. Par exemple, dans le cadre de ce projet doctoral, nous examinerons les liens entre les outils papier crayon du niveau 2 et la performance en situation authentique au niveau 4 de la pyramide.

2.2.2 Perspective développementale de la compétence clinique

Dans le but de clarifier les termes polysémiques de compétence et de niveau de compétence, il importe de bien situer le terme compétence en regard de ce qui est attendu d'un professionnel. Les compétences attendues d'un professionnel de la santé au terme de sa formation initiale sont définies selon les besoins de la société et selon le cadre légal de la profession (Fernandez, 2012). Des organismes réglementaires nationaux ou provinciaux sont en général responsables de l'application de ces normes minimale de compétences. Bien que ce profil de sortie au terme de

la professionnalisation soit explicite, le profil de développement des compétences⁸ qu'acquiert le professionnel en formation est mal connu. Or, la connaissance de ce profil de développement est souhaitable pour bien évaluer une compétence ou composante de compétence tout au long d'un parcours de professionnalisation universitaire (Tardif, 2006). En l'absence de ces modèles de développement spécifiques à chaque compétence et à chaque profession, certains modèles de développement de l'expertise peuvent aider à comprendre l'évolution des étudiants et possiblement à en guider l'évaluation. Par exemple le modèle de Dreyfus (Dreyfus & Dreyfus, 1986) qui a été utilisé dans le domaine de la médecine (Carraccio, Benson, Nixon, & Derstine, 2008; Khan & Ramachandran, 2012; ten Cate, Snell, & Carraccio, 2010) et des sciences infirmières (Benner & Ovion, 1995) peut nous aider à situer le niveau de l'étudiant dans son parcours de formation. Ce modèle décrit cinq niveaux de développement vers l'expertise. Le premier niveau (novice) est caractérisé par un étudiant qui possède des connaissances théoriques de son domaine, mais qui les applique de façon plutôt rigide, celui-ci ne possédant pas d'expérience clinique pour déterminer de l'applicabilité au contexte de ses différentes connaissances. Le novice peut également avoir tendance à centrer son attention sur des faits de moindre importance au détriment des éléments centraux du problème. Celui-ci doit aussi fréquemment utiliser certaines règles et algorithmes pour effectuer ses choix (Benner & Ovion, 1995). Le deuxième niveau décrit dans le modèle de Dreyfus est le niveau intermédiaire (ou débutant avancé), à ce niveau les futurs professionnels commencent à adapter les principes et théories en fonction des situations spécifiques, mais cette application demeure bien souvent rigide et centrée sur la règle plutôt que sur les particularités du cas. Leurs expériences cliniques commencent à leur permettre de voir l'importance de certains aspects du contexte, mais ils éprouvent encore souvent le besoin de valider leurs impressions.

Le niveau suivant est le niveau compétent. À ce niveau, l'individu est capable d'avoir une pratique sécuritaire et efficace. Ce niveau est en général le niveau ciblé d'entrée à la pratique par les organismes de réglementation. Le professionnel est alors en mesure d'interpréter les informations cliniques disponibles pour effectuer des choix réfléchis que ce soit au niveau diagnostique ou thérapeutique (Dreyfus & Dreyfus, 1986). Il est également en mesure d'ajuster

⁸ Modèles cognitifs de l'apprentissage selon Tardif (2006)

son intervention aux besoins spécifiques de chaque patient en tenant compte du contexte. Le professionnel de niveau compétent peut bien faire face à certaines situations plus complexes, mais éprouve encore de la difficulté à adapter son intervention à des situations inattendues. De plus, celui-ci n'a pas la vitesse d'exécution et le répertoire de ressources, par rapport aux individus des niveaux supérieurs (Benner, 1995).

Au niveau avancé, avec quelques années de pratique, le professionnel de la santé possède une meilleure compréhension des situations et fait preuve de plus de flexibilité dans l'application de la théorie. Il est en mesure d'avoir une vue d'ensemble de la situation clinique rendue possible par ses expériences précédentes. Il a recours de plus en plus en à l'intuition dans sa pratique (Dreyfus & Dreyfus, 1986).

Au dernier niveau, celui de l'expertise, l'individu maîtrise parfaitement toutes les compétences de son champ de compétence. Il exécute ses actions professionnelles avec rapidité et efficacité (Patel, Arocha, & Kaufman, 1999) et ses interventions sont ajustées en fonction des caractéristiques uniques du patient. Sa vaste expérience lui permet de différencier rapidement les informations importantes du contexte et considérer le problème de façon globale.

Selon le modèle original de Dreyfus (1986), qui a été au départ décrit dans le contexte des pilotes d'avion, l'expert agit surtout de façon intuitive et automatique, et ce, basé principalement sur des connaissances tacites. Cette vision des processus employés par les experts est qualifiée de simpliste par certains auteurs (English, 1993; Pena, 2010) et ne serait pas entièrement transférable dans le cas de la pratique clinique en sciences de la santé. Car tel que décrit dans littérature sur le raisonnement clinique, un individu utilise en général des processus analytiques et non analytiques de résolution de problème, quel que soit le niveau d'expertise du clinicien (Marcum, 2012). De plus l'automatisation des gestes, telle que décrite par Dreyfus, est non souhaitable et possiblement dangereuse dans le contexte des soins de santé (Pena, 2010).

Bien que controversé sous certains aspects, le modèle de Dreyfus permet de cibler les attentes selon le niveau de l'étudiant et de voir le niveau compétent comme une étape de développement vers l'expertise et non comme l'atteinte de l'objectif ultime. Contrairement à la conception quelquefois dichotomique de la compétence (compétent ou non compétent), celle-ci se

développe et se raffine tout au long de la carrière d'un individu (Scallon, 2004; Tardif, 2006). Au niveau de l'évaluation des compétences (ou des composantes de compétences), par exemple en situation de stages, les attentes doivent être clairement spécifiées selon le niveau de développement de l'étudiant. Cette tâche est bien sûr complexe, mais mérite qu'on la regarde attentivement pour améliorer nos pratiques évaluatives et la formation des professionnels en science de la santé.

La figure 1 schématise les différentes composantes du cadre conceptuel utilisé dans cette recherche à partir des différents modèles présents dans la littérature.

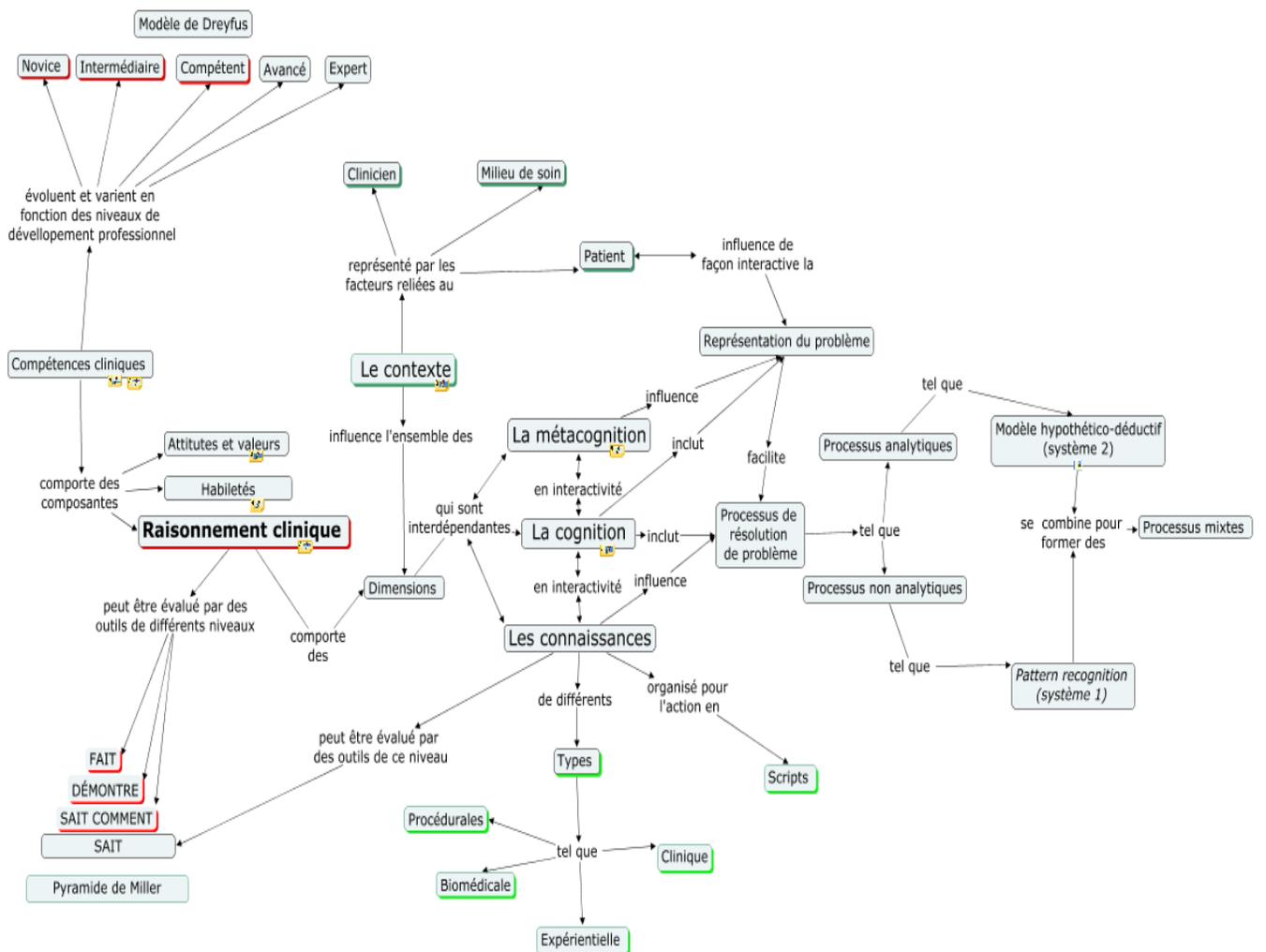


Figure 1 Cadre conceptuel inspiré des modèles de la littérature sur le raisonnement clinique

2.3 Validité et processus de validation des tests de raisonnement clinique

L'objectif de cette recherche nécessite le développement d'outils d'évaluation du raisonnement clinique applicable à la physiothérapie, nous discuterons, dans cette troisième partie, des choix effectués en lien avec le processus de validation de deux outils proposés à cet effet : le TCS et le PRC. Les différents courants de pensée contemporains ayant proposé des cadres de validation comportant des spécificités, nous essayerons de justifier nos choix dans ce domaine.

2.3.1 Évolution du concept de validité

Historiquement, la validité d'un test est liée à sa capacité de bien mesurer ce qu'il est attendu qu'il mesure. Bien que cette définition soit largement répandue, elle est peu explicite au niveau opérationnel. Dans le but de définir de façon opérationnelle cette validité, plusieurs cadres conceptuels ont été proposés. Dans les années cinquante, un comité conjoint formé de membres de *l'American Psychological Association* (APA), de *l'American Educational Research Association* (AERA) et du *National Council on Measurements Used in Education* (NCMUE) (APA, AERA, & NCME, 1954), ont reconnu l'existence de plusieurs types de validité dans le document *Technical Recommendations for Psychological Tests and Diagnostic Techniques*. Cette classification est demeurée présente également dans les publications subséquentes renommé *Standards for Educational and Psychological Tests and Manuals* (American Educational Research Association AERA, NCME, & APA, 1966). Dans ces documents de référence, les différents types de validité étaient définis de la façon suivante :

La validité de contenu (*content validity*) est déterminée par la correspondance entre le contenu du test et le domaine sur lequel on entend tirer des conclusions. Ainsi les items tests doivent être un échantillon représentatif du domaine testé;

La validité de construit (*construct validity*) est déterminée en recherchant quelles sont les informations obtenues à partir du test, de façon à établir dans quelle mesure certains concepts ou certaines hypothèses permettent de comprendre les résultats au test;

La validité de critère (*criterion-related validity*) est déterminée par la comparaison des résultats du test avec une ou plusieurs variables censées fournir une mesure du comportement évalué.

Ce type de validité peut être divisé en deux formes. La validité prédictive (*predictive validity*) étant la capacité du test de prédire un critère externe sur la base de ses scores. Quant au terme validité concomitante (*concurrent validity*) il est utilisé dans le cas où les résultats au test sont comparés au résultat d'un test de référence pour le critère étudié. Ces différents types de validité ont une pertinence en fonction de ce qui est recherché par le test. Par exemple, dans un test utilisé pour l'évaluation des apprentissages, l'accent doit être placé sur la validité de contenu. Par contre, un test de sélection de candidats doit mettre l'accent sur la validité de critère, puisque ce type de test désire faire le meilleur choix en fonction de la performance future des candidats. Selon ces documents qui servaient de guide au processus de validation des tests, il n'était pas attendu que toutes les formes soient nécessairement toutes documentées dans tous les cas.

Ce concept de validité en différentes composantes a graduellement évolué vers la conception unifiée de la validité de Messick (1989) qui a mis l'accent sur le construit et les conséquences de l'interprétation du test. Cette évolution vers un regroupement sous la validité de construit en tant que concept principal a été amorcé, en autres, par Cronbach et Meehl en 1955. Dans leur article, ces auteurs notaient que dans plusieurs situations de tests (de nature psychologique principalement) les arguments sur la nature du contenu et sur les éléments empiriques reliés aux critères n'étaient pas suffisants pour en assurer la validité. Des concepts abstraits (telle que la personnalité) avaient besoin d'un réseau dynamique de différentes variables en interaction entre elles, pour arriver à en saisir l'essence. Ce réseau qualifié de nomologique est en quelque sorte une représentation du concept à étudier qui ne peut être observé directement.

Un autre des courants qui est à l'origine de la conception unifiée de la validité attribuée à Messick (1989) est la volonté d'un regard holistique (ou à tout le moins intégré) du processus de validation des tests. Cette vision partagée par Cronbach est reflétée ici dans un extrait de ces textes :

« La validation d'un instrument de mesure requiert plusieurs types de preuves. Les différentes formes d'investigation ne sont pas des alternatives toutes égales entre elles. Ces différentes investigations se complètent plutôt. Les personnes travaillant à la validation de tests devraient penser à toutes les questions sous les trois rubriques

listées par le comité, en gardant en tête que l'importance des questions varie d'un test à l'autre. » (traduction libre de Cronbach & Thorndike, 1971, p. 445)⁹

Graduellement, Messick dans sa théorie de la validité, a unifié le concept sous la bannière de la validité de construit, car selon lui la preuve par le contenu et les critères contribue elle aussi à légitimer le construit qui est mesuré. De plus, celle-ci incorpore la pertinence et l'utilité des scores du test, les valeurs qui les transcendent et les conséquences sociales comme faisant partie intégrante de la validité d'un test. Cette approche de la validité a fait l'objet d'un certain consensus exprimé par la version de 1999 du *Standards for Educational and Psychological Tests and Manuals* (American Educational Research Association. AERA, APA, & NCME, 1999)

Ce concept unifié de validité a, par contre, été critiqué par plusieurs auteurs sur plusieurs aspects entre autres sur l'aspect central du construit. Le terme construit étant par définition en lien avec le réseau nomologique des différentes théories qui le concerne (Cronbach & Meehl, 1955). Celui-ci réfère entre autres à des aspects externes au test qui ne sont nécessairement central dans le contexte de l'éducation. Par exemple, quel est le réseau nomologique d'un test de géographie de 4^e année? À quelles autres connaissances ou habiletés doit-il être associé pour témoigner de sa valeur? Doit-il vraiment considérer le construit ou être principalement guidé par sa représentativité au contenu dicté par le programme d'enseignement? (Lissitz & Samuelson, 2007)

Ainsi, certains auteurs (Markus & Borsboom, 2013) considèrent que le concept unifié de validité de Messick n'est pas souhaitable au niveau conceptuel, ceux-ci ne voyant pas la validité du test que dans l'interprétation faite de celui-ci. Plusieurs autres auteurs considèrent que le processus de validation qu'implique la théorie unifiée de Messick est peu pratique pour les concepteurs de test. Par exemple, (Kane, 2013) questionne la lourdeur du processus qui peut être sans fin en mentionnant : « Si toutes les données sont potentiellement pertinentes à la validité, où devons-nous débiter et combien d'éléments de preuves sont suffisantes pour supporter une interprétation? »¹⁰ (traduction libre de Kane, 2013; p.8). Dans la même ligne de pensée, Newton

⁹ Voir annexe 10 pour texte original

¹⁰ Voir annexe 10 pour texte original

(Newton & Shaw, 2014) questionne l'inclusion de l'éthique dans la conception de la validité : « D'un côté, la théorie de la validité devenait si complexe qu'il devenait difficile de savoir comment la traduire en processus de validation particulièrement en incluant à la fois les aspects scientifiques et éthiques. » (traduction libre de Newton & Shaw, 2014, p. 135)¹¹

Reconnaissant les limites du cadre conceptuel de Messick (1989), certains auteurs ont proposé de simplifier le concept de validité et de proposer des composantes pouvant guider plus facilement les efforts liés aux activités de validation d'un test.

Lissitz & Samuelsen (2007) considèrent qu'il existe deux parties distinctes dans le développement d'un test. Une première partie qualifiée d'interne inclut le contenu, le processus latent, et la fidélité. La deuxième partie qualifiée d'externe, quant à elle, inclut le réseau nomologique, son utilité et l'impact (ou les conséquences du test). Selon ces auteurs, cette portion du focus d'investigation, bien qu'importante, ne devrait pas faire partie de la validité du test, qu'ils attribuent plutôt à la portion interne. De plus, ceux-ci divisent leur taxonomie des procédures d'évaluation des tests (figure 2) selon une perspective théorique et pratique, cette dernière comprend le contenu et la fidélité du test pour la portion interne, de même que son utilité et impact dans la portion externe. Dans la perspective théorique, Lissitz & Samuelsen (2007) incluent le processus latent (lié à la passation du test) pour la portion interne et les liens du score du test avec un réseau nomologique pour la portion externe du test.

¹¹ Voir annexe 10 pour texte original

		Perspective	
		Théorique	Pratique
Focus d'investigation	Interne	Processus latent	Contenu et fidélité
	Externe	Réseaux nomologiques	Utilité et impact

Figure 2. Taxonomie des procédures d'évaluation des tests selon Lissitz & Samuelsen (2007)

Ainsi selon ce cadre, le processus de validation d'un test devrait d'abord se concentrer sur la portion interne du test avant de se pencher sur les liens avec d'autres variables et comment seront interprétés les résultats de ce test (Lissitz & Samuelsen, 2007). Bien que primordial, ces aspects peuvent difficilement être étudiés simultanément et dans bien des cas seront considérés que lors d'études subséquentes. Cette perspective permet de distinguer l'interprétation des résultats du test et la justification de son utilisation telle que préconisée par certains auteurs (Cizek, 2012; Markus & Borsboom, 2013) afin d'opérationnaliser le processus de validation d'un test de façon plus réaliste pour le domaine de l'éducation.

2.3.2 Sources d'évidence de la qualité du contenu d'un test

Pour Lissitz & Samuelsen (2007), les sources d'évidence de validité du contenu d'un test peuvent prendre la forme d'une analyse de la représentativité du cours ou programme à couvrir par le test (table de spécification), d'un jugement d'experts sur la complexité cognitive des items et sur la qualité de construction de ces items ainsi que sur les mécanismes mis en place pour s'assurer de la qualité de la correction et de l'attribution du score.

Appliqué au contexte du raisonnement clinique en milieu académique, il pourrait s'agir de démontrer que le test : 1) reflète la prise de décision à partir de connaissances du domaine choisi;

2) que les items soient bâtis de façon à refléter de bonnes qualités techniques (en lien avec les consignes aux étudiants, la pertinence du format des items et clarté des libellés, etc.)

Ces sources d'évidences devraient provenir à la fois des étudiants qui ont fait le test et des experts qui l'ont évalué. De plus, l'étape de l'attribution du score devrait être faite selon les meilleures pratiques rapportées par la littérature pertinente à ces tests.

2.3.3 Sources d'évidence de la fidélité d'un test

Selon Lissitz & Samuelsen (2007), les sources d'évidence liée à la fidélité d'un test devraient examiner les biais possibles pouvant être reliés au genre ou à au groupe culturel des répondants. Si différentes formes du test existent, est-ce que celle-ci procure la même information? Si le test est répété à différentes occasions, donne-t-il le même résultat? Dans le cas où plusieurs correcteurs sont impliqués, est-ce que l'attribution des scores est consistante?

Dans le cas de l'évaluation du raisonnement clinique en milieu académique, il pourrait s'agir de démontrer que le test : 1) possède une bonne consistance interne; 2) qu'il existe un bon accord entre la correction sur des éléments de réponses ouvertes et des scores qui en découlent et : 3) qu'il n'existe pas de fonctionnement différentiel au niveau des items.

2.3.4 Sources d'évidence de la qualité du processus latent d'un test

En lien avec la qualité du processus latent évalué, Lissitz & Samuelsen (2007) suggèrent de s'assurer que les tâches correspondent bien aux connaissances ou habiletés évaluées chez les répondants. Par exemple, il pourrait être pertinent de vérifier :

Est-ce que le niveau de difficulté attendu des items s'est avéré fondé et est-ce que les leurres ont été efficaces? Est-ce que le regroupement de certaines tâches ou items donne une information unique.

Appliqué au contexte du raisonnement clinique en milieu académique, il pourrait s'agir : 1) de vérifier auprès d'experts de contenu que les items présentent un niveau de difficulté adéquat pour les étudiants qui seront évalués; 2) de vérifier la perception des étudiants sur le niveau de difficulté des items et du format; 3) d'explorer les processus cognitifs (protocole à voix haute) associés au processus de réponse aux items; 4) d'explorer le regroupement des scores aux items

en lien avec des aspects tels que l'évaluation et l'intervention. 5) de vérifier si les scores du test convergent ou divergent par rapport à des méthodes d'évaluation documentées

2.3.5 Sources d'évidence de l'utilité d'un test

L'utilité d'un test qui selon la perspective de Lissitz & Samuelsen (2007) est liée à la portion externe du test peut être mise en évidence par la possibilité de prédire un résultat à un test par le score d'un test d'un autre domaine ou dans le même domaine. Dans ce cas, si un test est appelé à en remplacer un autre, une analyse des coûts-bénéfices peut être impliquée. Une des utilités du test pourrait être également de servir à démontrer l'efficacité d'une intervention dans le domaine évalué par le test.

Appliqué au contexte du raisonnement clinique en milieu académique, il pourrait s'agir de déterminer si le test : 1) permet de dépister des difficultés qui surviendrait dans un contexte académique ou de stage; 2) permet d'évaluer l'effet d'un processus de remédiation pour des étudiants ayant de telles difficultés.

2.3.6 Sources d'évidence de l'impact d'un test

Aussi associée à la portion externe d'un test, l'analyse de son impact peut être documentée par la façon dont les résultats ont été interprétés et utilisée par les différentes parties prenantes et sur les changements dans les politiques institutionnelles liés à l'utilisation de ce test.

Dans le cas de l'évaluation du raisonnement clinique en milieu académique, il pourrait s'agir de déterminer si l'utilisation du test dans le programme entraîne des modifications : 1) au niveau de l'enseignement des professeurs et chargés de cours; 2) dans les stratégies d'apprentissage des étudiants.

2.3.7 Sources d'évidence du réseau nomologique d'un test

Dans une perspective théorique, liés à la portion externe du test, Lissitz & Samuelsen (2007) suggèrent d'évaluer le trait latent visé par le test par rapport à ce même trait latent évalué par un autre test. Il pourrait également être pertinent d'évaluer ce trait latent en lien avec certains autres traits apparentés au niveau conceptuel et de suivre son évaluation dans le temps. La comparaison

de ce trait dans des groupes manifestement différents peut être utilisée comme une source d'évidence du réseau nomologique dans cette perspective.

Dans le cas du TCS et des PRC, les sources d'évidence du réseau nomologique pourraient porter; 1) sur la comparaison des scores entre des groupes à différents niveaux de formation dans le programme; 2) sur l'évolution des scores à des passations répétées à différentes étapes du curriculum.

2.3.8 Synthèse du processus de validation

Bien que le cadre décrit par Lissitz & Samuelsen (2007) implique certains choix au niveau conceptuel qui ne font pas l'unanimité dans le domaine de la validité (Sireci, 2007), celui-ci incorpore la plupart des éléments du processus de validation rencontrés dans d'autres propositions comme celles de Messick et de Kane, par exemple.

Dans la littérature portant sur le processus de validation des outils TCS et PRC, la plupart des efforts ont porté sur la démonstration de la consistance interne de l'outil, de sa capacité à discriminer en fonction d'un niveau d'expertise et dans le cas du TCS sur les facteurs qui influencent l'attribution du score par le panel d'experts.

Cette recension des écrits nous a permis de bien définir le cadre conceptuel retenu pour ce projet doctoral et d'arbitrer les choix effectués. Ces choix incluent : une théorie du raisonnement clinique, un paradigme de l'évaluation, une approche de la validation et des outils pour récolter des données. L'ensemble de ces choix aura évidemment des conséquences sur le processus de validation des tests de raisonnement clinique. D'abord au niveau de la création des vignettes et des items, celles-ci devront permettre à l'étudiant de porter un jugement à partir de connaissances acquises, et ce, dans des contextes de pratiques variés. Ces contextes tout en étant spécifiques à la physiothérapie et à la place prépondérante du patient dans le raisonnement clinique pourront également incorporer des éléments liés au milieu de soin et au clinicien.

Ces choix reliés au cadre conceptuel auront également des implications sur l'interprétation de l'évaluation du raisonnement clinique faite à partir des formulaires de stages. L'interprétation de la performance de l'étudiant sera effectuée en fonction des attentes liées au niveau de

compétence attendue de l'étudiant en fonction de son niveau de formation et des particularités du raisonnement clinique en physiothérapie.

De façon à mieux comprendre les avantages et les limites des outils retenus pour répondre à nos questions de recherche, c'est-à-dire le TCS et les PRC, le prochain chapitre présentera une revue de littérature de ces outils.

CHAPITRE 3

REVUE DE LITTÉRATURE SUR LES OUTILS D'ÉVALUATION DU RAISONNEMENT CLINIQUE RETENUS

Le prochain chapitre traitera des outils d'évaluation clinique retenus dans le cadre de ce projet doctoral. Dans un premier temps, nous ferons une revue de littérature sur les éléments de preuve liés à la validité du test de concordance de script. Dans un deuxième temps, nous aborderons les éléments de preuve liés à la validité des problèmes de raisonnement clinique. Cette revue de littérature nous permettra de mieux connaître les deux outils qui seront utilisés pour nous permettre de poser un jugement sur le raisonnement clinique des étudiants en physiothérapie.

3.1 Test de concordance de script

Le TCS a fait l'objet d'une quantité importante d'études depuis son introduction, en effet plus de 80 études ont porté sur le TCS (Dory *et al.*, 2012). Plusieurs de ces études ont porté sur des éléments liés à la validité. Nous reprendrons ces éléments en nous basant sur le cadre de la validité de Lissitz & Samuelsen (2007) qui décrit des facteurs internes liés à la validité et des facteurs externes aux tests.

3.1.1 Sources d'évidence de la qualité du contenu du TCS

Plusieurs éléments d'évidence liés à la qualité du contenu d'un test peuvent être retrouvés dans la littérature concernant le TCS. Ces éléments incluent :

Composition du panel de référence : Selon les guides de construction du TCS, un panel comprenant des experts du domaine de référence est nécessaire pour élaborer des scores de performance au TCS. Par contre, la définition d'expertise étant hautement variable (Ericsson, 2007) celle-ci correspond souvent à une attribution de ce statut par une communauté dans un domaine spécifique. Dans le cas du panel pour le TCS, la définition d'expertise peut inclure un certain nombre d'années d'expérience et/ou une certification formelle des compétences combinée à une reconnaissance sociale du statut d'expert (Lubarsky *et al.*, 2011). La composition de panel a fait l'objet de plusieurs études compte tenu de l'importance de celui-ci

dans l'attribution du score (Charlin *et al.*, 2002; Gagnon, Charlin, Coletti, Sauve, & van der Vleuten, 2005; Gagnon, Lubarsky, Lambert, & Charlin, 2011). Ces études font ressortir, en effet, l'importance du choix des experts dans le panel, par exemple dans une étude qui comportait un panel de 45 experts, le regroupement de ceux-ci de façon aléatoire en 20 groupes de 15 participants ont permis de constater des variations importantes de l'ordre d'une taille d'effet au niveau des scores attribués (Gagnon *et al.*, 2011). De façon plus spécifique, un aspect comme le groupe culturel des membres du panel a été noté comme une variable pouvant influencer les scores au test. En effet, un groupe de chercheur (Sibert, Charlin, Corcos, Gagnon, Grise, *et al.*, 2002) a comparé les scores attribués par un panel composé d'urologues français aux scores attribués par un panel composé d'urologues québécois, et a démontré que les résidents en urologie et les étudiants en médecine québécois obtenaient un score supérieur avec le panel composé d'urologues québécois que lorsque le panel était composé d'urologues français.

Le nombre recommandé d'experts dans un panel de référence varie de 10, pour un test à enjeux faible à 20 pour un test à enjeux élevé (Fournier *et al.*, 2008; Lubarsky, Dory, Duggan, Gagnon, & Charlin, 2013). Un nombre plus élevé de membres du panel ayant peu d'impact sur les scores attribués (Gagnon *et al.*, 2005). Dans cette étude qui incluait 38 médecins dans le panel d'experts et 80 résidents comme participants, les auteurs ont mis en évidence l'effet du nombre d'experts dans le panel sur la consistance interne (alpha de Cronbach) et sur la corrélation entre les résultats attribués entre des panels de différentes tailles et le panel complet (n=38). Ils ont ainsi déterminé que dans leur échantillon, les scores obtenus à partir des réponses de 10 et 20 panélistes obtenaient un alpha de Cronbach de 0,70, 0,74 respectivement et de 0,76 pour les scores attribués par le panel complet. De même, la corrélation entre les scores obtenus avec le panel complet était de $r=0,95$ pour un panel de 10 membres et $r=0,98$ pour un panel de 20 membres.

Pour ce qui est de la passation du test par les membres du panel, il est recommandé dans un des guides de construction du TCS, que ceux-ci répondent aux items de façon indépendante, sans préparation, dans le même délai que celui donné aux personnes ciblées par le test (Lubarsky *et al.*, 2013).

Optimisation des réponses du panel : Une étude effectuée par Gagnon *et al.* (2011) a évalué l'effet de différentes méthodes d'optimisation des réponses du panel, ces méthodes incluaient le retrait de membres discordants et le retrait de réponses discordantes. Selon ces auteurs, à condition que le panel comporte 15 membres ou plus, l'élimination d'un ou deux membres discordants du panel avait peu d'influence sur la corrélation des scores (avec le résultat du test initial), la fidélité et la taille d'effet. De même, l'élimination de réponses discordantes (par rapport à la réponse modale) avait également peu d'influence sur ces mêmes paramètres si le panel est composé de plus de 15 membres (Gagnon *et al.*, 2011).

Attribution des scores : La méthode des scores combinée (*aggregate scoring method*) est la méthode d'attribution des scores présentement la plus utilisée dans la littérature sur le TCS. Celle-ci sera décrite en détail dans le prochain chapitre. Plusieurs recherches se sont penchées sur ce processus d'attribution des scores au TCS à partir d'une échelle de réponse de type Likert à cinq niveaux. L'utilisation de cette méthode repose sur la prémisse que dans les problèmes mal définis, des opinions divergentes d'experts représentent des points de vue riches qui permettent d'apprécier la complexité du problème. Cette variabilité des réponses permettrait de mieux discriminer l'expertise clinique entre des groupes de différents niveaux (Charlin *et al.*, 2006). Ainsi, selon l'étude de Charlin (2006) faite à partir d'un test de concordance comprenant 145 items, on a regroupé en trois catégories, les items de faible variabilité (*Mean square Error (MSE)* de moins de 0,49), de variabilité moyenne (*MSE* entre 0,5 et 1,0) et de variabilité plus forte (*MSE* plus de 1,0). À la suite de ce regroupement, les items ayant une variabilité élevée permettaient de séparer plus distinctement les groupes dans une analyse de variance, ces items ayant une taille d'effet de 1,5 par rapport à 0,56 pour les items de variabilité modérée et de 0,31 pour les items de variabilité faible. Il doit être noté cependant, que cette distinction entre les items de variabilité faible, modérée et forte n'est pas couramment rapportée dans la littérature et que nous n'avons pas trouvé d'autres éléments de preuve supportant cette catégorisation.

Optimisation des items du test : Tel que rapporté dans le tableau I, plusieurs niveaux d'optimisation des items ont été utilisés dans les études portant sur le TCS, certaines de celles-ci ont éliminé les items ayant un niveau de corrélation item-total de 0,1 et moins, alors que d'autres études ont utilisé un niveau de 0,05 et moins et que quelques-unes n'ont éliminé que

les items ayant une corrélation négative. Un grand nombre de celles-ci cependant n'ont pas spécifié le niveau retenu pour l'optimisation. Nonobstant le niveau d'optimisation retenu, il est recommandé de s'assurer de la représentativité du test suite à l'optimisation des items (Lubarsky *et al.*, 2013) en n'éliminant pas de façon automatique tous les éléments en lien avec un contenu spécifique sur la seule base d'une faible corrélation item-total. En moyenne, près de 25 % des items ont été éliminés suite au processus d'optimisation *a posteriori* à partir des réponses des participants selon la revue systématique de Dory *et al.* (2012).

Construction du test : Issus de plus d'une dizaine d'années d'utilisation du TCS dans différents domaines, plusieurs guides ont émis des recommandations en lien avec la construction de ce type de test (Dory *et al.*, 2012; Fournier *et al.*, 2008; Giet, Massart, Gagnon, & Charlin, 2013; Lubarsky *et al.*, 2013; Wan, 2015). Comme dans tout processus d'élaboration des items d'un test, il est recommandé d'abord de bien circonscrire le contenu évalué et d'avoir un échantillonnage adéquat compte tenu de l'importance relative des différents concepts évalués (Downing & Haladyna, 2006). Pour ce qui est des concepteurs de test dans le cas du TCS, il est recommandé de faire appel à des cliniciens actifs qui peuvent puiser au niveau de leur expérience pour favoriser l'authenticité des items. Le contenu des vignettes cliniques doit être bref, réaliste et concerner un problème mal défini. Pour optimiser le processus de création des cas et question le regroupement de concepteurs en équipe de 2 est recommandé (Lubarsky *et al.*, 2013). Ces concepteurs doivent décrire une situation clinique permettant d'élaborer une hypothèse (au niveau diagnostique, thérapeutique ou d'investigation) et déterminer les données à interpréter en lien avec cette situation. Au terme du processus de création des cas et questions, il est recommandé par la suite de faire appel à une équipe de deux ou trois réviseurs indépendants qui se prononceront sur la pertinence et la clarté des items. (Fournier *et al.*, 2008; Lubarsky *et al.*, 2013).

3.1.2 Sources d'évidence de la fidélité du TCS

Dans la grande majorité des recherches sur le TCS ayant utilisé la théorie classique des tests, une estimation de la fidélité est calculée à l'aide du coefficient alpha de Cronbach. Bien qu'il soit plus approprié de parler dans ce cas, de consistance interne qui est en fait une borne inférieure de la fidélité (Sijtsma, 2009). Ce coefficient permet d'apprécier jusqu'à quel point les

réponses aux items sont en fortes corrélations les unes avec les autres et de déterminer la capacité des scores du test à discriminer les individus à l'intérieur d'un groupe (Tavakol & Dennick, 2011).

Le TCS démontre, en général, une bonne consistance interne. Le tableau I présente la consistance interne de plusieurs études ayant utilisé le TCS comme outil d'évaluation, la plupart de ces études rapportent des coefficients au-dessus de 0.80 pour une heure de testing. La consistance interne est influencée par le nombre d'items (Panayides, 2013; Tavakol & Dennick, 2011), par contre, dans plusieurs cas, il n'est pas mentionné si la consistance interne est calculée par items ou par vignette. Comme mentionnée par Dory *et al.* (2012), les items de chaque vignette ne pouvant être considérés comme indépendants, le calcul de la consistance par items aurait tendance à surestimer celle-ci. De plus, il est rarement mentionné, si la consistance interne est calculée par groupe ou pour l'ensemble des groupes. Or selon Cook (2015), inclure des groupes ayant des différences connues dans une analyse de la fidélité peut surestimer le coefficient alpha de Cronbach, en augmentant la variance des scores au test. Il est donc recommandé d'effectuer les analyses de la fidélité en fonction des groupes pour lesquels l'outil est destiné (Cook, 2015). Selon notre revue de littérature, seule l'étude de Charlin *et al.* (1998) spécifie une valeur d'alpha de Cronbach différente pour les groupes de résidents et d'étudiants. Il est ainsi possible que le niveau de consistance interne soit à des niveaux très différents d'un groupe à l'autre ce qui pourrait affecter les inférences faites à partir des scores de chaque groupe. La consistance interne souhaitée telle qu'estimée par le coefficient alpha de Cronbach devrait être de plus de 0,75 selon les auteurs des guides de construction des TCS (Fournier *et al.*, 2008; Lubarsky *et al.*, 2013).

Seulement quatre études (Brailovsky, Charlin, Beausoleil, Cote, & van der Vleuten, 2001; Gagnon, Charlin, Lambert, Carriere, & van der Vleuten, 2009; Ramaekers, Kremer, Pilot, Beukelen, & Keulen, 2010; K. Tan, Tan, Kandiah, Samarasekera, & Ponnampereuma, 2014) ont utilisé la théorie de la généralisabilité pour estimer la fidélité des TCS. Les coefficients G ainsi obtenus variaient entre 0,54 et 0,88. Une de ces études a tenté de déterminer à partir des résultats de trois études précédemment publiées, le nombre optimal d'items par vignette. À partir des variances obtenues par item et par vignette, ils ont déterminé qu'il existe un effet plafond sur la

fidélité (coefficient G) au-delà de quatre questions par vignette (Gagnon *et al.*, 2009). Basés sur ces résultats, les auteurs recommandent de bâtir des tests comportant de 20 à 25 vignettes ayant entre 3 à 4 questions chacune de façon à obtenir un coefficient G au-delà de 0,80.

3.1.3 Sources d'évidence de la qualité du processus latent du TCS

Le TCS a démontré une bonne corrélation avec d'autres outils d'évaluation du raisonnement clinique dans le cadre du suivi d'un groupe de résidents en médecine familiale (Brailovsky *et al.*, 2001). Dans cette étude, les résultats aux TCS étaient modérément corrélés avec les résultats de deux autres épreuves ($r=0.451$, $p=0.013$; $r=0.447$, $p=0.015$, respectivement) portant sur le raisonnement clinique à la fin de la résidence. En contraste, la corrélation des résultats au TCS était moins élevée ($r=0,340$, $p=0,052$) avec les résultats des ECOS, principalement utilisés comme outils d'évaluation des habiletés cliniques dans le contexte de l'article.

Quelques auteurs se sont aussi penchés sur la relation entre le TCS et d'autres outils destinés à évaluer un contenu différent tel que des QCM pour les connaissances factuelles. Dans une étude dans le domaine de la médecine d'urgence Fournier *et al.* (2008) n'ont pas trouvé de corrélation significative ($r=0,016$, $p=0,59$) entre un test qui incluait 60 items QCM portant sur des connaissances factuelles et un TCS incluant 90 items dans le même domaine. Au cours d'une autre recherche sur ce thème, Collard *et al.* (2009) ont bâti un test de connaissance factuelle de type vrai ou faux et un TCS à partir autour d'une même table de spécification. Les auteurs ont trouvé une corrélation significative ($r=0,53$, $p<0,0001$) entre les résultats aux items de connaissances factuelles et les résultats du TCS, chez les étudiants en milieu de programme (an 3 et 4) et pas de corrélations ($r=0,07$, $p=0,64$) à ces mêmes épreuves, plus tard dans le programme (an 5 et 6). Ces résultats sont interprétés par les auteurs comme une possible diminution de l'importance des connaissances factuelles à mesure que celles-ci sont remplacées par l'expérience clinique.

3.1.4 Sources d'évidence de l'utilité du TCS

Dans le cadre de validité de Lissitz & Samuelsen (2007), l'utilité est un des facteurs externes au test, celui-ci réfère aux décisions qui peuvent être prises sur la base des scores à ce test. Jusqu'à maintenant, peu d'études sur le TCS ont porté sur cet aspect. Un des exemples de l'utilité du

TCS pour supporter des décisions à caractère pédagogique est rapporté par le programme de médecine de l'Université du Colorado (Guerrasio, Garrity, & Aagaard, 2014). En effet, dans ce programme, le TCS de même que l'observation directe et la révision des dossiers de stage sont utilisés pour déterminer quels sont les étudiants qui pourraient bénéficier de remédiation au niveau du raisonnement clinique.

3.1.5 Sources d'évidence de l'impact du TCS

Bien qu'il soit généralement admis que la façon d'évaluer les apprentissages peut influencer de façon directe l'apprentissage lui-même (Larsen, Butler, & Roediger, 2008) l'impact de l'utilisation du TCS sur les stratégies d'apprentissages est très peu documenté. Selon certains auteurs, les particularités du TCS pourraient influencer la perception des étudiants de la réalité clinique en leur faisant réaliser qu'il peut exister plusieurs bonnes réponses à un problème (Llorca, Roy, & Riche, 2003). Cet effet serait attribuable au processus d'attribution des scores au TCS qui contraste avec celui utilisé pour les questions à choix multiples traditionnelles où une seule bonne réponse est acceptée, par contre nous n'avons pas trouvé de données empiriques qui appuient cette hypothèse.

3.1.6 Sources d'évidence de la qualité du réseau nomologique du TCS

Une des sources d'évidence du réseau nomologique du TCS fréquemment utilisé repose sur sa capacité à discriminer le niveau d'expertise des participants. Ainsi, comme on peut le voir dans le tableau I, la plupart des études recensées ont déterminé des différences statistiquement significatives entre les résultats des groupes ayant des formations ou des niveaux d'expérience différents. Quelques-unes de celles-ci ont comparé les résultats d'étudiants et de résidents de différents niveaux entre eux sans inclure les experts, alors que d'autres les incluaient dans la comparaison. Par exemple, dans une des premières études utilisant les scores au TCS pour comparer le raisonnement clinique de certains groupes, Charlin *et al.* (1998) ont obtenu des scores statistiquement plus élevés ($p < 0,001$) pour les experts (score moyen : 45,3, $\pm 3,7$) que pour les étudiants en fin de cursus universitaire en médecine (score moyen : 35,8, $\pm 7,4$). Par contre, il n'y avait pas de différences significatives entre les scores des résidents (score moyen : 40,5, $\pm 7,1$) en gynécologie obstétrique et celui des étudiants ou celui des experts. L'importante

variabilité dans les scores des résidents répartis entre de la première à la cinquième année semble expliquer en partie ce résultat.

Plus récemment dans une étude (Nouh *et al.*, 2012) incluant 202 résidents en chirurgie de neuf universités canadiennes, les auteurs ont fait ressortir des différences significatives ($p=0,001$) entre les scores d'un TCS comportant 131 items après optimisation, des résidents juniors regroupés (R1 et R2) et des résidents seniors regroupés (R3, R4 et R5). Ces différences ont été constatées de façon uniforme à travers les différents milieux de formation à travers le Canada. Par contre, les différences rencontrées au niveau des scores individuels pour chacune des années de résidence (de R1 à R5) n'atteignent pas des niveaux significatifs ($p=0,001$). Ces résultats ont aussi été constatés par un groupe de chercheurs (Meterissian, Zabolotny, Gagnon, & Charlin, 2007) dans le domaine de la chirurgie également. Il est donc possible que les différences de raisonnement clinique, mis en évidence par le TCS soient détectables à partir d'un certain seuil de développement de script clinique.

Bien qu'une des prémisses du TCS repose sur la richesse et la nature mieux organisée des scripts des experts, des différences de scores entre les experts et les novices représentent une association qui n'implique pas nécessairement un lien de causalité entre score élevé et expertise, plusieurs autres liens explicatifs peuvent exister (Cook, 2015). Lorsque présente, les différences entre groupes supportent le réseau nomologique du trait latent et l'absence de différence le questionne.

Tableau I Études ayant comparé le niveau de raisonnement clinique de groupe à l'aide du TCS

Études Domaine	Groupes de participants	Différences statistiques entre les scores des groupes	Items retenus après optimisation Niveau de corrélation item-total utilisé	Fidélité Alpha de Cronbach (α) Coefficient G
(Charlin, Brailovsky, Leduc, & Blouin, 1998) Gynécologie-Obstétrique	Experts (n=15) Résidents (n=12) Étudiants (n=76)	Oui entre étudiants et experts	50 Non spécifié	$\alpha=0,81$ (non spécifié) $\alpha=0,79$ (non spécifié)

(Sibert, Charlin, Corcos, Gagnon, Lechevallier, <i>et al.</i> , 2002) Urologie	Experts (n=22) Résidents (n=25) Étudiants (n=23)	Oui entre tous les groupes	80 Non spécifié	$\alpha=0,79$ (non spécifié)
(Brazeau-Lamontagne, Charlin, Gagnon, Samson, & van der Vleuten, 2004) Radiologie	Résidents Juniors (n=20) Seniors (n=20) Étudiants (n=20)	Oui entre tous les groupes	145 Non spécifié	$\alpha=0,81$ (vignette)
(Meterissian <i>et al.</i> , 2007) Chirurgie	Résidents (n=36) Juniors (n=14) Seniors (n=22)	Oui entre résidents juniors et seniors	62 0,1	$\alpha=0,85$ (non spécifié)
(Lambert, Gagnon, Nguyen, & Charlin, 2009) Radio-oncologie	Experts (n=47) Résidents (n=38) Étudiants (n=70)	Oui entre tous les groupes	70 0,1	$\alpha=0,90$ (par items)
(Lubarsky <i>et al.</i> , 2009) Neurologie	Experts (n=16) Résidents (n=41) Étudiants (n=8)	Oui entre résidents experts et étudiants et experts	79 0,05	$\alpha=0,79$ (non spécifié)
(Carrière, Gagnon, Charlin, Downing, & Bordage, 2009) Pédiatrie	Résidents Résidents 1 (n=21) Résidents 2 (n=21) Séniors (n=11)	Oui entre résident 1 et senior	50 Items de corrélation négative	$\alpha=0,79$ (par item) $\alpha=0,77$ (par vignette)
(Humbert <i>et al.</i> , 2011) Médecine	Experts (n=30) Étudiants 2 ^e (n=411) Étudiants 4 ^e (n=70)	Oui entre tous les groupes	75 Non spécifié	$\alpha=0,73$ (non spécifié)
(Deschênes <i>et al.</i> , 2011) Sciences infirmières	Experts (n=12) Étudiants (n=30)	Oui	73 0,05	$\alpha=0,86$ (par item)
(Kania <i>et al.</i> , 2011) Otorhinolaryngologie	Experts (n=22) Résidents (n=22) Étudiants (n=21)	Oui entre tous les groupes	94 Non spécifié	$\alpha=0,95$ (non spécifié)

(Nouh <i>et al.</i> , 2012) Chirurgie	Résidents Juniors (n=96) Seniors (n=106)	Oui entre résidents junior et senior	131 Items de corrélacion négative	$\alpha=0,85$ (non spécifié)
(Boulouffe, Doucet, Muschart, Charlin, & Vanpee, 2013) Médecine d'urgence	Experts (n=12) Résidents (n=19) Étudiants (n=21)	Oui entre tous les groupes	64 Non spécifié	$\alpha=0,80$ (non spécifié)
(See, Tan, & Lim, 2014) Médecine interne	Experts (n=16) Résidents Juniors (n=10) Seniors (n=16)	Oui entre tous les groupes	Set 1 :96 Set 2 : 96 Non spécifié	$\alpha=0,79$ (non spécifié) $\alpha=0,89$ (non spécifié)
(Tan <i>et al.</i> , 2014) Neurologie	Experts (n=15) Résidents (n=37) Étudiants (n=52)	Oui entre tous les groupes	53 Non spécifié	Coefficient G= 0,72
(Ducos <i>et al.</i> , 2015) Anesthésiologie	Experts (n=10) Résidents Juniors (60) Seniors (47)	Oui entre tous les groupes	60 Non spécifié	$\alpha=0,63$ (non spécifié)

3.2 Problèmes de raisonnement clinique (PRC)

Par rapport au TCS, relativement peu d'études ont porté sur les PRC et sur les éléments liés à sa validité. La première recherche portant sur les PRC publiée par Groves en 2002 incluait deux séries de PRC comprenant 10 vignettes chacune, chaque vignette comprenait une quantité importante d'information conçue de façon à être associée à plusieurs diagnostics possibles. Pour chacun de ces diagnostics, il était par la suite demandé de les justifier sur la base des caractéristiques fournies. Chaque série de 10 vignettes nécessitait entre 90 à 120 minutes à compléter. Tout comme pour le TCS, le processus d'attribution des scores aux PRC fait appel à un panel d'experts, cette méthode d'attribution des scores sera décrite en détail dans le prochain chapitre. L'étude de Groves (2002) a comparé les résultats aux PRC de groupes d'étudiants en médecine (n=92) de 2^e, 3^e et de dernière année (6^e) avec les résultats du panel d'experts composé de 22 médecins de famille ayant en moyenne 21 ans d'expérience clinique. Leurs résultats suggéraient une amélioration graduelle des scores aux PRC du groupe d'étudiants de 2^e année

jusqu'au panel d'experts. Des différences significatives ont été notées entre tous les groupes sauf entre le groupe d'étudiants de 3^e année et de 6^e année.

Les PRC développés par (Groves *et al.*, 2002) indiquaient une bonne consistance interne (α variant entre 0,61 à 0,83) pour les deux séries de PRC développées, par contre il n'est pas spécifié si la consistance interne a été calculée par item ou par vignette. Dans cette même étude, les auteurs ont demandé aux différents groupes de répondre également à un test portant sur les qualités associées à la performance diagnostique, le *Diagnostic Thinking Inventory (DTI)* (Bordage, Grant, & Marsden, 1990). À partir des résultats aux PRC et au DTI, une corrélation modérée ($r=0,50$) a été établie entre les résultats des deux outils pour l'ensemble des groupes.

Dans une publication subséquente sur la valeur prédictive de certaines variables utilisées pour la sélection de candidats dans un programme de médecine, les auteurs (Groves, O'Rourke, & Alexander, 2003a) ont effectué des corrélations entre les résultats du PRC et les résultats d'entrevue et les résultats de tests de sélection (*Graduate Australian Medical School Admission Test*) chez 253 étudiants en médecine. Les corrélations ainsi obtenues étaient très faibles allant de $r=0,02$, $p=0,73$ à $r=0,11$, $p=0,08$. De plus, les résultats académiques des deux premières années du programme ont aussi été corrélés avec les résultats aux PRC donnant également un niveau de corrélation très faible ($r=0,10$, $p=0,10$).

Par la suite, quelques autres études ont repris les PRC développés initialement par Groves (2002) et les ont réutilisés dans d'autres contextes. Par exemple, dans le but d'évaluer l'effet d'un atelier portant sur l'apprentissage du raisonnement clinique à l'aide de script chez des étudiants en médecine de 4^e année, Lee *et al.* (2010) ont administré les PRC chez un groupe de participants ($n=24$) et chez un groupe contrôle ($n=29$) n'ayant pas suivi l'atelier. Dans cette étude, les étudiants ayant suivi l'atelier démontraient des scores plus élevés aux PRC que le groupe contrôle avec une taille d'effet importante de 1,24. Il est à noter cependant, que dans cette étude, les participants à l'atelier utilisaient une des deux versions des PRC pendant le déroulement de l'atelier et qu'un effet d'apprentissage à l'outil est possiblement un facteur pouvant avoir influencé les résultats.

Plus récemment, Groves *et al.* (2013) ont réutilisé les deux mêmes versions des PRC dans le but de comparer les résultats obtenus avec ce format d'items par rapport aux items de format TCS. À partir des thèmes contenus dans les vignettes PRC, 10 vignettes comportant 50 items de format TCS ont été conçues. Les PRC et les TCS ainsi obtenus ont donc été administrés à 119 étudiants en médecine de 3^e et 4^e année et à 12 médecins de famille. Les résultats aux PRC démontraient une augmentation graduelle des scores des étudiants de 3^e, 4^e et des médecins de famille pour les deux versions du test (PRC1 et PRC2), mais n'atteignaient une différence statistiquement significative que pour le PRC2. Du côté des résultats au TCS, on ne notait pas d'effets discriminants des scores. Par contre, on notait une corrélation modérée entre les scores au PRC et au TCS variant entre $r=0,46$ à $r=0,49$. La consistance interne variait entre $\alpha=0,36$ et $\alpha=0,60$ pour les TCS et entre $\alpha=0,56$ et $\alpha=0,61$ pour les PRC. Plusieurs faiblesses méthodologiques ont été remarquées dans cette étude, par exemple peu d'information est donnée sur le processus de construction du TCS à partir des PRC et il n'est pas mentionné comment les scores du TCS, nouvellement conçus, ont été attribués à partir des réponses des experts aux PRC de l'étude de Groves *et al.* (2002). On remarque également des tailles d'échantillon très faibles pour les groupes de médecins de famille (de $n=2$ à $n=8$).

Un autre groupe de chercheurs a également tenté de comparer les résultats des formats TCS et PRC, cette fois chez un groupe d'étudiants en dernière année de leur programme de médecine en Belgique. Dans ce but, Dory *et al.* 2014 ont repris les PRC développés initialement par Groves (2002), les ont traduits en français et ont construit des TCS de 20 vignettes à partir des mêmes tables de spécifications. Les scores de 16 médecins de famille et de 44 étudiants qui ont participé à cette recherche démontraient des différences significatives entre les deux groupes avec des tailles d'effet variant entre 0,45 et 0,48 pour les TCS et 0,67 et 0,73 pour les PRC. La consistance interne variait entre 0,66 et 0,79 pour les TCS et entre 0,49 et 0,76 pour les PRC. La corrélation entre les résultats des deux TCS était modérée avec un des PRC (Spearman $\rho=0.357$ et 0.358 ; $p=0,017$) et non significative avec le 2^e PRC (Spearman $\rho=0.126$ et 0.059 , $p=0.413$ et 0.705). De plus, les auteurs ont corrélé les résultats des PRC aux résultats des ECOS (évaluant les habiletés cliniques) et mentionnent n'avoir trouvé aucune corrélation sans toutefois préciser de valeur.

À notre connaissance, seulement un autre groupe de chercheurs a développé des vignettes PRC différentes de celles initialement développées par Groves (2002). Dans leur étude d'Amini *et al.* (2011) décrivaient l'utilisation de plusieurs outils d'évaluation du raisonnement clinique incluant des PRC et des TCS chez un groupe de 135 étudiants en médecine provenant de 45 universités iraniennes dans un contexte nommé *olympiade du raisonnement clinique*. Les résultats décrits dans cet article comprenaient des corrélations entre les différents outils et les résultats globaux du test. Dans le cas du PRC et du TCS, celles-ci étaient élevées à 0,74 ($p < 0,001$) et 0,67 ($p < 0,001$) respectivement. Par contre, la corrélation avec ces mêmes tests et les moyennes cumulatives académiques des étudiants étaient faibles à 0,27 ($p < 0,01$) et 0,13 ($p > 0,1$) pour les PRC et les TCS. La consistance interne rapportée pour le PRC était de $\alpha = 0,71$ et de $\alpha = 0,78$ pour le TCS. Bien que les auteurs aient mentionné avoir respecté les consignes quant à la construction des tests et l'attribution des scores, très peu de détails ont été donnés dans l'article pour nous permettre de constater le respect de ces consignes.

Globalement, en contraste avec le TCS, il existe très peu d'éléments de preuve pouvant nous guider dans la construction et l'attribution des scores des PRC. Jusqu'à maintenant, très peu de données empiriques, autres que l'étude initiale de Groves (2002) ont été publiées comme source d'évidence de la qualité de contenu relié au PRC. Au niveau des autres facteurs internes liés à la validité du PRC, les études publiées ont rapporté surtout des évidences liées à la fidélité du test, bien que l'on ait peu de contextes différents pour poser un jugement sur celle-ci, puisque presque toutes les études ont utilisé les mêmes vignettes développées par les concepteurs du PRC. Au niveau de ces données disponibles sur la fidélité des PRC, celles-ci sont en général plus faibles que pour le TCS ce qui peut possiblement impliquer une plus grande erreur de mesure dans le cas des PRC.

Toujours au niveau des facteurs internes, on note aussi la présence de quelques évidences liées à la qualité du processus latent découlant de corrélations effectuées entre les résultats du PRC et d'autres tests évaluant le raisonnement clinique ou un autre construit (connaissance factuelle, habileté clinique).

Au niveau des facteurs externes pouvant être considérés pour le PRC, il n'existe pas d'évidence de son impact et les quelques études (Groves *et al.*, 2003a; Lee, Ho, Groves, & Vlantis, 2010)

qui ont tenté de documenter l'utilité de ce test étaient probablement prématurées compte tenu du peu d'évidences disponibles sur les facteurs internes déterminant la validité du PRC. Les autres éléments d'évidence du PRC reliés aux facteurs externes ont porté sur la qualité du réseau nomologique, par le biais de la discrimination des scores chez des groupes ayant un niveau de développement présumé différent du trait latent étudié.

En conclusion, il apparaît clair que les deux outils en sont à des stades de développement différents, le TCS ayant bénéficié d'une assez grande quantité d'études dans différents domaines, alors que les études sur les PRC ont été très limitées et que très peu d'adaptation de l'outil a été effectuée dans des contextes différents. Ces différents niveaux d'évidences disponibles pour chacun des deux tests devront être pris en considération dans les prochaines parties de ce projet doctoral.

CHAPITRE 4

LA MÉTHODOLOGIE

Le présent chapitre aborde la méthodologie développée pour répondre aux deux questions spécifiques de recherche. Il comporte sept sections distinctes. La première section introduit la posture de recherche et le type d'étude alors que la deuxième présente les participants et le milieu de la collecte de données. La troisième section présente les étapes reliées au processus de validation des outils d'évaluation du raisonnement clinique en les plaçant en perspective avec le cadre conceptuel utilisé dans le chapitre deux. Nous avons choisi de présenter dans cette section les données résultant de la passation d'un questionnaire sur les outils, car nous croyons que cette information est un des éléments du contexte qui illustre la qualité du processus de collecte des données. De son côté, la quatrième section décrit les choix et les étapes reliées à la collecte des données dans les formulaires de stages. Dans cette section également nous avons choisi de présenter des résultats de l'accord sur la catégorisation. Comme décrit dans la section précédente, nous croyons que ces informations documentent le contexte de collecte de données et ne sont pas sujettes à des analyses nous permettant de répondre aux questions de recherche. La cinquième section aborde les stratégies d'analyse des données. En terminant, la sixième section présente les ressources humaines, matérielles et financières nécessaires à la réalisation de l'étude et la septième, les considérations éthiques.

4.1 Posture de recherche et type d'étude

La première portion de l'étude aborde l'évaluation du raisonnement clinique chez les étudiants en physiothérapie à l'aide d'outils d'évaluation papier crayon. Pour ce faire, deux outils d'évaluation du raisonnement clinique distincts ont été choisis. Ces deux outils abordent

l'évaluation en comparant le niveau de raisonnement clinique des étudiants à celui des experts par le biais des scripts.

La deuxième portion de l'étude fera une analyse de contenu des formulaires de stages des étudiants en physiothérapie afin de catégoriser leurs niveaux de raisonnement clinique en situation authentique. Cette portion est inspirée du courant interprétatif qui considère que le contexte est indissociable du savoir et que le jugement de l'expert permet de tenir compte de ce même contexte. Une recherche s'inscrivant dans cette optique essaiera de comprendre les phénomènes en fonction des acteurs en présence (dans ce cas les superviseurs cliniques).

Bien que chacune des portions de l'étude vise un objectif différent, la mise en commun des résultats permettra de mieux cerner l'objet de recherche. L'ensemble du projet utilise donc un devis globalement mixte qui s'inspire d'une vision pragmatique considérant que les points de vue des paradigmes qualitatifs et quantitatifs ne sont pas incompatibles et que différentes approches méthodologiques peuvent être utilisées selon les questions de recherches et les circonstances (Biesta, 2010). Les méthodes de recherche mixte, lorsqu'indiquées, ont comme principal avantage de nous permettre de mieux comprendre le problème de recherche que l'utilisation de méthode qualitative ou quantitative de façon isolée (Creswell, 2010). Cette approche mixte utilisant des outils de différentes natures est d'ailleurs préconisée en sciences de la santé pour en arriver à une évaluation juste de la compétence dans une vision contemporaine d'un programme d'évaluation (van der Vleuten & Schuwirth, 2005).

4.2 Description des participants et du milieu de collecte de donnée

4.2.1 Les étudiants et leur programme de formation

Les participants à l'étude sont des étudiants du programme de physiothérapie de la Faculté de médecine de l'Université de Sherbrooke. Ce programme de B.Sc./M.Sc. intégré est d'une durée de quatre ans et permet d'accéder à la pratique professionnelle. Le programme a été conçu en quatre phases (voir annexe 1). La phase 1 (durée de 1,5 an) est centrée sur l'acquisition de connaissances biomédicales et psychosociales et utilise la méthode pédagogique de l'apprentissage par problèmes. La phase 2 (durée de six mois) met l'accent pour sa part sur

l'évaluation clinique tout en incorporant graduellement des notions d'interventions. La phase 3 poursuit l'intégration des connaissances cliniques sur l'intervention dans les différents domaines de la physiothérapie. La phase 4, quant à elle, consiste en grande partie en de la formation clinique en stage. Les étudiants du programme ont six stages d'une durée totale de 1 200 heures.

Ce milieu de formation est particulièrement approprié dans le cadre de la présente recherche, puisque le programme met beaucoup l'accent sur le raisonnement clinique. En effet, la méthode pédagogique de l'apprentissage du raisonnement clinique (Chamberland, 1998) fait partie intégrante des méthodes d'enseignement à partir de la 2^e phase du programme (2^e semestre de la 2^e année) et se poursuit également dans d'autres phases du programme. Dans le cadre de l'étude, nous avons sollicité l'ensemble des 45 étudiants des cohortes de 2^e, 3^e et 4^e année du programme de physiothérapie de l'Université de Sherbrooke. Les étudiants de 1^{re} année n'ont pas été inclus dans l'étude, ceux-ci n'ayant pas suffisamment de ressources pour répondre aux items de raisonnement clinique des tests développés dans le cadre de cette recherche.

En tout, 91 étudiants du programme de physiothérapie de l'Université de Sherbrooke ont accepté de participer à l'étude. De ce nombre 33 étaient en deuxième année, 32 en troisième année et 26 en quatrième année. La proportion des participants de chaque cohorte représente entre 63 % et 79 % de chaque cohorte. La proportion d'hommes était de 25 sur 91 (27 %) pour les trois cohortes ce qui est très près de la proportion homme/femme de ces trois cohortes (24,1 %) dans le programme de physiothérapie de l'Université de Sherbrooke.

4.2.2 Panel d'experts

En ce qui a trait au panel d'experts, les physiothérapeutes recrutés devaient avoir plus de cinq ans d'expérience clinique dans le domaine musculosquelettique et avoir une expertise reconnue par les pairs dans ce domaine. Pour éviter que les conceptions des experts ne représentent que la vision du programme de l'Université de Sherbrooke, nous avons également recruté des experts qui sont associés aux autres universités du Québec et des cliniciens qui ne sont pas impliqués dans le milieu universitaire.

Le panel d'experts recrutés était composé de 19 physiothérapeutes ayant un minimum de cinq ans d'expérience cliniques et travaillant principalement dans le domaine musculosquelettique.

Comme décrit dans le tableau II, la grande majorité de ces experts ont plus de 15 ans d'expérience (15/19). Ce panel était composé de dix hommes et neuf femmes, la composition du panel à ce niveau est peu représentative de la proportion homme femme dans la profession, celle-ci étant estimée à 82 % de femme pour l'ensemble de la profession au Canada en 2013 (gouvernement du Canada). Par contre au niveau de la spécialisation clinique, notre échantillon se rapproche de la proportion d'hommes qui a complété le processus pour obtenir le *Fellowship* en thérapie manuelle musculosquelettique (44,4 %) au Québec.

Tableau II Années d'expérience du panel d'experts

Années d'expérience	Nombre d'experts	Pourcentage du panel
5 à 10 ans	2	10,5 %
10 à 15 ans	2	10,5 %
15 à 20 ans	7	36,9 %
Plus de 20 ans	8	42,1 %
Total	19	100 %

La majorité de ceux-ci œuvrait en milieu privé (14/18) et participait à l'enseignement du cursus universitaire de base et à la formation continue (11/19). La plupart des membres du panel travaillaient principalement en clinique, sauf un membre qui enseignait surtout à l'université. Au niveau de la formation des membres du panel, on note dans le tableau II que six participants ont une formation de baccalauréat en physiothérapie, que six ont complété une maîtrise en recherche et que sept ont le titre de *Fellow of the Canadian Academy of Manipulative Physiotherapy (FCAMT)* qui est une reconnaissance de leur expertise dans le domaine musculosquelettique en plus de leur formation de base.

Ce choix des experts correspond, selon nous, aux recommandations pragmatiques de Lubarsky (2013), et ce, dans le contexte de la physiothérapie musculosquelettique. En lien avec notre cadre de référence pour établir le processus de validation de nos tests, le choix des membres du panel de référence est un aspect important de la portion interne du test. Comme démontré par plusieurs études dans le cas du TCS (Charlin, Gagnon, Sauvé, & Coletti, 2007; Sibert, Charlin, Corcos, Gagnon, Grise, *et al.*, 2002), la façon de déterminer le panel a un effet déterminant sur

l'attribution des scores. Cet effet sur les scores du choix du panel pourrait être comparé à l'importance de la sélection des critères de correction dans un cadre plus conventionnel d'évaluation des apprentissages.

4.3 Tests d'évaluation du raisonnement clinique

4.3.1 Description des tests et du domaine évalué

Comme mentionné au chapitre 1, les outils de récolte de données qui ont été retenus pour évaluer le raisonnement clinique des étudiants du programme de physiothérapie sont le test de concordance de scripts (TCS) et les problèmes de raisonnement clinique (PRC). Basés sur la littérature médicale en évaluation, ces tests présentent un potentiel très intéressant pour évaluer le raisonnement clinique en physiothérapie. De plus, compte tenu des différences de format, ces tests pourraient présenter une certaine complémentarité entre eux dans l'évaluation du raisonnement clinique. Les questions des tests portent sur le domaine de la physiothérapie musculosquelettique, celui-ci a été choisi parce qu'il représente le domaine dans lequel les étudiants ont eu le plus grand nombre d'heures de formation dans le curriculum. Il a été décidé de s'en tenir à un domaine spécifique de la physiothérapie pour des questions de faisabilité, compte tenu de la nécessité de recourir à des groupes d'experts cliniques différents si d'autres domaines (pédiatrie, gériatrie, cardiorespiratoire, et neurologie) avaient été choisis. La volonté d'inclure la cohorte des étudiants de 2^e année (qui ont vu principalement des notions dans le domaine musculosquelettique) est également un facteur impliqué dans la décision.

4.3.1.1 Description du TCS

Comme dans l'exemple présenté à la figure 3, une vignette du TCS comprend habituellement quatre parties. La première partie consiste à présenter un scénario clinique authentique. Ce scénario ne donne habituellement pas suffisamment d'informations pour être en mesure de déterminer un diagnostic ou une intervention spécifique. La deuxième partie, l'item, propose un choix de diagnostic ou de traitements présentés sous forme d'hypothèse, énoncée sous la forme : « *vous pensez à* ». Pour ce qui est de la troisième partie, le participant doit envisager l'effet d'une nouvelle donnée sur l'hypothèse proposée. La quatrième partie, demande de faire un

choix, à partir d'une série d'énoncés sous forme d'échelle de type Likert, relativement à l'impact qu'a la nouvelle donnée sur l'hypothèse initiale (Fournier *et al.*, 2008).

Vignette 1

Votre patiente de 64 ans a fait une chute sur sa main droite après avoir glissé sur une plaque de glace. Elle vous consulte en accès direct.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)				
une fracture du scaphoïde	une bonne force à la préhension de la main	-2	-1	0	1	2
une fracture du radius	une coloration bleutée de la main	-2	-1	0	1	2
une fracture du radius	une diminution de 20 degrés en flexion et extension du poignet	-2	-1	0	1	2

-2 : éliminée ou presque; -1 : moins probable; 0 : ni plus, ni moins probable; 1 : plus probable; 2 : certaine ou presque

Figure 3 Exemple de vignette de test de concordance de script comprenant trois items

Les énoncés pour chaque niveau de l'échelle varient selon qu'il s'agisse d'un item de nature d'évaluation ou d'intervention (figure 4). À l'intérieur d'une vignette, trois ou quatre items différents sont présentés aux participants. Chaque item a été développé de façon à être indépendant un de l'autre, bien qu'ils soient issus du même contexte clinique.

Niveaux de l'échelle	Items évaluation	Items intervention
-2	éliminée ou presque	contre-indiqué ou presque
-1	moins probable	peu indiqué
0	ni plus, ni moins probable	ni plus, ni moins indiqué
1	plus probable	plus indiqué
2	certaine ou presque	complètement indiqué ou presque

Figure 4 Énoncés proposés pour chaque niveau de l'échelle de type Likert du TCS.

4.3.1.2 Description des PRC

Tout comme pour le TCS, les PRC incluent également un scénario clinique crédible, mais plus complet et élaboré que dans le cas de ce dernier. Comme on peut le voir dans la figure 5, les informations exhaustives du scénario du PRC sont destinées à permettre l'élaboration d'un diagnostic ou d'une intervention chez le participant. Cinq vignettes cliniques portaient sur le diagnostic et formulait la question ainsi : « *Basé sur les informations du cas, quel est le diagnostic (structure et processus pathologique) le plus probable?* » Par la suite, il est demandé de coter une série de caractéristiques tirées du scénario clinique qui appuient ou questionnent le diagnostic : « *En utilisant l'échelle ci-dessous, déterminez le poids relatif pour chacune des caractéristiques cliniques qui ont influencé votre jugement quant au diagnostic.* » Dans un deuxième temps, il est demandé au participant de refaire le même processus avec un autre diagnostic possible : « *Quel serait le 2^e diagnostic (structure et processus pathologique) le plus probable dans ce cas?* » et de coter à nouveau les caractéristiques du cas cette fois en fonction du nouveau diagnostic. Les scénarios cliniques ont donc été conçus de manière à avoir plus d'un diagnostic possible. Trois vignettes cliniques portaient sur une intervention ciblée d'exercice en demandant : « *Basé sur les informations du cas, quel serait l'exercice le plus pertinent que vous pourriez donner (nommez l'articulation ou le groupe musculaire ciblé et les paramètres de l'exercice)?* » Tout comme pour les vignettes portant sur le diagnostic, il est par la suite demandé de coter une série de caractéristiques tirées du scénario clinique qui appuient ou questionnent l'intervention choisie et de refaire le processus avec une deuxième intervention.

Vignette 1

Votre patient de 68 ans a des douleurs intermittentes (**D1**) et diffuses au niveau postérieur des cuisses et des mollets depuis environ 8 mois. La douleur s'est graduellement installée sans circonstances particulières. Il décrit ses douleurs aux membres inférieurs principalement comme une lourdeur. Aussi, il présente une douleur lombaire droite intermittente (**D2**) qui est stable depuis environ 7 ans. Ses conditions médicales incluent une athérosclérose modérée et une insuffisance rénale contrôlée médicalement. Le patient est fumeur (1 paquet par jour) depuis l'adolescence.

De plus, il est limité dans ses activités de la vie quotidienne qui impliquent de la marche (difficulté à faire plus de 15 minutes) depuis l'apparition de douleurs aux membres inférieurs. Il note une tendance à avoir de l'œdème aux chevilles en fin de journée et des pertes d'équilibre occasionnelles.

Observations générales :

Ambulation : sans particularité

Observations spécifiques :

Posture : pas de déviation du tronc, pas de spasme et pas d'atrophie musculaire visible

ROM debout	Actif
Flexion	40 cm doigts-sol Pas de douleur
Extension	Légère dlr D2 en fin AA Pas de limitation d'AA
Flexion latérale G et D	Légère dlr D2 en fin AA Pas de limitation d'AA
Rotation D et G	Légère dlr D2 en fin AA Pas de limitation d'AA

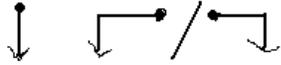
Examen neurologique :

- Dermatomes : bonne discrimination sensitive de L1 à S2
- ROT :
- Patellaire : vifs et symétriques
- Achilléen : non élicité à gauche, diminué à droite
- Extenseurs orteils : vifs et symétriques
- Myotomes : L1-S2 : pas de faiblesse ou de fatigabilité notée sauf pour S1 à gauche
- Clonus : absent
- Cutané plantaire : négatif

Tests spéciaux :

- SLR D et G : pas de douleur de consultation reproduite

Mouvements accessoires :



L4-L5 = reproduit dlr typique au dos D2 et hypomobilité modérée à ces niveaux.

Palpation : pas de douleur notée à la région lombosacrée et au niveau postérieur des membres inférieurs.

Pouls : pédieux et fémoral diminués bilatéralement

Basé sur les informations du cas, quel est le diagnostic (structure et processus pathologique) le plus probable pour **D1** (douleurs intermittentes et diffuses au niveau postérieur des cuisses et des mollets)?

Réponse :

En utilisant l'échelle suivante, déterminez le poids relatif pour chacune des caractéristiques cliniques qui ont influencé votre jugement quant au diagnostic.

+3 = supporte fortement +2 = supporte modérément +1 = supporte légèrement 0 = n'influence pas
 -3 = questionne fortement -2 = questionne modérément -1 = questionne légèrement

Caractéristiques cliniques

Poids relatifs

Âge du patient	
Type de douleur (D1 lourdeur aux membres inférieurs)	
Présence de douleur lombaire D2	
Apparition graduelle	
Athérosclérose modérée	
Fumeur	
Œdème aux chevilles	
Pas de limitations de mouvements lombaires	
Douleur lombaire en extension	
Faiblesse myotome S1 à gauche	
Hypomobilité à L4-L5	
Pouls diminué aux membres inférieurs	
Perte d'équilibre	
SLR négatif bilatéralement	

Quel serait le 2^e diagnostic (structure et processus pathologique) le plus probable dans ce cas pour **D1** (douleurs intermittentes et diffuses au niveau postérieur des cuisses et des mollets)?

Réponse :

En réutilisant l'échelle ci-dessous, déterminez le poids relatif pour chacune des caractéristiques cliniques qui ont influencé votre jugement quant au 2^e diagnostic le plus probable.

+3 = supporte fortement +2 = supporte modérément +1 = supporte légèrement 0 = n'influence pas
 -3 = questionne fortément -2 = questionne modérément -1 = questionne légèrement

Caractéristiques cliniques

Poids relatifs

Âge du patient	
Type de douleur (D1 lourdeur aux membres inférieurs)	
Présence de douleur lombaire D2	
Apparition graduelle	
Athérosclérose modérée	
Fumeur	
Œdème aux chevilles	
Pas de limitations de mouvements lombaires	
Douleur lombaire en extension	
Faiblesse myotome S1 à gauche	
Hypomobilité à L4-L5	
Pouls diminué aux membres inférieurs	
Perte d'équilibre	
SLR négatif bilatéralement	

Figure 5 Exemple de vignette de problèmes de raisonnement clinique

À la différence du TCS, les énoncés en lien avec chaque niveau de l'échelle étaient les mêmes pour les vignettes diagnostiques ou d'interventions (voir figure 6).

Niveaux de l'échelle	Énoncés
-3	questionne fortement
-2	questionne modérément
-1	questionne légèrement
0	n'influence pas
+1	supporte légèrement
+2	supporte modérément
+3	supporte fortement

Figure 6 Énoncés proposés pour chaque niveau de l'échelle de type Likert des PRC

4.3.2 Étapes du développement des tests de raisonnement clinique

Dans le but de produire deux formats de tests qui vont nous permettre d'évaluer le raisonnement clinique des étudiants en physiothérapie, nous avons suivi plusieurs étapes (Downing & Haladyna, 2006) qui sont résumées dans le tableau III.

Tableau III Étapes de développement des tests de raisonnement clinique

1	Construction d'un tableau de spécifications et sélection des situations cliniques
2	Rédaction des vignettes et des items du TCS et des PRC
3	Évaluation des scénarios et des items des TCS et des PRC
4	Administration pilote du TCS et des PRC

Nous reprendrons dans les prochaines parties chacune de celles-ci plus en détail.

4.3.2.1 Construction des tests

Comme recommandé par Charlin *et al.* (2002) Fournier *et al.* (2006) et Groves *et al.* (2002), le TCS et les PRC de la présente étude ont été montés par une équipe de deux personnes. Le chercheur principal de même qu'une chargée de cours ayant de l'expérience dans la rédaction de questions d'examen et connaissant bien le curriculum de physiothérapie ont élaboré les vignettes et les items pour le TCS et pour les PRC. Un processus itératif d'environ 50 heures par concepteur, réparti sur quatre semaines a permis de bâtir la 1^{re} version des deux tests. Les deux concepteurs étant également des cliniciens actifs, ceux-ci ont été en mesure de puiser à travers leurs expériences cliniques pour bâtir des scénarios réalistes inspirés des situations authentiques. Les concepteurs devaient également indiquer quelle était la meilleure réponse

selon eux. Durant cette période, de nombreuses discussions ont permis d'améliorer le travail de chacun des concepteurs. En accord avec notre cadre conceptuel du raisonnement clinique, les questions devaient porter sur des connaissances acquises (avant la fin de la 2^e année) et demander un jugement de l'étudiant en fonction du contexte.

Selon le cadre conceptuel sur la validité de Lissitz & Samuelsen (2007), un des aspects les plus importants du processus de validation dans un contexte éducationnel est la représentativité du contenu du test. Nous avons donc effectué le choix des thèmes des vignettes en fonction d'un tableau de spécifications (voir annexe 6) produit à partir des objectifs du cours *PHT 240 Évaluation en physiothérapie*. Les contextes des vignettes ont été choisis de manière à être représentatifs des 13 histoires de cas utilisées en termes de région anatomique et des conditions pathologiques couvertes par les objectifs du cours. Une proportion plus importante a été accordée à l'évaluation qu'à l'intervention selon les objectifs du cours PHT 240. Bien que le tableau de spécification ait été produit à partir d'un cours de la fin de la 2^e année, les connaissances acquises depuis le début du curriculum étaient aussi nécessaires à l'interprétation de données cliniques demandées par les tests. Le contenu des vignettes et les items ont été construits en voulant diversifier le niveau de difficulté. De plus, nous nous sommes assurés que le contexte clinique était accessible pour un étudiant de 2^e année, mais tout en gardant l'aspect ambigu des scénarios. L'aspect authentique des situations cliniques choisies a également été un critère important dans la production des vignettes et des items produits (Fournier *et al.*, 2008). Le chercheur principal était responsable de la sélection finale des vignettes et items. Le nombre de vignettes construites devait permettre de répondre aux tests en moins de 90 minutes chacun.

4.3.2.2 Rédaction des vignettes et des items du TCS

Plus spécifiquement pour la rédaction des vignettes du TCS nous nous sommes inspirés des recommandations de Fournier (2008) qui sont : 1. Trouver des contextes réalistes (inspirés de la réalité clinique des concepteurs). 2. Déterminer des hypothèses réalistes dans ces contextes. 3. Déterminer les données pertinentes à interpréter en fonction des hypothèses et des contextes. Ces recommandations ont été bien appliquées dans l'ensemble, mais une des difficultés notées par les concepteurs de questions du TCS était de vouloir donner beaucoup d'information dans

la vignette de mise en contexte ou de donner plus d'une nouvelle information à considérer en fonction de l'hypothèse.

Dans le but d'avoir une bonne représentativité du contenu du cours PHT 240 et d'obtenir une bonne consistance interne selon les recommandations de la revue systématique de la littérature de Dory *et al.*, (2012) présentée au chapitre 3, nous avons créé une première version du test qui comportait 85 items basés sur 26 vignettes. Selon les recommandations de ces auteurs, nous avons inclus de trois à quatre items par vignettes.

4.3.2.3 Rédaction des vignettes et des items des PRC

Bien que ce format a été utilisé jusqu'à maintenant pour le diagnostic exclusivement (Dory, Charlin, Vanpee, & Gagnon, 2014; Groves, Dick, McColl, & Bilszta, 2013; Groves *et al.*, 2003a; Groves *et al.*, 2002), nous avons décidé d'inclure dans la présente étude, des vignettes qui demandent d'élaborer un traitement et de déterminer comment les caractéristiques du cas appuient ou questionnent le traitement proposé. Compte tenu de l'éventail des traitements possibles en physiothérapie pour un même cas, nous avons convenu de limiter le choix d'un type de traitement, aux exercices spécifiques.

Les vignettes ont donc été construites de façon à offrir plusieurs possibilités à ce niveau également. Certains indices présents dans les vignettes et accessibles pour des étudiants de 2^e année permettaient d'identifier au moins deux exercices pertinents selon les concepteurs.

Tout comme pour le TCS, nous nous sommes assurés d'avoir la meilleure représentativité des objectifs du cours PHT 240 possible dans le contenu des vignettes des PRC (voir annexe 6). Les thèmes ciblés par les vignettes étaient les mêmes que ceux du TCS, mais en nombre plus restreint. Il a été prévu que les deux tests fassent appel à la même base de connaissance de façon à pouvoir comparer les deux formats. Comme recommandé par Groves *et al.* (2002), dans le but de pouvoir compléter le test en moins de 90 minutes et assurer une certaine représentativité de contenu, huit vignettes ont été créées. Contrairement au TCS peu de recommandations existent quant à la création optimale de vignettes de PRC pour les concepteurs de questions mis à part le côté réaliste des vignettes cliniques (Groves *et al.*, 2002).

4.3.2.4 Évaluations des vignettes et des items des tests de raisonnement clinique

Dans une optique de validation du contenu selon la perspective de Lissitz & Samuelsen (2007), nous avons voulu d'une part en déterminer la qualité et, d'autre part préciser le niveau de difficulté des items élaborés pour nos tests de raisonnement clinique. Nous voulions ainsi nous assurer que les tests permettaient de faire appel au raisonnement clinique sans que d'autres éléments viennent interférer. Cette première étape a d'abord été effectuée par un physiothérapeute enseignant du programme contenu impliqué dans le cours PHT 240 et dans le programme de façon générale, pour recueillir ses commentaires.

Les questions demandées à cet expert de contenu étaient :

Est-ce que le contexte des vignettes est accessible pour un étudiant de 2^e année en physiothérapie? Est-il crédible? Est-ce que les libellés sont clairs et bien construits?

Les réponses à ces questions ont été discutées lors d'une entrevue individuelle. Globalement, selon les commentaires recueillis de cet enseignant le niveau de difficulté des vignettes semblait adéquat, pour des étudiants de 2^e année en physiothérapie. Toujours selon cet enseignant, les informations contenues dans ces vignettes étaient crédibles cliniquement, quoiqu'un peu complexes et longues surtout au niveau des vignettes des PRC. Plusieurs points ont été discutés pour simplifier et raccourcir certaines vignettes. De plus, des suggestions et commentaires pour clarifier les libellés ont aussi été recueillis directement dans les textes des tests complétés par le physiothérapeute enseignant. À la suite de ces recommandations, plusieurs ajustements ont été faits dans les deux tests au niveau des libellés et de la vraisemblance des informations cliniques des cas. Aucune vignette n'a été retranchée à cette étape.

Comme suggéré dans plusieurs cadres de validation, dont celui de Lissitz & Samuelsen (2007), nous avons procédé à une autre étape du processus de validation du contenu en administrant le test à un groupe pilote. Deux physiothérapeutes (une physiothérapeute récemment graduée du programme de l'Université de Sherbrooke, une physiothérapeute ayant huit ans d'expérience et diplômée d'une autre université) ont donc été recrutées dans le but d'estimer le temps requis pour compléter le test et d'obtenir un échantillon de réponses possibles aux items.

Ces physiothérapeutes ont également été invitées à donner des commentaires sur le format et la clarté des vignettes, sur leur niveau de difficulté, et ce, directement dans le texte des tests et lors d'une entrevue individuelle.

À la suite de cette administration pilote, plusieurs commentaires ont été émis par les deux physiothérapeutes. Parmi les principales suggestions des participants à l'étude pilote, celles-ci recommandaient encore une fois de simplifier et raccourcir les vignettes et ont mentionné certaines ambiguïtés dans les informations cliniques données surtout au niveau des PRC. Nous avons donc à nouveau réajusté les libellés et les informations cliniques contenues dans les tests. De plus, la difficulté de répondre à des questions liées à l'intervention dans le format des PRC a été soulevée par les deux physiothérapeutes, et ce, même si nous avons restreint le choix des interventions à des exercices. Pour ce qui est du TCS, trois items ambigus ont été retirés de la version finale. Pour les PRC, plusieurs ajustements ont été faits pour diminuer la quantité d'information donnée dans les vignettes et ne garder que l'essentiel, le nombre de vignettes est resté à huit après cette phase de consultation. La durée de 90 minutes a été considérée comme suffisante pour compléter chacun des tests, le TCS étant le plus court des deux, celui-ci ayant été complété en moins d'une heure lors du pilote.

4.3.3 Collecte de données

Panel d'experts

La collecte de données des participants du panel d'experts s'est déroulée pendant les mois de mars et avril 2014. Ceux-ci ayant d'abord été recrutés par courriel en mars 2014 et informés au préalable des buts de l'étude et du temps requis nécessaire pour compléter les tests. Environ une semaine après que ceux-ci ont signifié leur accord par retour de leur formulaire de consentement, les participants experts ont reçu une version électronique des tests qu'ils pouvaient remplir à un moment qui leur convenait. Aucune limite de temps n'a été fixée, bien que la consigne était que normalement chaque test se complète en moins de 90 minutes chacun. Des instructions sur la façon de répondre aux questions apparaissaient sur la première page de chaque test, mais les participants du panel pouvaient rejoindre le chercheur par téléphone ou courriel s'ils avaient des questions sur les tests. Il leur était spécifié de ne pas utiliser de ressources externes (livres,

collègues, etc.) pour compléter les tests. Ils avaient une semaine pour compléter et retourner les tests par courriels, ce qui a été fait par tous les participants. Les participants avaient également un sondage à remplir sur leurs perceptions des tests et ceux-ci devaient fournir quelques informations sur leurs expériences cliniques et milieux de pratique (annexe 8). Aucun des participants (sauf un qui avait déjà utilisé le TCS) n'était familier avec les formats des tests de raisonnement clinique de l'étude.

Étudiants

Pour ce qui est de la collecte de donnée auprès des étudiants, celle-ci s'est déroulée en classe à un moment différent pour chaque cohorte. La collecte a eu lieu en mars 2014 pour la cohorte de 4^e année, en mai 2014 pour la cohorte de 2^e année et en juin 2014 pour la cohorte de 3^e année. L'horaire des stages des étudiants a été la contrainte majeure pour déterminer la date des collectes des données. Les participants ont été informés au préalable de l'étude lors d'une séance d'information donnée par le chercheur. Cette séance décrivait entre autres les buts de l'étude, les avantages et inconvénients d'y participer, de même que le temps requis nécessaire pour compléter les tests. Les étudiants n'étaient pas familiers avec les formats de question de type TCS ou PRC, bien qu'un format relativement similaire aux PRC est utilisé dans un contexte d'évaluation formative lors des séances d'apprentissage du raisonnement clinique. Lors des collectes en salle de classe, le chercheur était présent pour expliquer brièvement les formats des questions en reprenant les exemples qui apparaissaient sur les premières pages des tests. Les étudiants pouvaient poser des questions pendant le déroulement du test. Les étudiants avaient un maximum de 90 minutes pour compléter individuellement chacun des tests et pouvaient les compléter dans l'ordre de leurs choix. Les étudiants pouvaient prendre une pause entre les deux tests, mais celle-ci ne permettait pas d'obtenir plus de temps que les trois heures prévues pour les deux tests combinés. Les étudiants avaient par la suite, un sondage à remplir sur leurs perceptions des tests (annexe 7). Les étudiants n'obtenaient aucune rémunération en échange de leur participation, mais avaient la chance de participer à un tirage pour un iPad. De plus, les participants ont eu une séance de rétroaction pour chaque cohorte où les réponses des items ont été discutées en groupe.

4.3.4 Méthodes pour établir les scores

Score au TCS

Pour établir les scores des TCS nous avons utilisé la méthode des scores combinés (*aggregate scoring method*) qui est actuellement la plus utilisée dans le contexte du TCS (Dory *et al.*, 2012). Certaines critiques (Lineberry, Kreiter, & Bordage, 2013) de cette méthode sont présentes dans la littérature et nous y reviendrons plus en détail dans la discussion. Dans le cadre de cette étude, les scores au TCS ont été établis en deux étapes. D'abord, une pondération est établie pour chaque item à partir des réponses des experts du panel. Dans un deuxième temps on attribue un score à chaque participant à partir de la pondération des réponses du panel d'experts. Ainsi, les participants qui opteront pour la réponse choisie par le plus grand nombre d'experts (réponse modale) recevront 1 point. Les autres réponses choisies recevront une partie des points en fonction du nombre d'experts qui ont choisi cette réponse, par exemple à la question 1 (Tableau IV), la réponse modale est -1 (*moins probable*) douze experts l'ayant choisi, de leur côté, cinq experts ont choisis -2 (*éliminée ou presque*), ce qui donnera $5/12$ (0,42) pour cette réponse, la réponse 0 (*ni plus, ni moins probable*) ayant été choisi par 2 experts ce qui donnera $2/12$ (0,17). On n'accordera aucun point aux réponses qui n'ont pas été choisies par les experts. Le score total pour chaque participant est la somme de chaque item, ce score sera par la suite exprimé en pourcentage.

Le score attribué aux experts est calculé de la même façon, mais en retirant le profil de réponse de chaque expert des réponses du panel avant d'attribuer le pointage. Cette méthode permet de diminuer l'influence de son propre profil de réponse sur le score attribué à l'expert (Lambert *et al.*, 2009).

Tableau IV TCS (Item 1) Exemple de grille pour établir le score à un item

	-2	-1	0	1	2	
Nombres de réponses des experts	5	12	2	0	0	Réponse modale : - 1
Pointage	5/12	12/12	2\12	0\12	0\12	Nombre divisé par 12
Score pour l’item	0,42	1	0,17	0	0	Point attribué

Score au PRC

Pour établir les scores au PRC, nous avons suivi initialement la procédure décrite par Groves *et al.*, 2002. Nous reviendrons dans la discussion sur les difficultés inhérentes à cette procédure et aux raisons qui nous amènent à changer certains aspects de cette méthode. Tout comme pour le TCS, nous avons établi le score du PRC en fonction des réponses du panel d’experts. Par contre, à la différence du TCS, les PRC comportent deux parties distinctes qui contribuent en part égale au score selon Groves *et al.* 2002. La première partie est une réponse de type ouverte qui demande d’établir un diagnostic ou une intervention dans le contexte de la vignette clinique. Une fois les réponses des experts recueillies, celles-ci sont catégorisées selon une taxonomie clinique. Plusieurs termes diagnostiques doivent ainsi être considérés pour leur équivalence. Le tableau V présente les différents diagnostics par les experts pour la vignette 2. Le tableau VI illustre les regroupements de diagnostics, une fois cette catégorisation effectuée et le pointage attribué pour le choix de ces diagnostics.

Tableau V Énoncés des diagnostics des experts sur la vignette 2 (1^{re} partie)

- Tendon coiffe des rotateurs
- Capsulite phase douloureuse
- Déchirure coiffe des rotateurs
- Rupture partielle coiffe
- Bursite sous-acromiale D
- Bursopathie sous acromio-deltoïdienne en phase inflammatoire sub aiguë
- Tendinopathie de la coiffe des rotateurs
- Bursopathie sous-deltoïdienne

- Syndrome d'accrochage causant une tendinopathie aiguë de la coiffe des rotateurs
- Rupture de la coiffe des rotateurs
- Capsulite gléno-humérale
- Tendinose de la coiffe des rotateurs
- Déchirure de la coiffe des rotateurs, stade aigu
- Bursite aiguë
- Capsulite aiguë épaule droite
- Capsulite rétractile stade 1
- Syndrome d'abutement (tendinopathie aiguë coiffe des rotateurs)
- Tendinite de la coiffe des rotateurs

Les experts ont par la suite à déterminer un 2^e diagnostic plausible dans ce cas.

Énoncés des diagnostics des experts sur la vignette 2 (2^e partie)

- Atteinte inflammatoire subaiguë coiffe des rotateurs droite
- Déchirure incomplète, mais augmentée du supra-épineux
- Capsulite rétractile
- Syndrome d'accrochage causant bursite aiguë
- Capsulite de l'épaule
- Bursite sous acromiale
- Accrochage sous acromiale
- Bursite sous-acromiale aiguë
- Tendinalgie aiguë de la CDR
- Déchirure complète de la coiffe
- Bursite aiguë
- Déchirure complète coiffe des rotateurs
- Tendon coiffe des rotateurs (Déchirure complète)
- Bursite sous acromiale subaiguë
- Capsulite
- Bursite sous acromiale
- Tendinopathie de la coiffe des rotateurs D

En compilant l'ensemble des diagnostics pour la vignette 2, on obtient les catégories suivantes listées dans le tableau VI :

Tableau VI Catégorie de diagnostic émis par les experts pour la vignette 2 et pointage attribué pour ce diagnostic

PRC 2	Diagnostic émis	Nombre d'experts ayant choisi ce diagnostic	Pointage	Score attribué
	Bursopathie	11	11/11	1,00
	Déchirure coiffe des rotateurs	9	9/11	0,82
	Tendinopathie	8	8/11	0,73
	Capsulite	7	7/11	0,64

Pour qu'une catégorie de diagnostic soit retenue, nous avons convenu qu'il devait y avoir plus d'un expert mentionnant ce diagnostic. Le score de la deuxième portion des PRC est établi en fonction de la cotation des caractéristiques qui appuient ou questionnent le diagnostic ou le traitement que les experts ont choisi pour chacun des cas (Groves *et al.*, 2013).

Nous avons choisi de soumettre les caractéristiques pour qu'elles soient cotées individuellement (Groves *et al.*, 2013) plutôt que de demander aux participants de choisir et lister les caractéristiques qui remettent en question ou supportent leurs choix diagnostiques (ou thérapeutiques) à partir de la vignette comme dans la première version des PRC (Groves *et al.*, 2002). Ce choix a été fait pour limiter la difficulté de l'interprétation des réponses des participants.

Dans le tableau VII, on constate que parmi les onze experts qui ont suggéré la bursopathie comme diagnostic dans la vignette 2, six d'entre eux considèrent que l'âge de la patiente dans la vignette *n'influence pas* (0) leur diagnostic, ce qui est considéré comme la réponse modale pour cette caractéristique dans le profil bursopathie. Cette réponse donnera un point si elle est choisie par un participant. Trois experts ont, pour leur part, choisi que cette caractéristique *supportait légèrement* (1) et deux ont déterminé que cette caractéristique *supportait modérément* (2) leur diagnostic. De façon similaire à la façon que nous avons utilisée pour établir le score du TCS, le choix de la réponse modale du panel d'experts se verra attribuer un point et les autres

réponses choisies par le panel se verront attribuer un score en proportion du nombre d'experts qui ont choisi cette réponse.

Tableau VII Profil des réponses sur les caractéristiques de la bursopathie à la vignette 2 à partir des réponses des experts qui ont choisi ce diagnostic (n=11)

Caractéristiques cliniques	-3	-2	-1	0	1	2	3
Âge de la patiente	0	0	0	6	3	2	0
Intensité de la douleur	0	0	0	0	2	7	2
Apparition à la suite du grand ménage	0	0	0	0	2	3	6
Antécédent de déchirure de la coiffe des rotateurs	0	0	0	0	6	3	2
Patiente droitrière	0	0	0	1	5	3	2
Diabète de type 1	0	0	1	8	1	1	0
Insuffisance cardiaque modérée	0	0	2	9	0	0	0
Rythme scapulo-huméral inversé lors de l'élévation	0	0	0	1	5	4	1
Légère atrophie deltoïde droit	0	1	3	3	1	3	0
Patron de limitations des mouvements actifs et passifs	0	0	1	0	4	4	2
Les sensations de fin de mouvement (SFM)	0	0	0	0	1	4	6
Patron de réponses isométriques résistantes	0	1	1	0	3	5	1
Glissement inférieur soulage légèrement la douleur	0	0	0	0	3	7	1
Douleur à la palpation au pourtour de l'acromion	0	0	0	0	3	5	3

Le tableau VIII illustre le pointage attribué pour l'item 1 du profil bursopathie de la vignette 2. Comme on le voit dans ce tableau, le processus d'attribution des scores aux participants en fonction des réponses du panel d'experts est similaire à celui utilisé par le TCS.

Tableau VIII Pointage attribué pour l'item 1 du profil bursopathie de la vignette 2

PRC Item 1 de la vignette 2	-3	-2	-1	0	1	2	3
(bursopathie) Grille de correction							
Nombres de réponses des experts	0	0	0	6	3	2	0
Pointage	0	0	0	6/6	3/6	2/6	0
Score pour l'item	0	0	0	1	0,50	0,33	0

Pour chaque diagnostic nous avons établi le profil de réponse des caractéristiques associées et une grille de correction différente est utilisée en fonction du diagnostic émis par le participant. Le tableau IX illustre le profil de réponse pour le diagnostic capsulite choisi sept fois par le panel d'expert et le tableau X illustre le pointage attribué par l'item 1 dans le profil capsulite de la vignette 2.

Tableau IX Profil des réponses sur les caractéristiques de la capsulite à la vignette 2 à partir des réponses de sept experts qui ont choisi ce diagnostic n=7

Caractéristiques cliniques	-3	-2	-1	0	1	2	3
Âge de la patiente	0	0	0	1	1	3	2
Intensité de la douleur	0	0	0	0	2	4	1
Apparition à la suite du grand ménage	1	0	3	1	2	0	0
Antécédent de déchirure de la coiffe des rotateurs	0	0	1	0	2	4	0
Patiente droitière	0	0	0	4	2	1	0
Diabète de type 1	0	0	0	0	1	4	2
Insuffisance cardiaque modérée	0	0	0	5	0	2	0
Rythme scapulo-huméral inversé lors de l'élévation	0	0	0	0	3	3	1
Légère atrophie deltoïde droit	0	2	0	1	3	1	0
Patron de limitations des mouvements actifs et passifs	0	0	0	0	1	0	6
Les sensations de fin de mouvement (SFM)	0	0	1	1	0	3	2
Patron de réponses isométriques résistées	0	3	0	1	2	1	0
Glissement inférieur soulage légèrement la douleur	0	2	1	2	2	0	0
Douleur à la palpation au pourtour de l'acromion	1	0	1	2	3	0	0

Tableau X Pointage attribué pour l'item 1 du profil capsulite de la vignette 2

PRC Item 1 de la vignette 2	-3	-2	-1	0	1	2	3
(capsulite) Grille de correction							
Nombres de réponses des experts	0	0	0	1	1	3	2
Pointage	0	0	0	1/3	1/3	3/3	2/3
Score pour l'item	0	0	0	0,33	0,33	1	0,67

Le tableau XI résume l'attribution des scores pour le candidat 4U2 selon la méthode décrite par Groves *et al.* 2002. Le candidat 4U2 ayant choisi le diagnostic de bursopathie (réponse modale des experts), il se voit attribuer le score de 1/1 (100 %) pour la 1^{re} partie.

Par la suite, un score est attribué pour chacune des caractéristiques en fonction du profil choisi par le panel d'experts. Un 2^e score est attribué de la même façon pour la 2^e hypothèse diagnostique et ses caractéristiques. Le score total de la vignette 2 est calculé en effectuant la moyenne des 2 parties.

Tableau XI Exemple de calcul de la vignette 2 pour le candidat 4U2

Caractéristiques cliniques Diagnostic : bursopathie	réponse choisie	score attribué
1. Âge de la patiente	0	1,00
2. Intensité de la douleur	3	0,29
3. Apparition à la suite du grand ménage	3	1,00
4. Antécédent de déchirure de la coiffe des rotateurs	1	1,00
5. Patiente droitrière	3	0,40
6. Diabète de type 1	0	1,00
7. Insuffisance cardiaque modérée	0	1,00
8. Rythme scapulo-huméral inversé lors de l'élévation	-3	0,00
9. Légère atrophie deltoïde droit	-1	1,00
10. Patron de limitations des mouvements actifs et passifs	3	0,50
11. Les sensations de fin de mouvement (SFM)	-1	0,00
12. Patron de réponses isométriques résistantes	-1	0,20
13. Glissement inférieur soulage légèrement la douleur	3	0,14
14. Douleur à la palpation au pourtour de l'acromion	3	0,60
Score pour les caractéristiques du 1 ^{er} diagnostic		8,1/14=57,8 %
Score pour le 1 ^{er} diagnostic (bursopathie)		100 %

Caractéristiques cliniques	réponse choisie	score attribué
Diagnostic : déchirure de coiffe des rotateurs		
1. Âge de la patiente	2	0,00
2. Intensité de la douleur	2	0,00
3. Apparition à la suite du grand ménage	1	1,00
4. Antécédent de déchirure de la coiffe des rotateurs	3	0,00
5. Patiente droitrière	0	1,00
6. Diabète de type 1	-1	1,00
7. Insuffisance cardiaque modérée	0	1,00
8. Rythme scapulo-huméral inversé lors de l'élévation	2	0,00
9. Légère atrophie deltoïde droit	-2	1,00
10. Patron de limitations des mouvements actifs et passifs	-1	0,33
11. Les sensations de fin de mouvement (SFM)	-2	0,67
12. Patron de réponses isométriques résistantes	2	1,00
13. Glissement inférieur soulage légèrement la douleur	-1	0,250
14. Douleur à la palpation au pourtour de l'acromion	2	1,00
Score pour les caractéristiques du 2 ^e diagnostic	1	10,6/14= 75,7 %
Score pour le 2 ^e diagnostic (déchirure de coiffe des rotateurs)		82 %

Comme on le voit dans le tableau XII, de deux à quatre diagnostics/interventions ont été acceptés par vignette. Dans certains cas, le nombre de fois qu'un diagnostic a été choisi par les experts dépasse le nombre total d'experts, car certains d'entre eux ont choisi un diagnostic/intervention très similaire comme deuxième choix et celui-ci entrait dans la même catégorie lors du regroupement.

Tableau XII Tableau d'attribution des scores pour les réponses ouvertes du PRC

PRC	Diagnostic émis	Nombre d'experts ayant choisi ce diagnostic	Score attribué
1a	Claudication intermittente vasculaire	18	1
1b	Sténose spinale	17	0,94

2a	Bursopathie	11	1
2b	Déchirure coiffe des rotateurs	9	0,82
2c	Tendinopathie	8	0,73
2d	Capsulite	7	0,64
3a	Étirement/mobilité	20	1
3b	Renforcement	11	0,55
3c	Mobilité neurale	5	0,25
4a	Arthrose hanche	20	1
4b	Tendinopathie iliopsoas	11	0,55
5a	Composante cervicale\neurale	18	1
5b	Épicondylalgie/tendinopathie	14	0,78
5c	Défilé thoracique	3	0,17
6a	Déchirure ménisque interne	18	1,00
6b	Entorse du ligament latéral interne	16	0,89
7a	Exercices d'extension axiale/activation court fléchisseurs	12	1
7b	Étirement des sous-occipitaux	9	0,75
7c	Étirement du trapèze sup/angulaire gauche	8	0,67
8a	Renforcement moyen fessier	22	1
8b	Étirement de la bandelette ilio-tibiale	10	0,45

Pour chacune des vignettes, le même processus d'attribution des scores a été effectué. Le score total des PRC est la moyenne de toutes les vignettes (n=8). Le tableau XIII décrit un exemple de la compilation des scores des huit vignettes du PRC pour établir le score total du candidat 4U2. Comme on l'aperçoit dans le tableau, les scores des hypothèses et des caractéristiques de chaque hypothèse sont additionnés, ces deux composantes ayant une valeur dans le score final (Groves *et al.*, 2002).

Tableau XIII Compilation des scores des huit vignettes du PRC pour le candidat 4U2 selon la méthode préconisée par Groves (2002)

Vignette	Hypothèse 1	Hypothèse 1 caractéristiques	Hypothèse 2	Hypothèse 2 caractéristiques	Score total %
1	0,0	0,0	95,0	51,4	36,6
2	100	57,8	82,0	75,7	78,9
3	100	48,1	63,0	51,8	65,7
4	100	48,6	63,0	68,6	70,1
5	100	55,3	74,0	55,3	71,2
6	100	45,7	84,0	50,0	69,9
7	69,0	67,3	69,0	63,6	67,2
8	100	68,8	53,0	25,0	61,7
Score du candidat 4U2 aux PRC					65,2

En calculant la cohérence interne (alpha de Cronbach) des scores des 8 vignettes obtenus, nous avons obtenu des valeurs très faibles à 0,37 pour l'ensemble des trois cohortes. En reprenant les calculs de cohérence interne par items, on remarque que celle-ci est particulièrement faible (0,07) pour les 16 items reliés aux réponses ouvertes de diagnostic/intervention et très bonne (0,89) pour les 202 items de justification des caractéristiques. Considérant le fait que les items de réponses ouvertes de diagnostic/intervention sont un des aspects distinctifs de ce test, mais que l'erreur de mesure liée à ces items est élevée, nous avons fait le choix de leur attribuer une importance moins élevée que celle proposée par Groves *et al.* (2002). Nous avons donc établi le score des PRC en attribuant une valeur égale aux items de réponses ouvertes qui établissent un diagnostic (ou un traitement) et aux items de justifications de ces caractéristiques. Le tableau XIV démontre un exemple du calcul effectué pour établir le score du candidat 4U2 en compilant le résultat des 218 items.

Tableau XIV Compilation des scores des huit vignettes du PRC pour le candidat 4U2 en attribuant une valeur égale aux items de réponses ouvertes

Item 1	Item 2	Item 3	Item ...	Item 218	Total 218	sur	Score total %
0,17	1,00	,63	...	0,0	115,9		
Score du candidat 4U2 aux PRC							53,2

Les avantages et les limites de l'attribution des scores selon la méthode originale préconisée par Groves et les implications de notre choix d'adapter cette méthode seront traités dans le chapitre de la discussion des résultats.

Comme pour le TCS, le score attribué aux experts est calculé de la même façon, mais en retirant le profil de réponse de chaque expert des réponses du panel avant d'attribuer le pointage pour la portion justification. Cette attribution d'un score aux experts pour les deux tests nous permettra par la suite de déterminer si les scores rendent possible une discrimination des groupes. Dans le cadre conceptuel de Lissitz & Samuelsen (2007), des scores qui discriminent les groupes peuvent être considérés comme une des évidences du réseau nomologique du trait latent relié au test.

4.3.5 Temps de passation des tests de raisonnement clinique

Le tableau XV fait état du temps de passation moyen ainsi que le temps minimum et maximum utilisé par cohorte et par test. On note qu'en moyenne le temps de passation moyen se situait entre 41 et 48 minutes pour le TCS. Du côté des PRC, le temps de passation moyen se situait entre 69 et 74 minutes. Le temps de passation des tests est une des informations pertinentes en lien avec la qualité du processus de collecte de données. Les temps de passation observés lors de cette recherche se situent près des valeurs attendues pour le nombre d'items de chacun des deux formats de test (Dory *et al.*, 2014; Dory *et al.*, 2012; Groves *et al.*, 2013). De plus, les valeurs minimales observées des temps de passation des tests sont compatibles avec une passation effectuée selon l'effort attendu. Cette information était importante considérant la possibilité que certains étudiants participent à l'étude seulement dans une optique de gain secondaire (tirage).

Tableau XV Temps de passation moyen, minimum et maximum des tests par cohorte

Test	Temps de passation	2 ^e année (min.)	3 ^e année (min.)	4 ^e année (min.)
TCS	Moyenne	46 (±14)	41 (±11)	48 (±12)
	Min-Max	21-70	25-60	33-80
PRC	Moyenne	74 (±14)	69 (±12)	70 (±14)
	Min-Max	50-90	46-90	36-90

4.3.6 Résultats du questionnaire sur la perception et appréciation des tests de raisonnement clinique

La prochaine section expose les résultats du questionnaire anonyme rempli par les étudiants et les experts quant à la nature et la pertinence des formats des tests de raisonnement clinique. La totalité des cohortes de 2^e et 3^e année, 17 étudiants de 4^e année et 16 experts ont rempli le questionnaire. Ce questionnaire a été conçu de manière à nous informer sur la perception des participants au sujet de la qualité de la construction des tests et la représentativité du contenu. Les réponses des participants pourront ainsi être utilisées comme une source d'évidence de ces qualités en lien avec le processus de validation des tests.

4.3.6.1 Première question

Dans le but de vérifier la perception des étudiants quant aux liens entre les items des tests et les connaissances apprises dans le programme nous avons posé la question suivante : « *Est-ce que les scénarios cliniques étaient en lien avec les connaissances apprises dans le curriculum?* »

Comme le démontre le tableau XVI, la grande majorité (entre 91 et 100 % selon les cohortes) des étudiants ont trouvé que les items étaient très en lien ou complètement en lien avec les connaissances vues dans le curriculum. On ne note pas de différences importantes entre les cohortes à ce niveau.

La perception des étudiants quant au lien entre les questions des PRC et les connaissances est très similaire à la perception pour le TCS. Encore une fois la grande majorité (entre 91 et 100 % selon les cohortes) des étudiants ont trouvé que les items étaient très en lien ou complètement en lien avec les connaissances vues dans le curriculum. Le fait que 91 % des étudiants de 2^e

année mentionnent que les items étaient *très en lien ou complètement en lien* avec les connaissances vues dans le programme supporte le fait que le niveau du test était accessible pour un étudiant de 2^e année au niveau du contenu et de la complexité de la tâche.

Cette question n'a pas été incluse dans le questionnaire des experts étant donné que la plupart de ceux-ci ne connaissaient pas suffisamment le programme de physiothérapie de l'Université de Sherbrooke pour y répondre.

Tableau XVI Fréquences relatives des réponses à la question 1 du sondage aux étudiants

TCS	Pas du tout en lien	Un peu en lien	Moyennement en lien	Très en lien	Complètement en lien
<i>TCS (%)</i>					
<i>2^e année</i>			9	67	24
<i>3^e année</i>			3	63	34
<i>4^e année</i>				62	38
<i>PRC (%)</i>					
<i>2^e année</i>			9	66	25
<i>3^e année</i>			3	63	34
<i>4^e année</i>				56	44

Commentaires qualitatifs

Dans la section des commentaires qualitatifs en lien avec cette section, en général peu de commentaires ont été émis et ceux-ci supportaient, en général, le fait que les étudiants avaient des connaissances adéquates pour répondre aux items des PRC et du TCS.

Commentaire d'un étudiant de 2^e année : « Le cours d'évaluation en physiothérapie nous aidait à être préparés à ces questions. » Commentaire d'un étudiant de 3^e année : « Peu de choses qu'on n'avait pas vues, peut-être les bursites ».

4.3.6.2 Deuxième question

La deuxième question du questionnaire portait sur la difficulté de s'adapter à ce type d'item étant donné que les étudiants n'avaient pas expérimenté ces formats auparavant.

Plus précisément le libellé de la question était : « Est-ce que vous avez trouvé difficile de vous adapter à ce type de questions? »

Le tableau XVII démontre plus de variation des réponses que pour la première question du sondage. Globalement, la majorité des étudiants (entre 69 à 81 %) ont trouvé : *un peu ou moyennement difficile* de s'adapter au format TCS. Par contre, entre 13 et 18 % des étudiants *n'ont pas du tout trouvé difficile* de s'adapter à type d'item et entre 6 à 12 % des étudiants ont trouvé *très difficile* de s'y adapter. Un étudiant (3 %) de 3^e année a de son côté trouvé *extrêmement difficile* de s'adapter à ce type de question.

Encore une fois les résultats du questionnaire pour les PRC se rapprochaient de ceux pour le TCS. Entre 56 et 75 %, des répondants ont trouvé *un peu difficile ou moyennement difficile* de s'adapter à ce type d'item. De leur côté entre 13 et 25 % *ont trouvé pas du tout difficile* de s'y adapter et entre 13 et 19 % ont trouvé cette même tâche très difficile. Deux étudiants (6 %) de 3^e année ont trouvé *extrêmement difficile* de s'adapter à ce type d'item. La proportion des étudiants qui ont trouvé très difficile à extrêmement difficile de s'adapter semble un peu supérieur pour le PRC que le TCS, ce qui est surprenant étant donné qu'un format similaire est utilisé dans le programme de physiothérapie au niveau formatif et sommatif. Cette difficulté perçue par une certaine proportion des étudiants ajoute à la complexité cognitive de la tâche et pourrait être perçue comme une source de variance non souhaitable (Downing, 2005). Des séances de familiarisations formatives aux formats du TCS et des PRC pourraient probablement aider à ce niveau, nous reviendrons sur cette dimension dans la discussion.

Globalement, les experts partagent la perception des étudiants sur la difficulté de s'adapter à ces deux formats de question d'évaluation du raisonnement clinique, 75 % de ceux-ci ont trouvé *un peu ou moyennement difficile* de s'adapter au TCS et 63 % pour les PRC. Dans une proportion similaire aux étudiants, 19 % des experts *n'ont pas trouvé difficile du tout* de s'adapter au format du TCS contre 25 % pour le format des PRC. Deux experts ont trouvé, par contre, *très difficile* de s'adapter au format du PRC et un expert a trouvé *extrêmement difficile* de s'adapter au format du TCS.

Tableau XVII Fréquences relatives des réponses à la question 2 du sondage aux étudiants

	Pas du tout difficile	Un peu difficile	Moyennement difficile	Très difficile	Extrêmement difficile
TCS (%)					
<i>2^e année</i>	18	42	27	12	-
<i>3^e année</i>	18	31	38	9	3
<i>4^e année</i>	13	50	31	6	-
<i>Expert</i>	19	50	25	-	6
PRC (%)					
<i>2^e année</i>	19	31	31	19	
<i>3^e année</i>	25	34	22	13	6
<i>4^e année</i>	13	25	50	13	-
<i>Expert</i>	25	44	19	12	-

Commentaires qualitatifs

Pour cette question, un plus grand nombre de commentaires ont été recueillis pour le TCS, plusieurs commentaires portaient sur le format inhabituel des items : commentaire d'un étudiant de 2^e année : « *Très différent des examens habituels. Difficile de sélectionner une réponse qui correspond à ce que l'on veut réellement dire.* »

D'autres commentaires d'étudiants de 3^e année portent sur la particularité des libellés des items du TCS : « J'ai eu de la difficulté à coter lorsque je n'étais pas d'accord à la base avec le traitement ou diagnostic de la colonne de gauche » « J'avais de la difficulté à oublier l'hypothèse précédente. » « J'étais très tenté de répondre en accordant plus d'importance à la section ce que vous pensez qu'à la section ce que vous avez trouvé. »

Pour le PRC un étudiant de 3^e année commente une difficulté sur le format : « Difficile seulement lorsqu'une 2^e hypothèse ne me venait pas en tête, il devenait peu pertinent de remplir la grille sur une hypothèse floue. »

Pour le PRC quelques commentaires portaient sur la quantité d'information :

Commentaires d'étudiants de 3^e année : « Beaucoup plus long, je perds le focus », « Parfois trop d'informations, difficile de s'orienter. Plusieurs réponses possibles dans ma tête ».

Plusieurs étudiants ont trouvé que le format des PRC ressemblait à un outil utilisé pour l'évaluation formative dans le programme de physiothérapie.

Commentaires d'étudiants de 2^e année : « *Relativement semblable aux fiches que nous devons remplir durant la session.* » Un autre commentaire d'un étudiant de 3^e année soulève en plus une question pertinente à ce format d'item : « *Ressemble beaucoup à nos grilles, mais tout comme dans celles-ci, est-ce que ce qui questionne l'hypothèse 1 soutient l'hypothèse 2?* »

Du côté des experts, pas de commentaires spécifiques ont été émis sur la difficulté de s'adapter au TCS et aux PRC.

4.3.6.3 Troisième question

La dernière partie du questionnaire portait sur la pertinence des formats TCS et PRC pour évaluer le raisonnement clinique. « *Est-ce que vous trouvez ce type d'évaluation pertinent pour évaluer le raisonnement clinique?* »

Le tableau XVIII résume les réponses à la question 3 du sondage. Pour le format TCS, les étudiants et les experts trouvent, dans des proportions variant entre 69 % (4^e année) et 88 % (experts) que le format TCS soit *très ou extrêmement pertinent* pour évaluer le raisonnement clinique. Par contre, un expert (6 %) a trouvé le format TCS pas du tout pertinent dans le contexte de l'évaluation du raisonnement clinique. Quant à la pertinence du format PRC la perception des étudiants et des experts ressemblent à la perception pour le TCS; entre 66 % (2^e année) et 94 % (experts) trouvent que le PRC est *très ou extrêmement pertinent* pour évaluer le raisonnement clinique. Comme pour le TCS, un expert (la même personne) trouve que le format PRC *n'est pas du tout pertinent* pour évaluer le raisonnement clinique.

Tableau XVIII Fréquences relatives des réponses à la question 3 du sondage aux étudiants

	Pas du tout pertinent	Un peu pertinent	Moyennement pertinent	Très pertinent	Extrêmement pertinent
<i>TCS (%)</i>					
<i>2^e année</i>			24	55	21
<i>3^e année</i>		3	18	63	16
<i>4^e année</i>			31	69	
<i>Expert</i>	6	6		50	38

PRC (%)					
2^e année			34	47	19
3^e année			19	62	19
4^e année		6	25	63	6
Expert	6			63	31

Commentaires qualitatifs

Sur la pertinence du TCS et des PRC comme outils d'évaluation du raisonnement clinique plusieurs étudiants ont émis le souhait d'avoir une section qui permet de justifier ses réponses : « *L'approche est bonne, peut-être laisser place à une justification.* »; « *Ce serait bien de pouvoir justifier nos points de vue au besoin.* »

Du côté des experts, certains ont aussi exprimé ce souhait de laisser une place à la discussion de la réponse : « Il y a peu d'information dans ce type de questionnaire, un complément de discussion est essentiel selon moi parce qu'il n'y a pas qu'une bonne réponse et beaucoup de : ça dépend de... »; « Excellent pour les étudiants. Espace pour court narratif de raisonnement pourrait aider à mieux cerner lorsque les réponses proposées portent à interprétation »

Du côté des PRC, un étudiant de 3^e année commente sur l'échelle : « Il était difficile de noter de -3 à +3; trop de choix ». Un expert a également mentionné : « Pas certain que l'échelle avec 3 niveaux de + et 3 de - (7 niveaux en tout) soit la meilleure. J'en aurais mis 5 en tout -2,-1, 0, +1, +2... »

4.4 Analyse de contenu des formulaires de stage

L'analyse de contenu se définit comme un ensemble de méthode d'analyse de documents textuels permettant d'en dégager le sens (Mucchielli, 2006). Cet ensemble de méthodes systématiques de description du contenu des messages vise à obtenir des indicateurs permettant l'inférence à partir de ces messages (Bardin, 2007). Les indicateurs obtenus à partir de l'analyse de contenu peuvent être de nature qualitative ou quantitative (Mucchielli, 2006).

La première étape de l'analyse de contenu suggéré par Bardin (2007) est la préanalyse qui consiste à choisir les documents à soumettre à l'analyse et à déterminer les indicateurs qui

permettront l'interprétation. À cette étape, nous avons convenu de cibler les formulaires de stage (annexe 3) des étudiants ayant accepté de participer à l'étude et ayant fait au moins un stage dans le domaine musculosquelettique (88/91). Ces formulaires incluent, en général, une grande quantité d'informations sur le raisonnement clinique des stagiaires (voir ci-dessous). Le résumé des discussions téléphoniques entre les superviseurs et la coordonnatrice dans les cas problématiques a aussi été considéré pour une analyse de contenu, mais ces résumés n'étaient pas disponibles pour tous les étudiants ayant eu des difficultés de stage.

La deuxième étape de l'analyse de contenu qui correspond selon Bardin (2007) à l'exploitation du matériel s'est donc effectuée dans les formulaires de stage FOCUS. Dans un premier temps une analyse thématique de nature qualitative s'est effectuée en essayant de bien saisir les commentaires formulés par les superviseurs de stage de façon à pouvoir les rattacher ou non à un aspect du raisonnement clinique.

Les formulaires d'évaluation de stages FOCUS (annexe 3) sont basés sur les différentes compétences du programme de physiothérapie. Les compétences du programme qui correspondent aux rôles du profil de compétence canadien (Groupe national consultatif en physiothérapie, 2009) sont également incluses dans le formulaire (par exemple la compétence évaluation du programme fait partie du rôle d'expert du profil canadien). Pour chaque niveau de compétence (stage 1 à 6), un certain nombre d'indicateurs servent de balises pour déterminer les attentes à chacun des stages. Pour chacun des compétences/rôles, le superviseur doit déterminer à l'aide d'une échelle ordinale si le niveau de performance de l'étudiant « *dépasse les attentes* », est « *conforme aux attentes* », « *se conforme minimalement aux attentes* » ou est « *en deçà des attentes* ». Le niveau se conforme minimalement aux attentes a été ajouté en 2012-2013 en réponse à la réticence des superviseurs cliniques d'utiliser la catégorie « *est en deçà des attentes* ». Une section où le superviseur est appelé à donner des commentaires qualitatifs sur la performance de l'étudiant complète l'évaluation de chacun des rôles. Le formulaire se divise en une section d'évaluation à mi-stage et une section de fin de stage. Une rencontre est prévue entre le superviseur et l'étudiant à ces deux périodes de façon à permettre un échange sur le déroulement du stage. Le formulaire, tout comme le profil de compétence canadien est construit selon les rôles du physiothérapeute. Il n'y a donc pas de section du formulaire qui porte

spécifiquement sur le raisonnement clinique, mais celui-ci se retrouve incorporé dans plusieurs rôles. Pour cette raison, l'analyse de contenu des commentaires s'est avérée nécessaire pour pouvoir bien saisir le niveau du raisonnement clinique évalué en stage.

4.4.1 Élaboration d'une grille d'analyse et catégorisation des commentaires des superviseurs

Dans un premier temps, nous avons consulté le formulaire et les informations fournies aux superviseurs pour compléter l'évaluation. Après analyse des critères en lien avec l'ensemble des rôles, nous avons déterminé que les critères de performance associés avec les rôles d'expert étaient très en lien avec les concepts liés au raisonnement clinique de notre cadre d'analyse. Ces critères de performance varient à chaque stage pour permettre aux superviseurs de déterminer quels aspects incluent dans chaque rôle ainsi que le degré d'autonomie attendu pour ceux-ci. Nous avons par la suite discuté avec la coordonnatrice clinique qui nous a confirmé que la grande majorité des commentaires sur le raisonnement clinique se retrouve dans les rôles d'expert, par contre quelquefois on retrouve aussi selon elle, certains énoncés dans le rôle praticien érudit. Le critère de performance : *Démontre une capacité analytique permettant la résolution de problèmes multiples* associés au rôle de praticien érudit pour le stage 4, peut en effet être associé au raisonnement clinique tel qu'on le conçoit généralement.

À la suite de cette première phase d'analyse, nous avons été en mesure de cibler les énoncés qualitatifs qui touchent le raisonnement clinique dans les sections appropriées des formulaires de stages. L'utilisation de ces commentaires qualitatifs nous a permis d'exclure dans la mesure du possible les autres déterminants (habiletés psychomotrices ou interpersonnelles, etc.) de la performance (Khan & Ramachandran, 2012) qui ne sont pas directement reliés au raisonnement clinique. L'analyse de contenu effectué dans les formulaires de stage s'est effectuée à partir d'une grille d'analyse (annexe 9) qui a été développée selon les principales étapes du raisonnement clinique (Audétat, Laurin, *et al.*, 2011).

Pour chaque stage, les commentaires des superviseurs ont été analysés pour déterminer leurs liens avec les processus de raisonnement clinique. Les unités de codage de cette analyse thématique étaient les phrases utilisées par les superviseurs cliniques pour décrire le

raisonnement clinique des étudiants. Lorsqu'une phrase pouvait être associée à une des étapes de raisonnement clinique, ceux-ci étaient recopiés dans un tableur Excel. Lorsque l'énoncé correspondait à une difficulté, le degré de celle-ci était noté en fonction de l'occurrence des unités de codage, de l'importance des difficultés rapportées et aussi en fonction des contextes spécifiques qui sont les attentes de chaque stage. Par exemple, un stage de 2^e année comportant des attentes plus faibles qu'un stage de 4^e année, en général, les commentaires des superviseurs tenaient compte de cet aspect. Un autre aspect du contexte qui a été considéré dans l'analyse de contenu est la propension à énoncer les commentaires de façon à être constructif selon les principes de la rétroaction efficace. Cet aspect pouvait à l'occasion rendre plus difficile l'interprétation des commentaires qui portaient sur des difficultés, l'aspect constructif de ceux-ci ayant tendance à en diminuer la portée.

Lorsqu'un énoncé positif en lien avec une des étapes de raisonnement clinique était présent, celui-ci était également recopié dans la grille d'analyse. Les cotations « *au-delà des attentes* » ou « *sous les attentes* » présentes pour les rôles d'expert intervention ou évaluation étaient également notées dans la grille. De plus, lorsque les commentaires reliés au rôle de praticien érudit correspondaient au processus de raisonnement clinique ceux-ci étaient également inclus, de même que la cotation associée. Après avoir effectué l'analyse de contenu des étudiants ayant participé à l'étude et ayant effectué au moins un stage dans le domaine musculosquelettique, une classification basée sur les résultats d'analyse de cette grille a été produite (voir le tableau XIX). Après plusieurs ajustements, nous avons retenu cette dernière catégorisation qui tient compte de la diversité des résultats de stage observés et nous permet d'utiliser un jugement global de l'ensemble des stages pertinents (musculosquelettique) pour chaque étudiant.

Tableau XIX Catégorisation de l'évaluation du raisonnement clinique effectuée à partir des commentaires des formulaires FOCUS

Catégories	Critères
------------	----------

1. Difficulté importante de raisonnement clinique	Deux mentions de problèmes de raisonnement clinique (dans des stages différents) dans le FOCUS ou gravité importante d'un problème.
2. Difficulté mineure de raisonnement clinique	Une mention de problème de raisonnement clinique dans le FOCUS.
3. Selon les attentes	Raisonnement clinique adéquat (ou pas de mention spécifique) peut inclure une mention de raisonnement clinique supérieur.
4. Raisonnement clinique supérieur	Deux (ou plus) mentions de raisonnement clinique supérieur dans des stages différents (sans mention de difficulté).

Pour la catégorie 4 (*Raisonnement clinique supérieur*), si une mention de difficulté de raisonnement clinique était notée même en présence de deux mentions de raisonnement clinique supérieur, l'étudiant était classé dans la catégorie 3 (*selon les attentes*).

L'analyse catégorielle a donc été effectuée à partir de l'ensemble des énoncés des superviseurs cliniques sur le raisonnement clinique de chaque étudiant en situation de stage. Dans la majorité des cas, la cotation du rôle évalué (expert ou praticien érudit) allait dans la même direction que les commentaires qualitatifs de la même section, le raisonnement clinique étant un aspect important des rôles évalués. Par contre, dans plusieurs cas, la cotation était plus ou moins en accord avec la cotation émise particulièrement lorsque des difficultés de raisonnement clinique étaient mentionnées. Étant donné que certains étudiants de 3^e année et que la majorité les étudiants de 2^e année n'ont eu qu'un stage dans le domaine musculosquelettique, nous avons établi des critères différents pour ceux-ci (tableau XX).

Tableau XX Catégorisation de l'évaluation du raisonnement clinique effectuée à partir des commentaires des formulaires FOCUS pour les étudiants qui ont eu seulement un stage dans le domaine musculosquelettique

Catégories	Critères
1. Difficulté importante de raisonnement clinique	Une mention de problème de raisonnement clinique majeur dans le FOCUS.
2. Difficulté mineure de raisonnement clinique	Une mention de problème mineur de raisonnement clinique dans le FOCUS.
3. Selon les attentes	Raisonnement clinique adéquat (ou pas de mention spécifique).
4. Raisonnement clinique supérieur	Une mention de raisonnement clinique supérieur.

À partir de ces catégories, l'ensemble des commentaires contenus dans les formulaires de stages a été compilé pour tous les étudiants des cohortes de 2^e, 3^e et 4^e année au début de l'année 2015. Tous les stages ayant été effectués de mars 2012 à décembre 2014 par les étudiants participants à l'étude ont été inclus.

4.5 Stratégie d'analyse

Nos stratégies d'analyse des données ont été choisies en continuité avec la littérature sur les tests de raisonnement clinique, c'est-à-dire en utilisant la théorie classique de tests. En effet la grande majorité des études sur les TCS et les PRC en particulier ont utilisé ce cadre théorique de mesure (Dory *et al.*, 2012; Lineberry *et al.*, 2013). L'usage d'un modèle de mesure contemporain tel que la modélisation de Rasch a été envisagé, mais aurait nécessité un plus grand nombre de participants dans chaque cohorte, compte tenu du grand nombre d'items des deux tests.

Au départ, dans le but d'optimiser les items du TCS, nous avons calculé la corrélation item-total et le coefficient alpha de Cronbach pour estimer la cohérence interne des tests et de certaines

des dimensions des tests pour les étudiants et les experts. Les items ayant un niveau de corrélation item-total de moins de 0,05 ont été éliminés de façon à optimiser la consistance interne en essayant de ne pas affecter la représentativité des items. Les choix effectués quant à l'optimisation des items ont été faits principalement dans le but de pouvoir comparer les résultats de notre étude à la littérature existante sur le TCS. Ces choix seront décrits plus en détail dans le chapitre de l'analyse des résultats.

Dans le cas des PRC, nous avons eu recours à un codage des réponses aux questions ouvertes qui nous a permis ensuite de les regrouper en plusieurs catégories diagnostiques ou thérapeutiques. Ce regroupement a été effectué par le même codeur.

Pour ce qui est des résultats des participants aux différents tests, nous avons, au départ, utilisé les statistiques descriptives des groupes et avons exploré la normalité des distributions et leurs variances. Dans ce but, nous avons effectué une analyse des distributions des scores et de leurs écarts types, de l'asymétrie et de la kurtose des distributions de chaque cohorte et nous avons effectué les tests de Kolmogorov-Smirnov et celui de Shapiro-Wilk. Le test de Levene a été utilisé pour déterminer de l'égalité des variances. Étant donné que nos données remplissaient les conditions d'application, nous avons pu recourir à des outils paramétriques de type ANOVA pour pouvoir comparer les résultats des différents groupes. L'ampleur des différences entre les groupes a été évaluée à l'aide du coefficient d de Cohen. Nous avons choisi comme mesure de tendance centrale, la moyenne en complétant en certaines occasions par la médiane comme indication de la normalité de la distribution. Nous avons également fait le choix de présenter les résultats des tests en pourcentage de façon à en simplifier l'interprétation par le lecteur. Dans une optique d'utilité des tests dans le programme de physiothérapie, nous avons également effectué des corrélations entre les scores des tests. Ces valeurs pourraient possiblement nous indiquer si les deux tests procurent essentiellement la même information et ainsi nous aider à déterminer si l'utilisation des deux tests est complémentaire ou redondante.

De façon à explorer le lien potentiel entre les résultats de stage et les résultats aux tests de raisonnement clinique, ces derniers ont été calculés par catégorie de résultats de stage pour chaque cohorte et pour les trois cohortes. L'ensemble des stratégies d'analyses sera traité plus en détail dans le chapitre suivant.

Une triangulation des énoncés qualitatifs recueillis dans les formulaires de stage et les résultats aux tests de raisonnement clinique des candidats ayant eu des difficultés en stage ont également été produits. Cette mise en relation, pourrait nous renseigner sur certains aspects de l'utilité (Lissitz & Samuelsen, 2007) des tests de raisonnement clinique dans le programme de physiothérapie. La mise en commun de ces informations pourrait par exemple servir dans le dépistage de certains candidats qui pourraient bénéficier de remédiation au niveau du raisonnement clinique. La nature exploratoire du processus d'analyse de contenu de formulaires de stage limite les possibilités d'analyses statistiques, mais sa mise en relation avec les résultats aux tests de raisonnement clinique pourrait être un élément d'information du réseau nomologique de ces derniers selon le cadre conceptuel de Lissitz et Samuelsen (2007).

4.5.1 Logiciels utilisés pour l'analyse et la présentation des données

Les scores des TCS ont été calculés à l'aide du calculateur développé par le Centre de pédagogie appliquée aux sciences de la santé (CPASS) de l'Université de Montréal et les scores du PRC ont été aussi calculés à l'aide d'une adaptation de ce logiciel effectué par Robert Gagnon, psychométricien au CPASS. Les statistiques ont été obtenues avec l'aide du logiciel *Statistical Package for the Social Sciences (SPSS)* version 22 (2014) et AgreeStat (2013). Les logiciels Excel (2010) et SPSS ont été utilisés pour effectuer les graphiques.

4.6 Ressources humaines et financières

Le tableau XXI résume les différentes étapes qui ont nécessité des ressources humaines ayant été rémunérées dans le cadre du projet.

Tableau XXI Ressources humaines ayant été rémunérées dans le cadre de l'étude

Ressources humaines	Étapes
Chargée de cours	Conception des tests
Cliniciens	Commenter les vignettes et items
Experts	Attribution des scores
Assistant de recherche	Entrée de données informatiques
Coordonnatrice clinique	Analyse de contenu des formulaires de stage

Les différentes ressources ont été défrayées par un fonds de recherche attribué par la Faculté de médecine et des sciences de la santé de l'Université de Sherbrooke au doctorant. L'achat des iPad pour les tirages des étudiants a été payé par un fonds du directeur de recherche.

4.7 Considérations éthiques

Dans le cadre de la réalisation de cette étude, nous avons obtenu l'accord du comité d'éthique plurifacultaire de la recherche de l'Université de Montréal en janvier 2014 et à la suite d'un accord avec le comité d'éthique de la recherche en éducation et sciences sociales de l'Université de Sherbrooke, nous avons pu débiter la collecte de donnée auprès de nos différents groupes de participants. L'annexe 2 présente l'attestation de conformité éthique.

Les étudiants et les cliniciens furent invités à signer le formulaire de consentement approprié à leurs statuts (annexe 2). Dans celui-ci, il est clairement indiqué que les participants sont libres de participer ou non et qu'ils peuvent se retirer de l'étude à tout moment sans aucun préjudice, que la confidentialité leur est garantie et que l'anonymat sera préservé. Considérant que le chercheur principal est aussi professeur dans le programme, il est clairement indiqué dans les formulaires de consentement que le participant est libre d'accepter ou non, qu'aucune conséquence négative ne surviendra s'il refuse de participer et qu'il peut se désister en tout temps par la suite.

CHAPITRE 5

ANALYSE DE RÉSULTATS

Introduction

Ce chapitre est divisé en deux grandes parties. Dans un premier temps, nous présenterons les données issues de la passation des deux tests d'évaluation du raisonnement clinique et exposerons les étapes qui ont guidé le processus d'optimisation des items de ceux-ci. Dans un deuxième temps, nous ferons une mise en relation des résultats aux tests de raisonnement clinique et des résultats de stage. Cette partie mettra l'accent sur les étudiants ayant éprouvé des difficultés de raisonnement clinique en situation de stage dans une optique de dépistage de celles-ci. La discussion sur les résultats suivra dans le chapitre suivant.

5.1 Première partie analyse des scores des tests de raisonnement clinique

Dans cette première partie de la présentation des résultats aux tests de raisonnement clinique qui nous permettront de répondre à la première question spécifique de recherche, nous commencerons premièrement par présenter le résultat de l'analyse des items à partir des réponses du panel d'experts et deuxièmement à partir des réponses des cohortes d'étudiants pour chacun des deux tests, en expliquant les choix qui sous-tendent leurs optimisations. Par la suite, nous présenterons les résultats des différentes cohortes à ces tests dans des sections séparées. Cette partie présentera également une analyse descriptive et comparative des scores des deux tests ainsi qu'une corrélation entre les scores des deux instruments.

5.1.1 Résultats au test de concordance de script

La prochaine section présentera les résultats au premier test de raisonnement clinique utilisé dans cette étude : le test de concordance de script. Cette section sera divisée en quatre parties, la première portera sur les réponses du panel d'experts. La deuxième partie présentera les choix qui ont été effectués pour optimiser les items du test de concordance de script. De son côté, la troisième partie présentera les statistiques descriptives reliées aux résultats du test. Finalement, la quatrième partie présentera une analyse comparative des scores des différents groupes ayant participé à cette portion de l'étude.

5.1.1.1 Analyse des réponses du panel aux items du test de concordance de script

Dans un premier temps, nous avons procédé à l'analyse des réponses du panel d'expert puisque celles-ci seront utilisées pour établir les scores des différentes cohortes. Après observation des réponses du panel, nous n'avons pas constaté de panéliste fortement déviant (plus de deux écarts-types) sauf un (E4) qui était à cette limite. Le tableau XXII présente les résultats des différents experts ayant complété le TCS : le score minimum du panel est de 64,3 % pour l'expert E4 et le score maximum obtenu est de 86,7 % pour l'expert E13.

La moyenne des scores du groupe est de 76,7 % et l'écart type de 6,1. Le score de chaque expert a été obtenu en excluant leur profil de réponse du calcul.

Tableau XXII Tableau des scores des experts (calculé en excluant leur profil de réponse du calcul) au TCS (82 items)

Expert	Score (%)	Expert	Score (%)
E1	86,4	E11	75,7
E2	83,8	E12	75,6
E3	71,8	E13	86,7
E4	64,3	E14	71,3
E5	68,2	E15	80,6
E6	78,0	E16	73,8
E7	78,2	E17	81,4
E8	72,5	E18	77,0
E9	79,6	E19	71,2
E10	82,1		

Moyenne : 76,7; écart type : 6,1

À la lumière de ces résultats et selon les recommandations actuelles quant à la conception et l'optimisation des panels d'expert de TCS (Charlin *et al.*, 2006; Dory *et al.*, 2012; Gagnon *et al.*, 2011) mentionnées au chapitre 3, nous avons convenu de ne pas exclure de panéliste. Même en excluant les panélistes (E4 et E5) ayant obtenu les scores les plus faibles des panélistes, aucun changement notable n'aurait été observé dans les moyennes des scores attribuées aux trois cohortes d'étudiants et celle du panel d'expert. La moyenne de chacune des cohortes d'étudiants aurait diminué de 0,8 à 0,9 % et celle des experts aurait augmenté de 0,3 %.

L'erreur moyenne quadratique (*Mean square error*)¹² a été utilisée comme mesure de la variabilité des réponses du panel et est définie comme faible si moins de 0,5, modéré entre 0,5 et 1,0 et élevée si supérieur à 1,0 selon l'interprétation de Charlin *et al.* (2006). Ces auteurs recommandent, à cette étape, de questionner les items ayant une variabilité extrême, ceux-ci ayant possiblement un problème de conception. Le tableau XXIII présente le nombre d'items selon les trois catégories de variabilité ainsi que l'erreur quadratique moyenne de chaque item. Ce tableau nous permet de constater que la majorité des items (n=64) sont dans les catégories de variabilité faible à modéré et que 18 items sont dans la catégorie variabilité élevée. Bien que

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2$$

¹²

certaines des items ayant une variabilité élevée avaient une distribution clairement bimodale (item 3, 19 et 79, voir tableau XXIV), nous n'avons pas exclu d'items sur la base d'une telle distribution considérant qu'il s'agit de deux façons plausibles d'interpréter les données et qu'il n'y a pas d'évidence claire qu'une réponse soit meilleure que l'autre (Lubarsky *et al.*, 2013).

Tableau XXIII Profil de la variabilité des réponses aux items par le panel d'expert

82 items		Variabilité faible (< 0,5)		Variabilité modérée (0,5 à 1,0)		Variabilité élevée (>1,0)	
Nombre d'items		30		34		18	
Item	EQM	Item	EQM	Item	EQM	Item	EQM
1	0,36	22	0,91	43	1,21	64	0,83
2	0,32	23	0,61	44	0,56	65	0,81
3	1,23	24	1,26	45	0,44	66	0,49
4	0,49	25	1,02	46	0,43	67	0,60
5	0,58	26	0,48	47	0,58	68	0,34
6	0,27	27	2,01	48	0,81	69	0,72
7	0,50	28	0,33	49	0,48	70	0,59
8	0,47	29	0,59	50	1,67	71	0,37
9	0,29	30	0,60	51	0,81	72	0,76
10	1,06	31	0,92	52	0,93	73	0,47
11	0,62	32	0,45	53	0,58	74	0,22
12	0,36	33	1,04	54	1,15	75	0,18
13	0,23	34	0,56	55	1,15	76	0,36
14	0,36	35	0,71	56	0,58	77	0,82
15	0,37	36	0,71	57	0,62	78	1,44
16	0,50	37	1,27	58	0,92	79	2,09
17	0,36	38	0,51	59	1,01	80	0,20
18	0,69	39	0,58	60	2,13	81	1,58
19	1,69	40	0,58	61	0,36	82	0,25
20	0,72	41	1,26	62	0,95		
21	0,34	42	0,98	63	0,56		

En considérant les réponses des membres du panel aux 82 items du test, on remarque aussi que dans le cas de 13 items, la réponse d'un ou des experts se situe à plus de deux catégories de la réponse modale, comme le démontre le tableau XXIV qui suit.

Tableau XXIV Items ayant des réponses du panel à plus de 2 catégories du mode

Items	-2	-1	0	1	2	MSE
Item 3	1	5	1	11	1	1,23
Item 18	0	1	1	7	10	0,69
Item 19	6	6	2	4	1	1,69
Item 22	0	12	3	3	1	0,91
Item 33	1	0	1	5	12	1,04
Item 36	12	5	1	1	0	0,71
Item 41	3	9	3	3	1	1,26
Item 52	1	1	6	9	2	0,93
Item 54	2	10	2	3	1	1,15
Item 55	1	3	2	11	2	1,15
Item 62	1	6	1	2	0	0,95
Item 70	10	7	0	1	0	0,69
Item 79	3	0	1	8	6	2,09

L'exclusion des réponses qui sont à plus de deux catégories de la réponse modale change également très peu les moyennes des scores des cohortes d'étudiants. En excluant ces réponses, les scores auraient diminué de 0,1 à 0,2 % pour l'ensemble des groupes incluant celui des experts. En considérant ces résultats et les conclusions de Gagnon *et al.* (2011) et de Charlin *et al.* (2006), nous avons choisi de ne pas exclure des réponses individuelles des panélistes.

5.1.1.2 Analyse des réponses des étudiants aux items du test de concordance de script

Les corrélations item-total pour l'ensemble des trois cohortes sont présentées dans le tableau XXV. Nous constatons la présence de quatre corrélations item-total négatives et huit dépassants le 0,30. La grande majorité (59 %) des items ont une corrélation item-total qui se situe entre 0,1 et 0,29.

Tableau XXV Corrélations item-total pour l'ensemble des trois cohortes

Item	Corrélation item-total	Item	Corrélation item-total	Item	Corrélation item-total	Item	Corrélation item-total
1	0,15	22	0,13	43	0,26	64	0,20
2	-0,12	23	0,18	44	0,11	65	0,11
3	0,26	24	0,14	45	0,26	66	0,12
4	0,08	25	-0,04	46	0,13	67	0,21
5	0,15	26	0,16	47	0,23	68	0,15
6	0,39	27	0,24	48	0,20	69	0,30
7	0,16	28	0,42	49	0,20	70	-0,18
8	0,25	29	0,30	50	0,05	71	0,14
9	0,30	30	0,15	51	0,28	72	0,17
10	0,52	31	0,19	52	0,23	73	0,37
11	0,07	32	0,08	53	0,09	74	0,06
12	0,25	33	0,31	54	0,07	75	-0,02
13	0,17	34	0,14	55	0,24	76	0,18
14	0,41	35	0,31	56	0,09	77	0,24
15	0,06	36	0,04	57	0,32	78	0,23
16	0,13	37	0,00	58	0,18	79	0,03
17	0,24	38	0,32	59	0,06	80	0,28
18	0,22	39	0,10	60	0,05	81	0,26
19	0,00	40	0,12	61	0,14	82	0,27
20	0,13	41	0,28	62	0,00		
21	-0,07	42	0,24	63	0,09		

Nous avons par la suite procédé à une optimisation des items du TCS selon les recommandations de Dory *et al.* (2012) Dans un premier temps, nous avons choisi un seuil d'exclusion à ($r \leq 0,05$), comme utilisé par plusieurs des auteurs ayant rapporté une optimisation des items du TCS comme décrit dans le chapitre 3. Cette première optimisation permettait de conserver 71 items avec un coefficient alpha de Cronbach de 0,68 pour l'ensemble des trois cohortes. Par contre, la consistance interne de la cohorte de 3^e année demeurait à un niveau très faible à 0,41, (voir tableau XXVII) ce qui nous a amené à refaire une optimisation en considérant les

particularités des réponses de ce groupe et à éliminer certains items qui semblaient poser problème spécifiquement pour ce groupe tout en considérant l'effet de leur exclusion sur la consistance interne des tests des autres cohortes. Le tableau XXVI rapporte la corrélation item-total pour la cohorte de 3^e année. On constate que les corrélations se distinguent des autres cohortes pour plusieurs items. Il est possible que certaines connaissances ou concept aient été moins bien couverts pour cette cohorte. Nous reviendrons sur cet aspect dans le chapitre suivant.

Tableau XXVI Corrélations item-total pour la cohorte de 3^e année

Item	Corrélation	Item	Corrélation	Item	Corrélation	Item	Corrélation
	item-total		item-total		item-total		item-total
1	0,04	22	0,04	43	0,43	64	-0,10
2	-0,26	23	0,16	44	0,08	65	0,41
3	0,34	24	0,22	45	0,26	66	-0,15
4	0,10	25	0,19	46	0,03	67	-0,12
5	0,06	26	0,18	47	0,23	68	-0,14
6	0,40	27	0,05	48	0,09	69	0,50
7	0,30	28	0,18	49	0,09	70	-0,30
8	0,20	29	0,16	50	0,16	71	-0,22
9	0,12	30	0,45	51	0,10	72	0,40
10	0,50	31	0,33	52	0,26	73	0,37
11	0,39	32	-0,15	53	0,08	74	0,31
12	0,42	33	0,09	54	-0,01	75	0,02
13	0,09	34	0,33	55	0,42	76	-0,07
14	0,20	35	0,23	56	0,19	77	0,29
15	0,00	36	0,12	57	0,13	78	0,35
16	0,19	37	-0,24	58	0,15	79	0,01
17	0,15	38	0,10	59	-0,26	80	0,14
18	0,18	39	-0,12	60	0,07	81	0,12
19	-0,14	40	-0,26	61	0,10	82	0,20
20	0,15	41	0,06	62	-0,01		
21	0,11	42	0,27	63	0,10		

La version finale retenue pour les analyses comportait 67 items. Les items 2, 15, 19, 37, 39, 40, 59, 62, 64, 68, 70, 71, 75, 76, 79 ont ainsi été exclus dans la version finale du test. L'optimisation du test n'a pas éliminé de vignettes cliniques qui sont demeurées stables à 26, ce qui implique que tous les contextes de la table de spécification de départ demeurent présents dans la version finale à 67 items.

Le tableau XXVII résume les valeurs du coefficient alpha de Cronbach pour les différentes cohortes et les différentes versions du TCS. Le fait d'exclure 15 items du test a permis de hausser le coefficient alpha de Cronbach de 0,62 à 0,68 pour les 3 groupes d'étudiants, se rapprochant ainsi des niveaux recommandés dans la littérature (Dory *et al.*, 2012; Tavakol & Dennick, 2011). L'alpha a été calculé en ne considérant que les groupes d'étudiants. L'inclusion du groupe d'experts aurait pour effet d'augmenter ce coefficient à 0,74. Le calcul de l'alpha est souvent effectué par item, mais il a été mentionné dans la littérature (Dory *et al.*, 2012) qu'il serait probablement plus pertinent de le calculer par vignette dans le cas du TCS considérant la dépendance des items d'une même vignette. Le calcul de l'alpha par vignette (sans le groupe d'expert) donne un résultat très similaire à 0,67.

Tableau XXVII Coefficient alpha pour les différentes versions du test et les différentes cohortes

Groupe	Alpha 82 items	Alpha 71 items	Alpha 67 items
2^e année (n=32)	0,68	0,74	0,69
3^e année (n=33)	0,21	0,41	0,59
4^e année (n=26)	0,59	0,63	0,65
Expert (n=19)	0,68	0,69	0,67
3 cohortes (n=91)	0,62	0,68	0,68

Une relecture des items exclus a permis de mettre en relief des ambiguïtés dans la formulation de certains items exclus pouvant être à l'origine de la faible corrélation item-total.

Comme on le voit dans le tableau XXVIII, l'optimisation des items a légèrement diminué les moyennes des scores des étudiants et augmenté les écarts types dans chaque groupe, avec un effet plus important au niveau de la cohorte de 3^e année.

Tableau XXVIII Effet de l'optimisation sur les moyennes et écarts types

Groupes	TCS (67 items)	Écart type
2^e année (n=32)	-1,0 %	+1,0 %
3^e année (n=33)	-0,7 %	+2,3 %
4^e année (n=26)	-0,7 %	+1,2 %
Expert	-0,3 %	+0,5 %

5.1.1.3 Analyse descriptive de la distribution des scores du test de concordance de script

Nous présenterons dans cette partie les résultats des cohortes d'étudiants au TCS après optimisation.

On constate en regardant le tableau XXIX et la figure 7 que la moyenne des scores des experts au TCS est plus élevée que les moyennes des résultats des étudiants de 4^e année, de 3^e année et de 2^e année. On constate aussi que les moyennes des scores des étudiants sont plus élevées en fonction de leur niveau de formation dans le programme. La moyenne des scores des étudiants de 2^e année est de 64,3 %, de 66,2 % en 3^e année, de 70,4 % en 4^e année et de 76,4 % pour les membres du panel d'experts. Les valeurs des médianes qui sont très près de celles moyennes pour les différents groupes reflètent la symétrie des distributions. En regardant les distributions des scores des différentes cohortes, on note un peu plus de variation chez la cohorte de 2^e année que chez les autres cohortes comme démontré par la comparaison de l'écart type et la variance de ces groupes. La distribution des scores s'étend de 45,6 à 78,7 % pour la cohorte de 2^e année, de 47,0 à 74,4 % pour la cohorte de 3^e année et de 58,8 à 84,3 % pour celle des étudiants de 4^e année. Du côté du panel d'experts les scores varient de 63,9 à 88,3 %. On note deux valeurs extrêmes dans la cohorte de 2^e année et une pour la cohorte de 3^e, ces cas extrêmes influencent la moyenne à la baisse et l'écart type à la hausse dans le cas de la cohorte de 2^e année. Ces cas pourraient représenter les étudiants ayant des difficultés au niveau du raisonnement clinique potentiellement dépistés par le TCS. On peut noter aussi en considérant l'étendue des scores de chaque cohorte que plusieurs étudiants ont des scores supérieurs à certains membres du panel

d'experts. Cette observation a été relevée par certains auteurs (Lineberry *et al.*, 2013) et nous y reviendrons plus en détail dans le chapitre sur la discussion des résultats. Comme on le voit dans le tableau XXIX et dans la figure 8 les niveaux d'asymétrie et de kurtose des distributions se situent entre -0,80 et 0,81. Généralement, des niveaux d'asymétrie et de kurtose se situant entre -2,0 et 2,0 sont considérés compatibles avec une distribution normale (George, 2003).

Tableau XXIX Statistiques descriptives des scores au TCS optimisé (26 vignettes, 67 items)

	Étudiants 2^e année (n=32)	Étudiants 3^e année (n=33)	Étudiants 4^e année (n=26)	Experts (n=19)
Moyenne (%)	64,3	66,2	70,4	76,4
Score Min- Max	45,6-78,7	47,0-77,4	58,8-84,3	63,9-88,3
Médiane	65,9	66,6	70,5	76,2
Écart-type	7,5	6,8	6,9	6,6
Variance	56,2	45,9	47,8	43,0
Asymétrie	-0,59	-0,68	0,12	-0,01
Kurtose	0,30	0,81	-0,80	-0,42

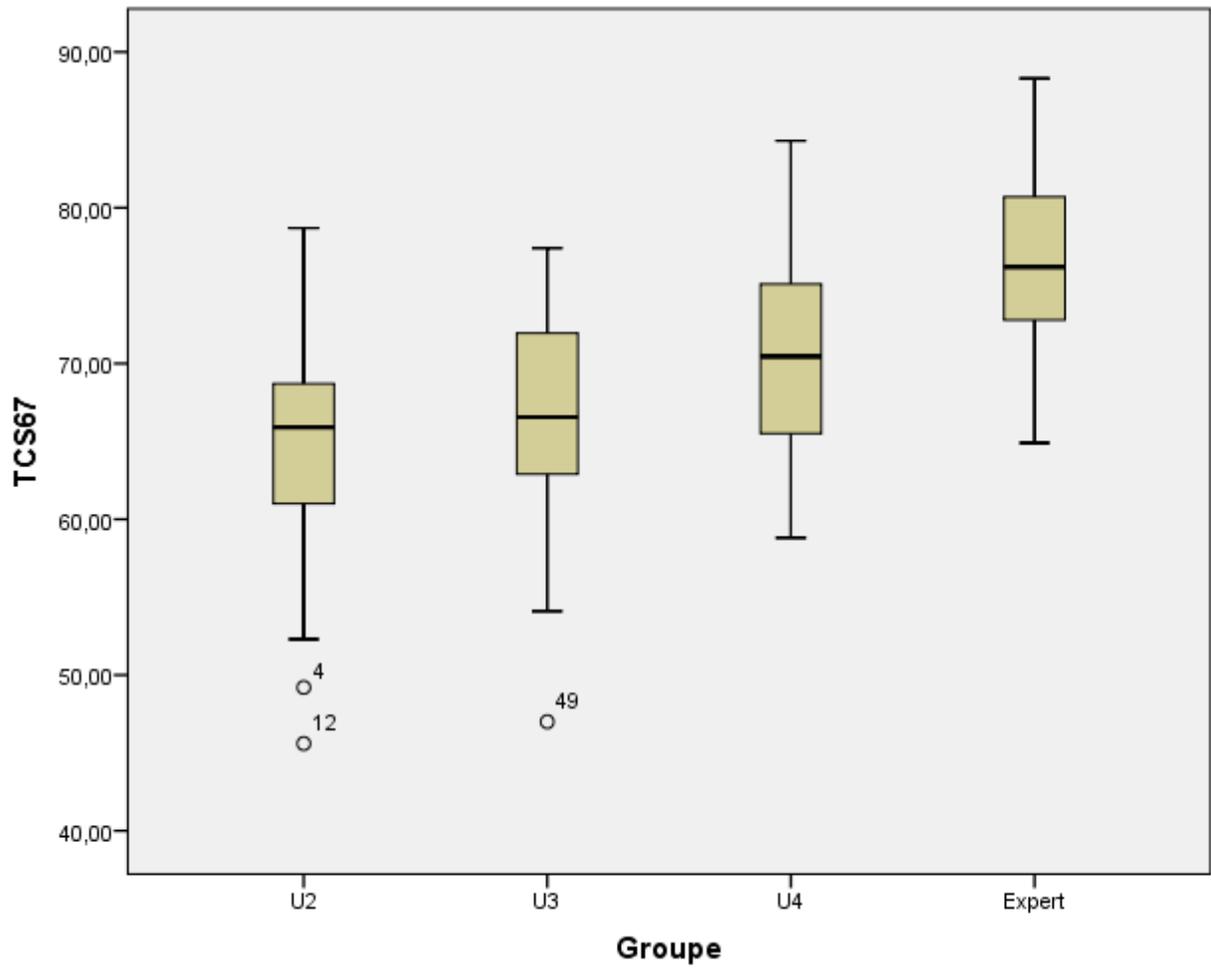


Figure 7 Distribution des scores au TCS des cohortes d'étudiants et du panel d'experts.

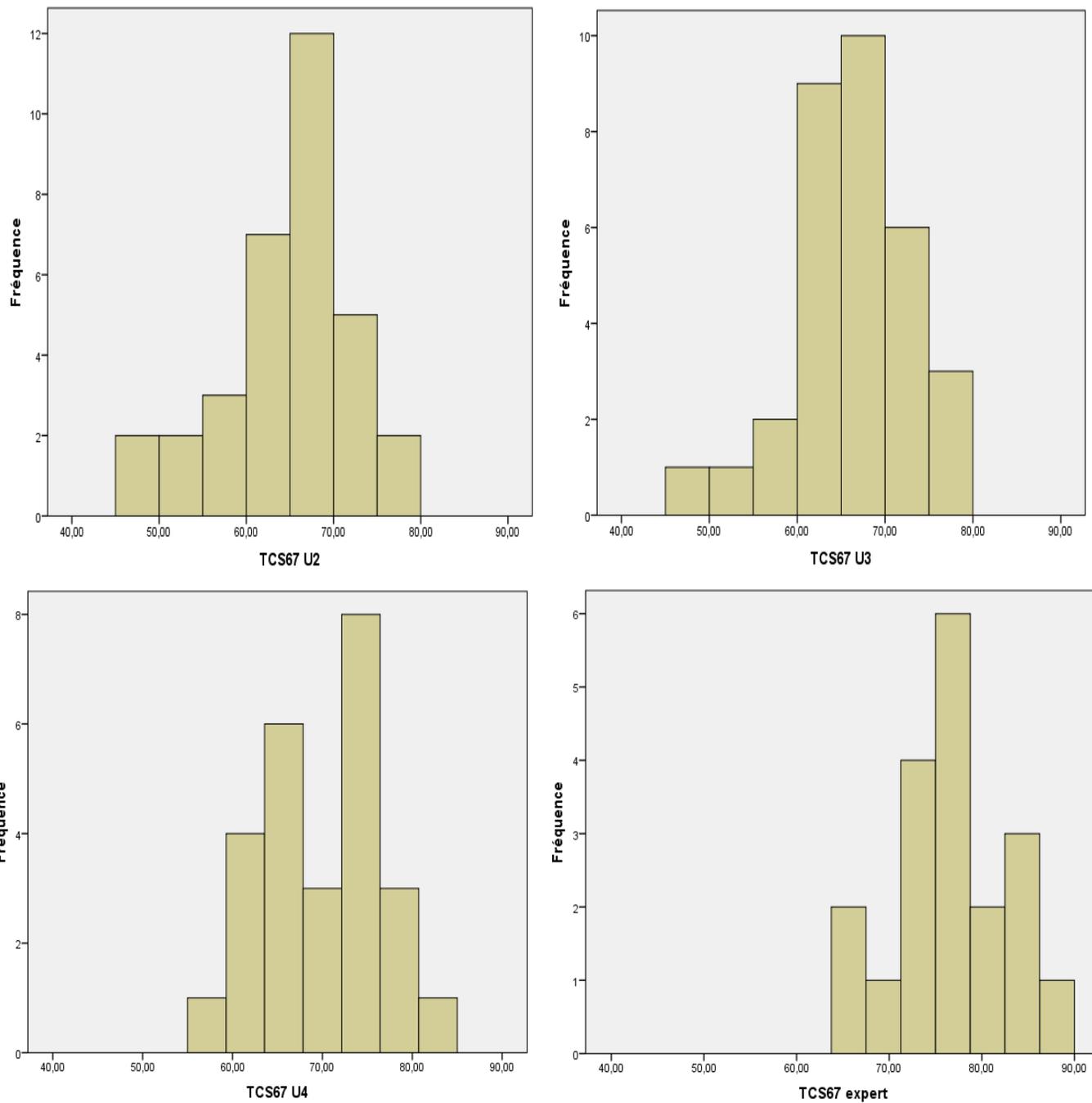


Figure 8 Histogramme de fréquence des scores au TCS pour chacune des cohortes et du panel

Afin de pouvoir attester de la normalité de la distribution, nous avons effectué le test de Kolmogorov-Smirnov (corrigé par le coefficient de Lilliefors) et celui de Shapiro-Wilk pour les 110 observations (voir tableau XXX). Nous avons répété le processus pour la distribution des scores des quatre groupes. Les résultats du test de Kolmogorov-Smirnov et de Shapiro-Wilk, de la totalité des observations ainsi que celle de chacune des distributions des trois groupes d'étudiants et du panel d'experts, supportent l'hypothèse que les scores proviennent d'une population normalement distribuée avec une probabilité supérieure au niveau alpha choisi (0,05). Dans le but de pouvoir utiliser des tests paramétriques de type ANOVA pour comparer les moyennes de nos groupes, nous avons de plus utilisé le test de Levene pour déterminer de l'égalité des variances. Celui-ci démontre qu'il n'y a pas de différence significative entre les variances des différents groupes : $F(3,106)=1,59, p \geq 0,05$

Tableau XXX Test de Kolmogorov-Smirnov et Shapiro-Wilk sur la normalité des distributions

Groupe	Kolmogorov-Smirnov ^b			Shapiro-Wilk		
	Statistiques	ddl	Sig.	Statistiques	ddl	Sig.
TCS67 Expert	,107	19	,200*	,975	19	,868
U2	,136	33	,124	,965	33	,363
U3	,078	32	,200	,966	32	,387
U4	,105	26	,200*	,974	26	,732

b. Correction de signification de Lilliefors

5.1.1.4 Analyse comparative de la distribution des scores du test de concordance de script

Nous avons analysé les scores des quatre groupes de participants et selon l'hypothèse nulle, la distribution des scores serait la même pour les différents groupes. Dans le but de vérifier cette hypothèse, nous avons utilisé un test d'analyse de variance avec une correction post-hoc de Bonferroni. Comme le démontre le tableau XXXI, il existe des différences significatives entre le groupe d'experts et tous les groupes d'étudiants.

Au niveau des groupes d'étudiants, seule la cohorte des étudiants de 4^e année a une moyenne significativement différente ($p > 0,05$) de la cohorte des étudiants de 2^e année.

Tableau XXXI ANOVA des scores du TCS Test Post Hoc : Bonferroni

(I) Groupe	(J) Groupe	Différence moyenne (I-J)	Erreur std.	Signification	Intervalle de confiance à 95 %	
					Borne inférieure	Borne supérieure
Expert	U2	12,13*	2,01	,000	6,71	17,55
	U3	10,15*	2,02	,000	4,69	15,59
	U4	6,01*	2,11	,032	,33	11,69
U2	Expert	-12,13*	2,01	,000	-17,55	-6,71
	U3	-1,99	1,74	1,000	-6,66	2,68
	U4	-6,12*	1,84	,007	-11,05	-1,18
U3	Expert	-10,14*	2,03	,000	-15,74	-5,19
	U2	1,98	1,74	1,000	-2,68	6,66
	U4	-4,13	1,85	,165	-9,10	,84
U4	Expert	-6,01*	2,11	,032	-11,69	-,33
	U2	6,12*	1,84	,007	1,18	11,05
	U3	4,13	1,85	,165	-,84	9,10

*. La différence moyenne est significative au niveau 0,05.

La taille de l'effet peut nous donner une appréciation de l'importance de la différence entre les groupes. Comme on l'observe dans le tableau XXXII, la taille de l'effet maximale se situe entre le groupe de 2^e année et le groupe d'expert à $d=1,71$. Selon Cohen (1988), on peut qualifier d'importante une taille d'effet supérieur à 0,8, ce qui est le cas des différences qui sont statistiquement significatives entre les groupes (Cohen, 1988).

Tableau XXXII Taille d'effet de la différence des scores du TCS entre les groupes

Cohen d	Expert	2 ^e année	3 ^e année	4 ^e année
Expert	-	1,71	1,52	0,89
2 ^e année	1,71	-	0,28	0,85
3 ^e année	1,52	0,39	-	0,61
4 ^e année	0,89	0,85	0,61	-

Les résultats au TCS nous permettent ainsi de répondre partiellement à notre première question spécifique de recherche. Basés sur les résultats des trois cohortes à ce test, il existe donc une différence statistiquement significative entre le raisonnement clinique (tel qu'évalué par le TCS) entre les étudiants novices (2^e année) et les étudiants finissants de 4^e année du programme de physiothérapie, et ce, avec une taille d'effet importante. Par contre, il n'existe pas de différence significative entre le score moyen de la cohorte de 3^e année et des deux autres cohortes à ce test. Nous reviendrons plus en détail sur ces résultats et leurs implications dans le chapitre de la discussion des résultats.

5.1.2 Résultats aux problèmes de raisonnement clinique

La prochaine section présentera les résultats au deuxième test de raisonnement clinique utilisé dans cette étude : les problèmes de raisonnement clinique. Cette section sera aussi divisée en quatre parties, la première partie présentera les réponses du panel aux PRC. La deuxième partie portera sur la consistance interne des PRC basée sur les réponses des étudiants aux items. Il est à noter que contrairement à ce que nous avons fait pour le TCS, nous ne ferons pas d'analyse d'items considérant que pour chaque vignette il existe différents profils de réponse des items reliés à chaque diagnostic et à chaque intervention. La troisième partie, quant à elle, présentera les statistiques descriptives reliées aux résultats des PRC. Finalement la 4^e partie présentera une analyse comparative des scores des différents groupes ayant complété les PRC.

5.1.2.1 Analyse des réponses du panel aux PRC

Le tableau XXXIII présente les résultats des différents experts ayant complété le PRC, le score minimum du panel est de 49,8 pour l'expert E19 et le score maximum obtenu est de 82,8 % pour l'expert E1. La moyenne pour l'ensemble du panel est de 71,0 et un écart type de 7,2. Étant donné que le résultat de l'expert E19 est à plus de deux écarts types de la moyenne du groupe d'expert, et bien en deçà de la moyenne des cohortes d'étudiants celui-ci pourrait être considéré comme un panéliste déviant. Considérant que cet expert a possiblement eu des difficultés au niveau de la compréhension des consignes, nous avons décidé de l'exclure des résultats. En excluant ce panéliste (E19) la moyenne de chacune des cohortes d'étudiants diminue de 0,4 à 0,6 % et celle des experts augmente de 0,9 % à 71,0 ($\pm 6,9$). Toutes les réponses de ce panéliste ont donc été exclues des tableaux d'attribution des scores.

Tableau XXXIII Tableau des scores des experts au PRC (8 vignettes, 218 items)

Expert	Score (%)	Expert	Score (%)
E1	82,8	E11	64,8
E2	78,4	E12	78,3
E3	68,5	E13	61,6
E4	73,6	E14	66,6
E5	80,9	E15	68,5
E6	79,5	E16	69,7
E7	64,4	E17	71,9
E8	60,5	E18	75,6
E9	64,3	E19	49,4
E10	70,3		

Moyenne : 71,0; écart type : 6,9 (après exclusion du panéliste 19)

5.1.2.2 Analyse des réponses des étudiants aux items des problèmes de raisonnement clinique

Comme mentionné précédemment, nous n'avons pas exclu d'item (comme pour le TCS) sur la base d'une faible corrélation item-total étant donné que pour chaque vignette il existe différents profils de réponse des items en fonctions des diagnostics et des interventions. Dans ce contexte, il aurait été difficile d'exclure un item dans un profil, mais pas dans un autre pour la même vignette. Le tableau XXXIV nous permet de constater les valeurs élevées (de 0,85 à 0,92) des différents coefficients de consistance interne alpha de Cronbach obtenu à partir des scores des PRC aux 218 items des PRC pour chaque cohorte.

Comme nous l'avons fait dans le cas du TCS, nous avons également refait le calcul de l'alpha de Cronbach par vignette (sans le groupe d'experts) en incluant les items de réponses ouvertes et en leur attribuant une pondération équivalente aux items sur les caractéristiques, ce qui donne un résultat plus faible à $\alpha=0,54$. Ce résultat est probablement relié, en partie, au faible nombre de vignettes (n=8) pour les PRC. Nous reviendrons dans le chapitre de la discussion sur les conséquences et implications de la méthode de calcul de la consistance interne.

Tableau XXXIV Coefficient alpha par items sur les scores des PRC (218 items pour les différentes cohortes)

Groupes	PRC Alpha
2 ^e année (n=32)	0,92
3 ^e année (n=33)	0,88
4 ^e année (n=26)	0,85
Expert (n=18)	0,91
3 cohortes (n=91)	0,88

5.1.2.3 Analyse descriptive de la distribution des scores des problèmes de raisonnement clinique

Le tableau XXXV qui fait état des statistiques descriptives des résultats aux PRC et les figures 9 et 10 confirment que les moyennes des groupes d'étudiants sont relativement similaires entre elles et ne se distinguent pas selon le niveau de formation de ceux-ci. Il apparaît même que la moyenne des étudiants de 2^e année (59,1 %) est légèrement supérieure à celle des étudiants de 3^e année (58,0 %) et est très près de celle de la cohorte des étudiants de 4^e année (59,3 %). Par contre la moyenne des résultats du groupe d'experts (71,0 %) est nettement supérieure à celles des groupes d'étudiants. La figure 9 démontre également une valeur extrême basse dans la cohorte de 2^e année. Il est à noter que cette valeur n'est pas associée à un candidat ayant eu une valeur extrême basse au TCS. Comme pour les scores des TCS, on constate que les valeurs des médianes sont très près de celles des moyennes pour les différents groupes reflétant ainsi la symétrie des distributions. Les niveaux d'asymétrie et de kurtose des distributions des différentes cohortes, bien que légèrement plus grande à celles des scores des TCS sont également compatibles avec une distribution normale des scores.

Tableau XXXV Statistiques descriptives des scores au PRC (8 vignettes, 218 items)

	Étudiants 2^e année (n=32)	Étudiants 3^e année (n=33)	Étudiants 4^e année (n=26)	Experts (n=18)
Moyenne (%)	59,1	58,0	59,3	71,0
Score Min-Max	35,7-79,5	45,4-72,3	46,9-70,9	58,9-82,8
Médiane	60,6	58,4	58,7	70,2
Écart-type	9,2	7,6	6,7	7,2
Variance	83,9	57,8	44,4	51,8
Asymétrie	-0,54	0,07	0,03	0,06
Kurtose	0,92	-1,03	-0,58	-1,06

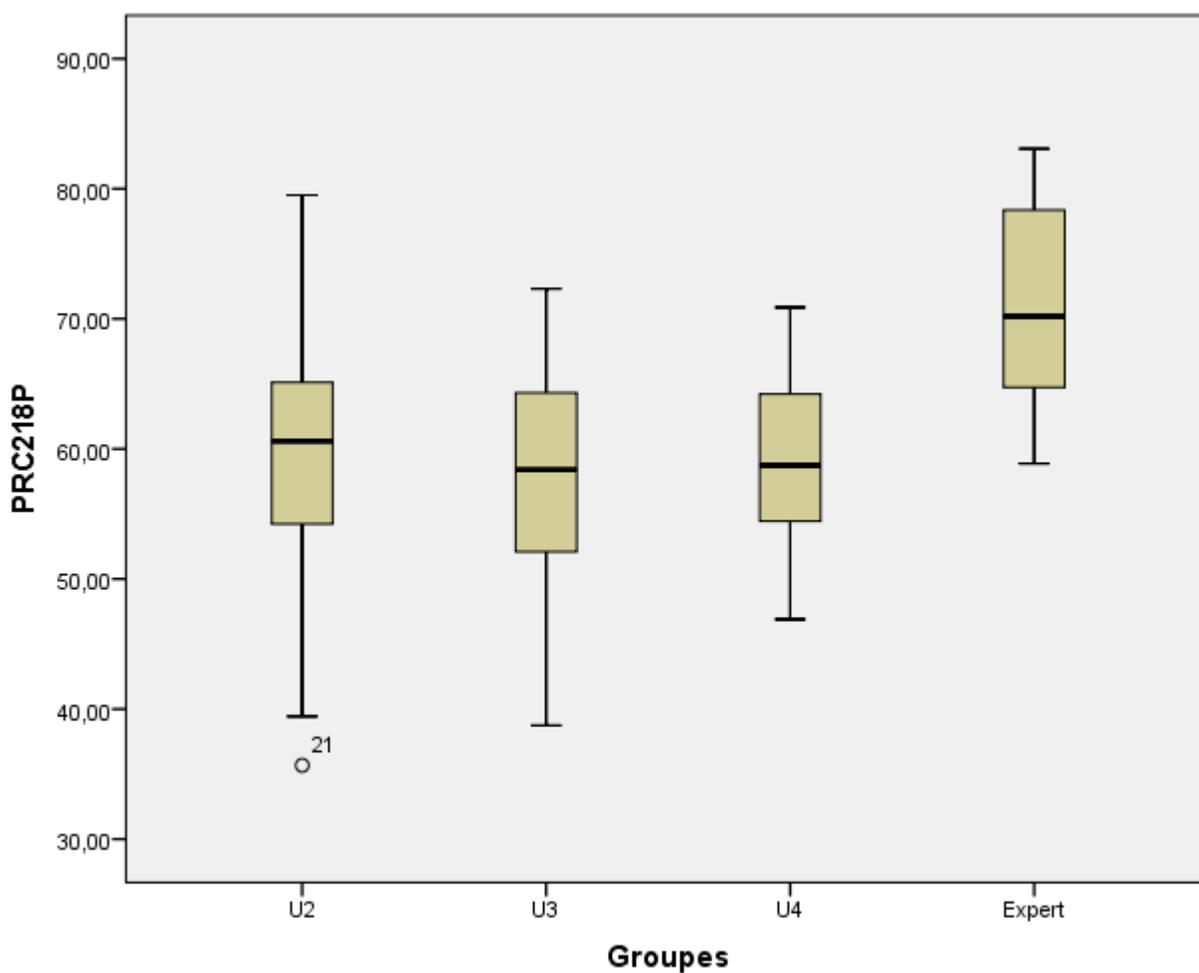


Figure 9 Distribution des scores des PRC pour chaque groupe

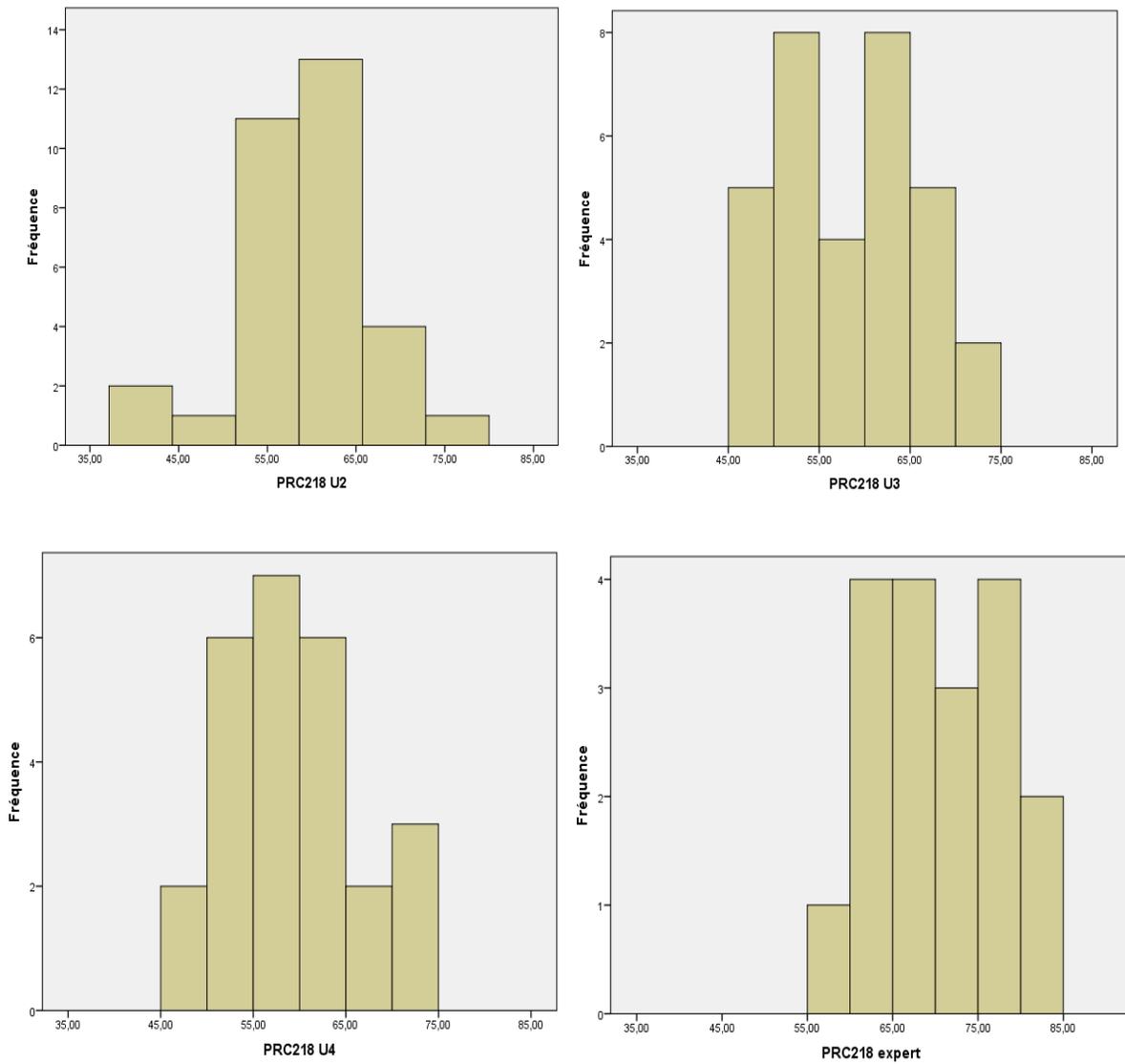


Figure 10 Histogramme des scores des PRC pour chacune des cohortes et du groupe d'expert

Tel qu'effectué pour les résultats du TCS et comme l'illustre le tableau XXXVI, nous avons effectué le test de Kolmogorov-Smirnov (corrigé par le coefficient de Lilliefors) et celui de Shapiro-Wilk pour la distribution des scores de chacun des groupes. Les résultats du test de Kolmogorov-Smirnov et de Shapiro-Wilk, de la totalité des observations ainsi que celle de chacune des distributions des trois groupes d'étudiants et du panel d'experts, soutiennent le fait que les scores proviennent d'une population distribuée normalement avec une probabilité supérieure au niveau alpha de 0,05.

Tableau XXXVI Test de Kolmogorov-Smirnov et Shapiro-Wilk sur la normalité des distributions

Groupe	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistiques	ddl	Sig.	Statistiques	ddl	Sig.
PRC218 expert	,125	18	,200*	,966	18	,717
2	,110	33	,200*	,964	33	,330
3	,115	32	,200*	,960	32	,282
4	,069	26	,200*	,975	26	,759

* Il s'agit de la borne inférieure de la vraie signification.

a. Correction de signification de Lilliefors

Nous avons de plus utilisé le test de Levene pour déterminer de l'égalité des variances. Celui-ci démontre qu'il n'y a pas de différence significative entre les variances des différents groupes : $F(3,106)=0,25 > 0,05$.

5.1.2.4 Analyse comparative de la distribution des scores des problèmes de raisonnement clinique

Les conditions d'applications des tests paramétriques utilisés étant respectées, nous avons analysé les scores des quatre groupes de participants en testant l'hypothèse nulle, soit que la distribution des scores serait la même pour les différents groupes. De façon à vérifier cette hypothèse, nous avons utilisé un test d'analyse de variance avec une correction post-hoc de Bonferroni. Ce test illustré dans le tableau XXXVII démontre que les résultats des tests des experts sont plus élevés que celui des différents groupes d'étudiants de façon statistiquement

significative ($p > 0,05$). Par contre, la comparaison des moyennes des différents groupes d'étudiants entre elles ne démontre pas de différence significative.

Tableau XXXVII ANOVA sur les scores des PRC Test post hoc : Bonferroni

(I) Groupe	(J) Groupe	Différence moyenne (I-J)	Erreur standard	Sig.	Intervalle de confiance à 95 %	
					Borne inférieure	Borne supérieure
Expert	2	11,91952*	2,29930	,000	5,7363	18,1027
	3	12,96350*	2,31194	,000	6,7463	19,1807
	4	11,70101*	2,40606	,000	5,2307	18,1713
2	Expert	-11,91952*	2,29930	,000	-18,1027	-5,7363
	3	1,04398	1,94683	1,000	-4,1914	6,2793
	4	-,21851	2,05772	1,000	-5,7521	5,3150
3	Expert	-12,96350*	2,31194	,000	-19,1807	-6,7463
	2	-1,04398	1,94683	1,000	-6,2793	4,1914
	4	-1,26249	2,07184	1,000	-6,8340	4,3090
4	Expert	-11,70101*	2,40606	,000	-18,1713	-5,2307
	2	,21851	2,05772	1,000	-5,3150	5,7521
	3	1,26249	2,07184	1,000	-4,3090	6,8340

*. La différence moyenne est significative au niveau 0.05.

Le tableau XXXVIII présente la taille d'effet des différences entre le groupe d'experts et les groupes d'étudiants qui se situent entre 1,45 pour les étudiants de la cohorte de 2^e année et de 1,75 pour les étudiants de la cohorte de 3^e année. Cette taille d'effet importante entre les groupes d'experts et les groupes d'étudiants se situe au niveau de celle du TCS

Tableau XXXVIII Taille d'effet de la différence des scores du PRC entre les groupes d'experts et les cohortes d'étudiants

<i>Cohen d</i>	<i>Expert</i>	<i>2^e année</i>	<i>3^e année</i>	<i>4^e année</i>
<i>Expert</i>	x	1,45	1,75	1,69

En lien avec notre première question spécifique de recherche et basé sur ces résultats, il semble que les PRC ne nous permettent pas de distinguer le niveau de développement de raisonnement clinique des étudiants en physiothérapie. Nous reviendrons dans la discussion sur les différentes raisons pouvant être à l'origine de ces résultats.

5.1.3 Corrélation entre les scores du test de concordance de script et des problèmes de raisonnement clinique

Les corrélations de Pearson furent calculées pour évaluer les liens entre les deux tests de raisonnement clinique optimisés. Comme on le voit dans les diagrammes de dispersion de la figure 11 et dans les données du tableau XXXIX, les corrélations sont faibles entre les résultats des étudiants aux PRC et au TCS. Ces corrélations varient entre 0,152 pour les résultats de la cohorte de 3e année à 0,314 pour les résultats de la cohorte de 2e année. Pour l'ensemble des trois cohortes d'étudiants, la corrélation s'élève à 0,231, ce qui est statistiquement significatif ($p=0,027$) mais relativement bas considérant que les deux tests ont été conçus à partir des objectifs du même cours du programme de physiothérapie. Il est à noter que la corrélation est aussi très faible entre les résultats des experts aux deux tests ($r=0,061$). Ces résultats, bien que surprenants, peuvent laisser entrevoir que les deux tests évaluent des aspects différents du raisonnement clinique. Nous reviendrons sur cet aspect dans le prochain chapitre.

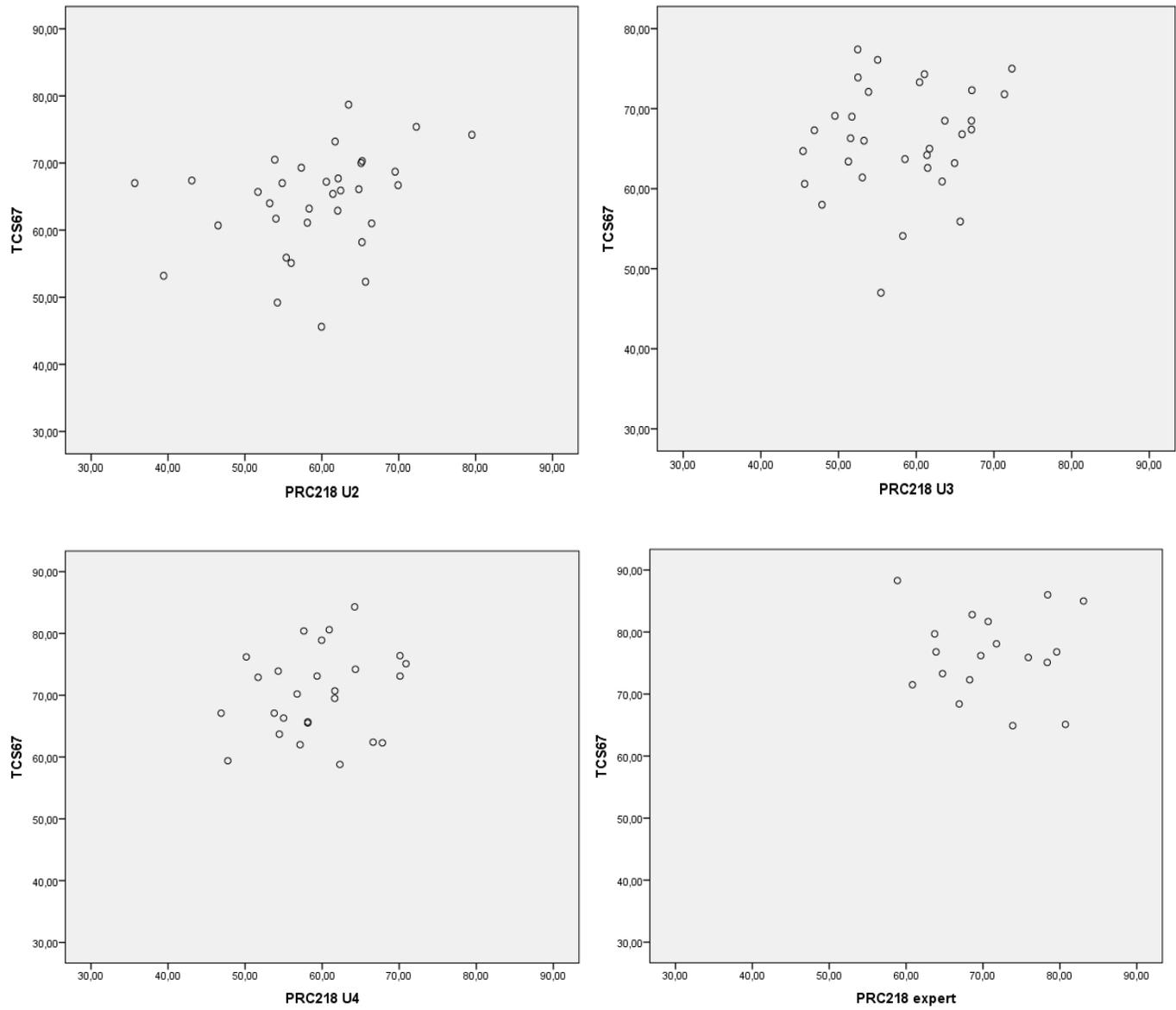


Figure 11 Diagramme de dispersion des scores au TCS et aux PRC

Tableau XXXIX Corrélation entre les scores du test de concordance de script et des problèmes de raisonnement clinique

Groupe	Pearson r	NS (bilat.)
U2	0,314	0,075
U3	0,152	0,407
U4	0,221	0,278
3 cohortes	0,231	0,027
Experts	0,061	0,811

5.2 Deuxième partie : mise en relation entre les résultats aux tests de raisonnement clinique et les observations de stage

Dans cette dernière partie de la présentation des résultats, nous explorerons les liens entre les résultats des tests de raisonnement clinique et les résultats de la catégorisation basée sur le raisonnement clinique observé en stage qui nous permettront de répondre à la deuxième question spécifique de recherche. De plus, nous effectuerons une triangulation des résultats de stage et des résultats aux tests pour les étudiants qui se sont démarqués positivement ou négativement aux tests de raisonnement clinique et lors des situations de stage.

5.2.1 Catégorisation des résultats de stage

La prochaine partie décrira les résultats de la catégorisation des résultats de stage effectués à partir des commentaires portant sur le raisonnement clinique des étudiants des différentes cohortes. Cette catégorisation nous permettra par la suite d'explorer les liens potentiels entre les résultats aux tests de raisonnement clinique et l'évaluation du raisonnement clinique observée en stage. Cette partie présentera aussi les résultats d'une étude qui porte sur l'accord entre la catégorisation effectuée par deux évaluateurs.

Comme nous l'avons décrit précédemment, il n'y a pas de portion raisonnement clinique clairement identifiée dans les formulaires d'évaluation des stages des étudiants en physiothérapie. Par contre, comme mentionné par la coordonnatrice des stages et à la suite de notre observation de ces formulaires, il apparaît clair que des commentaires liés à l'évaluation du raisonnement clinique apparaissent dans le rôle d'expert (évaluation et intervention) et dans de rares occasions dans le rôle de praticien érudit (voir annexe FOCUS). Nous avons donc

procédé à une analyse de contenu des formulaires de stages effectués en musculosquelettique (à plus de 50 %) de tous les étudiants ayant accepté de participer à l'étude et ayant eu au moins un stage dans le domaine musculosquelettique (n=88). Comme présenté dans le tableau XL, en fonction de la cohorte, d'un à quatre stages ont été analysés. Trois étudiants de 2^e année n'ont pas eu un stage dans le domaine musculosquelettique et n'ont donc pas pu faire l'objet d'une catégorisation en fonction de leurs résultats de stage. Les 30 autres étudiants de cette cohorte ont effectué un stage ayant plus de 50 % de contenu musculosquelettique. Dans la cohorte de 3^e année, 27 des 32 (84 %) étudiants ont complété plus de deux stages en musculosquelettique. Cette proportion atteint 96 % (25 étudiants sur 26) en 4^e année, la majorité (77 %) de ceux-ci ayant complété trois ou quatre stages dans le domaine musculosquelettique. En tout, 183 formulaires de stages ont été analysés pour les trois cohortes.

Tableau XL Nombres de formulaires de stages analysés par cohorte

Nombre de stages analysé par étudiant	0	1	2	3	4
2^e année (n=30)	3	30	-	-	-
3^e année (n=76)	-	5	14	9	4
4^e année (n=77)	-	1	5	14	6
Nombre total des formulaires analysés (n=183)					

Lors de l'analyse de contenu, nous avons à déterminer si la difficulté de raisonnement clinique identifiée dans les commentaires des superviseurs était mineure ou majeure selon l'importance et la fréquence des commentaires émis au sujet du raisonnement clinique en fonction du contexte.

Par exemple : le commentaire émis au sujet du candidat 4U3 : « *Besoin encadrement fréquent pour orienter\optimiser les traitements* » a été codé comme difficulté mineure pour un stage de 3^e année selon la catégorie « *Diagnostic final et élaboration d'un plan d'intervention* ».

Commentaires au sujet de l'étudiant 12U2 : « *Besoin d'aide pour prioriser les points à évaluer, interprète les données avec aide ++, besoin d'aide pour identifier les déficiences, doit améliorer esprit analytique* » a été codé comme une difficulté majeure dans la catégorie : « *Raffinement et traitement des hypothèses* », pour un stage de 2^e année.

De même, pour déterminer si un étudiant pouvait être catégorisé comme ayant un raisonnement clinique au-delà des attentes, celui-ci devait avoir une cotation au-delà des attentes dans le rôle expert (évaluation ou intervention) ainsi qu'un commentaire écrit qui devait confirmer que le raisonnement clinique faisait partie des aspects évalués pour poser le jugement global. Par exemple pour l'étudiant 22U4 dans un stage de 4^e année le commentaire : « *démontre une capacité d'analyse qui lui permet de déterminer l'orientation de ses traitements* » nous a permis de confirmer que la justesse du raisonnement clinique avait été considérée pour donner une cotation d'au-delà des attentes dans le rôle expert (intervention) pour cet étudiant.

Même si un étudiant pouvait avoir des commentaires dans plus d'une catégorie de difficulté de raisonnement clinique, nous cotions la principale, donc les nombres du tableau XLI identifient des étudiants différents. On remarque dans le tableau XLI que le nombre de commentaires en lien avec le raisonnement clinique est relativement faible dans la cohorte de 2^e année ceux-ci n'ayant qu'un seul stage évalué. On note une fréquence plus élevée de difficulté de raisonnement clinique observée par les superviseurs cliniques dans la cohorte de 3^e année où 12 mentions de difficultés de raisonnement clinique (dans des stages différents) ont été notées (12/76 stages évalués en 3^e année). Dans la cohorte de 4^e année, on note six difficultés de raisonnements cliniques commentées par les superviseurs cliniques pour une proportion de 6/77 stages évalués, ce qui est nettement inférieur à la cohorte des étudiants de 3^e année. De façon générale, on remarque que le type de difficulté de raisonnement clinique noté se retrouve à peu près également distribué selon les trois étapes identifiées dans la taxonomie des difficultés de raisonnement clinique (Audétat, Laurin *et al.*, 2011). Vingt difficultés de raisonnement clinique déterminées à partir des commentaires des superviseurs ont été notées, dans 15 des 20 stages, les superviseurs avaient cotés « *est en deçà des attentes* » ou « *se conforme minimalement aux attentes* » à la fin du stage pour le rôle expert en évaluation et/ou intervention.

Tableau XLI Nombre et type de difficulté de raisonnement clinique notés lors de l'analyse de contenu des formulaires de stage

Nombre de difficultés de raisonnement clinique identifiées selon étapes	Génération d'hypothèses et récolte de données	Raffinement et traitement des hypothèses	Diagnostic final et élaboration d'un plan d'intervention	Nombre de stages observé
2^e année	-	2 (1 majeur)	-	30
3^e année	4	4	4 (1 majeur)	76
4^e année	1	2	3	77

5.2.1.1 Évaluation de l'accord sur les catégorisations

Dans le but de vérifier l'accord sur cette catégorisation, nous avons demandé à une 2^e personne, la coordonnatrice des stages cliniques du programme de physiothérapie, de refaire l'analyse de contenu des formulaires de stage. La coordonnatrice des stages cliniques du programme de physiothérapie a pour rôle, entre autres, de lire l'ensemble des formulaires FOCUS et était donc très à l'aise d'accomplir cette tâche. Après lui avoir expliqué le but de cette partie de l'étude, la taxonomie des étapes de raisonnement clinique et les catégories à utiliser, celle-ci a effectué une première analyse pilote des formulaires de stage d'un étudiant. Cette première étape a permis de répondre aux questions de la coordonnatrice clinique et de s'assurer de sa compréhension du processus. Par la suite, une analyse de contenu de l'ensemble des formulaires de stages de 12 étudiants par cohorte a été effectuée par la coordonnatrice clinique sans avoir d'information sur les résultats de la première analyse de contenu et de la catégorisation qui en a découlé. Cette analyse correspond à une analyse de type catégorielle (Bardin, 2007) qui prend en considération la totalité du texte analysé plutôt que des unités de codage.

Le tableau XLII présente la cotation des deux évaluateurs sur la catégorisation pour l'échantillon de douze participants par cohorte. Sur les 36 étudiants ayant eu deux catégorisations de leurs résultats de stage, quatre de ceux-ci ont eu une catégorisation différente par le 2^e évaluateur. Les quatre désaccords étaient entre les catégories 2 et 3. Le coefficient kappa (non pondéré) de l'accord entre les deux évaluateurs est de 0,78 et le kappa pondéré est de 0,92 ce qui est qualifié d'excellent (Landis & Koch, 1977).

Tableau XLII Cotation des deux évaluateurs sur les catégories de stage

	Éval.1	Éval. 2		Éval. 1	Éval. 2		Éval. 1	Éval. 2
1U2	2	2	1U3	3	3	1U4	4	4
2U2	3	3	2U3	3	3	2U4	3	3
3U2	3	3	3U3	3	3	3U4	4	4
4U2	3	3	4U3	2	3	4U4	3	3
5U2	3	3	5U3	4	4	5U4	2	3
6U2	3	3	6U3	2	2	6U4	3	3
7U2	4	4	7U3	3	3	7U4	4	4
8U2	4	4	8U3	3	3	8U4	3	3
9U2	N/A*	N/A*	9U3	3	2	9U4	2	2
10U2	4	4	10U3	1	1	10U4	3	3
11U2	3	3	11U3	4	4	11U4	3	2
12U2	1	1	12U3	2	2	12U4	4	4

* Pas de stage en musculosquelletique

5.2.2 Résultats des tests de raisonnement clinique par catégorie

Dans cette partie les moyennes des scores des étudiants aux tests de raisonnement clinique seront présentées selon les catégories observées à partir des formulaires de stage.

Le tableau XLIII présente les moyennes des résultats aux tests de raisonnement clinique selon les catégories pour la cohorte de 2^e année. Bien que le nombre d'étudiants soit faible dans les catégories 1 et 2 pour la cohorte des étudiants de 2^e année, on peut quand même observer une amélioration graduelle des scores au TCS et dans une moindre mesure pour les PRC, qui concordent avec une amélioration du raisonnement clinique observée dans les stages. Le résultat au TCS du candidat ayant connu des difficultés majeures de raisonnement clinique en stage étant de 45,6 % par rapport à la moyenne de 63,2 % des candidats ayant été selon les attentes (catégorie 3) à ce niveau. Les étudiants ayant été placés dans la catégorie au-delà des attentes (catégorie 4) au niveau du raisonnement clinique ont en moyenne des résultats de 3,7 % au TCS et 0,4 % au PRC supérieur aux étudiants de la catégorie selon les attentes.

Tableau XLIII Scores aux tests de raisonnement clinique par catégorie pour la cohorte de 2^e année

Catégorisation de stage		PRC218	TCS67
1	Score	59,95	45,6
	N	1	1
	Écart type	.	.
2	Score	65,21	58,2
	N	1	1
	Écart type	.	.
3	Score (moyen)	58,40	63,2
	N	13	13
	Écart type	11,33	7,14
4	Score (moyen)	58,8	66,9
	N	15	15
	Écart type	8,59	8,42
Total	Score (moyen)	59,1	64,3
	N	30*	30*
	Écart type	9,20	7,50

*3 étudiants de cette cohorte n'ont pas eu de stage correspondant aux critères d'inclusion

Dans le tableau XLIV, on peut observer que pour la cohorte de 3^e année, les étudiants (n=3) de la catégorie 1 se distinguent par des scores au TCS de 8,6 % et de 9,3 % aux PRC plus faibles que les étudiants de la catégorie 3 et de 13,8 % plus faibles que la catégorie 4 aux PRC. On note également une progression des scores au TCS en fonction des résultats de stage de la catégorie 1 à 3. Par contre, pour la catégorie 4, on note une légère baisse des scores présente pour les six étudiants ayant été classés de cette catégorie.

Tableau XLIV Scores aux tests de raisonnement clinique par catégorie pour la cohorte de 3^e année

Catégorisation de stage		PRC218	TCS67
1	Score (moyen)	50,21	58,90
	N	3	3
	Écart type	4,95	11,14
2	Score (moyen)	52,47	64,26
	N	5	5
	Écart type	4,64	6,99
3	Score (moyen)	58,87	68,22
	N	18	18
	Écart type	7,60	5,89
4	Score (moyen)	64,05	65,65
	N	6	6
	Écart type	4,62	5,23
Total	Score (moyen)	58,03	66,24
	N	32	32
	Écart type	7,60	6,78

Comme on l'observe dans le tableau XLV, il n'y avait pas d'étudiant dans le groupe 1 pour la cohorte de 4^e année. Par contre, dans le cas des PRC on semble observer une augmentation graduelle des scores selon les résultats de stage pour les 3 catégories présentes. Les étudiants de la catégorie 4 avaient une moyenne de 8,0 % plus élevée au PRC que les étudiants de la catégorie 2. Du côté du TCS, les scores à ce test ne semblent pas différencier les catégories de résultats de stage pour la cohorte des étudiants de 4^e année.

Tableau XLV Scores aux tests de raisonnement clinique par catégorie pour la cohorte de 4^e année

Catégorisation de stage		PRC218	TCS67
2	Score (moyen)	58,26	72,00
	N	4	4
	Écart type	5,29	7,15
3	Score (moyen)	56,92	69,73
	N	16	16
	Écart type	5,93	6,43
4	Score (moyen)	66,29	71,03
	N	6	6
	Écart type	4,66	8,95
Total	Score (moyen)	59,29	70,38
	N	26	26
	Écart type	6,66	6,91

Le tableau XLVI combine les résultats des trois cohortes. On note de façon générale, une augmentation des résultats aux deux tests selon la catégorie des résultats de stage, mais cette tendance est moins marquée pour les catégories 2, 3 et 4 aux scores du TCS.

Tableau XLVI Scores aux tests de raisonnement clinique par catégorie pour les trois cohortes

Catégorisation de stage		TCS67	PRC218
1	Score (moyen)	55,58	52,64
	N	4	4
	Écart type	11,27	6,33
2	Score (moyen)	66,75	56,06
	N	10	10
	Écart type	7,91	6,13
3	Score (moyen)	67,34	58,14
	N	46	46
	Écart type	6,83	8,27
4	Score (moyen)	67,52	61,40
	N	28	28
	Écart type	6,73	7,64
Total	Score (moyen)	66,70	58,77
	N	88	88
	Écart type	7,44	7,90

Considérant que la catégorisation de stage de la cohorte de 2^e année n'a été effectuée que sur les résultats d'un stage et que les attentes de ce stage sont en général assez limitées, nous avons

poursuivi notre exploration des résultats en excluant cette cohorte. Comme le démontre le tableau XLVII produit à partir des résultats des 58 étudiants de 3^e et 4^e année, il apparaît que les résultats aux PRC sont assez discriminants en fonction des résultats de stage. Les étudiants de la catégorie 1 ont en moyenne un score de 8 % inférieur à la moyenne des étudiants de 3^e et 4^e année. De leur côté, les étudiants de la catégorie 4 ont un résultat supérieur de 6 % à la moyenne des étudiants de 3^e et 4^e année. Pour les résultats du TCS, seule la catégorie 1 semble avoir des résultats différents des autres catégories, la catégorie 1 étant à environ 10 % de la moyenne de l'ensemble des étudiants de 3^e et 4^e.

Tableau XLVII Scores aux tests de raisonnement clinique par catégorie pour les cohortes de 3^e et 4^e année

Catégorisation de stage		TCS67	PRC218
1	Score (moyen)	58,90	50,21
	N	3	3
	Écart type	11,15	4,95
2	Score (moyen)	67,7	55,04
	N	9	9
	Écart type	7,76	5,53
3	Score (moyen)	69,00	58,04
	N	33	33
	Écart type	6,18	6,92
4	Score (moyen)	68,18	64,39
	N	13	13
	Écart type	7,23	5,21
Total	Score (moyen)	69,18	58,59
	N	58	58
	Écart type	7,09	7,16

Cette exploration des scores au TCS et aux PRC nous suggère qu'il existe un lien entre les résultats des tests de raisonnement clinique et ce qui est observé en milieu de stage.

Ce lien semble plus important pour les PRC dans l'ensemble des catégories et pour le TCS dans la catégorie 1.

5.2.3 Triangulation entre les résultats de stage et les résultats aux tests de raisonnement clinique

Dans cette section nous allons explorer le profil des étudiants ayant éprouvé des difficultés au niveau de leur raisonnement clinique en stage.

Nous avons rassemblé les informations en lien avec les quatre étudiants ayant eu des difficultés importantes au niveau du raisonnement clinique en stage (catégorie 1) dans le tableau XLVIII. Dans une optique curriculaire de dépistage de difficulté de raisonnement clinique, nous avons ciblé la catégorie 1, celle-ci pouvant potentiellement faire l'objet de remédiation. La mise en relation des informations de ce tableau nous indique que trois des quatre étudiants de cette catégorie ont au moins un des tests qui est à plus de deux écarts types sous la moyenne de leur groupe (score Z). Il est à noter qu'à deux occasions, il s'agit du TCS et une occasion pour les PRC, mais que les deux tests semblent très peu corrélés au niveau de leurs scores. Pour l'étudiant de 3^e année 25U3, les deux tests sont près de la moyenne du groupe. Il est à noter que cet étudiant a connu aussi des problèmes de communication dans ses stages (deux cotations sous les attentes à ce niveau dans les stages où les difficultés de raisonnement clinique ont été notées). Il est possible que cette difficulté puisse avoir influencé la perception globale du superviseur et les commentaires mentionnés dans le formulaire de stage au niveau du raisonnement clinique.

Tableau XLVIII Triangulation des informations sur les étudiants ayant eu des difficultés en stage au niveau du raisonnement clinique

Étudiant	Énoncés tirés du formulaire de stage :	Types de difficulté	Degré	Score Score Z TCS67	Score Score Z PRC218	Cat. de stage
-----------------	-----------------------------------------------	----------------------------	--------------	----------------------------	-----------------------------	----------------------

12U2 (stage 2)	<i>« Besoin d'aide pour prioriser les points à évaluer, interpréter les données avec aide ++, besoin d'aide pour identifier les déficiences, doit améliorer esprit analytique »</i>	Raffinement et traitement des hypothèses	Majeur	45,6 % -2,49	60,0 % 0,10	1
10U3 (stage 2)	<i>« Besoin d'aide pour réajuster son évaluation »</i>	Génération d'hypothèses et récolte de données	Mineur	60,6 % -,83	38,7 % -2,37	1
10U3 (stage 4)	<i>« Difficulté à savoir quels éléments devant se retrouver dans le A (analyse), Nécessite aide pour émettre une impression clinique, besoin d'assistance modérée pour faire varier le plan de traitement »</i>	Diagnostic final et élaboration d'un plan d'intervention	Majeur			
16U3 (stage 3)	<i>« Analyse les informations recueillies avec assistance, doit intégrer davantage RC pour sélectionner les éléments à évaluer »</i>	Raffinement et traitement des hypothèses	Mineur	47,0 % -2,84	55,5 % -0,29	1
16U3 (stage 4)	<i>« Évaluation à optimiser pour faciliter son analyse clinique »</i>	Génération d'hypothèses et récolte de données	Mineur			
25U3 (stage 4)	<i>« Approfondir RC de sorte à identifier adéquatement les objectifs »</i>	Diagnostic final et élaboration d'un plan d'intervention	Mineur	69,1 % 0,42	49,5 % -1,03	1
25U3 (stage 5)	<i>« Besoin aide pour trouver des modalités en lien avec son plan de traitement »</i>	Diagnostic final et élaboration d'un plan d'intervention	Mineur			

5.2.4 Profil des étudiants ayant des scores qui se sont démarqués positivement ou négativement aux tests de raisonnement clinique

Dans cette section nous allons tenter d'explorer le profil des étudiants qui se sont démarqués au niveau de leurs résultats aux tests de raisonnement clinique. Il est à noter qu'il n'y a pas d'autre étudiant de 3^e et 4^e année qui se retrouve avec des résultats de tests de raisonnement clinique à plus de deux écarts types sous la moyenne de leur groupe à part les deux étudiants (10U3 et 16U3) identifiés dans le tableau XLVIII. Par contre, au niveau de la cohorte de 2^e année, trois étudiants ont eu des scores à un des tests de raisonnement clinique à plus de deux écarts types sous la moyenne de leurs groupes. Le tableau XLIX illustre le profil des trois étudiants ayant eu au moins un des tests de raisonnement clinique à près de deux écarts types sous la moyenne de leur groupe. Ces étudiants ont eu un premier stage sans particularité au niveau du raisonnement clinique ou des autres aspects évalués.

Tableau XLIX Profil des étudiants à moins de deux écarts types de la moyenne de leurs groupes aux tests de raisonnement clinique

Étudiants	Score Z TCS	Score Z PRC	Catégorisation de stage
2U2	,37	-2,55	3
4U2	-2,01	-,53	3
6U2	-1,47	-2,14	3

Comme on le voit dans le tableau L, peu d'étudiants se sont démarqués par des scores élevés aux tests de raisonnement clinique. Pour l'ensemble des trois cohortes, un étudiant a eu un score de deux écarts types ou plus au-delà de la moyenne de leurs groupes à un des tests de raisonnement clinique. Cet étudiant était dans la catégorie 4 basée sur leur raisonnement clinique en stage.

Tableau L Étudiants ayant eu un score à plus de deux écarts types de la moyenne de leurs groupes à un des tests de raisonnement clinique

Étudiant	Score Z TCS	Score Z PRC	Catégorisation de stage
12U4	2,01	0,74	4

En relation avec notre deuxième question spécifique de recherche, il semble exister un lien entre les résultats aux tests de raisonnement clinique et les observations en situation authentique de pratique. Les étudiants ayant les meilleurs résultats en stage présentent en moyenne, des meilleurs résultats aux tests de raisonnement clinique. Cette observation semble s'appliquer aux scores des PRC pour l'ensemble des catégories alors que le TCS cette observation semble ne s'appliquer que pour la catégorie 1.

Bien que basé sur un nombre restreint de cas ayant eu des difficultés de stage, il semble que la combinaison des deux tests puisse faciliter le dépistage des étudiants pouvant présenter des difficultés de raisonnement clinique en stage. Dans le prochain chapitre, nous discuterons des présents résultats et de leurs implications.

CHAPITRE 6

DISCUSSION DES RÉSULTATS

Nous présenterons dans ce chapitre la discussion des résultats décrits au chapitre précédent. Ce chapitre sera divisé en fonction des deux questions spécifiques de recherche et des éléments principaux des résultats qui nous permettent de répondre à ces questions. Dans le but de soutenir ces résultats, nous reviendrons sur les éléments pertinents liés au cadre de validation des outils utilisés. Ces éléments permettent d'apprécier, non seulement la rigueur utilisée, mais également l'ampleur de la tâche qui entoure le processus de validation de ces outils. De plus, nous aborderons les principales limites de cette recherche de même que les aspects qui nous incitent à une certaine prudence dans l'interprétation de ces résultats et dans l'utilisation potentielle des outils développés. Nous terminerons en proposant des pistes de recherches qui découlent de cette étude doctorale.

6.1 Première question de recherche

La première question spécifique était énoncée de la façon suivante :

Est-ce que les tests de concordance des scripts (TCS) et les problèmes de raisonnement clinique (PRC) peuvent nous permettre de distinguer le niveau de développement du raisonnement clinique des étudiants universitaires en physiothérapie?

6.1.1 Résultats en lien avec le TCS

Selon les résultats décrits dans le chapitre précédent, le TCS tel que conçu dans la présente étude permet de distinguer le niveau de formation des étudiants novices (2^e année) et avancés (4^e année) dans le programme de physiothérapie de l'Université de Sherbrooke. De plus, les scores attribués à partir de cet outil nous permettent de distinguer le niveau de raisonnement clinique des experts de celui de l'ensemble des cohortes d'étudiants.

Ce résultat obtenu à partir des scores du TCS de cette étude est très similaire aux autres études qui ont mis en perspective les scores au TCS en fonction du niveau de formation. En général, le TCS discrimine bien selon le niveau de formation ou d'expertise des participants. Par contre, la

majorité (Boulouffe *et al.*, 2013; Lubarsky *et al.*, 2009; Nouh *et al.*, 2012; See *et al.*, 2014) des études ont utilisé le TCS dans un contexte post-gradué (résidence médicale) en incluant un groupe d'étudiants en formation de base. Par exemple, dans les études de Lambert *et al.* (2009) et Lubarsky *et al.* (2009), les scores du TCS permettent de discriminer les étudiants, des résidents et des spécialistes dans le contexte de la radio-oncologie et de la neurologie. La présente étude est, à notre connaissance, la première à avoir utilisé le TCS avec trois cohortes de participants de niveaux différents en formation de base. Une des retombées attendues de ce projet doctoral est de pouvoir utiliser les outils TCS et PRC dans le cadre du curriculum de physiothérapie de l'Université de Sherbrooke. Bien que la capacité de distinguer le niveau des étudiants soit importante, d'autres aspects liés à la validité sont aussi importants pour soutenir l'utilisation du TCS comme outil d'évaluation du raisonnement clinique dans ce contexte.

En se basant sur le modèle de Lissitz et Samuelsen (2007), les sources d'évidence de la validité du test sont le processus latent, la représentativité du contenu du test et la fidélité. Pour ce qui est du processus latent et de la représentativité du contenu du test du TCS, ceux-ci reposent en partie sur l'évaluation du niveau de difficulté des items, de leur clarté et de leur pertinence à évaluer le raisonnement clinique. Cette étape préliminaire effectuée par un physiothérapeute enseignant et deux physiothérapeutes cliniciennes nous a permis de réajuster le niveau de difficulté et la clarté des énoncés des vignettes et des items avant de soumettre une version finale aux étudiants. Les étudiants des différentes cohortes, lors du sondage effectué après le test, ont confirmé que le niveau des items était approprié selon eux et reflétait des connaissances acquises au cours du programme de formation. Du côté de la difficulté de s'adapter au format du TCS, certains auteurs (Duggan & Charlin, 2012; van den Broek, van Asperen, Custers, Valk, & Ten Cate, 2012) ont mentionné que le format du TCS est souvent mal perçu par les étudiants au niveau de la formation de base, ceux-ci éprouvant de la difficulté à s'adapter au format. Nos résultats sont similaires, environ 40 % des étudiants ont trouvé moyennement ou très difficile de s'adapter à ce format, ce qui peut avoir nui à la précision de la mesure et conséquemment à la capacité de l'instrument de discriminer le niveau de l'étudiant. Nous croyons, par contre, que cette impression n'est pas une entrave majeure à l'utilisation de ce format en formation de base et que globalement comme rapporté dans la littérature la perception des étudiants du format TCS est positive (Cobb, Brown, Hammond, & Mossop, 2015). Néanmoins, une familiarisation en

contexte formatif serait souhaitable avant d'utiliser ce format au niveau sommatif. L'utilisation du TCS en formation de base pourrait permettre l'acquisition de compétences cliniques en développant la capacité de poser un jugement en utilisant ses connaissances dans différents contextes. Le fait que plusieurs bonnes réponses sont acceptées dans le TCS pourrait ainsi familiariser les étudiants avec le processus de prise de décision en situation authentique mal définie.

Dans le but d'assurer une bonne représentativité des items du TCS, nous avons bâti une table de spécification basée sur le cours de 2^e année *PHT 240 Évaluation en physiothérapie*. Bien que la table de spécification se veuille la plus représentative des concepts principaux du cours, certains éléments du cours de huit crédits n'ont pas été inclus dans le but de pouvoir compléter le TCS en une heure ou moins. Un TCS plus long serait assurément plus représentatif de l'ensemble du contenu du cours PHT 240. Considérant que le temps moyen de passation a été d'environ 45 minutes pour l'ensemble des cohortes, le nombre de vignettes et d'items pourrait possiblement être doublé dans une version subséquente pour améliorer cette représentativité.

Selon Lissitz et Samuelsen (2007), l'attribution des scores fait partie des sources d'évidence de la qualité du contenu d'un test. Pour ce qui est de l'attribution des scores au TCS, nous avons suivi la majorité des recommandations de la littérature sur le sujet (Dory *et al.*, 2012; Fournier *et al.*, 2008; Lubarsky *et al.*, 2013). Bien que plusieurs études ont documenté la façon d'attribuer les scores du TCS, certaines controverses persistent, par exemple, certains auteurs (Lineberry *et al.*, 2013) conçoivent difficilement que plusieurs bonnes réponses puissent exister à un item et qu'un élément ne peut à la fois supporter et questionner une hypothèse. Pour certains des items du TCS nous avons constaté une distribution bimodale des réponses des experts. Nous nous sommes alors questionnés sur la possibilité d'exclure ces questions, sur la base d'un problème de clarté ou de construction de la question. Mais en observant les réponses des étudiants, celles-ci avaient également une distribution bimodale reflétant ainsi qu'il peut avoir plus d'une façon de percevoir l'énoncé de l'item, autant pour l'expert que pour l'étudiant, sans qu'il y ait nécessairement une bonne et une mauvaise conception. Par définition, le TCS repose sur des problèmes mal définis (Charlin, Roy, Brailovsky, Goulet, & van der Vleuten, 2000) et il peut être considéré acceptable (Charlin *et al.*, 2006) dans ce contexte d'avoir une grande variabilité

dans les réponses allant possiblement même jusqu'à des distributions bimodales. Dans le but de mieux comprendre, les différentes conceptions reliées à un item, il pourrait être intéressant, particulièrement dans un contexte formatif, de demander aux membres du panel d'experts de justifier leurs réponses et d'afficher celles-ci aux étudiants. Ces explicitations pourraient ainsi permettre aux étudiants de mieux saisir les différents points de vue en présence et comparer leurs propres jugements à celui d'experts dans le domaine.

D'un point de vue d'une utilisation au niveau sommatif, certains auteurs ont suggéré qu'il pourrait être pertinent, lors de la construction du test, de veiller à avoir une répartition plus équilibrée des réponses modales sur l'échelle de type Likert (See *et al.*, 2014). Le format de question, étant par définition mal défini, il est possible que les questions ayant des réponses aux extrémités de l'échelle de type Likert soient peu utilisées ce qui pourrait influencer la façon de répondre des participants à une épreuve sommative. Dans l'étude de See *et al.* (2014) le fait de recoder en évitant les réponses extrêmes a contribué à augmenter le score d'un écart type chez les participants ayant eu les scores les plus faibles. Dans le but de diminuer la possibilité que de telles stratégies soient efficaces, il serait souhaitable de pouvoir optimiser les réponses à l'étape de conception des questions de façon à introduire a priori une répartition optimale des items d'un TCS selon les choix proposés dans l'échelle de réponse (See *et al.*, 2014). Nous ne croyons pas que cette stratégie ait influencé les réponses des participants dans la présente étude, considérant qu'il n'y avait pas d'enjeux de notes pour les étudiants, mais il serait souhaitable de tenir compte de cette recommandation dans une version ultérieure du TCS, si elle est utilisée à des fins sommatives. Des résultats préliminaires (Dumas, Blais, & Charlin, 2014) supportent le fait que les réponses des concepteurs de TCS puissent guider à priori un tel choix des items.

Au niveau de la fidélité du test comme source d'évidence de la validité, nous avons obtenu des niveaux de consistance interne (alpha de Cronbach variant entre 0,59 et 0,69) qui se situent légèrement sous les niveaux habituellement rapportés dans la littérature (Dory *et al.*, 2012; Tavakol & Dennick, 2011). Ce résultat s'explique, en partie, par le faible nombre d'items retenus dans la version finale du TCS. Ce niveau de consistance interne implique une erreur de mesure un peu plus élevée qui aurait pu influencer la capacité du test à discriminer les différents niveaux des étudiants. Comme proposé précédemment, l'augmentation du nombre d'items

devrait avoir un effet positif au niveau de la consistance interne. Selon la formule de Spearman–Brown, l’augmentation à 150 items du TCS pourrait augmenter la valeur de l’alpha de Cronbach à 0,82. Lors de l’analyse de la consistance interne, nous avons constaté un niveau globalement faible et des variations importantes d’une cohorte à l’autre. Dans un premier temps, nous avons effectué une optimisation des items ayant une valeur de la corrélation item-test inférieure à 0,05 pour les trois cohortes (TCS 71), comme recommandé par la littérature (Dory *et al.*, 2012). Mais en procédant ainsi le coefficient alpha de la cohorte de 3^e année demeurait quand même à un niveau très faible. Nous avons donc refait une optimisation en essayant de garder un seuil minimal du niveau alpha (0,6) pour chacune des cohortes, ce qui a conduit à des choix différents d’items dans la version retenue (TCS 67). Il est difficile à ce stade de déterminer la raison qui pourrait expliquer ce profil de réponse différent de la cohorte de 3^e année. Des entrevues avec des membres de cette cohorte auraient pu nous informer sur la façon dont certains items ont été interprétés spécifiquement par cette cohorte. L’utilisation subséquente du TCS dans le programme pourra nous renseigner s’il s’agit d’une particularité ponctuelle ou non du profil de réponses de cette cohorte de 3^e année.

Plusieurs limites du coefficient alpha de Cronbach comme mesure de la fidélité ou de la consistance interne ont été rapportées dans la littérature de façon générale (Green & Yang, 2009; Panayides, 2013; Sijtsma, 2009) et dans le contexte spécifique du TCS (Lineberry *et al.*, 2013). Par exemple, une des prémisses de l’utilisation de ce coefficient est que chaque item contribue de façon égale à mesurer ce même trait latent (tau-équivalent). Lorsque ces conditions ne sont pas respectées et en présence de multidimensionalité, le coefficient alpha de Cronbach sous-estime la fidélité (Graham, 2006; Schmitt, 1996). Bien que ce coefficient ait été largement utilisé dans le contexte du TCS, peu d’études, à notre connaissance ont exploré l’unidimensionalité des TCS conçus et le fait que les données rencontrent les prémisses du modèle tau-équivalent. L’utilisation d’analyse factorielle et de modélisation en équations structurelles pourrait nous permettre de mieux comprendre ces aspects et déterminer si le coefficient alpha est approprié comme mesure de la fidélité pour le TCS (Graham, 2006; Green & Yang, 2009). Cependant lorsqu’il y a peu de participants le recours à ces modèles est problématique, il serait donc nécessaire de recruter des étudiants de plusieurs programmes en physiothérapie pour se conformer aux besoins de ces modèles.

Le support de réponse de type Likert comportant cinq niveaux peut aussi être remis en question dans le contexte d'une évaluation des apprentissages en formation de base. Il est possible, particulièrement pour les étudiants de la cohorte de 2^e et 3^e année, que la différence entre les niveaux de support de l'énoncé soit difficile à interpréter pour eux. Par exemple, il est vraisemblable que les énoncés : plus probable et presque certain (pour les items reliés à l'évaluation) ou plus indiqué ou complètement indiqué (pour les items reliés à l'intervention) fassent appel à un niveau de développement des scripts cliniques trop élevés pour ces cohortes considérant le peu d'exposition clinique de ceux-ci. Le niveau des scripts des étudiants de 2^e et 3^e année ne leur permettrait pas de détecter des différences dans le poids de l'énoncé, mais permettrait de discerner, à tout le moins, une direction (supporte ou remet en question). Une analyse des processus cognitifs associés à la sémantique du support de réponse qui utiliserait un protocole à voix haute pourrait nous guider sur cette question en comparant la façon dont les différentes cohortes s'y prennent pour répondre aux items. Possiblement un support de réponse spécifiquement associé au niveau de développement des scripts des participants pourrait améliorer le processus de collecte de donnée du TCS.

6.1.2 Résultats en lien avec le PRC

Dans le cas des PRC, en réponse à notre première question spécifique de recherche, nous n'avons pas été en mesure de distinguer le niveau de développement du raisonnement clinique entre les cohortes d'étudiants en physiothérapie à l'aide des scores attribués à partir de l'outil PRC. Par contre, les résultats des PRC ont été en mesure de discriminer entre le niveau de raisonnement clinique des experts et des étudiants.

Plusieurs facteurs peuvent être à l'origine de cette absence de discrimination des scores qui diffère des résultats de Groves *et al.* (2002, 2013) et Dory *et al.* (2014). Dans un premier temps nous avons constaté un faible niveau de cohérence interne des scores des PRC en calculant ceux-ci selon la méthode préconisée par Groves *et al.* (2002). En effectuant les calculs de cohérence interne par item, nous avons découvert qu'une grande partie de ce problème résidait dans les items à réponses ouvertes. En considérant qu'il y avait peu d'information pertinente pouvant nous guider dans la littérature existante sur les PRC, nous avons fait le choix à ce moment de conserver ces items puisqu'ils représentent l'aspect distinctif de ce test et que c'est sur la base

de ces réponses ouvertes (diagnostique ou thérapeutique) que les profils des justifications des caractéristiques sont bâtis. Nous avons par contre, réduit considérablement le poids relatif de ces items (de 50 % à $16/218=7,3$ %) dans le calcul du score total, pour diminuer l'erreur de mesure associée à ces items, ce qui n'avait pas été fait précédemment dans les études portant sur les PRC. Ces difficultés avec les réponses ouvertes, n'ayant pas été mentionnées lors des études faites dans le domaine de la médecine, il est possible que des différences existent au niveau du diagnostic en médecine et en physiothérapie. En effet, même s'il existe des similitudes dans le processus diagnostique entre le physiothérapeute et le médecin, il y a aussi des différences importantes. Bien que nous ayons centré les vignettes sur le diagnostic d'une structure lésée, cet aspect ne représente qu'une partie du diagnostic du physiothérapeute, car en pratique une bonne partie de ce diagnostic du physiothérapeute est de cibler et prioriser les déficiences et les incapacités du patient (Ordre professionnel de la physiothérapie, 2012). Cette adaptation des PRC pour l'intervention nous a fait réaliser la difficulté d'utiliser ce format à d'autres fins que pour le diagnostic, cette difficulté n'ayant pas été documentée dans la littérature jusqu'à maintenant, celle-ci peut certainement limiter son utilisation dans les différentes sciences de la santé.

Comme mentionné précédemment, à l'heure actuelle il n'existe pas de guide de construction pour les PRC et les études existantes sont peu explicites quant à plusieurs aspects importants de conception du test et de l'attribution des scores. Ainsi pour le calcul de la consistance interne du test, il n'est pas mentionné si le calcul est effectué par item ou par vignette (Groves *et al.*, 2013; Groves *et al.*, 2002). À partir du choix que nous avons effectué de pondérer les items à réponse ouverte de la même façon que les items de justification, nous avons considéré tous les items sur un pied d'égalité et avons calculé la consistance interne par item. Cette méthode de calcul de l'alpha de Cronbach (à partir des 218 items) peut possiblement surestimer la valeur de cohérence interne puisque le nombre d'items est un facteur important mentionné par plusieurs auteurs (Panayides, 2013; Streiner, 2003; Tavakol & Dennick, 2011). La valeur de l'alpha de Cronbach calculée sur la base des huit vignettes est plus basse (0,54) que celle calculée par items (0,88), par contre ces deux valeurs représentent l'étendue des indices rapportés dans la littérature. Dans la première étude publiée sur les PRC qui incluait 10 vignettes, les niveaux de consistance interne (alpha de Cronbach) rapportés variaient de 0,63 à 0,81 (Groves *et al.*, 2002).

Dans une deuxième étude utilisant les mêmes vignettes, mais avec une autre cohorte de médecins et d'étudiants (Groves *et al.*, 2013), le coefficient alpha se situait entre 0,56 et 0,61. De son côté, Dory *et al.* (2014) rapportent des valeurs de consistance interne (alpha de Cronbach) de 0,49 à 0,76 pour des PRC comportant aussi 10 vignettes.

Contrairement au TCS, le format des PRC ne permet pas une optimisation en enlevant a posteriori certains items puisque ceux-ci sont reliés à un profil spécifique. Donc, une des seules façons d'influencer la consistance interne du test a posteriori est d'enlever une ou des vignettes, ce qui peut avoir pour effet de nuire à la validité de contenu du test. Or, dans notre étude, le faible nombre de contextes des PRC était déjà une menace à la représentativité du test par rapport au contenu du cours PHT 240 sur lequel est construite la table de spécification du test. Donc, même si certaines vignettes avaient posé problème au niveau de la corrélation item-total, il aurait été difficile de les enlever pour ne pas nuire à la validité de contenu du test.

Une autre menace à la validité de contenu et à l'utilité des PRC est la difficulté de les utiliser en contexte d'intervention. Nous avons construit au départ trois vignettes qui portaient sur l'intervention de façon générale, mais de façon à pouvoir gérer la grande quantité de réponses possibles, nous avons ciblé seulement les exercices comme type d'intervention accepté. Les vignettes ont donc été bâties de façon à avoir quelques réponses possibles à ce niveau. Lors de l'analyse des résultats de ces vignettes portant sur l'intervention, nous avons quand même constaté qu'il était très difficile de regrouper les exercices en catégorie pouvant nous permettre de construire le score. Cette difficulté était présente autant pour les experts que pour les étudiants. Nous avons noté également que plusieurs étudiants avaient obtenu 0 % pour certaines de ces vignettes parce qu'ils avaient répondu un diagnostic au lieu d'un exercice ou que leurs réponses ne correspondaient pas à un des choix du panel. De plus, certains étudiants (et experts) avaient proposé un 2^e exercice très similaire au 1^{er} compromettant ainsi une caractéristique de l'outil qui est d'être en mesure de proposer un 2^e diagnostic (ou traitement dans ce cas) différent. Pour éviter cette difficulté, il aurait été nécessaire de clarifier les consignes du test et de le souligner verbalement lors de la passation du test. Une mise à l'essai du format du test avant la passation officielle aurait pu également être utile pour minimiser les difficultés rencontrées par les étudiants avec ce format. Car même si les étudiants ont déjà utilisé dans leur formation une

grille d'analyse de cas qui ressemblait aux PRC, plus d'étudiants ont mentionné avoir eu de la difficulté à s'adapter aux PRC qu'au TCS. Il aurait été intéressant de vérifier auprès des étudiants à l'aide d'une entrevue semi-structurée les raisons pour lesquelles les étudiants ont eu de la difficulté à s'adapter à ce format. Quelques pistes d'explication ressortent des commentaires qualitatifs récoltés lors du questionnaire : premièrement la grande quantité d'information de la vignette, qui représente plus d'une page de données cliniques, semble poser problème pour certains étudiants. Deuxièmement, un autre aspect mentionné par les étudiants et les experts est l'échelle à sept niveaux qui est possiblement trop détaillée. Pour les cohortes d'étudiants de 4^e année et les experts, une échelle de type Likert comportant cinq niveaux comme pour le TCS serait probablement suffisante. La différence entre les niveaux légèrement et modérément ou modérément et fortement est possiblement trop mince pour être discerné clairement même par des cliniciens ayant une bonne expérience clinique. Pour les cohortes de 2^e et 3^e année, encore une fois, étant donné que leurs scripts cliniques ne sont pas encore solidement formés, une échelle à trois niveaux pourrait être suffisante. De la même façon que nous le suggérons pour le TCS, une analyse des processus cognitifs avec un protocole à voix haute nous permettrait probablement de déterminer quelles sont les caractéristiques optimales du support de réponse en fonction du niveau de l'étudiant.

L'attribution des scores est une autre des difficultés rencontrées avec le format PRC pouvant influencer la fidélité du test. Dans un premier temps, nous avons regroupé les réponses des experts à la question ouverte en profil diagnostique selon une taxonomie clinique. Par contre, en pratique clinique, plusieurs termes existent pour désigner une entité diagnostique et une analyse de l'équivalence des différents termes doit être faite. Compte tenu de la difficulté de catégoriser ces termes, il aurait été souhaitable de recourir à un autre évaluateur pour confirmer les catégories établies à partir des réponses des experts et des étudiants. Malheureusement, le temps et les ressources disponibles ne nous ont pas permis de réaliser cette étape, ce qui est une limite de cette partie de l'étude.

De façon générale, tout le processus d'attribution des scores des PRC est complexe et inclut potentiellement plusieurs sources d'erreur. Cet outil ayant fait l'objet de peu d'étude empirique sur des aspects importants de sa validité, il serait recommandé de documenter des éléments

comme l'effet de la composition du panel, le processus de catégorisation des réponses des experts et le choix de l'échelle de réponse. Il est donc probablement prématuré, à l'heure actuelle, d'utiliser cet outil en formation de base à des fins sommatives avant d'avoir plus d'études sur ses caractéristiques métriques.

6.1.3 Comparaison entre les deux outils

Un élément surprenant de nos résultats est le niveau de corrélation (Pearson $r=0,231$) entre les scores des PRC et du TCS. Cette corrélation peut être considérée faible étant donné que les deux outils ont été construits pour nous permettre de poser un jugement sur l'habileté à interpréter les données cliniques et qu'ils ont été conçus à partir des mêmes objectifs d'apprentissage du cours PHT 240. Une faible corrélation entre les scores des deux outils a également été notée dans l'étude de Dory *et al.* (2014) (Spearman $\rho=0,059$ à $0,357$) alors qu'une corrélation modérée ($r=0,44$ à $0,49$) était présente dans l'étude de Groves *et al.* (2013). Il est à noter que ces deux études ont également utilisé les mêmes thèmes cliniques pour les vignettes PRC et TCS de façon à pouvoir comparer les deux outils. Tout comme les auteurs de ces deux études, nous considérons que cette faible corrélation est due, en partie, au fait que les deux outils permettent un accès à des composantes différentes du raisonnement clinique.

Bien qu'à première vue le TCS et les PRC partagent plusieurs points en commun, entre autres, au niveau de l'interprétation des caractéristiques et du poids relatif des éléments du script clinique ainsi que de l'attribution du score par un panel d'experts. Par contre, il y a aussi plusieurs points qui diffèrent. Par exemple, un des éléments qui différencie les PRC du TCS est le fait que l'étudiant puisse énoncer ses propres hypothèses dans une réponse ouverte. Cette caractéristique des PRC en fait un outil plus près de la situation authentique où le clinicien doit généralement lui-même construire ses hypothèses et les justifier à l'aide des informations disponibles. En fonction du niveau de l'étudiant et de son expérience clinique, l'attribution d'une hypothèse diagnostique ou de traitement à partir de la vignette peut faire appel à des processus cognitifs analytiques, non analytiques ou une combinaison des deux (Durning *et al.*, 2015).

Dans le cas du TCS, les participants ne formulent pas leurs hypothèses ce qui fait que dans certains cas le participant doit analyser une information en lien avec une hypothèse à laquelle il ne souscrit pas d'emblée. Cette activation induite d'un script dans le TCS peut être perçue comme perturbante par certains participants. Comme mentionné précédemment, il pourrait être pertinent d'analyser les différences, à l'aide d'un protocole à voix haute, des processus cognitifs impliqués lorsque l'hypothèse est donnée par le participant ou lorsqu'elle est incluse dans l'item.

Comme mentionné précédemment, une autre différence notée entre les deux outils est la grande quantité d'information associée aux PRC. Cette quantité d'information peut être vue comme un avantage pour se rapprocher de la complexité de la situation authentique, par contre, celle-ci peut limiter le nombre de contextes d'évaluation possible et ainsi nuire à la représentativité du test. Donc, en termes du nombre de contextes pris en considération (par le biais des vignettes) dans la même période de temps, le TCS présente un avantage. En lien avec cet avantage, certains auteurs (Groves *et al.*, 2013) ont proposé, par souci d'efficacité, d'utiliser le TCS comme outil de dépistage de difficulté de raisonnement clinique préalablement à l'administration du PRC.

Donc, en réponse à la première question spécifique de recherche, nos résultats indiquent que les scores du TCS de notre étude peuvent nous permettre de distinguer le niveau de développement du raisonnement clinique de certains groupes d'étudiants en physiothérapie (cohorte de 2^e année et 4^e année) alors que les scores des PRC de notre étude ne nous le permettent pas pour l'ensemble des trois cohortes.

6.2 Deuxième question de recherche

La deuxième question était énoncée de la façon suivante :

Est-ce qu'il y a une relation entre les résultats aux tests de raisonnement clinique et les résultats de stage?

Globalement les résultats de notre étude nous indiquent qu'il existerait un lien entre l'évaluation de la performance de raisonnement clinique des étudiants en physiothérapie lors des stages et leurs résultats au TCS et aux PRC. De façon surprenante, compte tenu des limites de l'outil, ce

lien apparaît plus fort avec les PRC qu'avec le TCS. En effet pour les PRC, la moyenne des scores augmente selon le niveau de performance du raisonnement clinique évalué en contexte de stage. Dans le cas du TCS, les scores des étudiants ayant eu des difficultés importantes de raisonnement clinique en stage sont plus faibles que les autres catégories, mais on note peu de différence entre les trois autres catégories.

En général, les résultats académiques et les résultats en stage sont peu ou pas corrélés autant en médecine (Hamdy *et al.*, 2006) qu'en physiothérapie (Rheault & Shafernich-Coulson, 1988). Cette différence peut probablement être expliquée par le fait que la performance en stage est modulée par plusieurs autres composantes de la compétence que ce qui est traditionnellement évalué dans la portion académique d'un curriculum de formation en sciences de la santé. Bien que le raisonnement clinique soit relié aux connaissances et à d'autres composantes de compétence, nous avons cru qu'il est possible de cibler spécifiquement cette composante en situation de stage pour la mettre en relation avec des outils d'évaluation du raisonnement clinique. À notre connaissance, notre étude est la première à avoir mis en relation la composante raisonnement clinique de la performance en milieu authentique de stage et des tests de raisonnement clinique papier crayon. Pour être en mesure d'effectuer cette étude, nous avons dû utiliser une méthodologie qui nous a permis de porter un jugement global sur la performance de raisonnement clinique de l'étudiant à partir des commentaires écrits des superviseurs cliniques puisque les formulaires ne comportaient pas une cotation spécifique sur le raisonnement clinique. Nous croyons qu'en plus de cibler le raisonnement clinique, cette méthodologie à l'avantage de pouvoir corriger jusqu'à un certain point, certaines critiques de l'évaluation en situation de stage. En effet, plusieurs auteurs (Albanese, 1999; Dudek, Marks, & Regehr, 2005; Hatala & Norman, 1999) ont noté qu'en situation de stage les superviseurs vont hésiter à donner une cotation qui pourrait entraîner des conséquences négatives chez l'étudiant (*failure to fail*). Ce phénomène semble répandu et peut compromettre la valeur de l'évaluation en situation de stage (Albanese, 1999). Dans notre échantillon, nous avons noté à plusieurs reprises (n=5) que la cotation dans les rôles experts et la nature des commentaires qualificatifs n'étaient pas toujours cohérentes quand ceux-ci étaient négatifs, mais qu'une bonne cohérence existait quand ces derniers étaient positifs. Cette non-concordance des cotations et des commentaires négatifs a été également notée par d'autres auteurs (Schwind, Williams,

Boehler, & Dunnington, 2004). Ceux-ci ont noté que les commentaires qualitatifs identifiaient mieux les déficiences des étudiants que les cotations quantitatives. Les cinq cas où notre jugement de la performance de l'étudiant diffère de la cotation du superviseur dans les rôles experts représentent la moitié des cas notés de difficulté mineure de raisonnement clinique (n=10) rencontrés dans notre échantillon.

Une autre des critiques fréquemment rencontrées dans la littérature sur l'évaluation en situation de stage porte sur la forte subjectivité et la faible fidélité inter évaluateurs du jugement posé, celui-ci pouvant être influencé par plusieurs biais (Williams, Klamen, & McGaghie, 2003, Downing, 2005; Daelmans *et al.*, 2005; Kogan, Hess, Conforti, & Holmboe, 2010; McGill, van der Vleuten, & Clarke, 2011). Pour atténuer ce problème, il est en général recommandé d'avoir un plus grand échantillonnage d'évaluateurs et de contextes différents (van der Vleuten & Schuwirth, 2005; Williams *et al.*, 2003). Dans le cas des étudiants de la cohorte de 2^e année, ceux-ci n'ont complété qu'un seul stage et par le fait même n'ont été évalués que par un superviseur (la majorité du temps). Nous avons donc dû adapter les critères de nos catégories pour cette cohorte pour tenir compte de cette particularité. Nous avons également refait une analyse séparée pour les cohortes de 3^e et 4^e année, les évaluations des membres de ces cohortes reposant sur un jugement basé sur plusieurs stages, nous croyons que celui-ci représente mieux la performance des étudiants.

Bien qu'elle puisse potentiellement corriger certaines lacunes de l'évaluation en situation authentique, la catégorisation utilisée dans cette méthodologie comporte elle aussi certaines limites. En effet, cette catégorisation repose sur une observation indirecte qui dépend de la qualité des notes du superviseur et sur l'interprétation de celle-ci par une tierce personne. Or, cette interprétation peut aussi incorporer ses propres biais dans le jugement qu'elle porte. Reconnaisant cette difficulté potentielle, nous avons ajouté une étape impliquant un deuxième observateur qui a refait une catégorisation à partir des commentaires des superviseurs cliniques et avons démontré qu'un accord élevé entre deux évaluateurs peut être obtenu dans cette tâche. Bien qu'elle ne soit pas une assurance de l'absence de biais pouvant être associé à celle-ci, cette étape supplémentaire indique qu'une information relativement claire sur le raisonnement

clinique des étudiants en stage peut être dégagée à partir des notes qualitatives des superviseurs cliniques.

Cette relation observée entre les résultats de stage et les tests de raisonnement clinique (pour les PRC et le TCS) suggère que ceux-ci pourraient être utilisés dans une optique de dépistage de problème de raisonnement clinique pour les étudiants en physiothérapie. Dans notre échantillon, trois des quatre cas identifiés comme ayant eu des problèmes de raisonnement clinique majeur en stage ont eu des scores à un des tests de raisonnement clinique à plus de deux écarts types de leurs cohortes de référence. Par contre, encore une fois, la corrélation entre les deux outils est faible et suggère d'utiliser les deux formats de façon complémentaire dans une optique de dépistage.

Compte tenu de la nature de cette partie de l'étude et que la relation observée repose sur un nombre limité de cas ayant des difficultés de raisonnement clinique en stage, il faudrait être prudent dans l'interprétation des tests à cette étape et d'autres études seront nécessaires pour préciser la nature de cette relation. Néanmoins, le recours au TCS pour dépister des problèmes de raisonnement clinique est déjà une pratique du programme de médecine de l'Université du Colorado où cet outil est utilisé (en combinaison avec les évaluations en milieu de stage) pour dépister et également réévaluer les candidats après une remédiation (Guerrasio & Aagaard, 2014).

Donc, en réponse à notre deuxième question spécifique, nos résultats nous suggèrent qu'il existe un lien entre l'évaluation de la performance de raisonnement clinique des étudiants en physiothérapie lors des stages et leurs scores au TCS et aux PRC.

6.3 Limites de l'étude

Dans les sections précédentes, nous avons abordé certaines des limites associées aux outils eux-mêmes et au processus de catégorisation à partir des formulaires de stage. Ils existent également d'autres limites qui s'appliquent aussi à l'ensemble de cette recherche.

Premièrement, au niveau de la taille de l'échantillon, il aurait été souhaitable d'obtenir la participation de l'ensemble des étudiants des cohortes concernées. Bien que la proportion des

étudiants ayant participé à l'étude représente plus de 60 % des étudiants, il est possible que plusieurs étudiants ayant des difficultés académiques et/ou de stage, aient choisi ne pas participer à l'étude par crainte d'un jugement et de conséquences négatives à leurs égards. Même s'il a été spécifié clairement dans le formulaire de consentement que les informations les concernant seraient traitées de manière confidentielle, cette crainte peut avoir contribué à la non-participation de certains étudiants ayant des difficultés académiques et/ou de stage. Cette possible sous-représentation des étudiants ayant des difficultés semble plus présente dans la cohorte des étudiants de 4^e année. De façon générale pour les trois cohortes, les résultats de cette portion de l'étude, malgré le faible nombre d'étudiants ayant eu des difficultés de raisonnement clinique en stage, nous suggèrent de poursuivre l'investigation dans des études subséquentes.

Deuxièmement, il est possible que certains étudiants aient participé à l'étude dans le seul but de faire partie du tirage et que ceux-ci n'aient pas répondu au meilleur de leurs capacités. Même si cette possibilité existe, nous ne croyons pas qu'un nombre significatif d'étudiants ait agi de la sorte compte tenu de la culture de performance des étudiants en sciences de la santé en général et du fait que les étudiants ayant obtenu des résultats faibles aux tests avaient aussi pour la plupart des difficultés de raisonnement clinique observées en stage. Jusqu'à un certain point, le temps de passation des tests peut aussi nous rassurer sur cette possibilité.

Troisièmement, il est possible qu'il y ait eu un effet de contamination entre les cohortes puisque les tests ont été administrés à des périodes différentes pour chacune des cohortes. Nous ne croyons pas que cet aspect a eu un effet significatif sur la qualité des données, étant donné qu'il n'y avait pas d'enjeux académiques pour les étudiants. Par contre, il aurait été préférable d'administrer les tests aux trois cohortes de façon simultanée pour éviter que les items des tests puissent être discutés d'une cohorte à l'autre.

Quatrièmement, de façon à favoriser la participation des experts, l'administration des tests à ceux-ci ne s'est pas effectuée de façon aussi contrôlée que pour les cohortes d'étudiants. En effet, contrairement aux étudiants, les experts pouvaient répondre aux tests aux moments qui leur convenaient et les retourner en formats électroniques. Malgré les consignes reliées au temps de passation suggéré et à la non-utilisation de ressources externes, il est possible que certains experts n'aient pas respecté ces consignes. Encore une fois, les enjeux liés à la performance

lors du test n'étant pas majeurs, il est peu probable que cet aspect ait eu un impact sur les réponses du panel d'experts.

Cinquièmement au niveau du processus de développement des tests, les concepteurs et les personnes impliquées dans la relecture des vignettes et des items n'avaient pas d'expérience avec les outils TCS et PRC. Il est donc possible que le peu d'expérience, avec les formats TCS et PRC, des personnes impliquées dans les différentes étapes de conception des tests ait influencé la qualité de construction des tests. Si nous en avons eu la possibilité, il aurait été préférable de recourir à des personnes ayant une expertise de contenu en physiothérapie et une connaissance des formats TCS et PRC.

Sixièmement, le fait que les tests de raisonnement clinique de la présente étude ayant été construits à partir d'une table de spécification d'un cours de 2^e année du programme de l'Université de Sherbrooke limite la possibilité de généraliser les conclusions à d'autres programmes. Des études impliquant d'autres programmes universitaires en physiothérapie seraient donc nécessaires pour confirmer si nos conclusions s'appliquent à d'autres contextes de formation. L'organisation des contenus pouvant varier d'un programme à l'autre, il est possible que les tests développés dans cette étude ne puissent être administrés à des étudiants de 2^e année d'autres programmes universitaires, mais pourrait possiblement être utilisé à d'autres moments du curriculum.

6.4 Pistes de recherche

Une des pistes de recherche prioritaire devrait être de revoir le processus de mise au point des instruments pour le rendre plus rigoureux et systématique. Principalement au niveau des PRC, des efforts devraient être faits pour clarifier et possiblement simplifier l'attribution des scores. Compte tenu des résultats de notre étude, il pourrait être pertinent de poursuivre certaines collectes de données avec des objectifs légèrement différents. Une piste de recherche, en référence à la proposition de Lissitz et Samuelsen (2007) de considérer le réseau nomologique des tests de raisonnement clinique qui a été peu documentée jusqu'à maintenant consisterait à étudier de manière longitudinale l'évolution des scores au TCS et aux PRC chez une même cohorte d'étudiants. Cette recherche qui nécessiterait une collecte de donnée supplémentaire

pourrait nous permettre de constater une évolution du raisonnement clinique à travers un curriculum de formation. En lien avec nos résultats, une période de deux ans entre les deux administrations du test serait probablement nécessaire pour apprécier une différence. De plus, il pourrait être intéressant de mettre en parallèle l'évolution des scores de raisonnement clinique avec l'évaluation en situation de stage à travers le curriculum, de façon à faire ressortir l'aspect dynamique de cette composante majeure de la compétence du physiothérapeute. Dans la même optique, il pourrait être pertinent d'étudier l'évolution des scores des tests après une période de remédiation du raisonnement clinique le tout en parallèle, encore une fois, avec l'évaluation de stage post remédiation. Ces études de la relation entre les scores des tests de raisonnement clinique et l'évaluation de stage dans plusieurs contextes différents pourraient nous permettre de mieux comprendre certains aspects de la validité et de l'utilité de ces tests. Dans le cadre de cette étude, nous avons été surpris de constater une faible corrélation entre les scores des deux formats de tests d'évaluation du raisonnement clinique. Afin de mieux comprendre les processus de réponse reliés à ces deux formats, il pourrait être approprié d'utiliser un protocole de pensée à voix haute chez les étudiants pendant l'administration du TCS et des PRC. L'utilisation d'un protocole de pensée à voix haute pourrait ainsi nous donner accès aux pensées et aux processus cognitifs de raisonnement clinique mobilisés par les étudiants et ainsi mieux saisir les différences et les similarités entre ces deux formats.

Au cours de cette recherche, nous avons exploré l'évaluation du raisonnement clinique en situation de stage et avons été confrontés à certaines limites de celle-ci dans ce contexte. Une des limites observées dans cette étude et qui est fréquemment mentionnées dans la littérature relève de la difficulté à donner une évaluation négative à un étudiant. Une piste de recherche que nous avons explorée dans notre méthodologie et qu'il pourrait être pertinent d'étudier plus en profondeur est le recours aux commentaires qualitatifs de l'évaluation de stage plutôt qu'aux seules cotations pour déterminer du niveau d'atteinte de la compétence. Une co-construction de l'évaluation de stage avec une tierce personne pourrait ainsi être effectuée lorsqu'il y a non-concordance entre la cotation et les commentaires qualitatifs observés. Cette approche qui pourrait s'appliquer au raisonnement clinique et aux diverses compétences évaluées en stage pourrait possiblement permettre de diminuer les biais attribués aux jugements du superviseur et améliorer ainsi globalement la qualité des évaluations dans ce contexte.

Au niveau de l'analyse des résultats, nous avons traité nos données de façon traditionnelle avec la théorie classique des tests, mais en considérant les limitations inhérentes à ce modèle, il pourrait être pertinent de refaire les analyses avec d'autres modèles de mesure. Par exemple, le fait de recourir à la modélisation de Rasch pourrait nous permettre d'optimiser les tests de raisonnement clinique en considérant les items, les membres du panel d'experts et les étudiants qui ne se conforment pas au modèle. Ils pourraient ainsi être très intéressants de comparer les versions et les résultats des tests obtenus avec les deux modèles de mesure. Par contre, le modèle de Rasch est plus restrictif quant à son utilisation comme par exemple au niveau de la non-indépendance des items dans les vignettes et implique un nombre plus élevé de participants ce qui n'est pas possible dans le programme de physiothérapie de l'Université de Sherbrooke, mais pourrait l'être dans d'autres programmes de formation en sciences de la santé ayant des cohortes plus nombreuses.

CONCLUSION

Cette recherche visait à répondre à une question générale de recherche et à deux questions spécifiques portant sur l'évaluation du raisonnement clinique dans un contexte de formation professionnelle en physiothérapie. A la lumière des réponses données aux deux questions spécifiques de recherche, nous allons tenter de répondre à notre question générale de recherche : est-ce que le TCS et les PRC développés en contexte médical permettent d'évaluer le raisonnement clinique en physiothérapie?

D'abord, nous croyons que les tests standardisés d'évaluation du raisonnement clinique tels que le TCS et les PRC originellement développés pour le domaine de la médecine peuvent nous aider à porter un jugement sur le raisonnement clinique des étudiants en physiothérapie. Malgré certaines différences, entre autres reliées à la conception et à l'importance du diagnostic entre la physiothérapie et la médecine, il apparaît que ce type de modalité d'évaluation peut être adapté au domaine de la physiothérapie. Nous croyons que les limites observées des outils ne sont pas reliées spécifiquement au contexte de cette profession, mais démontrent, surtout dans le cas du PRC, qu'il reste encore beaucoup de travail de développement à faire sur l'outil et qu'il est probablement prématuré d'utiliser cet outil à des fins sommatives présentement.

Dans une optique d'utilisations de ces outils en formation universitaire, il est important de garder en tête que ces deux outils ne permettent de poser un jugement que sur des aspects limités du raisonnement clinique et qu'ils doivent être utilisés dans une perspective globale d'un programme d'évaluation pour pouvoir apprécier l'ensemble du raisonnement clinique d'un étudiant. Dans cette perspective globale, on doit également inclure l'évaluation du raisonnement clinique en situation authentique de stage et bien que ce type d'évaluation pose aussi certains défis, elle est indéniablement un aspect central d'un programme d'évaluation en sciences de la santé.

Apprécier en profondeur le raisonnement clinique d'un étudiant en physiothérapie que ce soit en utilisant des tests standardisés tels que le TCS ou les PRC ou bien en situation de stage demeure toutefois difficile, car le cheminement utilisé par l'étudiant pour répondre à une

question ou résoudre un problème clinique n'est pas toujours explicite. Pour accéder à ce cheminement, il peut être nécessaire de retracer les étapes du processus et ce qui a possiblement conduit à des erreurs de raisonnement clinique. Cette façon de faire peut non seulement aider à mieux évaluer l'étudiant, mais peut aussi créer une situation d'évaluation qui soutient l'apprentissage. Le raisonnement clinique étant indissociable des connaissances qui le composent, il peut donc être profitable d'utiliser ces situations d'évaluations du raisonnement clinique, pour donner une rétroaction sur le raisonnement clinique ainsi que sur les connaissances et sur la mobilisation efficace de celles-ci en cohérence avec les principes de l'approche par compétences.

À notre connaissance, cette recherche a été la première à développer des tests de formats TCS et PRC pour le domaine de la physiothérapie. Bien que le TCS ait été utilisé dans plusieurs autres domaines de la santé, les PRC n'avaient été utilisés jusqu'à maintenant que pour la médecine. Dans le but d'adapter les PRC aux caractéristiques de la physiothérapie, nous avons inclus des vignettes portant sur l'intervention dans ce domaine, alors que précédemment seul l'aspect du diagnostic avait été abordé avec ce format.

L'originalité de cette recherche repose également sur l'utilisation d'une méthodologie nous permettant de mettre en relation les observations sur le raisonnement clinique en situation authentique de stage et les résultats à des tests standardisés de raisonnement clinique. Cette recherche est aussi une des seules à avoir mis en relation ces éléments d'information sur le raisonnement clinique à des fins de dépistage de difficulté de raisonnement clinique. Bien que cet aspect demeure embryonnaire, nous croyons qu'il existe un potentiel important dans cette fonction des tests standardisés pour améliorer le raisonnement clinique de certains étudiants en difficulté.

En plus des aspects novateurs de cette thèse, nous croyons de plus que l'utilisation d'un cadre de validation pour développer nos tests de raisonnement clinique a contribué à améliorer la rigueur du processus. Ce cadre nous a amenés à considérer globalement tous les aspects du développement d'un test et de nous questionner sur les aspects à privilégier. À partir de ces choix, nous nous sommes efforcés de poser un regard critique sur les tests produits sans tenir pour acquise la qualité des instruments développés. L'analyse de plusieurs versions des tests

produits nous a permis de réaliser l'importance du processus de validation sur la qualité des outils permettant d'inférer sur le raisonnement clinique des étudiants. Dans cette optique, nous avons eu à ajuster certaines façons de faire, entre autres au niveau de l'optimisation du TCS et de l'attribution des scores des PRC. Nous sommes persuadés que ces ajustements ont permis d'améliorer la qualité des outils produits, mais des études subséquentes seront nécessaires pour déterminer de leurs applicabilités dans d'autres contextes.

Au terme de cette recherche, nous croyons que nos résultats peuvent alimenter la réflexion concernant l'évaluation du raisonnement clinique dans le domaine des sciences de la santé en général et plus particulièrement en ce qui a trait à la formation universitaire des étudiants en physiothérapie.

RÉFÉRENCES

- AERA, A. E. R. A., APA, A. P. A., & NCME, N. C. o. M. i. E. (1999). *Standards for educational and psychological testing*: Amer Educational Research Assn.
- AERA, A. E. R. A., NCME, N. C. o. M. i. E., & APA, A. P. A. (1966). *Standards for educational and psychological tests and manuals*: American Psychological Association.
- Aiken, A. (2012). Improved use of allied health professionals in the health care system: The case of the advanced practice physiotherapist in orthopedic care. *World Hospitals and Health Services*, 48(1), 28.
- Albanese, M. (1999). Rating educational quality: factors in the erosion of professional standards *Academic Medicine*, 74(6), 652-658.
- Allal, L. (1999). Acquisition et évaluation des compétences en situation scolaire. In J. D. E. Ollagnier (Ed.), *L'énigme de la compétence en éducation* (pp. 77-94). Bruxelles: De Boeck.
- APA, A. P. A., AERA, A. E. R. A., & NCME, N. C. o. M. i. E. (1954). *Technical recommendations for psychological tests and diagnostic techniques* (Vol. 51): American Psychological Association.
- Audétat, M. C., Faguy, J., Jacques, A., Blais, J.-G., & Charlin, B. (2011). Étude exploratoire des perceptions et pratiques de médecins cliniciens enseignants engagés dans une démarche de diagnostic et de remédiation des lacunes du raisonnement clinique. *Pédagogie Médicale*, 12(1), 7-16.
- Audétat, M. C., Laurin, S., & Sanche, G. (2011). Aborder le raisonnement clinique du point de vue pédagogique I. Un cadre conceptuel pour identifier les problèmes de raisonnement clinique. *Pédagogie Médicale*, 12(4), 223-229.
- Baillargeon, N. (2009). *Contre la réforme. La dérive idéologique du système d'éducation québécois*. Montréal: Les Presses de l'Université de Montréal.
- Bardin, L. (2007). *L'analyse de contenu*: Paris, Presses Universitaires de France.
- Benner, P., & Ovion, L. (1995). *De novice à expert: excellence en soins infirmiers*: InterEditions.
- Bennett, R. E. (1993). On the meaning of constructed response In R. E. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement: issues in constructed response performance testing and portfolio assessment* (pp. 1-27). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Beullens, J., Struyf, E., & Van Damme, B. (2005). Do extended matching multiple-choice questions measure clinical reasoning? *Med Educ*, 39(4), 410-417.
- Beullens, J., Struyf, E., & Van Damme, B. (2006). Diagnostic ability in relation to clinical seminars and extended-matching questions examinations. *Med Educ*, 40(12), 1173-1179.
- Biesta, G. (2010). Pragmatism and the philosophical foundations of mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Sage handbook of mixed methods in social & behavioral research*. Thousand Oaks, CA: SAGE Publications.

Black, P. (1995). Curriculum and assessment in science education: The policy interface. *International Journal of Science Education*, 17(4), 453-469.

Blais, J. G. (2012). Réflexions sur le déploiement de la quantification et de la mesure en éducation. Paper presented at the 34e Conférence ADMEE, Montréal.

Bloom, B. S. (1973). Handbook on formative and summative evaluation of student learning (2 ed.). New York: McGraw-Hill.

Bordage, G., Grant, J., & Marsden, P. (1990). Quantitative assessment of diagnostic ability. *Med Educ*, 24(5), 413-425.

Boshuizen, H. C., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science*, 16, 153-184.

Boulouffe, C., Doucet, B., Muschart, X., Charlin, B., & Vanpee, D. (2013). Assessing clinical reasoning using a script concordance test with electrocardiogram in an emergency medicine clerkship rotation. *Emerg Med J*. doi:10.1136/emmermed-2012-201737

Boursicot, K., Etheridge, L., Setna, Z., Sturrock, A., Ker, J., Smee, S., & Sambandam, E. (2011). Performance in assessment: consensus statement and recommendations from the Ottawa conference. *Med Teach*, 33(5), 370-383. doi:10.3109/0142159X.2011.565831

Brailovsky, C., Charlin, B., Beausoleil, S., Cote, S., & Van der Vleuten, C. (2001). Measurement of clinical reflective capacity early in training as a predictor of clinical reasoning performance at the end of residency: an experimental study on the script concordance test. *Med Educ*, 35(5), 430-436.

Brazeau-Lamontagne, L., Charlin, B., Gagnon, R., Samson, L., & van der Vleuten, C. (2004). Measurement of perception and interpretation skills during radiology training: utility of the script concordance approach. *Med Teach*, 26(4), 326-332.

Brooks, L. R., LeBlanc, V. R., & Norman, G. R. (2000). On the difficulty of noticing obvious features in patient appearance. *Psychol Sci*, 11(2), 112-117.

Carraccio, C. L., Benson, B. J., Nixon, L. J., & Derstine, P. L. (2008). From the educational bench to the clinical bedside: translating the Dreyfus developmental model to the learning of clinical skills. *Acad Med*, 83(8), 761-767. doi:10.1097/ACM.0b013e31817eb632

Carrière, B., Gagnon, R., Charlin, B., Downing, S., & Bordage, G. (2009). Assessing clinical reasoning in pediatric emergency medicine: validity evidence for a Script Concordance Test. *Annals of emergency medicine*, 53(5), 647-652.

Case, S. M., Swanson, D. B., & Ripkey, D. R. (1994). Comparison of items in five-option and extended-matching formats for assessment of diagnostic skills. *Acad Med*, 69(10 Suppl), S1-3.

Chamberland, M. (1998). Les Séances d'Apprentissage du Raisonnement Clinique (ARC): un exemple d'activité pédagogique contextualisée adaptée aux stages cliniques en Médecine. *Annales de médecine interne*, 149, 479-484.

Chamberland, M. (2006). Une vision multidimensionnelle du raisonnement clinique dans la pratique. Université de Sherbrooke. Sherbrooke, Qc.

Charlin, B. (2011). Aborder le raisonnement clinique d'un point de vue pédagogique. *Pédagogie Médicale*, 12(4), 197-198.

Charlin, B., Bordage, G., & van der Vleuten, C. P. (2003). L'évaluation du raisonnement clinique. *Pédagogie Médicale*, 4(1), 42-52.

Charlin, B., Boshuizen, H. P., Custers, E. J., & Feltovich, P. J. (2007). Scripts and clinical reasoning. *Med Educ*, 41(12), 1178-1184.

Charlin, B., Brailovsky, C., Leduc, C., & Blouin, D. (1998). The diagnosis script questionnaire: a new tool to assess a specific dimension of clinical competence. *Advances in health sciences education*, 3(1), 51-58.

Charlin, B., Desaulniers, M., Gagnon, R., Blouin, D., & van der Vleuten, C. (2002). Comparison of an aggregate scoring method with a consensus scoring method in a measure of clinical reasoning capacity. *Teach Learn Med*, 14(3), 150-156.

Charlin, B., Gagnon, R., Pelletier, J., Coletti, M., Abi-Rizk, G., Nasr, C., . . . van der Vleuten, C. (2006). Assessment of clinical reasoning in the context of uncertainty: the effect of variability within the reference panel. *Med Educ*, 40(9), 848-854. doi:10.1111/j.1365-2929.2006.02541.x

Charlin, B., Gagnon, R., Sauvé, E., & Coletti, M. (2007). Composition of the panel of reference for concordance tests: Do teaching functions have an impact on examinees' ranks and absolute scores? *Med Teach*, 29(1), 49-53.

Charlin, B., Lubarsky, S., Millette, B., Crevier, F., Audetat, M. C., Charbonneau, A., . . . Bourdy, C. (2012). Clinical reasoning processes: unravelling complexity through graphical representation. *Med Educ*, 46(5), 454-463.

Charlin, B., Roy, L., Brailovsky, C., Goulet, F., & van der Vleuten, C. (2000). The Script Concordance test: a tool to assess the reflective clinician. *Teach Learn Med*, 12(4), 189-195.

Christensen, N. (2009). *Development of Clinical Reasoning Capability In Student Physical Therapists*. (Ph.D.), University of South Australia, Adelaide.

Cizek, G. J. (2012). Defining and distinguishing validity: Interpretations of score meaning and justifications of test use. *Psychological methods*, 17(1), 31.

Cleather, J. (1995). Head, heart and hands: the story of physiotherapy in Canada: Canadian Physiotherapy Association.

Cobb, K. A., Brown, G., Hammond, R., & Mossop, L. H. (2015). Students' Perceptions of the Script Concordance Test and Its Impact on Their Learning Behavior: A Mixed Methods Study. *Journal of veterinary medical education*, 42(1), 45-52.

Code des professions. (2010a). Règlement sur certaines activités professionnelles pouvant être exercées en physiothérapie

Retrieved from

http://www2.publicationsduquebec.gouv.qc.ca/dynamicSearch/telecharge.php?type=2&file=%2F%2FM_9%2FM9R4.htm.

Code des professions. (2010b). Règlement sur les catégories de permis délivrés par l'Ordre professionnel de la physiothérapie du Québec

Retrieved from

http://www2.publicationsduquebec.gouv.qc.ca/dynamicSearch/telecharge.php?type=2&file=%2F%2FC_26%2FC26R196_1.htm.

Cohen, J. (1988). *Statistical power for the social sciences*. Hillsdale, NJ: Laurence Erlbaum and Associates.

Cohen, L. J., Fitzgerald, S. G., Lane, S., & Boninger, M. L. (2005). Development of the seating and mobility script concordance test for spinal cord injury: obtaining content validity evidence. *Assist Technol*, 17(2), 122-132. doi:10.1080/10400435.2005.10132102

Cohen, L. J., Fitzgerald, S. G., Lane, S., Boninger, M. L., Minkel, J., & McCue, M. (2009). Validation of the seating and mobility script concordance test. *Assist Technol*, 21(1), 47-56.

Cook, D. A. (2015). Much ado about differences: Why expert-novice comparisons add little to the validity argument. *Advances in health sciences education*, 20(3), 829-834.

Coulet, J. C. (2011). La notion de compétence : un modèle pour décrire, évaluer et développer les compétences *Le travail humain*, 74(1), 1-30.

Cox, M., Irby, D. M., Cooke, M., Irby, D. M., Sullivan, W., & Ludmerer, K. M. (2006). American medical education 100 years after the Flexner report. *New England journal of medicine*, 355(13), 1339-1344.

Creswell, J. W. (2010). Mapping the developing landscape of mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Sage handbook of mixed methods in social & behavioral research*. Thousand Oaks, CA: SAGE Publications.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, 52(4), 281.

Cronbach, L. J., & Thorndike, R. (1971). Educational measurement. *Test validation*, 443-507.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of educational research*, 58(4), 438-481.

Cruz, E. B., Moore, A. P., & Cross, V. (2012). A qualitative study of physiotherapy final year undergraduate students' perceptions of clinical reasoning. *Man Ther*, 17(6), 549-553. doi:10.1016/j.math.2012.05.013

Daelmans, H. E., van der Hem-Stokroos, H. H., Hoogenboom, R. J., Scherpbier, A. J., Stehouwer, C. D., & van der Vleuten, C. P. (2005). Global clinical performance rating, reliability and validity in an undergraduate clerkship. *Neth J Med*, 63(7), 279-284.

Dalton, M., Davidson, M., & Keating, J. (2011). The Assessment of Physiotherapy Practice (APP) is a valid measure of professional competence of physiotherapy students: a cross-sectional study with Rasch analysis. *J Physiother*, 57(4), 239-246. doi:10.1016/S1836-9553(11)70054-6

- Desbiens, J. F. (2005). Le behaviorisme et l'approche scientifique de l'enseignement. In C. Gauthier & M. Tardif (Eds.), *La pédagogie-Théories et pratiques de l'antiquité à nos jours* (2 ed., pp. 290-308). Montréal, Qc: Gaëtan Morin
- Deschênes, M. F., Charlin, B., Gagnon, R., & Goudreau, J. (2011). Use of a script concordance test to assess development of clinical reasoning in nursing students. *J Nurs Educ, 50*(7), 381-387.
- Desmeules, F. (2011). Les pratiques avancées en physiothérapie : une revue systématique de la littérature. Montréal.
- Dory, V., Charlin, B., Vanpee, D., & Gagnon, R. (2014). Multifaceted assessment in a family medicine clerkship: a pilot study. *Fam Med, 46*(10), 755-760.
- Dory, V., Gagnon, R., Vanpee, D., & Charlin, B. (2012). How to construct and implement script concordance tests: insights from a systematic review. *Med Educ, 46*(6), 552-563. doi:10.1111/j.1365-2923.2011.04211.x
- Downing, S. M. (2005). Threats to the validity of clinical teaching assessments: what about rater error? *Med Educ, 39*(4), 353-355. doi:10.1111/j.1365-2929.2005.02138.x
- Downing, S. M., & Haladyna, T. M. (2006). *Handbook of test development*. Lawrence Erlbaum Associates Publishers.
- Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: the power of human intuition and expertise in the era of the computer*. Free Press. New York, NY: The Free Press.
- Ducos, G., Lejus, C., Sztark, F., Nathan, N., Fourcade, O., Tack, I., . . . Minville, V. (2015). The Script Concordance Test in anesthesiology: Validation of a new tool for assessing clinical reasoning. *Anaesthesia critical care & pain medicine, 34*(1), 11-15.
- Dudek, N. L., Marks, M. B., & Regehr, G. (2005). Failure to fail: the perspectives of clinical supervisors. *Acad Med, 80*(10 Suppl), S84-87.
- Duggan, P., & Charlin, B. (2012). Summative assessment of 5th year medical students' clinical reasoning by script concordance test: requirements and challenges. *BMC Med Educ, 12*(1), 29. doi:10.1186/1472-6920-12-29
- Dumas, J. P., Blais, J. G., & Charlin, B. (2014). *Test de concordance de script (TCS) : est ce que les concepteurs de questions sont en accord avec les membres du panel?*. Paper presented at the 2e congrès international sur le raisonnement clinique, Montréal, Canada
- Durning, S. J., Artino, A. R., Jr., Pangaro, L. N., van der Vleuten, C., & Schuwirth, L. (2010). Perspective: redefining context in the clinical encounter: implications for research and training in medical education. *Acad Med, 85*(5), 894-901. doi:10.1097/ACM.0b013e3181d7427c
- Durning, S. J., Artino, A. R., Jr., Schuwirth, L., & van der Vleuten, C. (2013). Clarifying Assumptions to Enhance Our Understanding and Assessment of Clinical Reasoning. *Acad Med*. doi:10.1097/ACM.0b013e3182851b5b

- Durning, S. J., Dong, T., Artino, A. R., van der Vleuten, C., Holmboe, E., & Schuwirth, L. (2015). Dual processing theory and experts' reasoning: exploring thinking on national multiple-choice questions. *Perspect Med Educ*, 4(4), 168-175.
- Edwards, I., Jones, M., Carr, J., Braunack-Mayer, A., & Jensen, G. M. (2004). Clinical reasoning strategies in physical therapy. *Phys Ther*, 84(4), 312-330.
- Eisenberg, N. R. (2012). Post-structural conceptualizations of power relationships in physiotherapy. *Physiother Theory Pract*, 28(6), 439-446. doi:10.3109/09593985.2012.692585
- Elstein, A. S., Schulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving. An analysis of clinical reasoning*. Cambridge: Harvard University Press.
- Elstein, A. S., & Schwartz, A. (2002). Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ*, 324(7339), 729-732.
- English, I. (1993). Intuition as a function of the expert nurse: a critique of Benner's novice to expert model. *J Adv Nurs*, 18(3), 387-393.
- Ericsson, K. A. (2007). An expert-performance perspective of research on medical expertise: the study of clinical performance. *Med Educ*, 41(12), 1124-1130. doi:10.1111/j.1365-2923.2007.02946.x
- Eva, K. W. (2005). What every teacher needs to know about clinical reasoning. *Med Educ*, 39(1), 98-106.
- Farmer, E. A., & Page, G. (2005). A practical guide to assessing clinical decision-making skills using the key features approach. *Med Educ*, 39(12), 1188-1194.
- Fernandez, N., Dory, V., Ste-Marie, L. G., Chaput, M., Charlin, B., & Boucher, A. (2012). Varying conceptions of competence: an analysis of how health sciences educators define competence. *Med Educ*, 46(4), 357-365. doi:10.1111/j.1365-2923.2011.04183.x
- Fleming, M. H. (1991). Clinical reasoning in medicine compared with clinical reasoning in occupational therapy. *Am J Occup Ther*, 45(11), 988-996.
- Fournier, J., Demeester, A., & Charlin, B. (2008). Script concordance tests: guidelines for construction. *BMC Med Inform Decis Mak*, 8, 18.
- Frank, J. R., & Danoff, D. (2007). The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. *Med Teach*, 29(7), 642-647. doi:10.1080/01421590701746983
- Gagnon, R., Charlin, B., Coletti, M., Sauve, E., & van der Vleuten, C. (2005). Assessment in the context of uncertainty: how many members are needed on the panel of reference of a script concordance test? *Med Educ*, 39(3), 284-291. doi:10.1111/j.1365-2929.2005.02092.x
- Gagnon, R., Charlin, B., Lambert, C., Carriere, B., & Van der Vleuten, C. (2009). Script concordance testing: more cases or more questions? *Advances in health sciences education*, 14(3), 367-375.
- Gagnon, R., Lubarsky, S., Lambert, C., & Charlin, B. (2011). Optimization of answer keys for script concordance testing: should we exclude deviant panelists, deviant responses, or neither? *Adv Health Sci Educ Theory Pract*, 16(5), 601-608. doi:10.1007/s10459-011-9279-2

- George, D. (2003). *SPSS for windows step by step: A simple study guide and reference*, 17.0 update, 10/e: Pearson Education
- Gerard, F.-M., & Roegiers, X. (2010). Curriculum et évaluation : des liens qui ne seront jamais assez forts. In M. P. Alves & M. A. (Eds.), *Évaluation et curriculum*. Bruxelles: De Boeck.
- Gerard, F. M. (2002). L'indispensable subjectivité de l'évaluation. *Antipodes*, 156, 26-34.
- Giet, D., Massart, V., Gagnon, R., & Charlin, B. (2013). Le test de concordance de script en 20 questions. *Pédagogie Médicale*, 14(1), 39-48.
- Goulet, F., Jacques, A., Gagnon, R., Charlin, B., & Shabah, A. (2010). Poorly performing physicians: does the Script Concordance Test detect bad clinical reasoning? *J Contin Educ Health Prof*, 30(3), 161-166.
- Gouvernement du Canada. Caractéristiques de la profession de physiothérapeute. Retrieved from http://www.servicecanada.gc.ca/fra/qc/emploi_avenir/statistiques/3142_shtml
- Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability what they are and how to use them. *Educational and psychological measurement*, 66(6), 930-944.
- Green, S. B., & Yang, Y. (2009). Commentary on coefficient alpha: A cautionary tale. *Psychometrika*, 74(1), 121-135.
- Groupe national consultatif en physiothérapie. (2009). Profil des compétences essentielles des physiothérapeutes au Canada. Toronto.
- Groves, M., Dick, M. L., McColl, G., & Bilszta, J. (2013). Analysing clinical reasoning characteristics using a combined methods approach. *BMC Med Educ*, 13, 144. doi:10.1186/1472-6920-13-144
- Groves, M., O'Rourke, P., & Alexander, H. (2003a). The clinical reasoning characteristics of diagnostic experts. *Med Teach*, 25(3), 308-313.
- Groves, M., O'Rourke, P., & Alexander, H. (2003b). Clinical reasoning: the relative contribution of identification, interpretation and hypothesis errors to misdiagnosis. *Med Teach*, 25(6), 621-625.
- Groves, M., Scott, I., & Alexander, H. (2002). Assessing clinical reasoning: a method to monitor its development in a PBL curriculum. *Med Teach*, 24(5), 507-515.
- Gruppen, L. D., & Frohna, A. Z. (2002). Clinical reasoning. In G. R. Norman, C. P. van der Vleuten, & D. I. Newdle (Eds.), *International Handbook of Research in Medical Education* (pp. 205-230). Great Britain: Kluwer Academic Publisher.
- Gruppen, L. D., Mangrulkar, R. S., & Kolars, J. C. (2012). The promise of competency-based education in the health professions for improving global health. *Hum Resour Health*, 10(1), 43. doi:10.1186/1478-4491-10-43
- Guerrasio, J., & Aagaard, E. M. (2014). Methods and outcomes for the remediation of clinical reasoning. *J Gen Intern Med*, 29(12), 1607-1614. doi:10.1007/s11606-014-2955-1

- Guerrasio, J., Garrity, M. J., & Aagaard, E. M. (2014). Learner deficits and academic outcomes of medical students, residents, fellows, and attending physicians referred to a remediation program, 2006–2012. *Academic Medicine, 89*(2), 352-358.
- Hamdy, H., Prasad, K., Anderson, M. B., Scherpbier, A., Williams, R., Zwierstra, R., & Cuddihy, H. (2006). BEME systematic review: predictive values of measurements obtained in medical schools and future performance in medical practice. *Med Teach, 28*(2), 103-116.
- Hatala, R., & Norman, G. R. (1999). In-training evaluation during an internal medicine clerkship. *Academic Medicine, 74*(10), S118-120.
- Heemskerk, L., Norman, G., Chou, S., Mintz, M., Mandin, H., & McLaughlin, K. (2008). The effect of question format and task difficulty on reasoning strategies and diagnostic performance in Internal Medicine residents. *Adv Health Sci Educ Theory Pract, 13*(4), 453-462.
- Hendrick, P., Bond, C., Duncan, E., & Hale, L. (2009). Clinical reasoning in musculoskeletal practice: students' conceptualizations. *Phys Ther, 89*(5), 430-442.
- Higgs, J., & Jones, M. (2000). *Clinical Reasoning in the Health Professions*. Oxford: Butterworth-Heinemann.
- Higgs, J., Jones, M. A., Loftus, S., & Christensen, N. (2008). *Clinical Reasoning in the Health Professions* (3 ed.). Sydney: Elsevier.
- Higgs, J., Jones, M. A., & Titchen, A. (2008). Knowledge, reasoning and evidence for practice. In J. Higgs, M. A. Jones, S. Loftus, & N. Christensen (Eds.), *Clinical reasoning in the health professions* (3rd ed.). Sydney: Elsevier.
- Hirtt, N. (2009). L'approche par compétences : une mystification pédagogique. *L'école démocratique, 39*, 1-34.
- Holdar, U., Wallin, L., & Heiwe, S. (2013). Why Do We Do As We Do? Factors Influencing Clinical Reasoning and Decision-Making among Physiotherapists in an Acute Setting. *Physiother Res Int*. doi:10.1002/pri.1551
- Holmboe, E. S., Hawkins, R. E., & Huot, S. J. (2004). Effects of training in direct observation of medical residents' clinical competence: a randomized trial. *Ann Intern Med, 140*(11), 874-881.
- Hrynchak, P., Takahashi, S. G., & Nayer, M. (2014). Key-feature questions for assessment of clinical reasoning: a literature review. *Med Educ, 48*(9), 870-883. doi:10.1111/medu.12509
- Huba, M. E., & Freed, J. E. (2000). *Learner-Centered Assessment on College Campuses. Shifting the Focus from Teaching to Learning*. Boston, MA: Allyn and Bacon.
- Humbert, A. J., Johnson, M. T., Miech, E., Friedberg, F., Grackin, J. A., & Seidman, P. A. (2011). Assessment of clinical reasoning: A Script Concordance test designed for pre-clinical medical students. *Med Teach, 33*(6), 472-477. doi:10.3109/0142159X.2010.531157
- Jensen, G. M., Gwyer, J., & Shepard, K. F. (2000). Expert practice in physical therapy. *Phys Ther, 80*(1), 28-43. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10623958>

- Jette, D. U., Ardleigh, K., Chandler, K., & McShea, L. (2006). Decision-making ability of physical therapists: physical therapy intervention or medical referral. *Phys Ther*, *86*(12), 1619-1629. doi:10.2522/ptj.20050393
- Jette, D. U., Bertoni, A., Coots, R., Johnson, H., McLaughlin, C., & Weisbach, C. (2007). Clinical instructors' perceptions of behaviors that comprise entry-level clinical performance in physical therapist students: a qualitative study. *Phys Ther*, *87*(7), 833-843. doi:10.2522/ptj.20070003
- Jones, M. A. (2008). *Clinical reasoning : understanding the problem through diagnostic and narrative reasoning*. Montreal.
- Jonnaert, P. (2009). *Compétences et constructivisme : Un cadre théorique*. Bruxelles: De Boeck Université.
- Josephson, I., Bulow, P., & Hedberg, B. (2011). Physiotherapists' clinical reasoning about patients with non-specific low back pain, as described by the International Classification of Functioning, Disability and Health. *Disabil Rehabil*, *33*(23-24), 2217-2228. doi:10.3109/09638288.2011.563819
- Joughin, G. (2009). *Assessment, learning and judgement in higher education*: Springer.
- Jouquan, J. (2002). L'évaluation des apprentissages des étudiants en formation médicale initiale. *Pédagogie Médicale*, *3*, 38-52.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1-73.
- Kania, R. E., Verillaud, B., Tran, H., Gagnon, R., Kazitani, D., Huy, P. T. B., . . . Charlin, B. (2011). Online script concordance test for clinical reasoning assessment in otorhinolaryngology: the association between performance and clinical experience. *Archives of Otolaryngology–Head & Neck Surgery*, *137*(8), 751-755.
- Khan, K., & Ramachandran, S. (2012). Conceptual framework for performance assessment: Competency, competence and performance in the context of assessments in healthcare - Deciphering the terminology. *Med Teach*. doi:10.3109/0142159X.2012.722707
- Khatti, N., & Sweet, D. (1996). Assessment Reform: Promises and Challenges. In K. M.B. & M. R. (Eds.), *Performance Assessment. Promises, Problems and Challenges*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Kogan, J. R., Hess, B. J., Conforti, L. N., & Holmboe, E. S. (2010). What drives faculty ratings of residents' clinical skills? The impact of faculty's own clinical skills. *Academic Medicine*, *85*(10), S25-S28.
- Kogan, J. R., Holmboe, E. S., & Hauer, K. E. (2009). Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *JAMA*, *302*(12), 1316-1326.
- Lambert, C., Gagnon, R., Nguyen, D., & Charlin, B. (2009). The script concordance test in radiation oncology: validation study of a new tool to assess clinical reasoning. *Radiat Oncol*, *4*, 7. doi:10.1186/1748-717X-4-7

- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174.
- Larsen, D. P., Butler, A. C., & Roediger, H. L., 3rd. (2008). Test-enhanced learning in medical education. *Med Educ*, 42(10), 959-966. doi:10.1111/j.1365-2923.2008.03124.x
- Le Boterf, G. (2007). *Professionnaliser : Le modèle de la navigation professionnelle*. Paris: Eyrolles.
- Lee, A., Ho, A. M., Groves, M., & Vlantis, A. C. (2010). Using illness scripts to teach clinical reasoning skills to medical students. *Fam Med*, 42(4), 255-261.
- Legendre, M. F. (2004). Approches constructivistes et nouvelles orientations curriculaires: d'un curriculum fondé sur l'approche par objectifs à un curriculum axé sur le développement de compétences. In P. Jonnaert & D. Masciotra (Eds.), *Constructivisme : choix contemporains : Hommage à Ernst Von Glaserfeld*. (Vol. 120, pp. 53-91). Québec: Presse de l'Université du Québec.
- Legendre, M. F. (2008). Un regard socioconstructiviste sur la participation des savoirs à la construction du lien social. *Éducation et francophonie*, 26(2), 63-79.
- Legendre, R. (2005). *Dictionnaire actuel de l'éducation* (3 ed.). Montréal: Guérin.
- Lekkas, P., Larsen, T., Kumar, S., Grimmer, K., Nyland, L., Chipchase, L., . . . Finch, J. (2007). No model of clinical education for physiotherapy students is superior to another: a systematic review. *Aust J Physiother*, 53(1), 19-28.
- Lievens, F., & Patterson, F. (2011). The validity and incremental validity of knowledge tests, low-fidelity simulations, and high-fidelity simulations for predicting job performance in advanced-level high-stakes selection. *J Appl Psychol*, 96(5), 927-940. doi:10.1037/a0023496
- Lineberry, M., Kreiter, C. D., & Bordage, G. (2013). Threats to validity in the use and interpretation of script concordance test scores. *Med Educ*, 47(12), 1175-1183. doi:10.1111/medu.12283
- Lissitz, R. W., & Samuelsen, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational researcher*, 36(8), 437-448.
- Llorca, G., Roy, P., & Riche, B. (2003). Evaluation de résolution de problèmes mal définis en éthique clinique: variation des scores selon les méthodes de correction et les caractéristiques des jurys. *Pédagogie Médicale*, 4(2), 80-88.
- Loftus, S., & Smith, M. (2008). A history of clinical reasoning research. In J. Higgs, M. A. Jones, S. Loftus, & N. Christensen (Eds.), *Clinical reasoning in the health profession* (3 ed.). Sydney: Elsevier.
- Lubarsky, S., Chalk, C., Kazitani, D., Gagnon, R., & Charlin, B. (2009). The Script Concordance Test: a new tool assessing clinical judgement in neurology. *Can J Neurol Sci*, 36(3), 326-331.
- Lubarsky, S., Charlin, B., Cook, D. A., Chalk, C., & van der Vleuten, C. P. (2011). Script concordance testing: a review of published validity evidence. *Med Educ*, 45(4), 329-338.
- Lubarsky, S., Dory, V., Duggan, P., Gagnon, R., & Charlin, B. (2013). Script concordance testing: from theory to practice: AMEE guide no. 75. *Med Teach*, 35(3), 184-193. doi:10.3109/0142159X.2013.760036

- Lugtenberg, M., Zegers-van Schaick, J. M., Westert, G. P., & Burgers, J. S. (2009). Why don't physicians adhere to guideline recommendations in practice? An analysis of barriers among Dutch general practitioners. *Implement Sci*, 4, 54. doi:10.1186/1748-5908-4-54
- Mamede, S., Schmidt, H. G., Rikers, R. M., Penaforte, J. C., & Coelho-Filho, J. M. (2007). Breaking down automaticity: case ambiguity and the shift to reflective approaches in clinical reasoning. *Med Educ*, 41(12), 1185-1192. doi:10.1111/j.1365-2923.2007.02921.x
- Marcum, J. A. (2012). An integrated model of clinical reasoning: dual-process theory of cognition and metacognition. *J Eval Clin Pract*, 18(5), 954-961. doi:10.1111/j.1365-2753.2012.01900.x
- Markus, K. A., & Borsboom, D. (2013). *Frontiers of test validity theory: Measurement, causation, and meaning*: Routledge.
- Martinez, M. E. (1991). A Comparison of Multiple Choice and Constructed Figural Response Items. *Journal of Educational Measurement*, 28(2), 131-145.
- May, S., Greasley, A., Reeve, S., & Withers, S. (2008). Expert therapists use specific clinical reasoning processes in the assessment and management of patients with shoulder pain: a qualitative study. *Aust J Physiother*, 54(4), 261-266.
- McCoubrie, P. (2004). Improving the fairness of multiple-choice questions: a literature review. *Med Teach*, 26(8), 709-712.
- McGill, D. A., van der Vleuten, C. P., & Clarke, M. J. (2011). Supervisor assessment of clinical and professional competence of medical trainees: a reliability study using workplace data and a focused analytical literature review. *Adv Health Sci Educ Theory Pract*, 16(3), 405-425. doi:10.1007/s10459-011-9296-1
- Meterissian, S., Zabolotny, B., Gagnon, R., & Charlin, B. (2007). Is the script concordance test a valid instrument for assessment of intraoperative decision-making skills? *The American journal of surgery*, 193(2), 248-251.
- Morcke, A. M., Dornan, T., & Eika, B. (2012). Outcome (competency) based education: an exploration of its origins, theoretical basis, and empirical evidence. *Adv Health Sci Educ Theory Pract*. doi:10.1007/s10459-012-9405-9
- Morissette, D. (1984). *La mesure de l'évaluation en enseignement*. Ste-Foy (Québec): Les Presses de l'Université Laval.
- Moulton, C. A., Regehr, G., Mylopoulos, M., & MacRae, H. M. (2007). Slowing down when you should: a new model of expert judgment. *Acad Med*, 82(10 Suppl), S109-116.
- Mucchielli, R. (2006). *L'analyse de contenu: des documents et des communications; avec un plan d'autoformation et des exercices corrigés*: ESF éd.
- Nendaz, M., Charlin, B., Leblanc, V., & Bordage, G. (2005). Le raisonnement clinique: données issues de la recherche et implications pour l'enseignement. *Pédagogie Médicale*, 6(4), 235-254.
- Newton, P., & Shaw, S. (2014). *Validity in educational and psychological assessment*: Sage.

Nguyen, D.-Q., & Blais, J.-G. (2007). Approche par objectifs ou approche par compétences? Repères conceptuels et implications pour les activités d'enseignement d'apprentissage et d'évaluation au cours de la formation clinique. *Pédagogie Médicale*, 8(4), 232-251.

Norman, G. R. (2005). Research in clinical reasoning: past history and current trends. *Med Educ*, 39(4), 418-427.

Norman, G. R., Brooks, L. R., Regehr, G., Marriott, M., & Shali, V. (1996). Impact of feature interpretation on medical student diagnostic performance. *Acad Med*, 71(1 Suppl), S108-109.

Norman, G. R., & Eva, K. W. (2010). Diagnostic error and clinical reasoning. *Med Educ*, 44(1), 94-100.

Nouh, T., Boutros, M., Gagnon, R., Reid, S., Leslie, K., Pace, D., . . . Meterissian, S. H. (2012). The script concordance test as a measure of clinical reasoning: a national validation study. *Am J Surg*, 203(4), 530-534. doi:10.1016/j.amjsurg.2011.11.006

Ordre professionnelle de la physiothérapie. (2012). Définition de la physiothérapie. Retrieved from <https://oppq.qc.ca/grand-public/definitions-et-presentation/>

Page, G., & Bordage, G. (1995). The Medical Council of Canada's key features project: a more valid written examination of clinical decision-making skills. *Acad Med*, 70(2), 104-110.

Palmer, E. J., & Devitt, P. G. (2007). Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? Research paper. *BMC Med Educ*, 7, 49.

Palmer, E. J., Duggan, P., Devitt, P. G., & Russell, R. (2010). The modified essay question: its exit from the exit examination? *Med Teach*, 32(7), e300-307.

Panayides, P. (2013). Coefficient Alpha: Interpret With Caution. *Europe's Journal of Psychology*, 9(4), 687-696.

Patel, V. L., Arocha, J. F., & Kaufman, D. R. (1999). Expertise and tacit knowledge in medicine. In R. Sternberg & J. A. Horvath (Eds.), *Tacit knowledge in professional practice: Researcher and practitioner perspectives* (pp. 75-99). Mahwah, NJ: Lawrence Erlbaum Associates.

Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science*, 10, 91-116.

Patel, V. L., Groen, G. J., & Arocha, J. F. (1990). Medical expertise as a function of task difficulty. *Memory & cognition*, 18(4), 394-406.

Patel, V. L., & Kaufman, D. L. (2000). Clinical reasoning and biomedical knowledge : Implication for teaching. In H. M. & M. Jones (Eds.), *Clinical reasoning in the health professions* (2e ed ed.). Oxford: Butterworth-Heinemann.

Pelaccia, T., Tardif, J., Triby, E., & Charlin, B. (2011). An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory. *Med Educ Online*, 16.

Pelgrim, E. A., Kramer, A. W., Mokkink, H. G., van den Elsen, L., Grol, R. P., & van der Vleuten, C. P. (2011). In-training assessment using direct observation of single-patient encounters: a literature review. *Adv Health Sci Educ Theory Pract*, 16(1), 131-142.

- Pell, G., Fuller, R., Homer, M., & Roberts, T. (2010). How to measure the quality of the OSCE: A review of metrics - AMEE guide no. 49. *Med Teach*, 32(10), 802-811.
- Pena, A. (2010). The Dreyfus model of clinical problem-solving skills acquisition: a critical perspective. *Med Educ Online*, 15. doi:10.3402/meo.v15i0.4846
- Perrenoud, P. (2002). D'une métaphore à l'autre: transférer ou mobiliser ses connaissances? *Raisons éducatives*, 45-60.
- Perrenoud, P. (2004). Évaluer des compétences *L'éducateur*(Mars), 8-11.
- Perrenoud, P. (2005). Entrevue avec Philippe Perrenoud. Bulletin du Centre de recherche interuniversitaire sur la formation et la profession enseignante, 11(1), 5-15.
- Prégent, R., Bernard, H., & Kozanitis, A. (2009). *Enseigner à l'université dans une approche-programme*. Montreal,: Presses internationales polytechniques.
- Ramaekers, S., Kremer, W., Pilot, A., Beukelen, P. v., & Keulen, H. v. (2010). Assessment of competence in clinical reasoning and decision-making under uncertainty: the script concordance test method. *Assessment & Evaluation in Higher Education*, 35(6), 661-673.
- Resnick, L. B., & Resnick, D. L. (1992). Assessing the Thinking Curriculum: New Tools for Educational Reform. In B. R. G. a. M. C. O'Connor (Ed.), *Changing Assessments: Alternative Views of Aptitude, Achievement and Instruction* (pp. 37-77). Boston: Kluwer Academic Publishers.
- Rheault, W., & Shafernich-Coulson, E. (1988). Relationship between academic achievement and clinical performance in a physical therapy education program. *Phys Ther*, 68(3), 378-380.
- Rikers, R. M., Schmidt, H. G., & Moulart, V. (2005). Biomedical knowledge: encapsulated or two worlds apart? *Applied Cognitive Psychology*, 19(2), 223-231.
- Roach, K. E., Frost, J. S., Francis, N. J., Giles, S., Nordrum, J. T., & Delitto, A. (2012). Validation of the Revised Physical Therapist Clinical Performance Instrument (PT CPI): Version 2006. *Phys Ther*, 92(3), 416-428. doi:10.2522/ptj.20110129
- Sackett, D. L., Strauss, S. E., & Richardson, W. S. (2000). *Evidence based medicine: how to practice and teach EBM* (2nd ed.). Edinburgh: Churchill Livingstone.
- Scallon, G. (2000). *L'évaluation formative* (É. d. R. Pédagogiques Ed.). Bruxelles: Éditions du Renouveau Pédagogique.
- Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétences* St-Laurent, (Québec): Éditions du Renouveau Pédagogique.
- Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. (1990). A cognitive perspective on medical expertise: theory and implication. *Acad Med*, 65(10), 611-621.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological assessment*, 8(4), 350.
- Schön, D. A. (1983). *The reflective practitioner: how professionals think in action*. London: Temple Smith

- Schuwirth, L. W., & van der Vleuten, C. P. (2004). Different written assessment methods: what can be said about their strengths and weaknesses? *Med Educ*, 38(9), 974-979.
- Schuwirth, L. W., & van der Vleuten, C. P. (2011). General overview of the theories used in assessment: AMEE Guide No. 57. *Med Teach*, 33(10), 783-797. doi:10.3109/0142159X.2011.611022
- Schuwirth, L. W., Verheggen, M. M., van der Vleuten, C. P., Boshuizen, H. P., & Dinant, G. J. (2001). Do short cases elicit different thinking processes than factual knowledge questions do? *Med Educ*, 35(4), 348-356.
- Schwind, C. J., Williams, R. G., Boehler, M. L., & Dunnington, G. L. (2004). Do individual attendings' post-rotation performance ratings detect residents' clinical performance deficiencies? *Academic Medicine*, 79(5), 453-457.
- See, K. C., Tan, K. L., & Lim, T. K. (2014). The script concordance test for clinical reasoning: re-examining its utility and potential weakness. *Med Educ*, 48(11), 1069-1077. doi:10.1111/medu.12514
- Shavelson, R. J., Baxter, G. P., & Gao, X. (1993). Sampling Variability of Performance Assessments. *Journal of Educational Measurement*, 30(3), 215-232.
- Sibert, L., Charlin, B., Corcos, J., Gagnon, R., Grise, P., & van der Vleuten, C. (2002). Stability of clinical reasoning assessment results with the Script Concordance test across two different linguistic, cultural and learning environments. *Med Teach*, 24(5), 522-527. doi:10.1080/0142159021000012599
- Sibert, L., Charlin, B., Corcos, J., Gagnon, R., Lechevallier, J., & Grise, P. (2002). Assessment of clinical reasoning competence in urology with the script concordance test: an exploratory study across two sites from different countries. *Eur Urol*, 41(3), 227-233. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12180220>
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107-120.
- Simmons, B. (2010). Clinical reasoning: concept analysis. *J Adv Nurs*, 66(5), 1151-1158. doi:10.1111/j.1365-2648.2010.05262.x
- Sireci, S. G. (2007). On validity theory and test validation. *Educational researcher*, 36(8), 477-481.
- Stratford, P., & Pierce-Fenn, H. (1985). Modified essay question. *Phys Ther*, 65(7), 1075-1079.
- Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of personality assessment*, 80(1), 99-103.
- Swanson, D. B., Holtzman, K. Z., & Allbee, K. (2008). Measurement characteristics of content-parallel single-best-answer and extended-matching questions in relation to number and source of options. *Acad Med*, 83(10 Suppl), S21-24.
- Swanson, D. B., Holtzman, K. Z., Clauser, B. E., & Sawhill, A. J. (2005). Psychometric characteristics and response times for one-best-answer questions in relation to number and source of options. *Acad Med*, 80(10 Suppl), S93-96.

- Tan, C. (1992). An evaluation of the use of continuous assessment in the teaching of physiology. *Higher education, 23*(3), 255-272.
- Tan, K., Tan, N., Kandiah, N., Samarasekera, D., & Ponnampereuma, G. (2014). Validating a script concordance test for assessing neurological localization and emergencies. *European Journal of Neurology, 21*(11), 1419-1422.
- Tanggaard, L., & Elmholdt, C. (2008). Assessment in practice: An inspiration from apprenticeship. *Scandinavian Journal of Educational Research, 52*(1), 97-116.
- Tardif, J. (1993). L'évaluation dans le paradigme constructiviste. In R. Hivon (Ed.), *L'évaluation des apprentissages. Réflexions, nouvelles tendances et formation*. (pp. 27-53). Sherbrooke: Editions CRP.
- Tardif, J. (2006). *L'évaluation des compétences. Documenter le parcours de développement*. Montréal (Québec): Les Éditions de la Chenelière
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education, 2*, 53.
- ten Cate, O., Snell, L., & Carraccio, C. (2010). Medical competence: the interplay between individual ability and the health care environment. *Med Teach, 32*(8), 669-675. doi:10.3109/0142159X.2010.500897
- Thomas, A. (2014). *Le raisonnement clinique des superviseurs cliniques en Ergothérapie et physiothérapie*. Paper presented at the 2e conférence internationale sur le raisonnement clinique.
- Tsai, T. C., Chen, D. F., & Lei, S. M. (2012). The ethics script concordance test in assessing ethical reasoning. *Med Educ, 46*(5), 527.
- Tversky, A., & Kahnemann, D. (1982). Judgments of and by representativeness. In D. Kahnemann, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge, England: Cambridge University Press.
- van den Broek, W. E., van Asperen, M. V., Custers, E., Valk, G. D., & Ten Cate, O. T. (2012). Effects of two different instructional formats on scores and reliability of a script concordance test. *Perspect Med Educ, 1*(3), 119-128. doi:10.1007/s40037-012-0017-0
- van der Vleuten, C. P., & Schuwirth, L. W. (2005). Assessing professional competence: from methods to programmes. *Med Educ, 39*(3), 309-317. doi:10.1111/j.1365-2929.2005.02094.x
- Van der Vleuten, C. P., Schuwirth, L. W., Scheele, F., Driessen, E. W., & Hodges, B. (2010). The assessment of professional competence: building blocks for theory development. *Best Pract Res Clin Obstet Gynaecol, 24*(6), 703-719.
- van der Vleuten, C. P. M., Norman, G. R., & Schuwirth, L. W. T. (2008). Assessing clinical reasoning. In J. Higgs, M. A. Jones, S. Loftus, & N. Christensen (Eds.), *Clinical reasoning in the health professions* (3 ed.). Sydney: Elsevier.
- Verheyden, G., Handgraaf, M., Demirci, A., & Grüneberg, C. (2011). The Future of Physiotherapy Education: Towards a Translational Model of Learning Complex Skills. *Physiother. Res. Int., 16*, 187-190.

- Victor-Chmil, J. (2013). Critical thinking versus clinical reasoning versus clinical judgment: differential diagnosis. *Nurse Educ*, 38(1), 34-36. doi:10.1097/NNE.0b013e318276dfbe
- Vienneau, R. (2011). *Apprentissage et enseignement : théories et pratique* (2 ed.). Montréal: Chenelière éducation.
- Vissers, D., van Daele, U., & Denekens, J. (2012). Competency-based physiotherapy education: The need for a European view? *Med Teach*. doi:10.3109/0142159X.2012.737069
- Wan, S. H. (2015). Using the script concordance test to assess clinical reasoning skills in undergraduate and postgraduate medicine. *Hong Kong Med J*, 21(5), 455-461. doi:10.12809/hkmj154572
- Ward, W. C. (1982). A comparison of free-response and multiple-choice forms of verbal aptitude tests. *Applied Psychological Measurement*, 6(1), 1-11.
- Weekley, J. A., & Ployhart, R. E. (2006). *Situational judgment tests: Theory, measurement, and application*. Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers.
- Wiggins, G. P. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70, 703-713.
- Williams, R. G., Klamen, D. A., & McGaghie, W. C. (2003). Cognitive, social and environmental sources of bias in clinical performance ratings. *Teach Learn Med*, 15(4), 270-292.
- Woods, N. N. (2007). Science is fundamental: the role of biomedical knowledge in clinical reasoning. *Med Educ*, 41(12), 1173-1177. doi:10.1111/j.1365-2923.2007.02911.x
- Woods, N. N., Brooks, L. R., & Norman, G. R. (2007). The role of biomedical knowledge in diagnosis of difficult clinical cases. *Adv Health Sci Educ Theory Pract*, 12(4), 417-426. doi:10.1007/s10459-006-9054-y
- Woods, N. N., Neville, A. J., Levinson, A. J., Howey, E. H., Oczkowski, W. J., & Norman, G. R. (2006). The value of basic science in clinical diagnosis. *Acad Med*, 81(10 Suppl), S124-127.

ANNEXES

ANNEXE 1 CURRICULUM DU PROGRAMME DE PHYSIOTHÉRAPIE DE L'UNIVERSITÉ DE SHERBROOKE



Curriculum du programme de physiothérapie

	Septembre					Octobre					Novembre				Décembre			Janvier					Février				Mars					Avril				Mai				Juin		
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	1	2	3	2	3	4	5	1	2	3	4	1	2	3	4	5	1	2	3	4	1	2	3	4	1	2		
PHASE 1 : L'ACQUISITION DE CONNAISSANCES EN SCIENCES BIOMÉDICALES ET PSYCHOSOCIALES																																										
An 1 (38)	IPR		Fond. I			Sc	Rel	biol.			CDV			Stage I PHT100 1			Ex.	ECOS-RPP			Syst. Nerveux					Sc. psychiques	Rel	App. locomoteur					Ex.	Fond. II				SPR				Ex.
Session A= 13	RDP 100 1 cr		RDP 110 3 cr			RDP 120 4 cr			MDS 100 4 cr			Stage I PHT100 1			Ex.	ECOS-RPP			MDS 102 5 cr					MDS 103 4 cr				MDS 101 4 cr					Ex.	RDP 111 2 cr				RDP 140 4 cr				Ex.
Session H= 13	Habilités cliniques en réadaptation I RDP 130 - 3 cr																																									
Session É= 12	Approche en physiothérapie I PHT 110 - 2 cr																																									
	Réflexion sur sa pratique professionnelle I PHT 115 - 1 cr																																									
PHASE 1 (suite)																			PHASE 2 : L'ÉVALUATION																							
An 2 (36)	Sc psycho		Douleur			Cardio I			Ex.	Rel	Cardio II			Dig-uri-endo			Ex.	Recherch I			Évaluation en physiothérapie					Ex.	Rel	Stage II - PHT202 - 5 cr					AICC RDP210 2 cr				Ex.					
Session A= 13	RDP 160 2 cr		RDP 180 2 cr			RDP 150 2 cr			Ex.	Rel	PHT 120 2 cr			PHT 160 3 cr			Ex.	RDP 190 2 cr			PHT 240 8 cr					Ex.	Rel	Stage II - PHT202 - 5 cr					AICC RDP210 2 cr				Ex.					
Session H= 13	Habilités cliniques en réadaptation II RDP 220 - 1 cr																																									
	Approche en physiothérapie II PHT 150 - 2 cr																																									
	Habilités cliniques en physiothérapie I PHT 140 - 1 cr																																									
	Réflexion sur sa pratique professionnelle II PHT 230 - 1 cr																																									
PHASE 3 : L'INTERVENTION PHYSIOTHÉRAPEUTIQUE																																										
An 3 (37)	Neurologie (7 sem)							Ex.	Rel	Musculo-squelettique (7 sem)							Ex.	Cardio-respiratoire					Rééd. Pér. I	Ex.	Rel	*Cours à option 1- 2 cr					PHT720 - 8 cr (2010-2011)				Ex.							
Session A= 13	PHT 320 6 cr							Ex.	Rel	PHT 300 6 cr							Ex.	PHT 700 4 cr (5 sem)					PHT 745 1 cr	Ex.	Rel	(PHT756 ou PHT765 ou PHT785)					Stage III				Ex.							
Session H= 10	Habilités cliniques en physio III PHT 310 2 cr																																									
	Habilités cliniques en physio IV PHT 330 2 cr																																									
	HC Physio V PHT 710 - 1 cr																																									
	HC Physio VI PHT 715 - 2 cr																																									
Session É= 12	Recherche en réadaptation II RDP 710 - 2 cr																																									
	Réflexion sur sa pratique professionnelle III PHT 755 - 1 cr																																									
PHASE 4 : L'INTÉGRATION CLINIQUE																																										
An 4 (33)	Stage IV (7 sem)							Int. II - 1 cr			Stage V (7 sem)							Int. III - 1 cr			Stage VI (7 sem) sur plage de 8 semaines								Intégration - IV - 2 cr				Ex.									
Session A= 13	PHT 760 - 7 cr							Int. II - 1 cr			PHT 780 - 7 cr							Int. III - 1 cr			PHT 800 - 7 cr								PHT830 (5 semaines 2010-2011)				Ex.									
Session H= 13	Recherche en réadaptation III RDP 730 - 4 cr																																									
	Rôles connexes en réadaptation RDP 740 - 2 cr																																									

ANNEXE 2 LETTRE D'APPROBATION PAR LE COMITÉ D'ÉTHIQUE ET FORMULAIRE DE CONSENTEMENT



Comité plurifacultaire d'éthique de la recherche

9 janvier 2014

Monsieur Jean-Pierre Dumas
Candidat au doctorat
Administration et fondements de l'éducation - Faculté des sciences de l'éducation

OBJET: Reconnaissance d'une approbation éthique

M. Jean-Pierre Dumas,

Le *Comité plurifacultaire d'éthique de la recherche (CPER)* a étudié le projet de recherche intitulé « L'évaluation du raisonnement clinique en physiothérapie » et a délivré le certificat d'éthique demandé suite à la satisfaction des exigences précédemment émises. Vous trouverez ci-joint une copie numérisée de votre certificat; copie également envoyée à votre directeur/directrice de recherche et à la technicienne en gestion de dossiers étudiants (TGDE) de votre département.

Notez qu'il y apparaît une mention relative à un suivi annuel et que le certificat comporte une date de fin de validité. En effet, afin de répondre aux exigences éthiques en vigueur au Canada et à l'Université de Montréal, nous devons exercer un suivi annuel auprès des chercheurs et étudiants-chercheurs.

De manière à rendre ce processus le plus simple possible et afin d'en tirer pour tous le plus grand profit, nous avons élaboré un court questionnaire qui vous permettra à la fois de satisfaire aux exigences du suivi et de nous faire part de vos commentaires et de vos besoins en matière d'éthique en cours de recherche. Ce questionnaire de suivi devra être rempli annuellement jusqu'à la fin du projet et pourra nous être retourné par courriel. La validité de l'approbation éthique est conditionnelle à ce suivi. Sur réception du dernier rapport de suivi en fin de projet, votre dossier sera clos.

Il est entendu que cela ne modifie en rien l'obligation pour le chercheur, tel qu'indiqué sur le certificat d'éthique, de signaler au CPER tout incident grave dès qu'il survient ou de lui faire part de tout changement anticipé au protocole de recherche.

Nous vous prions d'agréer, Monsieur, l'expression de nos sentiments les meilleurs,

Pierre Lapointe, Président
Comité plurifacultaire d'éthique de la recherche (CPER)
Université de Montréal
PL/RS/rs

c.c. Gestion des certificats, BRDV
Jean-Guy Blais, Professeur titulaire, Administration et fondements de l'éducation, Faculté des sciences de l'éducation
Dr. Bernard Charlin, Professeur titulaire et directeur du secteur recherche et développement, CPASS, département de chirurgie, Faculté de médecine
Lucie Lefrançois
p.j. Certificat CPER-13-121-D

3744 Jean-Brillant, B-430-8
C.P. 6128, succ. Centre-ville
Montréal QC H3C 3J7
www.cper.umontreal.ca

Téléphone : 514-343-6111 poste 1896
cper@umontreal.ca

Comité plurifacultaire d'éthique de la recherche

CERTIFICAT D'APPROBATION ÉTHIQUE

Le Comité plurifacultaire d'éthique de la recherche (CER), selon les procédures en vigueur, en vertu des documents qui lui ont été fournis, a examiné le projet de recherche suivant et conclu qu'il respecte les règles d'éthique énoncées dans la Politique sur la recherche avec des êtres humains de l'Université de Montréal.

Projet	
Titre du projet	L'évaluation du raisonnement clinique en physiothérapie
Étudiant requérant	Jean-Pierre Dumas Candidat au doctorat, Administration et fondements de l'éducation - Faculté des sciences de l'éducation Université de Montréal
Financement	
Organisme	MELS
Programme	Fonds de recherche résiduel du directeur suite à un contrat avec le ministère qui concernait un autre projet de recherche.
Titre de l'octroi si différent	--
Numéro d'octroi	--
Chercheur principal	--
No de compte	--

Approbation reconnue	
Approbation émise par	oui
Certificat:	s.o.

MODALITÉS D'APPLICATION

Tout changement anticipé au protocole de recherche doit être communiqué au CER qui en évaluera l'impact au chapitre de l'éthique.

Toute interruption prématurée du projet ou tout incident grave doit être immédiatement signalé au CER.

Selon les règles universitaires en vigueur, un suivi annuel est minimalement exigé pour maintenir la validité de la présente approbation éthique, et ce, jusqu'à la fin du projet. Le questionnaire de suivi est disponible sur la page web du CER.

Pierre Lapointe, Président
Comité plurifacultaire d'éthique de la recherche
Université de Montréal

9 janvier 2014
Date de délivrance

1 février 2015
Date de fin de validité

adresse postale
3744 Jean-Brillant, B-430-8
C.P. 6128, succ. Centre-ville
Montréal QC H3C 3J7
www.cer.umontreal.ca

Téléphone : 514-343-6111 poste 1896
cer@umontreal.ca

FORMULAIRE DE CONSENTEMENT (version étudiants)

Titre de la recherche : Évaluation du raisonnement clinique en physiothérapie
Volet 1 : tests de raisonnement clinique

Chercheur : Jean-Pierre Dumas, étudiant au doctorat, Département d'administration et fondements de l'éducation, Faculté des sciences de l'éducation, Université de Montréal

Directeurs de recherche : Jean-Guy Blais, professeur titulaire, Département d'administration et fondements de l'éducation, Faculté des sciences de l'éducation, Université de Montréal.

Bernard Charlin, professeur titulaire, Département de chirurgie, Faculté de médecine, Université de Montréal.

Co-chercheur : Annick Bourget, professeur adjointe, École de réadaptation, Université de Sherbrooke

Source de financement : Fonds de recherche résiduel du directeur suite à un contrat avec le ministère de l'éducation.

A) RENSEIGNEMENTS AUX PARTICIPANTS

1. Objectifs de la recherche.

Ce projet de recherche vise à développer des outils qui devraient nous permettre de mieux évaluer le raisonnement clinique des étudiants en physiothérapie.

2. Participation à la recherche

Nous vous demanderons de répondre à des questions portant sur des situations de pratique en physiothérapie musculosquelettique. Nous vous demanderons aussi de nous donner l'autorisation (en signant le présent formulaire de consentement) d'utiliser les données recueillies lors des tests et d'avoir accès à vos formulaires de stages FOCUS dans le but d'effectuer une forme de validation des tests de raisonnement clinique avec les situations authentiques. Votre participation ou le refus de participer au projet n'aura aucune influence sur vos résultats scolaires.

3. Confidentialité

Les renseignements que vous nous donnerez demeureront confidentiels. Chaque participant à la recherche se verra attribuer un numéro et seules le chercheur et/ou la personne mandatée à cet effet auront la liste des participants et des numéros qui leur auront été attribués. De plus, les renseignements seront conservés dans un classeur sous clé situé dans le bureau fermé du chercheur principal. Aucune information permettant de vous identifier d'une façon ou d'une autre ne sera publiée. Ces renseignements personnels seront détruits 7 ans après la fin du projet. Seules les données ne permettant pas de vous identifier seront conservées après cette date, le temps nécessaire à leur utilisation.

Page 1 de 3

4. Avantages et inconvénients

Les inconvénients sont liés au temps consacré à la participation (environ 2h30). Toutefois, votre participation est, pour vous, une occasion d'apprentissage tout en contribuant à l'avancement des connaissances sur l'évaluation du raisonnement clinique et potentiellement d'améliorer nos pratiques à ce niveau. Une rétroaction de groupe sera offerte sur les résultats aux tests et une rétroaction individuelle sera aussi possible.

5. Droit de retrait

Votre participation est entièrement volontaire. Votre participation n'aura pas d'impact sur l'évaluation d'activité pédagogique. Vous êtes libre de vous retirer en tout temps sur simple avis verbal, sans préjudice et sans devoir justifier votre décision. Si vous décidez de vous retirer de la recherche, vous pouvez communiquer avec le chercheur, au numéro de téléphone indiqué ci-dessous. Si vous vous retirez de la recherche, les renseignements qui auront été recueillis au moment de votre retrait seront détruits.

6. Compensation

Aucune compensation financière ne sera versée pour votre participation à la recherche. Par contre un tirage d'un iPad (16Go écran Réтина) sera effectué le 11 juin 2014 parmi les étudiants de la cohorte de 3^e année ayant participé à cette étude (probabilités de gagner : environ 1 sur 40). Le tirage aura lieu à l'École de réadaptation.

7. Diffusion des résultats

Un rapport sera transmis aux participants décrivant les conclusions générales de cette recherche au cours de l'année prochaine, lorsque les analyses auront été effectuées.

B) CONSENTEMENT

Je déclare avoir pris connaissance des informations ci-dessus, avoir obtenu les réponses à mes questions sur ma participation à la recherche et comprendre le but, la nature, les avantages, les risques et les inconvénients de cette recherche.

Après réflexion et un délai raisonnable, je consens à participer à cette étude. Je sais que je peux me retirer en tout temps, sur simple avis verbal, sans aucun préjudice.

Je consens à ce que les données anonymisées recueillies dans le cadre de cette étude soient utilisées pour des projets de recherche subséquents de même nature, conditionnellement à leur approbation éthique et dans le respect des mêmes principes de confidentialité et de protection des informations.

Oui Non

Signature : _____ Date : _____
Nom : _____ Prénom : _____

Je déclare avoir expliqué le but, la nature, les avantages, les risques et les inconvénients de l'étude et avoir répondu au meilleur de ma connaissance aux questions posées.

Signature du chercheur
(ou de son représentant) : _____ Date : _____
Nom : _____ Prénom : _____

Si vous avez des questions concernant ce projet de recherche, n'hésitez pas à communiquer avec Jean-Pierre Dumas candidat au doctorat aux coordonnées indiquées ci-dessous.

Courriel : jean-pierre.dumas@usherbrooke.ca

Téléphone : 819-820-6868 ext 12904

Pour toute information d'ordre éthique concernant les conditions dans lesquelles se déroule votre participation à ce projet, vous pouvez contacter le coordonnateur du Comité plurifacultaire d'éthique de la recherche (CPER) par courriel : cper@umontreal.ca ou par téléphone au (514) 343-6111 poste 1896. Pour plus d'information sur vos droits comme participants, vous pouvez consulter le portail des participants de l'Université de Montréal à l'adresse suivante : <http://recherche.umontreal.ca/participants>.

Toute plainte relative à votre participation à cette recherche peut être adressée à l'ombudsman de l'Université de Montréal, au numéro de téléphone (514) 343-2100 ou à l'adresse courriel suivante: ombudsman@umontreal.ca (l'ombudsman accepte les appels à frais virés).

Un exemplaire du formulaire d'information et de consentement signé doit être remis au participant

FORMULAIRE DE CONSENTEMENT (version clinicien expert)

Titre de la recherche : Évaluation du raisonnement clinique en physiothérapie
Volet 1 : tests de raisonnement clinique

Chercheur : Jean-Pierre Dumas, étudiant au doctorat, Département d'administration et fondements de l'éducation, Faculté des sciences de l'éducation, Université de Montréal

Directeurs de recherche : Jean-Guy Blais, professeur titulaire, Département d'administration et fondements de l'éducation, Faculté des sciences de l'éducation, Université de Montréal.

Bernard Charlin, professeur titulaire, Département de chirurgie, Faculté de médecine, Université de Montréal.

Co-chercheur : Annick Bourget, professeure adjointe, École de réadaptation, Université de Sherbrooke

Source de financement : Fonds de recherche résiduel du directeur suite à un contrat avec le ministère de l'éducation.

A) RENSEIGNEMENTS AUX PARTICIPANTS

1. Objectifs de la recherche.

Ce projet de recherche vise à développer des outils qui devraient nous permettre de mieux évaluer le raisonnement clinique des étudiants en physiothérapie.

2. Participation à la recherche

Dans le cadre d'un processus de validation d'un test de raisonnement clinique, nous vous demanderons de répondre à des questions portant sur des situations de pratique en physiothérapie musculosquelettique. Vos réponses serviront à établir le barème de correction des tests que des étudiants auront à effectuer.

3. Confidentialité

Les renseignements que vous nous donnerez demeureront confidentiels. Chaque participant à la recherche se verra attribuer un numéro et seules le chercheur et/ou la personne mandatée à cet effet auront la liste des participants et des numéros qui leur auront été attribués. De plus, les renseignements seront conservés dans un classeur sous clé situé dans le bureau fermé du chercheur principal. Aucune information permettant de vous identifier d'une façon ou d'une autre ne sera publiée. Ces renseignements personnels seront détruits 7 ans après la fin du projet. Seules les données ne permettant pas de vous identifier seront conservées après cette date, le temps nécessaire à leur utilisation

Page 1 de 3

4. Avantages et inconvénients

Les inconvénients sont liés au temps consacré à la participation (environ 3h). Toutefois, votre participation est, pour vous, une occasion d'apprentissage tout en contribuant à l'avancement des connaissances sur l'évaluation du raisonnement clinique et potentiellement d'améliorer nos pratiques à ce niveau.

5. Droit de retrait

Votre participation est entièrement volontaire. Vous êtes libre de vous retirer en tout temps sur simple avis verbal, sans préjudice et sans devoir justifier votre décision. Si vous décidez de vous retirer de la recherche, vous pouvez communiquer avec le chercheur, au numéro de téléphone indiqué ci-dessous. Si vous vous retirez de la recherche, les renseignements qui auront été recueillis au moment de votre retrait seront détruits.

6. Compensation

Une compensation financière de cent (100 \$) dollars sera versée pour votre participation à la recherche.

7. Diffusion des résultats

Un rapport sera transmis aux participants décrivant les conclusions générales de cette recherche au cours de l'année prochaine, lorsque les analyses auront été effectuées.

B) CONSENTEMENT

Je déclare avoir pris connaissance des informations ci-dessus, avoir obtenu les réponses à mes questions sur ma participation à la recherche et comprendre le but, la nature, les avantages, les risques et les inconvénients de cette recherche.

Après réflexion et un délai raisonnable, je consens à participer à cette étude. Je sais que je peux me retirer en tout temps, sur simple avis verbal, sans aucun préjudice.

Je consens à ce que les données anonymisées recueillies dans le cadre de cette étude soient utilisées pour des projets de recherche subséquents de même nature, conditionnellement à leur approbation éthique et dans le respect des mêmes principes de confidentialité et de protection des informations

Oui Non

Signature : _____ Date : _____
Nom : _____ Prénom : _____

Je déclare avoir expliqué le but, la nature, les avantages, les risques et les inconvénients de l'étude et avoir répondu au meilleur de ma connaissance aux questions posées.

Signature de la chercheure

(ou de son représentant) :

Date :

Nom :

Prénom :

Si vous avez des questions concernant ce projet de recherche, n'hésitez pas à communiquer avec Jean-Pierre Dumas candidat au doctorat aux coordonnées indiquées ci-dessous.

Courriel : jean-pierre.dumas@usherbrooke.ca

Téléphone : 819-820-6868 ext 12904

Pour toute information d'ordre éthique concernant les conditions dans lesquelles se déroule votre participation à ce projet, vous pouvez contacter le coordonnateur Comité plurifacultaire d'éthique de la recherche (CPER) par courriel : cper@umontreal.ca ou par téléphone au (514) 343-6111 poste 1896. Pour plus d'information sur vos droits comme participants, vous pouvez consulter le portail des participants de l'Université de Montréal à l'adresse suivante : <http://recherche.umontreal.ca/participants>.

Toute plainte relative à votre participation à cette recherche peut être adressée à l'ombudsman de l'Université de Montréal, au numéro de téléphone (514) 343-2100 ou à l'adresse courriel suivante : ombudsman@umontreal.ca (l'ombudsman accepte les appels à frais virés).

Un exemplaire du formulaire d'information et de consentement signé doit être remis au participant

ANNEXE 3 FORMULAIRE D'ÉVALUATION DE STAGE FOCUS



Formulaire d'observation des compétences- UdeS FOCUS-PHYSIOTHÉRAPIE

2012-2013

IDENTIFICATION

Nom du stagiaire : _____

Établissement : _____

Adresse du point de service : _____

Visée du Stage 6

Ce dernier stage vient compléter la formation clinique de l'étudiant avant qu'il entre en pratique en tant que physiothérapeute. Pendant ce stage, l'étudiant devrait démontrer un bon niveau d'autonomie et une capacité de leadership dans l'accomplissement de ses différentes tâches et responsabilités, autant au niveau des activités cliniques avec les clients qu'au niveau des rôles connexes. Il est également attendu que le stagiaire démontre une consolidation de son identité professionnelle en tant que physiothérapeute.

Le niveau de supervision initial est d'une intensité légère évoluant vers une autonomie en fin de stage auprès de tout type de clientèle. Au cours du stage, la capacité de prise en charge de clients évolue en qualité, quantité et efficacité. À la fin du stage 6, le niveau de productivité attendu s'apparentera à 90% de la charge de travail d'un physiothérapeute.

Les activités suivantes sont à réaliser durant le Stage 6

- Effectuer la prise en charge d'une clientèle en physiothérapie;
- Effectuer une présentation orale des résultats de son projet de recherche de type séminaire (30 minutes)

Autres activités réalisées durant ce stage

Rôle : EXPERT (ÉVALUATION)	
Progression	Critères de performance
Stage 1	<p>Mène l'entrevue sous supervision totale.</p> <p>S'assure de recueillir les préoccupations et de clarifier les attentes du client.</p>
Stage 2	<p>Obtient le consentement du client par rapport aux services offerts.</p> <p>Se prépare adéquatement avant l'évaluation initiale : lectures préparatoires, revue du dossier, plan d'évaluation, etc.</p> <p>Retrouve les informations écrites pertinentes dans le dossier du client avec assistance.</p> <p>Consulte son superviseur afin de valider sa démarche d'évaluation.</p> <p>Mène l'entrevue d'une manière structurée.</p> <p>Choisit les outils de mesures à utiliser avec assistance pour une clientèle orthopédique.</p> <p>Effectue les étapes de l'évaluation pour les clientèles orthopédiques (périphérique) sous supervision modérée.</p> <p>Respecte le protocole d'utilisation de chacun des outils de mesure.</p> <p>Interprète les données recueillies avec assistance.</p> <p>Identifie une liste de problèmes pour le client.</p>
Stage 3	<p>Retrouve les informations écrites pertinentes dans le dossier du client de façon autonome.</p> <p>Mène l'entrevue physiothérapique d'une façon efficace et autonome.</p> <p>Choisit des outils de mesure pertinents sous supervision modérée pour toutes les clientèles.</p> <p>Effectue les étapes de l'évaluation pour les clientèles orthopédique (périphérique et vertébral) neurologique et cardiorespiratoire sous supervision modérée.</p> <p>Analyse l'ensemble des données en effectuant une démarche de raisonnement clinique sous supervision modérée.</p> <p>Adapte son plan d'évaluation en fonction de facteurs émergents avec assistance.</p>
Stage 4	<p>Choisit des outils de mesure pertinents et justifie leur utilisation sous supervision légère.</p> <p>Effectue les étapes de l'évaluation d'une façon organisée et cohérente pour toutes les clientèles sous supervision légère.</p> <p>Explicite son raisonnement clinique et démontre une capacité de proposer des stratégies d'évaluation alternatives.</p> <p>Adapte son plan d'évaluation en fonction de facteurs émergents sous supervision modérée.</p> <p>S'assure d'avoir en main tous les renseignements utiles à l'analyse sous supervision légère.</p> <p>Recherche d'une façon appropriée les renseignements manquants auprès d'un tiers.</p>
Stage 5	<p>Applique les techniques d'évaluation connues ou administre des outils de mesure connus de façon optimale.</p> <p>Fait preuve d'un raisonnement clinique analytique tenant compte de multiples facteurs et résultats possibles.</p> <p>Réalise une démarche d'évaluation complète, incluant le rapport, dans des délais raisonnables.</p> <p>Sait trier les informations accessoires des informations pertinentes en fonction du contexte du client.</p> <p>Utilise des méthodes d'enquête appropriées, de manière efficace et éthique, tout au long du processus d'évaluation d'une façon autonome.</p> <p>Établit un pronostic pour le client avec assistance</p>
Stage 6	<p>Réalise une démarche d'évaluation complète et appropriée.</p> <p>Intègre des recommandations à des organismes tiers et fait état d'un plan de suivi s'il y a lieu.</p> <p>Réalise une démarche d'évaluation de façon autonome à une vitesse jugée productive.</p> <p>Établit un pronostic pour le client en se basant sur les données probantes et les meilleures pratiques.</p>

Rôle : EXPERT (ÉVALUATION)

Compétences clés du physiothérapeute

- Consulte le client afin d'obtenir des renseignements sur sa santé, ses antécédents en la matière, les interventions déjà subies et les résultats qui en ont découlé.
- Recueille des données d'évaluation pertinentes pour les besoins des clients et pour la pratique en physiothérapie.
- Analyse les constatations de cette évaluation.
- Élabore un diagnostic et un pronostic en physiothérapie.

Mi stage

Selon le niveau de compétence attendu en stage 6, l'étudiant :

- Dépasse les attentes
- Progrès selon les attentes
- Progrès avec difficulté
- Est en deçà des attentes
- Compétence non explorée pendant le stage

Commentaires formatifs

Fin de stage

Selon le niveau de compétence attendu en stage 6, l'étudiant :

- Dépasse les attentes
- Se conforme aux attentes
- Se conforme minimalement aux attentes
- Est en deçà des attentes
- Compétence non explorée pendant le stage

Commentaires formatifs

Rôle : EXPERT (INTERVENTION)	
Progression	Critères de performance
Stage 1	Non évalué
Stage 2	<p>Applique les connaissances pertinentes des sciences physiques, sociales, psychosociales, environnementales et biomédicales fondamentales dans la pratique. Effectue une l'analyse de la condition du client et propose des modalités thérapeutiques pertinentes (clientèle orthopédique) avec assistance.</p> <p>Observe ou expérimente l'application de modalités thérapeutiques sous supervision totale.</p> <p>Réalise, après démonstration de la part du superviseur, une partie du plan d'intervention.</p> <p>Veille au confort du client durant les interventions.</p> <p>Est attentif à son ergonomie de travail.</p>
Stage 3	<p>Identifie les modalités thérapeutiques pertinentes à utiliser pour les clientèles orthopédique, neurologique et cardiorespiratoire.</p> <p>Sélectionne des modalités thérapeutiques favorisant l'atteinte des objectifs du plan d'intervention.</p> <p>Utilise les modalités thérapeutiques simples de façon autonome.</p> <p>Expérimente l'application de modalités thérapeutiques complexes sous supervision.</p> <p>Utilise des stratégies de rétroaction pour guider les apprentissages du client.</p> <p>Mène à terme un plan d'intervention complet sous supervision modérée et avec assistance occasionnelle.</p> <p>Évalue et documente l'efficacité de ses interventions.</p> <p>Crée une alliance thérapeutique avec engagement et motivation mutuels.</p> <p>Accompagne le client dans son vécu.</p>
Stage 4	<p>Sélectionne et applique des modalités thérapeutiques variées et appropriées sous supervision légère.</p> <p>Priorise les cibles d'intervention d'une façon appropriée en incorporant la perspective du client.</p> <p>Intègre les données probantes et les guides de pratiques dans la démarche d'intervention.</p> <p>Détermine les objectifs pour le plan d'intervention sous supervision légère.</p> <p>Progresses le niveau de difficulté de la modalité thérapeutique en fonction de l'évolution de la condition du client sous supervision modérée.</p> <p>Explore des méthodes permettant d'optimiser la motivation chez les clients.</p>
Stage 5	<p>Sélectionne des modalités thérapeutiques variées favorisant l'atteinte des objectifs du plan d'intervention de façon autonome.</p> <p>Réalise le plan d'intervention en impliquant activement le client et les acteurs concernés.</p> <p>Prévoit et planifie le congé de ses clients sous supervision modérée.</p> <p>Anticipe l'évolution attendue de ses clients et l'oriente vers des ressources au besoin.</p> <p>Utilise des techniques permettant de développer la motivation du client.</p> <p>Au besoin, consulte efficacement, adéquatement et au moment opportun d'autres professionnels de la santé en vue d'optimiser les services offerts aux clients.</p>
Stage 6	<p>Fait preuve d'expertise en menant à terme l'intervention efficacement.</p> <p>Ajuste son plan de traitement en fonction de l'efficacité de ses interventions d'une façon autonome.</p> <p>Établit un échéancier réaliste pour le plan de traitement.</p> <p>Assure la coordination du congé ou du suivi de ses clients d'une façon autonome.</p> <p>Fait preuve de raisonnement clinique tenant compte de multiples variables et/ou solutions possibles dans ses interventions auprès des clients.</p>

Rôle : EXPERT (INTERVENTION)	
Compétences clés du physiothérapeute <ul style="list-style-type: none">• Élabore et recommande une stratégie d'intervention.• Met en œuvre ces interventions.• Évalue l'efficacité de ses interventions.• Complète les services de physiothérapie.	
Mi stage <div style="border: 1px solid black; padding: 5px;"><p>Selon le niveau de compétence attendu en stage 6, l'étudiant :</p><ul style="list-style-type: none"><input type="checkbox"/> Dépasse les attentes<input type="checkbox"/> Progrès selon les attentes<input type="checkbox"/> Progrès avec difficulté<input type="checkbox"/> Est en deçà des attentes <input type="checkbox"/> Compétence non explorée pendant le stage</div> <p>Commentaires formatifs</p>	Fin de stage <div style="border: 1px solid black; padding: 5px;"><p>Selon le niveau de compétence attendu en stage 6, l'étudiant :</p><ul style="list-style-type: none"><input type="checkbox"/> Dépasse les attentes<input type="checkbox"/> Se conforme aux attentes<input type="checkbox"/> Se conforme minimalement aux attentes<input type="checkbox"/> Est en deçà des attentes <input type="checkbox"/> Compétence non explorée pendant le stage</div> <p>Commentaires formatifs</p>

Rôle : COMMUNICATEUR	
Progression	Critères de performance
Stage 1	<p>Communique de façon appropriée avec les clients. S'exprime clairement. Utilise un langage adapté au contexte. Éveille la confiance, le respect et favorise l'expression du vécu chez le client. Démontre une attitude de soutien, d'écoute et de compréhension. Fait preuve de jugement et de discernement quant au moment opportun pour discuter et au contenu abordé. Écrit lisiblement.</p>
Stage 2	<p>S'approprie le vocabulaire du domaine de la réadaptation et de la santé. Adapte sa communication en présence de barrières à la communication (ex. : langage, perte auditive, incapacité à communiquer verbalement). Écoute de manière active en utilisant diverses stratégies de communication : reformulation, clarification, vérification-synthèse, questions ouvertes/fermées. Crée un climat permettant au client de se confier. Est sensible aux indices de communication non verbale. Use de discrétion (ton, endroit, présences) de manière à respecter la confidentialité avec son client. Transmet de l'information de manière respectueuse et réfléchie. Donne des instructions et des consignes orales de façon claire. Rédige les notes au dossier en conformité avec les exigences de l'OPPQ avec assistance.</p>
Stage 3	<p>Explique au client les résultats de son évaluation et de son plan d'intervention sous supervision. Ajuste son mode d'interaction en fonction de l'état émotionnel du client. Vulgarise l'information communiquée au client. Sélectionne les informations pertinentes à transmettre. Synthétise les informations communiquées. Exprime son opinion physiothérapique d'une façon structurée et concise à ses collègues. Transmet les informations écrites, électroniques, téléphoniques en respectant les règles de confidentialité. Rédige les notes au dossier en conformité avec les attentes du milieu clinique et les exigences de l'OPPQ sous supervision modérée.</p>
Stage 4	<p>Explique au client le plan d'intervention dans un langage clair et d'une façon autonome. Accompagne le client dans son vécu affectif sans être déstabilisé. Ceme la bonne information à transmettre selon le destinataire visé. Emploie des approches qui encouragent la participation du client à la prise de décision. Exprime son opinion physiothérapique d'une façon structurée et concise en équipe interdisciplinaire. Communique devant un auditoire avec aisance. Utilise les technologies de l'information et de communication de façon appropriée (courriel, présentations audio-visuelles, etc.). Rédige les notes au dossier de façon structurée, concise et complète sous supervision légère.</p>
Stage 5	<p>Gère les situations conflictuelles ou difficiles avec son client. Fait preuve de diplomatie en modulant le contenu et le ton de sa communication en présence de perception divergente d'une situation clinique. Aide les différentes parties à communiquer ouvertement et à tenir compte des autres opinions. Documente la conclusion et la cessation des interventions et diffuse l'information et les recommandations pour les étapes suivantes. Rédige des notes au dossier adéquates de façon autonome. Utilise des stratégies de négociation.</p>
Stage 6	<p>Fait preuve d'introspection au niveau de ses propres dimensions émotives et de leur impact sur la relation thérapeutique. Gère les situations conflictuelles ou difficiles qui surviennent dans l'environnement social du client.</p>

Rôle : COMMUNICATEUR	
Compétences clés du physiothérapeute <ul style="list-style-type: none">• Grâce à une communication efficace, crée, construit et entretient une relation professionnelle fondée sur la confiance et sur l'éthique.• Repère, analyse, note, applique, relaie et partage l'information.• Utilise efficacement les communications verbales, non verbales, écrites et électroniques.	
Mi stage	Fin de stage
<p>Selon le niveau de compétence attendu en stage 6, l'étudiant :</p> <ul style="list-style-type: none"><input type="checkbox"/> Dépasse les attentes<input type="checkbox"/> Progrès selon les attentes<input type="checkbox"/> Progrès avec difficulté<input type="checkbox"/> Est en deçà des attentes <input type="checkbox"/> Compétence non explorée pendant le stage	<p>Selon le niveau de compétence attendu en stage 6, l'étudiant :</p> <ul style="list-style-type: none"><input type="checkbox"/> Dépasse les attentes<input type="checkbox"/> Se conforme aux attentes<input type="checkbox"/> Se conforme minimalement aux attentes<input type="checkbox"/> Est en deçà des attentes <input type="checkbox"/> Compétence non explorée pendant le stage
Commentaires formatifs	Commentaires formatifs

Rôle : COLLABORATEUR	
Progression	Critères de performance
Stage 1	<p>S'informe des rôles et responsabilités des autres intervenants impliqués dans les soins aux clients.</p> <p>Agit de façon respectueuse avec le personnel de soutien et les autres professionnels de la santé.</p> <p>Communique de façon appropriée avec le superviseur et les collègues de travail.</p>
Stage 2	<p>Reconnaît la contribution des autres professionnels de la santé.</p> <p>Ajuste sa communication en fonction de son interlocuteur (famille, collègue, etc.).</p> <p>Démontre une ouverture à collaborer.</p> <p>Travaille en collaboration avec le personnel de soutien.</p> <p>Démontre une compréhension des rôles et des responsabilités des membres de l'équipe.</p>
Stage 3	<p>Reconnaît, décrit et respecte la diversité des rôles et les responsabilités d'autres professionnels membres de l'équipe.</p> <p>Comprend le rôle, la contribution et la nature des interventions des autres professionnels de la santé.</p> <p>Vulgarise l'information communiquée à la famille ou à l'entourage social du client.</p> <p>Communique avec la famille et l'entourage du client de façon sensible à leurs besoins.</p> <p>Démontre une compréhension du travail en équipe interdisciplinaire.</p> <p>Crée un climat favorable à la collaboration.</p> <p>Contribue aux activités de l'équipe interdisciplinaire dans les limites de son rôle de stagiaire.</p> <p>Collabore à l'élaboration d'un plan de soins interdisciplinaire.</p> <p>S'initie au travail intradisciplinaire (TRP) et à la réglementation encadrant les responsabilités et les pratiques professionnelles.</p>
Stage 4	<p>Informe et intègre les personnes clés de l'environnement social dans le plan d'intervention avec assistance.</p> <p>Explique la contribution de la physiothérapie à l'intérieur d'une équipe interdisciplinaire en tenant compte des besoins spécifiques du client.</p> <p>Reconnaît les situations qui nécessitent de consulter les autres professionnels de la santé.</p> <p>Sollicite l'avis des autres professionnels de la santé (médecin, infirmière, orthophoniste, spécialiste, etc.).</p> <p>Ajuste son plan d'intervention physiothérapique en fonction du plan d'intervention interdisciplinaire.</p> <p>Contacte les agents payeurs ou les mandataires au besoin (SAAQ, CSST, assureur, etc.) avec assistance.</p> <p>Contribue à une dynamique d'équipe positive.</p> <p>Reconnaît les situations pouvant créer des tensions entre les membres de l'équipe.</p>
Stage 5	<p>Travaille avec les membres de l'équipe en utilisant la prise de décision partagée afin de répondre aux besoins du client.</p> <p>Fait participer le client à titre de membre actif aux démarches consultatives lorsque cela est possible.</p> <p>S'intègre activement dans une équipe interdisciplinaire.</p> <p>Informe efficacement et au moment opportun les membres de l'équipe et les partenaires concernant son apport spécifique pour le client.</p> <p>Démontre une capacité à faire des compromis.</p>
Stage 6	<p>Planifie la demande de collaboration à faire auprès de différents organismes.</p> <p>Travaille avec les membres de l'équipe pour évaluer, planifier et offrir aux clients une approche intégrée des services.</p> <p>Négocie avec les collaborateurs pour amener une prise de décision centrée sur le client.</p> <p>Assume différents rôles à l'intérieur d'une équipe de travail (leader, animateur, secrétaire, etc.).</p> <p>Défend son jugement professionnel en situation de divergence d'opinions.</p> <p>Gère les divergences d'opinions ou les conflits relatifs au fonctionnement d'un groupe de manière professionnelle et éthique.</p> <p>Démontre des habiletés de médiation et de négociation (advocacy).</p> <p>Prend l'initiative de contacter les professionnels de la santé (médecin, infirmière, orthophoniste, etc.), lorsque pertinent.</p>

Rôle : COLLABORATEUR	
Compétences clés du physiothérapeute <ul style="list-style-type: none">• Établit et maintient des relations interprofessionnelles qui favorisent une collaboration efficace, centrée sur le client.• Collabore avec les autres en vue de prévenir, de gérer et de dénouer les conflits.	
Mi stage	Fin de stage
<p>Selon le niveau de compétence attendu en stage 6, l'étudiant :</p> <ul style="list-style-type: none"><input type="checkbox"/> Dépasse les attentes<input type="checkbox"/> Progrès selon les attentes<input type="checkbox"/> Progrès avec difficulté<input type="checkbox"/> Est en deçà des attentes <input type="checkbox"/> Compétence non explorée pendant le stage	<p>Selon le niveau de compétence attendu en stage 6, l'étudiant :</p> <ul style="list-style-type: none"><input type="checkbox"/> Dépasse les attentes<input type="checkbox"/> Se conforme aux attentes<input type="checkbox"/> Se conforme minimalement aux attentes<input type="checkbox"/> Est en deçà des attentes <input type="checkbox"/> Compétence non explorée pendant le stage
Commentaires formatifs	Commentaires formatifs

Rôle : GESTIONNAIRE	
Progression	Critères de performance
Stage 1	Est ponctuel et a une tenue vestimentaire appropriée. Démontre un intérêt pour le fonctionnement de l'établissement de santé. Fait preuve de curiosité pour connaître l'utilité des ressources matérielles. Prépare et range les équipements selon les indications fournies par le superviseur.
Stage 2	Prévoit et organise les ressources matérielles requises pour ses évaluations. Gère les ressources matérielles pour ses clients : prépare, range, nettoie les équipements. Gère son temps de façon efficiente en présence d'un seul client. S'initie aux activités indirectes (dossiers, réunions, statistiques, appels téléphoniques, etc.) sous supervision totale.
Stage 3	Connait et agit en tenant compte des politiques et procédures de l'établissement. Planifie l'horaire de travail avec assistance. Organise et effectue sa journée de travail en fonction des tâches à accomplir. Gère son temps efficacement de façon à intégrer à son horaire la diversité de tâches cliniques en lien avec ses clients. Prévoit et organise toutes les ressources matérielles requises pour ses interventions. Réalise des activités indirectes (dossiers, réunions, statistiques, appels téléphoniques, etc.) sous supervision modérée. Établit les priorités et gère son temps afin d'établir un équilibre entre les services offerts aux clients et les exigences en matière de pratique.
Stage 4	Gère son temps de façon à intégrer la diversité des tâches cliniques d'une manière efficiente (activités directes et indirectes). Gère son temps de façon efficiente en superposant la fin du traitement d'un client et le début d'un autre. Prévoit et organise les ressources matérielles requises pour les évaluations et les interventions de façon autonome. Participe à la gestion de la liste d'attente en collaboration avec les collègues. S'initie à des activités administratives (facturation, budget, comités, etc.). Répartit les services thérapeutiques tout en établissant un équilibre entre les besoins des clients et les ressources disponibles. Reconnaît les activités à déléguer. Travaille en intradisciplinarité (avec TRP) dans un climat favorisant la collaboration sous supervision modérée. Reconnaît les activités à déléguer.
Stage 5	Participe à des activités autres que cliniques : promotion de la santé, tâches administratives, projets cliniques, ou autres. Planifie et exécute son horaire de travail de façon autonome à l'intérieur de ses heures de travail incluant les formations, les réunions, la recherche, etc. Gère son temps de façon efficiente en présence de plus d'un client. Fait preuve de souplesse dans la réorganisation des activités cliniques en présence d'imprévus. Détermine l'admissibilité, la pertinence et le degré de priorité de la demande de service. Planifie l'horaire et la supervision d'un stagiaire 1 de l'Université de Sherbrooke*. Établit les priorités efficacement et adéquatement face aux besoins multiples et concurrentiels des clients. Délègue certaines tâches au personnel de soutien et en assure le suivi sous supervision totale. Travaille en intradisciplinarité (TRP) de façon efficace et prévoit la contribution de chacun dans le plan de traitement.
Stage 6	Gère l'ensemble des activités cliniques directes et indirectes de façon autonome et efficiente. Planifie et exécute son horaire de travail de façon autonome à l'intérieur de ses heures de travail incluant des activités de promotion de la santé, des tâches administratives, des projets cliniques, de la formation, des réunions, de la recherche, etc. Établit un équilibre entre les activités liées au travail, les activités extérieures et les priorités personnelles. Assume des tâches administratives selon les besoins du milieu clinique. Reconnaît le conflit potentiel entre les ressources limitées et la présence de demandes de services, et propose une gestion éthique et équitable. S'informe des processus d'évaluation et d'amélioration des processus systémiques de maintien de la qualité, comme les initiatives sur la sécurité des clients. Délègue des activités thérapeutiques au personnel de soutien et fait le suivi sous supervision modérée.

Rôle : GESTIONNAIRE

Compétences clés du physiothérapeute

- Assure efficacement la gestion de sa pratique individuelle.
- Administre et supervise le personnel participant à la prestation de soins en physiothérapie.
- Participe à des activités qui contribuent à une pratique sûre et efficace de la physiothérapie.

Mi stage

Selon le niveau de compétence attendu en stage 6, l'étudiant :

- Dépasse les attentes
- Progrès selon les attentes
- Progrès avec difficulté
- Est en deçà des attentes
- Compétence non explorée pendant le stage

Commentaires formatifs

Fin de stage

Selon le niveau de compétence attendu en stage 6, l'étudiant :

- Dépasse les attentes
- Se conforme aux attentes
- Se conforme minimalement aux attentes
- Est en deçà des attentes
- Compétence non explorée pendant le stage

Commentaires formatifs

Rôle : DÉFENSEUR	
Progression	Critères de performance
Stage 1	Non évalué
Stage 2	Questionne le client sur ses habitudes de vie ou comportements de santé. Reconnaît les facteurs de risques et les déterminants de la santé. Effectue ou réfère le client à des dépistages pertinents. Reconnaît à quelle étape du processus de changement de comportement le client se situe. Identifie les facilitateurs et obstacles au changement de comportement de son client.
Stage 3	Fait la promotion de saines habitudes de vie avec sa clientèle.
Stage 4	Agit de façon à défendre les intérêts des individus recevant des services de physiothérapie. Définit les déterminants de la santé des clients recevant des services, notamment les barrières empêchant l'accès aux services et aux ressources. Réfère le client vers des ressources communautaires pouvant contribuer à son bien-être. Sensibilise l'environnement social en lien avec les différences liées aux capacités.
Stage 5	Aide le client, selon le cas, dans ses efforts pour obtenir le financement et les services requis pour atteindre les résultats déterminés dans le plan d'action. Représente les intérêts de son client lorsqu'il interagit avec les autres acteurs concernés. Agit à titre de représentant du client auprès de divers mandataires (SAAQ, CSST, Anciens Combattants du Canada, etc.) sous supervision. Est sensible à et gère les questions d'éthique inhérentes à la défense des droits du client, en promouvant l'altruisme, l'intégrité, l'autonomie, et la justice sociale.
Stage 6	Accompagne le client, son représentant ou le demandeur de services dans les démarches auprès des organismes qui peuvent répondre aux besoins du client. Gère le conflit entre son rôle d'agent de changement auprès du client et celui de gestionnaire de ressources et de services limités.

Rôle : DÉFENSEUR

Compétences clés du physiothérapeute

- Travaille en collaboration en vue d'identifier et de réagir aux besoins et aux préoccupations en matière de santé de ses clients individuels, de la population en général et de la collectivité; assure la promotion de ces besoins et préoccupations.

Mi stage

Selon le niveau de compétence attendu en stage 6, l'étudiant :

- Dépasse les attentes
- Progrès selon les attentes
- Progrès avec difficulté
- Est en deçà des attentes
- Compétence non explorée pendant le stage

Commentaires formatifs

Fin de stage

Selon le niveau de compétence attendu en stage 6, l'étudiant :

- Dépasse les attentes
- Se conforme aux attentes
- Se conforme minimalement aux attentes
- Est en deçà des attentes
- Compétence non explorée pendant le stage

Commentaires formatifs

Rôle : PRATICIEN ÉRUDIT	
Progression	Critères de performance
Stage 1	<p>Partage les connaissances acquises au cours de sa formation.</p> <p>Pose à son superviseur des questions et propose des hypothèses pour y répondre.</p> <p>Explore les documents disponibles dans le milieu clinique.</p>
Stage 2	<p>Complète son contrat d'apprentissage et son journal de bord quotidiennement.</p> <p>Fait preuve d'initiative dans la recherche de solutions ou de réponses à ses questions.</p> <p>Démontre de l'initiative pour communiquer ses apprentissages et connaissances à son superviseur.</p> <p>Fait des démarches pour acquérir les informations nécessaires pour la compréhension de sa clientèle.</p> <p>Utilise différents modèles théoriques dans sa pratique et les explique à son superviseur.</p> <p>Procède à l'auto-évaluation de sa compétence et le communique à son superviseur.</p>
Stage 3	<p>Enseigne ou transmet des connaissances à son client et à son entourage.</p> <p>Reconnaît les difficultés d'apprentissage chez ses clients et adapte des stratégies d'enseignement en conséquence.</p> <p>Identifie et discute des écarts qui existent entre les connaissances, les habiletés cliniques et les compétences en milieu clinique par celles enseignées à l'université.</p> <p>Reconnaît le besoin d'approfondir ses connaissances sur un sujet précis.</p> <p>Applique des méthodes d'apprentissage autonomes pour répondre aux questions et résoudre des problèmes.</p> <p>Explore les écrits scientifiques afin d'approfondir ses connaissances en lien avec la physiothérapie.</p> <p>Effectue une évaluation régulière de ses besoins personnels en matière d'apprentissage.</p>
Stage 4	<p>Démontre une capacité analytique permettant la résolution de problèmes multiples.</p> <p>Recherche des données probantes en lien avec sa clientèle et partage cette information avec le superviseur.</p> <p>Contribue aux échanges scientifiques informels dans le milieu clinique.</p> <p>Fait des démarches pour acquérir des informations au sujet de conditions de santé moins couramment rencontrées.</p> <p>Présente à ses collègues une étude de cas ou une critique d'un article scientifique.</p> <p>Choisit des stratégies d'enseignement et du contenu efficaces afin de faciliter l'apprentissage chez des tiers.</p> <p>Évalue de façon critique les données probantes afin de répondre à des questions cliniques.</p>
Stage 5	<p>Intègre et applique les informations des écrits scientifiques pour répondre à de nouvelles situations cliniques.</p> <p>Présente un plan d'intervention réalisé avec sa clientèle en intégrant une évaluation critique des données probantes.</p> <p>Développe du matériel didactique pouvant rehausser la qualité des informations communiquées aux clients et leur entourage.</p> <p>Intègre les nouveaux apprentissages et les nouvelles données probantes dans la pratique.</p>
Stage 6	<p>Collabore à la diffusion de communications scientifiques ou de projets de recherche.</p> <p>Utilise des stratégies pédagogiques pour rehausser la qualité de ses exposés scientifiques.</p> <p>Démontre des techniques efficaces de présentations ou de conférences et d'exposés.</p> <p>Porte un jugement critique (auto-évaluation) sur la construction de sa pratique professionnelle et sur le développement de ses compétences.</p> <p>Établit son plan de perfectionnement professionnel en fonction de ses intérêts, points forts et points à améliorer.</p>

Rôle : PRATICIEN ÉRUDIT

Compétences clés du physiothérapeute

- A recours à une approche réfléchie de la pratique.
- Incorpore l'éducation permanente et ses propres expériences aux pratiques exemplaires.
- Participe à de la recherche.

Mi stage

Selon le niveau de compétence attendu en stage 6, l'étudiant :

- Dépasse les attentes
- Progrès selon les attentes
- Progrès avec difficulté
- Est en deçà des attentes
- Compétence non explorée pendant le stage

Commentaires formatifs

Fin de stage

Selon le niveau de compétence attendu en stage 6, l'étudiant :

- Dépasse les attentes
- Se conforme aux attentes
- Se conforme minimalement aux attentes
- Est en deçà des attentes
- Compétence non explorée pendant le stage

Commentaires formatifs

Rôle : PROFESSIONNEL	
Progression	Critères de performance
Stage 1	<p>Agit de façon sécuritaire pour sa personne, les clients et ses collègues. Respecte la confidentialité et le secret professionnel. Fait preuve de curiosité pour comprendre les situations cliniques observées. Fait preuve d'humilité. Reçoit des directives, commentaires et recommandations et y répond positivement.</p>
Stage 2	<p>Établit et maintient une relation professionnelle avec son client (distanciation émotionnelle). Prend en compte les émotions, les valeurs et le contexte socioculturel du client et évite les jugements de valeur. Adopte les comportements professionnels appropriés, y compris l'honnêteté, l'intégrité, l'engagement, la compassion, le respect, et l'altruisme. Reconnaît ses limites personnelles ainsi que les limites de son expertise et consulte au besoin. Recherche et respecte les connaissances et l'expérience de son superviseur et des autres intervenants rencontrés. Reçoit la critique constructive de façon appropriée et adapte ses savoirs, savoir-faire et savoir être en conséquence. Connaît les codes de déontologie ainsi que les exigences de la profession et de la loi. Explique dans des termes vulgarisés le rôle du physiothérapeute. Demande de l'assistance, lorsque nécessaire.</p>
Stage 3	<p>Assure la responsabilité de ses attitudes, gestes et comportements et, si nécessaire, les modifie pour se conformer aux attentes du milieu. Gère les informations provenant d'un tiers (famille, proche aidant, etc.) en conformité avec les exigences légales. Reconnaît et répond de façon appropriée aux questions d'éthique qui se posent dans la pratique. Respecte les normes professionnelles et les normes relatives à la réglementation de la pratique. Explique la contribution possible de la physiothérapie pour le client.</p>
Stage 4	<p>Gère les informations provenant d'un tiers selon les règles déontologiques. Démonstre une appropriation de la culture physiothérapique au regard du langage, des agirs et du raisonnement. Explique à son client la contribution de la physiothérapie pour sa condition.</p>
Stage 5	<p>Fait la promotion de la physiothérapie auprès du client et de son réseau social. Prend des initiatives professionnelles avec l'accord du superviseur. Réfère la personne à un autre professionnel si la condition de la personne est hors de son champ de compétences.</p>
Stage 6	<p>Fait preuve de jugement et d'affirmation dans l'exercice de ses responsabilités professionnelles. Adopte une attitude qui témoigne d'une identité professionnelle et d'une assurance professionnelle certaine. Se montre sensible à la diversité et aux questions de pouvoir qui entrent en jeu dans une relation professionnelle. Fait les démarches pour s'inscrire à l'Ordre professionnel de la physiothérapie du Québec (OPPQ). S'engage à exercer la physiothérapie de manière compétente. Réalise son rôle de physiothérapeute.</p>

Rôle : PROFESSIONNEL

Compétences clés du physiothérapeute

- Adopte une conduite respectant les exigences juridiques et éthiques.
- Respecte l'individualité et l'autonomie du client.
- Contribue au développement de la profession de physiothérapeute.

Mi stage

Selon le niveau de compétence attendu en stage 6, l'étudiant :

- Dépasse les attentes
- Progrès selon les attentes
- Progrès avec difficulté
- Est en deçà des attentes
- Compétence non explorée pendant le stage

Commentaires formatifs

Fin de stage

Selon le niveau de compétence attendu en stage 6, l'étudiant :

- Dépasse les attentes
- Se conforme aux attentes
- Se conforme minimalement aux attentes
- Est en deçà des attentes
- Compétence non explorée pendant le stage

Commentaires formatifs

COMMENTAIRES FORMATIFS	
MI-STAGE	
Points forts	Cibles de travail pour la fin du stage
FIN DE STAGE	
Points forts	Cibles de travail pour le prochain stage

Absence(s) :	Nombre : _____	Motif :
Retard(s) :	Nombre : _____	Motif :

Description de la clientèle vue en stage

Mission

- | | | |
|-----------------------------|--------------------------------|--------------------------------------|
| <input type="checkbox"/> CH | <input type="checkbox"/> CLSC | <input type="checkbox"/> CHSLD |
| <input type="checkbox"/> CR | <input type="checkbox"/> PRIVÉ | <input type="checkbox"/> Autre _____ |

Programme particulier

(si applicable)

- | | | |
|------------------------------------------|-------------------------------|-----------------------------------------|
| <input type="checkbox"/> Centre de jour | <input type="checkbox"/> URFI | <input type="checkbox"/> Autre : _____ |
| <input type="checkbox"/> Hôpital de jour | <input type="checkbox"/> UCDG | <input type="checkbox"/> Non applicable |

Population

- | | | |
|---------------------------------------------|----------------------------------------------|-----------------------------------------------|
| <input type="checkbox"/> Pédiatrie (0-1 an) | <input type="checkbox"/> Pédiatrie (1-5 ans) | <input type="checkbox"/> Pédiatrie (5-17 ans) |
| <input type="checkbox"/> Adulte | <input type="checkbox"/> Aîné | |

Clientèle

- | | |
|-------------------------------------------------------------------------|---------------------|
| <input type="checkbox"/> Orthopédie/rhumatologie (externe) | Pourcentage : _____ |
| <input type="checkbox"/> Orthopédie/rhumatologie (hospitalisée) | Pourcentage : _____ |
| <input type="checkbox"/> Blessé orthopédique grave/trauma | Pourcentage : _____ |
| <input type="checkbox"/> Neurologie (AVC et autres atteintes centrales) | Pourcentage : _____ |
| <input type="checkbox"/> Neurologie (TCC) | Pourcentage : _____ |
| <input type="checkbox"/> Neurologie (blessés médullaires) | Pourcentage : _____ |
| <input type="checkbox"/> Cardiorespiratoire | Pourcentage : _____ |
| <input type="checkbox"/> Perte d'autonomie/atteintes évolutives | Pourcentage : _____ |
| <input type="checkbox"/> Amputés | Pourcentage : _____ |
| <input type="checkbox"/> Plaies/Plastie/Brûlés | Pourcentage : _____ |
| <input type="checkbox"/> Déficience intellectuelle - TED | Pourcentage : _____ |
| <input type="checkbox"/> Rééducation périnéale | Pourcentage : _____ |
| <input type="checkbox"/> Autre : _____ | Pourcentage : _____ |

ÉVALUATION SOMMATIVE DU STAGE 6

(À compléter à la fin du stage)

1. Développement des rôles professionnels

Pour les rôles observés durant ce stage, placer un trait sur l'échelle visuelle analogue.

Débutant : le stagiaire qui est exposé pour la première fois à une expérience clinique.

Prêt pour la pratique : le stagiaire qui est compétent pour travailler comme physiothérapeute.

Expert

Évaluation

↑ _____ ↑
Débutant Prêt pour la pratique

Expert

Intervention

↑ _____ ↑
Débutant Prêt pour la pratique

Communicateur

↑ _____ ↑
Débutant Prêt pour la pratique

Collaborateur

↑ _____ ↑
Débutant Prêt pour la pratique

Gestionnaire

↑ _____ ↑
Débutant Prêt pour la pratique

Défenseur

↑ _____ ↑
Débutant Prêt pour la pratique

Praticien érudit

↑ _____ ↑
Débutant Prêt pour la pratique

Professionnel

↑ _____ ↑
Débutant Prêt pour la pratique

2. L'inventaire des expériences cliniques reflète les expériences rencontrées pendant le stage

J'ai vérifié l'inventaire des expériences cliniques de l'étudiant : _____
(Initiales du superviseur)

3. Recommandation d'échec ou de réussite du superviseur

Réussite

Échec*

* S'il y a lieu, veuillez joindre les documents suivants :

Annexe F : Documentation des difficultés quant au développement d'une compétence

Annexe G : Formulaire de déclaration d'incident critique

SIGNATURES			
<u>SUPERVISEUR</u>			
_____ Superviseur	_____ Signature	_____ Date	_____ N° OPPQ
_____ Superviseur	_____ Signature	_____ Date	_____ N° OPPQ
<u>STAGIAIRE</u>			
<input type="checkbox"/> Je suis d'accord avec mon évaluation de stage.			
<input type="checkbox"/> Je ne suis pas d'accord avec mon évaluation de stage (apposer vos initiales).			
Commentaires :			
_____ Stagiaire	_____ Signature	_____ Date	_____ N° Matricule

Retourner le formulaire original dans l'enveloppe-réponse ci-jointe (signer l'enveloppe pour authentifier le document). L'enveloppe incluant le formulaire peut être retournée à l'École par l'étudiant ou peut être postée à :

Paule Desrochers
 Commis aux stages, École de réadaptation
 Faculté de médecine et des sciences de la santé
 Université de Sherbrooke
 3001, 12^e Avenue Nord Sherbrooke (Québec) J1H 5N4

Pour le stagiaire : conserver une copie de ce formulaire dans votre portfolio (avec votre auto-évaluation de stage)

Pour toute question, veuillez contacter :
 Anne-Marie Côté, pht, M.Sc.
 Coordinatrice de la formation clinique en physiothérapie
 819 820-6868 poste 72908
 Anne-Marie.Cote3@USherbrooke.ca

NOTES

NOTES

ANNEXE 4 TEST DE CONCORDANCE DE SCRIPT

Nom :

TESTS DE CONCORDANCE DE SCRIPTS

INSTRUCTIONS

Ce test contient 26 vignettes cliniques de 3 à 4 items chacune. Certaines vignettes portent sur des décisions à caractère diagnostique et d'autres à caractère thérapeutique.

1. Chacune des vignettes contient des informations générales et incomplètes sur un patient.
2. Dans la première partie de chaque item, nous vous suggérons une alternative de traitement ou une hypothèse diagnostique plausible. Par exemple, à la vignette 1, sous la première colonne, si vous pensiez à :

EXEMPLE DE VIGNETTE CLINIQUE :

Un patient de 56 ans consulte en physiothérapie à la suite de douleurs cervicales sévères en lien avec un accident d'auto survenu hier.

Si vous pensiez recommander...	et que vous trouvez...	votre recommandation de traitement devient* (encerclez)
un collet cervical,	des mouvements limités à moins de 10 degrés dans toutes les directions,	-2 -1 0 1 2
de revoir son médecin pour une radiographie,	une diminution du réflexe bicipital à droite,	-2 -1 0 1 2

*

-2 : contre-indiqué ou presque; -1 : peu indiqué; 0 : ni plus, ni moins indiqué;
1 : plus indiqué; 2 : complètement indiqué ou presque

3. Dans la 2^e colonne, une information additionnelle sur le patient vous est fournie. Ce que l'on cherche à savoir, c'est de quelle façon cette nouvelle information influencera l'alternative de traitement ou l'hypothèse diagnostique initialement considérée (dans la 1^{re} colonne).
4. Une échelle est fournie au bas de chaque vignette. Vous encerclez la réponse qui correspond le plus à votre opinion.
5. Chaque item se réfléchit en fonction de la vignette et est indépendant du contenu des autres items. Par exemple, dans la vignette ci-dessous, le patient de l'item B n'aura pas nécessairement un collet cervical comme traitement

Vignette 1

Votre patiente de 64 ans a fait une chute sur sa main droite après avoir glissé sur une plaque de glace. Elle vous consulte en accès direct.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)
1. une fracture du scaphoïde	une bonne force à la préhension de la main	-2 -1 0 1 2
2. une fracture du radius	une coloration bleutée de la main	-2 -1 0 1 2
3. une fracture du radius	une diminution de 20 degrés en flexion et extension du poignet	-2 -1 0 1 2

*

-2 : éliminée ou presque; -1 : moins probable; 0 : ni plus, ni moins probable;
1 : plus probable; 2 : certaine ou presque

Vignette 2

Votre patient est un menuisier de 34 ans qui a des douleurs au niveau latéral du coude droit depuis plus de 6 semaines.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)
4. une épicondylite latérale	une flexibilité normale des extenseurs du poignet droit	-2 -1 0 1 2
5. une épicondylite latérale	un engourdissement du pouce et de l'index à droite	-2 -1 0 1 2
6. une épicondylite latérale	une faiblesse importante de la préhension de la main droite	-2 -1 0 1 2

*

-2 : éliminée ou presque; -1 : moins probable; 0 : ni plus, ni moins probable;
1 : plus probable; 2 : certaine ou presque

Vignette 3

Votre patient de 41 ans a des douleurs diffuses au niveau lombo-sacré gauche depuis 3 semaines, et ce, après avoir pelleté plusieurs heures.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)				
7. une protusion discale	SLR négatif bilatéralement	-2	-1	0	1	2
8. une protusion discale	une diminution du réflexe achilléen à droite	-2	-1	0	1	2
9. une protusion discale	la présence de spasmes paravertébraux à gauche au niveau lombo-sacré	-2	-1	0	1	2
10. une protusion discale	la présence d'un arc de douleur en flexion	-2	-1	0	1	2

*

-2 : éliminée ou presque; -1 : moins probable; 0 : ni plus, ni moins probable;

1 : plus probable; 2 : certaine ou presque

Vignette 4

Votre patient de 71 ans a des douleurs diffuses au niveau postérieur de la cuisse droite, depuis plus de 6 mois, ce qui entraîne des difficultés à marcher plus de 10 minutes.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)				
11. une sténose spinale	une diminution de lordose lombaire en position debout, mais une lordose normale en position couchée	-2	-1	0	1	2
12. une sténose spinale	des réflexes ostéotendineux normaux aux membres inférieurs	-2	-1	0	1	2
13. une sténose spinale	une augmentation de la douleur à la cuisse droite en position assise	-2	-1	0	1	2
14. une sténose spinale	SLR négatif bilatéralement	-2	-1	0	1	2

*

<p>-2 : éliminée ou presque; -1 : moins probable; 0 : ni plus, ni moins probable; 1 : plus probable; 2 : certaine ou presque</p>

Vignette 5

Votre patient de 51 ans a des douleurs diffuses au niveau cervical droit et à l'épaule droite depuis plus de 9 mois. Ces douleurs sont constantes, mais augmentent surtout en fin de journée. Ce patient prend environ 8 acétaminophènes par jour pour soulager sa douleur.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)
15. une douleur d'origine facettaire	une augmentation de douleur cervicale en extension et rotation cervicale vers la droite	-2 -1 0 1 2
16. une sténose foraminale droite	des tests de mise en tension neuroméningés (ULNT) qui sont négatifs	-2 -1 0 1 2
17. une douleur référée du foie	la prise quotidienne d'alcool (2 à 3 verres) depuis quelques années	-2 -1 0 1 2

*

<p>-2 : éliminée ou presque; -1 : moins probable; 0 : ni plus, ni moins probable; 1 : plus probable; 2 : certaine ou presque</p>

Vignette 6

Votre patiente de 68 ans a des douleurs diffuses et des raideurs au niveau de la main droite lorsqu'elle fait de la couture surtout depuis 3 mois.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)
18. une consultation pour une orthèse	des engourdissements de la main qui surviennent la nuit	-2 -1 0 1 2
19. des mobilisations spécifiques des articulations interphalangiennes	des déformations en col de cygne	-2 -1 0 1 2
20. des exercices d'étirements statiques pour augmenter la mobilité des doigts	de l'œdème diffus à toute la main droite	-2 -1 0 1 2

*

<p>-2 : contre-indiqué ou presque; -1 : peu indiqué; 0 : ni plus, ni moins indiqué;</p> <p>1 : plus indiqué; 2 : complètement indiqué ou presque</p>

Vignette 7

Votre patiente de 74 ans a subi une fracture du col du radius à droite il y a 7 semaines. Son plâtre a été enlevé il y a 1 semaine.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)
21. des mobilisations spécifiques au niveau de l'articulation radio carpienne	une diminution importante de la mobilité en flexion et en extension du poignet	-2 -1 0 1 2
22. des exercices d'étirement statique en extension du poignet maintenus pendant 10 secondes	un engourdissement intermittent surtout le matin des 2 ^e et 3 ^e doigts à droite	-2 -1 0 1 2
23. des exercices de renforcement des extenseurs du poignet	une douleur au repos à 4/10	-2 -1 0 1 2

*

-2 : contre-indiqué ou presque; -1 : peu indiqué; 0 : ni plus, ni moins indiqué;

1 : plus indiqué; 2 : complètement indiqué ou presque

Vignette 8

Votre patiente de 46 ans a des douleurs au niveau lombo-sacré gauche depuis 2 ans. Elle a été en arrêt de travail environ 1 an pour ce problème, mais elle est maintenant de retour au travail.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)
24. des exercices de renforcements des extenseurs lombaires	un score élevé de catastrophisation	-2 -1 0 1 2
25. des exercices de renforcements des extenseurs lombaires	un faible sentiment d'efficacité personnelle	-2 -1 0 1 2
26. un programme pour augmenter l'endurance à la marche	la présence de douleur après 10 min de marche	-2 -1 0 1 2

*

<p>-2 : contre-indiqué ou presque; -1 : peu indiqué; 0 : ni plus, ni moins indiqué; 1 : plus indiqué; 2 : complètement indiqué ou presque</p>

Vignette 9

Votre patiente de 66 ans a des douleurs vives au niveau lombo-sacré gauche qui irradient dans le membre inférieur gauche jusqu'au mollet depuis 2 semaines.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)
27. des tractions mécaniques	une augmentation de la douleur lombaire et une diminution dans le mollet pendant la traction	-2 -1 0 1 2
28. des exercices d'extension lombaire	la présence de discarthrose modérée sur la radiographie	-2 -1 0 1 2
29. des exercices d'extension lombaire	la présence de SLR positif à gauche	-2 -1 0 1 2

*

-2 : contre-indiqué ou presque; -1 : peu indiqué; 0 : ni plus, ni moins indiqué;

1 : plus indiqué; 2 : complètement indiqué ou presque

Vignette 10

Votre patiente de 28 ans est une joueuse de tennis de compétition qui a développé une épicondylalgie depuis 6 mois.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)
30. des exercices d'étirements des extenseurs du poignet	des douleurs à 3/10 pendant ces exercices	-2 -1 0 1 2
31. des frictions transverses sur l'origine des extenseurs communs du poignet	que la patiente souffre de diabète de type 1	-2 -1 0 1 2
32. un bracelet épicondylien	que la patiente a reçu une injection de cortisone il y a 2 semaines	-2 -1 0 1 2

*

-2 : contre-indiqué ou presque; -1 : peu indiqué; 0 : ni plus, ni moins indiqué;
1 : plus indiqué; 2 : complètement indiqué ou presque

Vignette 11

Votre patiente âgée de 36 ans, gauchère, a des douleurs diffuses au niveau postérieur gauche du cou à la fin de sa journée de travail à l'ordinateur. Ces douleurs sont apparues de façon graduelle depuis trois semaines.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)				
33. des exercices d'extension axiale	des muscles sous occipitaux rétractés à droite	-2	-1	0	1	2
34. des exercices d'étirement du trapèze gauche	que l'épaule gauche est plus basse que la droite	-2	-1	0	1	2
35. des exercices de renforcements des extenseurs cervicaux	un point gâchette actif au niveau du trapèze gauche	-2	-1	0	1	2

*

-2 : contre-indiqué ou presque; -1 : peu indiqué; 0 : ni plus, ni moins indiqué;

1 : plus indiqué; 2 : complètement indiqué ou presque

Vignette 12

Vous évaluez une adjointe administrative de 36 ans qui a une douleur au mollet droit depuis 10 jours, et ce, à la suite d'un match d'*ultimate frisbee*. Pendant le match, elle pense avoir reçu un coup de pied au niveau du mollet. Elle a eu de la difficulté à terminer le match et n'a pas été en mesure de jouer depuis.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)
36. une déchirure complète du tendon d'Achille	un test de Thompson négatif avec force gastrocnémien à 3- au bilan musculaire	-2 -1 0 1 2
37. un claquage musculaire gastrocnémien	un SLR reproduit dlr de consultation à 70 degrés, pas de douleur avec flexion dorsale de la cheville debout (test genou au mur)	-2 -1 0 1 2
38. une tendinite du tendon Achille	qu'elle joue sur une surface synthétique pour la première fois avec des souliers non appropriés	-2 -1 0 1 2

*

<p>-2 : éliminée ou presque; -1 : moins probable; 0 : ni plus, ni moins probable; 1 : plus probable; 2 : certaine ou presque</p>

Vignette 13

Une femme de 26 ans vous consulte pour des douleurs à l'épaule gauche qui sont apparues pendant son déménagement, sans incident particulier, il y a 1 semaine. La douleur a débuté graduellement pendant qu'elle transportait des boîtes, mais s'est intensifiée ++ le lendemain.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)
39. une bursite sous deltoïdienne	une abduction limitée par la douleur à 60 degrés en actif et passif avec peu de limitation de RE et RI à 0 degré d'abduction	-2 -1 0 1 2
40. une bursite sous deltoïdienne	un patron capsulaire de restriction de mouvement (avec sensation de fin de mouvement vide)	-2 -1 0 1 2
41. une tendinopathie du supra-épineux	des mouvements résistés de l'épaule en position neutre tous forts et douloureux	-2 -1 0 1 2

*

-2 : éliminée ou presque; -1 : moins probable; 0 : ni plus, ni moins probable;

1 : plus probable; 2 : certaine ou presque

Vignette 14

Vous évaluez un planchiste débutant de 34 ans qui vous consulte pour une douleur au genou droit depuis une chute il y a 1 semaine.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)
42. une déchirure du ménisque interne	une perte de 5 degrés de flexion et d'extension sans douleur et un SFM étirement tissus mous	-2 -1 0 1 2
43. une déchirure du ménisque interne	un œdème sévère a/n du genou	-2 -1 0 1 2
44. une déchirure du ménisque interne	une douleur au mouvement résisté isométrique de flexion du genou	-2 -1 0 1 2
45. une déchirure grade 2 du ligament latéral interne	une AA complète en extension sans douleur	-2 -1 0 1 2

*

-2 : éliminée ou presque; -1 : moins probable; 0 : ni plus, ni moins probable;

1 : plus probable; 2 : certaine ou presque

Vignette 15

Vous évaluez un moniteur de rafting de 22 ans pour des douleurs et une sensation d'instabilité au genou G survenu à la suite d'un accident de rafting il y a 10 jours. Le patient mentionne que sa jambe G est restée coincée sous le banc de son embarcation pendant que celle-ci était débalancée par des vagues violentes.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)
46. une déchirure complète du LCP	qu'il est capable de marcher sans aide technique avec boiterie minime sans dérobade	-2 -1 0 1 2
47. une déchirure complète du LCA	la force des extenseurs du genou à 3-/5 sans douleur	-2 -1 0 1 2
48. une déchirure grade 2 du LLE	un test de stress en varus + (augmentation de jeu articulaire et douleur) à 30 degrés de flexion, mais négatif en extension	-2 -1 0 1 2

*

-2 : éliminée ou presque; -1 : moins probable; 0 : ni plus, ni moins probable;

1 : plus probable; 2 : certaine ou presque

Vignette 16

Votre patient de 24 ans a subi une chirurgie de reconstruction du ligament croisé antérieur (approche avec tendon ischio-jambier) au genou droit il y a 2 semaines.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)
49. une mobilisation de cicatrice	une rougeur et une chaleur localisées au pourtour de la cicatrice	-2 -1 0 1 2
50. des exercices de pompage à la cheville pour augmenter la circulation sanguine	une coloration et une température normale du genou et du mollet droits	-2 -1 0 1 2
51. une mise en charge selon tolérance	un œdème modéré à sévère du genou droit	-2 -1 0 1 2

*

-2 : contre-indiqué ou presque; -1 : peu indiqué; 0 : ni plus, ni moins indiqué;

1 : plus indiqué; 2 : complètement indiqué ou presque

Vignette 17

Votre patient de 27 ans, athlète en judo au niveau national, a subi une entorse du genou droit, il y a 2 jours, en pivotant sur son tibia qui est resté fixe. Sa douleur est environ 4/10 au repos et il ne peut faire de mise en charge complète sur son genou.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)
52. revoir son médecin pour une résonance magnétique	un œdème modéré au genou droit	-2 -1 0 1 2
53. une utilisation de béquilles	un <i>quad lag</i> de 5 degrés	-2 -1 0 1 2
54. des exercices d'AA autoassistés pour maintenir la flexion du genou	une flexion active et passive limitée à 125 degrés avec une sensation de fin de mouvement vide	-2 -1 0 1 2

*

-2 : contre-indiqué ou presque; -1 : peu indiqué; 0 : ni plus, ni moins indiqué;

1 : plus indiqué; 2 : complètement indiqué ou presque

Vignette 18

Une étudiante en physiothérapie se plaint de douleur à la face supérieure de l'épaule droite, et ce, après être tombée avec le bras étendu vers l'avant lors d'un match de frisbee il y a 10 jours. La douleur est surtout présente lors de l'élévation du bras.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)
55. une tendinopathie du supra-épineux	que le mouvement résisté de rotation externe est fort et non douloureux à 0, mais fort et douloureux à 90 d'abduction	-2 -1 0 1 2
56. une déchirure du labrum	un arc de douleur de 100 à 120 degrés d'abduction	-2 -1 0 1 2
57. une atteinte de la capsule postérieure	une douleur aux mouvements actifs en fin de flexion et RI	-2 -1 0 1 2

*

<p>-2 : éliminée ou presque; -1 : moins probable; 0 : ni plus, ni moins probable; 1 : plus probable; 2 : certaine ou presque</p>

Vignette 19

Une patiente de 28 ans est recommandée en physiothérapie pour un syndrome de la bandelette ilio-tibiale droite après avoir augmenté son entraînement de course à pied.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)
58. des étirements de la bandelette ilio-tibiale en position du test d'Ober	une légère bursite trochantérienne à droite	-2 -1 0 1 2
59. des mobilisations médiales soutenues de la rotule 10 x 10 secondes	une douleur à la palpation du rétinaculum latéral	-2 -1 0 1 2
60. un automassage de la bandelette (en distal) avec une balle de tennis	des tests de compression de Noble positif	-2 -1 0 1 2
61. une application de TENS 15 minutes 3x/jour	aucune dlr au repos, mais une douleur qui persiste 30 min après une séance de course à pied (intensité de 2/10)	-2 -1 0 1 2

*

-2 : contre-indiqué ou presque; -1 : peu indiqué; 0 : ni plus, ni moins indiqué;

1 : plus indiqué; 2 : complètement indiqué ou presque

Vignette 20

Après avoir raté une marche il y a 7 jours, une femme de 39 ans est référée en physiothérapie à la suite d'une entorse du ligament latéral interne du genou. Cette blessure lui occasionne douleur et enflure. Elle n'a pas d'antécédent d'entorse au genou.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)				
62. une chaleur au genou 15 minutes 2 à 3X/jour	un œdème léger à modéré a/n du genou	-2	-1	0	1	2
63. des exercices de demi-squat dos au mur (0° à 45° flexion)	une douleur avec mouvement actif d'extension	-2	-1	0	1	2
64. un ultrason a/n du ligament latéral interne	une hyperesthésie cutanée à la partie médiale du genou droit	-2	-1	0	1	2

*

-2 : contre-indiqué ou presque; -1 : peu indiqué; 0 : ni plus, ni moins indiqué;

1 : plus indiqué; 2 : complètement indiqué ou presque

Vignette 21

Vous évaluez un retraité actif de 72 ans qui se plaint de douleur à la face antérieure (D1) et postérieure (D2) de l'épaule droite depuis 3 mois. La douleur a débuté graduellement après avoir entrepris un programme d'exercices incluant des push-ups. Il a consulté en médecine sportive où il a reçu un diagnostic de tendinite du supra-épineux. À l'examen physique, vous notez une faiblesse indolore à 4/5 du biceps à droite.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)
65. un renforcement de la longue portion du biceps avec <i>theraband</i> le bras le long du corps	un test d'accrochage (Hawkins Kennedy) positif	-2 -1 0 1 2
66. des étirements en flexion horizontale	une raideur de la capsule postérieure de l'épaule droite	-2 -1 0 1 2
67. des exercices de renforcement du supra-épineux	que le RX démontre une légère calcification du tendon supra-épineux	-2 -1 0 1 2

*

-2 : contre-indiqué ou presque; -1 : peu indiqué; 0 : ni plus, ni moins indiqué;

1 : plus indiqué; 2 : complètement indiqué ou presque

Vignette 22

Une patiente de 36 ans est recommandée en physiothérapie pour une douleur localisée à la hanche droite, qui a débuté graduellement depuis le début de sa participation à des cours de Zumba, il y a 4 semaines.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)
68. une bursite trochantérienne subaiguë de la hanche	une absence d'œdème à la hanche	-2 -1 0 1 2
69. une bursite trochantérienne subaiguë de la hanche	une douleur avec les mouvements résistés isométriques de la hanche (surtout en RE et abd)	-2 -1 0 1 2
70. une bursite trochantérienne subaiguë de la hanche	que la patiente est capable de dormir plusieurs heures en décubitus latéral droit	-2 -1 0 1 2

*

<p>-2 : éliminée ou presque; -1 : moins probable; 0 : ni plus, ni moins probable; 1 : plus probable; 2 : certaine ou presque</p>

Vignette 23

Vous évaluez un sprinter de 26 ans pour des douleurs intermittentes à la face postéro-médiale du genou gauche, ayant débuté graduellement depuis 6 semaines. La douleur est surtout provoquée par les accélérations, mais il a un inconfort constant même avec le jogging lent, surtout pendant la phase d'attaque.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)
71. une tendinopathie des ischio-jambiers	une force des ischio 5/5 sans douleur	-2 -1 0 1 2
72. une tendinopathie du semi-tendineux	un œdème localisé a/n de la patte d'oie	-2 -1 0 1 2
73. Une tendinopathie des ischio-jambiers	un œdème important a/n du creux poplité	-2 -1 0 1 2

*

<p>-2 : éliminée ou presque; -1 : moins probable; 0 : ni plus, ni moins probable; 1 : plus probable; 2 : certaine ou presque</p>

Vignette 24

Votre patient haltérophile de 24 ans a des douleurs à l'épaule droite depuis près de 6 mois. Il présente plusieurs signes d'accrochage sous-acromial à l'examen clinique, incluant un arc de douleur entre 90 et 110 degrés d'abduction.

Si vous pensiez ...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)
74. que la déficience principale est un manque de contrôle moteur	un décollement du bord interne de l'omoplate lors de l'abduction du bras présent bilatéralement	-2 -1 0 1 2
75. que la déficience principale est une raideur de la capsule postérieure	une diminution importante du glissement postérieur gléno-humérale	-2 -1 0 1 2
76. que la déficience principale est un manque de contrôle moteur	une douleur en abduction résistée à 90 degrés, mais pas de douleur à 0 degré	-2 -1 0 1 2

*

-2 : éliminée ou presque; -1 : moins probable; 0 : ni plus, ni moins probable;

1 : plus probable; 2 : certaine ou presque

Vignette 25

Vous évaluez M. Lafond, 61 ans, pour des douleurs à l'aine droite depuis environ 4 mois. Les douleurs sont apparues graduellement, mais elles ont augmenté depuis 3 semaines et limitent la pratique du ski de fond. Il n'y a pas eu d'événement traumatique. La douleur était initialement localisée à l'aine droite, mais il ressent aussi une douleur à la fesse droite qui peut irradier à l'aspect postérieur jusqu'à mi-cuisse, et ce, depuis le début de la saison de ski de fond (style classique).

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)
77. une arthrose de la hanche	que tous les mouvements isométriques résistés de la hanche sont douloureux	-2 -1 0 1 2
78. une arthrose de la hanche	que le mouvement de flexion de la hanche est à 90/95 avec dlr et SFM vide, autres mouvements légèrement limités	-2 -1 0 1 2
79. une arthrose de la hanche	un SFM en extension et flexion = bloc osseux	-2 -1 0 1 2

*

-2 : éliminée ou presque; -1 : moins probable; 0 : ni plus, ni moins probable;

1 : plus probable; 2 : certaine ou presque

Vignette 26

Une patiente de 48 ans consulte pour des douleurs à la partie antérieure du genou droit qui sont apparues il y a 4 mois. Cette patiente est femme de ménage à temps plein et fait de la marche 30 minutes par jour.

Si vous pensiez à...	et que vous trouvez...	cette nouvelle donnée rend l'hypothèse diagnostique* (encerclez)
80. un syndrome fémoro-rotulien	une force des abducteurs de la hanche droite à 4/5	-2 -1 0 1 2
81. une bursite prépatellaire	un œdème léger avec signe de la vague positif	-2 -1 0 1 2
82. une tendinite du tendon rotulien	une douleur en montant les escaliers	-2 -1 0 1 2

*

-2 : éliminée ou presque; -1 : moins probable; 0 : ni plus, ni moins probable;
1 : plus probable; 2 : certaine ou presque

FIN DES TESTS DE CONCORDANCE DE SCRIPTS

82 items (15 vignettes 48 items DX 11 vignettes et 34 items RX)

Items Dx 1 à 17, 36 à 48, 55 à 57, 68 à 82

Items RX 18 à 35, 49 à 54, 58 à 67

ANNEXE 5 PROBLÈMES DE RAISONNEMENT CLINIQUE

Nom :

Consignes :

PROBLÈME DE RAISONNEMENT CLINIQUE (diagnostic)

Après avoir lu la vignette, vous devez mentionner le diagnostic (structure et processus pathologique) qui vous apparait le plus probable.

Exemple : entorse du ligament croisé antérieur grade 2

Pour **ce diagnostic**, vous devez donner un poids relatif aux caractéristiques cliniques énoncées dans le tableau à l'aide de l'échelle suivante :

+3 = supporte fortement +2 = supporte modérément +1 = supporte légèrement 0 = n'influence pas
-3 = questionne fortemen -2 = questionne modérément -1 = questionne légèrement

Caractéristiques cliniques **Poids relatifs**

Âge de la patiente	0
Mécanisme de blessure	+2

Si votre **diagnostic se révèle faux**, quel serait le 2^e diagnostic (structure et processus pathologique) le plus probable dans ce cas?

Exemple : déchirure méniscale

En réutilisant l'échelle ci-dessus, vous devez par la suite redonner un poids relatif pour chacune des caractéristiques en fonction du nouveau diagnostic.

Caractéristiques cliniques **Poids relatifs**

Âge de la patiente	+1
Mécanisme de blessure	-1

Consignes :

PROBLÈME DE RAISONNEMENT CLINIQUE (traitement)

Après avoir lu la vignette, vous devez mentionner quel serait l'exercice le plus pertinent que vous pourriez donner (nommez l'articulation ou le groupe musculaire ciblé et les paramètres de l'exercice).

Exemple : renforcement excentrique du biceps (3 séries de 10 répétitions avec poids selon tolérance)

En utilisant l'échelle ci-dessous, déterminez le poids relatif pour chacune des caractéristiques cliniques qui ont influencé votre jugement afin de déterminer l'exercice le plus pertinent.

+3 = supporte fortement +2 = supporte modérément +1 = supporte légèrement 0 = n'influence pas
-3 = questionne fortemen -2 = questionne modérément -1 = questionne légèrement

Caractéristiques cliniques Poids relatifs

La chronicité des symptômes	+3
Le type de sport	+1

Basé sur les informations du cas, quel serait le 2^e exercice le plus pertinent que vous pourriez donner (nommez l'articulation ou le groupe musculaire ciblé et les paramètres de l'exercice)?

Exemple : étirement du biceps (3 x 30 secondes)

En réutilisant l'échelle ci-dessus, déterminez le poids relatif pour chacune des caractéristiques cliniques qui ont influencé votre jugement afin de déterminer le traitement le plus pertinent.

Caractéristiques cliniques Poids relatifs

La chronicité des symptômes	+2
Le type de sport	-1

Nom :

Vignette 1

Votre patient de 68 ans a des douleurs intermittentes (**D1**) et diffuses au niveau postérieur des cuisses et des mollets depuis environ 8 mois. La douleur s'est graduellement installée sans circonstances particulières. Il décrit ses douleurs aux membres inférieurs principalement comme une lourdeur. Aussi, il présente une douleur lombaire droite intermittente (**D2**) qui est stable depuis environ 7 ans. Ses conditions médicales incluent une athérosclérose modérée et une insuffisance rénale contrôlée médicalement. Le patient est fumeur (1 paquet par jour) depuis l'adolescence.

De plus, il est limité dans ses activités de la vie quotidienne qui impliquent de la marche (difficulté à faire plus de 15 minutes) depuis l'apparition de douleurs aux membres inférieurs. Il note une tendance à avoir de l'œdème aux chevilles en fin de journée et des pertes d'équilibre occasionnelles.

Observations générales :

Ambulation : sans particularité

Observations spécifiques :

Posture : pas de déviation du tronc, pas de spasme et pas d'atrophie musculaire visible

ROM debout	Actif
Flexion	40 cm doigts-sol Pas de douleur
Extension	Légère dlr D2 en fin AA Pas de limitation d'AA
Flexion latérale G et D	Légère dlr D2 en fin AA Pas de limitation d'AA
Rotation D et G	Légère dlr D2 en fin AA Pas de limitation d'AA

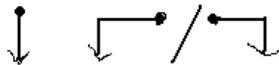
Examen neurologique :

- Dermatomes : bonne discrimination sensitive de L1 à S2
- ROT :
 - Patellaire : vifs et symétriques
 - Achilléen : non élicité à gauche, diminué à droite
 - Extenseurs orteils : vifs et symétriques
- Myotomes : L1-S2 : pas de faiblesse ou de fatigabilité notée sauf pour S1 à gauche
- Clonus : absent
- Cutané plantaire : négatif

Tests spéciaux :

- SLR D et G : pas de douleur de consultation reproduite

Mouvements accessoires :



L4-L5 = reproduit dlr typique au dos D2 et hypomobilité modérée à ces niveaux.

Palpation : pas de douleur notée à la région lombosacrée et au niveau postérieur des membres inférieurs.

Pouls : pédieux et fémoral diminués bilatéralement

Basé sur les informations du cas, quel est le diagnostic (structure et processus pathologique) le plus probable pour **D1** (douleurs intermittentes et diffuses au niveau postérieur des cuisses et des mollets)?

Réponse : radiculopathie S1 Gauche

En utilisant l'échelle suivante, déterminez le poids relatif pour chacune des caractéristiques cliniques qui ont influencé votre jugement quant au diagnostic.

+3 = supporte fortement +2 = supporte modérément +1 = supporte légèrement 0 = n'influence pas
 -3 = questionne fortement -2 = questionne modérément -1 = questionne légèrement

Caractéristiques cliniques

Poids relatifs

Âge du patient	
Type de douleur (D1 lourdeur aux membres inférieurs)	
Présence de douleur lombaire D2	
Apparition graduelle	
Athérosclérose modérée	
Fumeur	
Cedème aux chevilles	
Pas de limitations de mouvements lombaires	
Douleur lombaire en extension	
Faiblesse myotome S1 à gauche	
Hypomobilité à L4-L5	
Pouls diminué aux membres inférieurs	
Perte d'équilibre	
SLR négatif bilatéralement	

Quel serait le 2^e diagnostic (structure et processus pathologique) le plus probable dans ce cas pour **D1** (douleurs intermittentes et diffuses au niveau postérieur des cuisses et des mollets)?

Réponse : sciatalgie d'origine extra-segmentaire (syndrome du piriforme)

En réutilisant l'échelle ci-dessous, déterminez le poids relatif pour chacune des caractéristiques cliniques qui ont influencé votre jugement quant au 2^e diagnostic le plus probable.

+3 = supporte fortement +2 = supporte modérément +1 = supporte légèrement 0 = n'influence pas
 -3 = questionne fortement -2 = questionne modérément -1 = questionne légèrement

Caractéristiques cliniques

Poids relatifs

Âge du patient	
Type de douleur (D1 lourdeur aux membres inférieurs)	
Présence de douleur lombaire D2	
Apparition graduelle	
Athérosclérose modérée	
Fumeur	
Œdème aux chevilles	
Pas de limitations de mouvements lombaires	
Douleur lombaire en extension	
Faiblesse myotome S1 à gauche	
Hypomobilité à L4-L5	
Pouls diminué aux membres inférieurs	
Perte d'équilibre	
SLR négatif bilatéralement	

Vignette 2

Votre patiente de 58 ans a des douleurs vives à 5/10 au repos à la partie antérieure de l'épaule droite depuis 3 semaines. La douleur a commencé après avoir fait le grand ménage du printemps. Elle a un antécédent de douleur à l'épaule droite à la suite d'une déchirure partielle de la coiffe des rotateurs il y a 4 ans. Cet épisode avait laissé de légères douleurs résiduelles. La patiente est droitrière et très limitée dans ses activités (AVQ). Elle ne peut pas se coucher du côté droit. Ses conditions médicales incluent un diabète de type 1 contrôlé par insuline et de l'insuffisance cardiaque modérée.

Observations générales :

- Rythme scapulo-huméral inversé lors de l'élévation
- Légère atrophie deltoïde droit

Amplitude articulaire et bilan musculaire

Articulation	ROM	BILAN ARTICULAIRE (A/P)			RÉSISTÉ
		Gauche (A/P)	Droite (A/P) SFM		Droite
Épaule	Flexion	165°/170°	70°/90°	vide	Faible et douloureux
	Extension	40°/45°	30°/35°	capsulaire	Fort et douloureux
	Abduction	165°/170°	70°/90°	vide	Faible et douloureux
	Adduction	Normal	Normal	capsulaire	Fort et douloureux
	R.E. (à 45° abd)	80°/85°	20°/35°	vide	Faible et douloureux
	R.I. (à 45° abd)	70°/75°	35°/40°	vide	Faible et douloureux

Articulations adjacentes :

ROM cervical : normal et sans douleur

ROM coude droit : normal et sans douleur

Mouvements accessoires gléno-huméraux :

Le glissement inférieur soulage légèrement la douleur (de 5/10 à 3/10) et les autres mouvements sont non évalués, car trop douloureux.

Évaluation neurologique :

- Dermatome (C4-T1) : normal
- Réflexes ostéotendineux (C6, C7) : normal
- Myotome (C4-T1) : N/E

Palpation : à droite, douleur au pourtour de l'acromion

Basé sur les informations du cas, quel est le diagnostic (structure et processus pathologique) le plus probable?

Réponse : tendinopathie de la coiffe des rotateurs

En utilisant l'échelle ci-dessous, déterminez le poids relatif pour chacune des caractéristiques cliniques qui ont influencé votre jugement quant au diagnostic.

+3 = supporte fortement +2 = supporte modérément +1 = supporte légèrement 0 = n'influence pas
 -3 = questionne fortement -2 = questionne modérément -1 = questionne légèrement

Caractéristiques cliniques

Poids relatifs

Âge de la patiente	
Intensité de la douleur	
Apparition à la suite du grand ménage	
Antécédent de déchirure partielle de la coiffe des rotateurs	
Patiente droitrière	
Diabète de type 1	
Insuffisance cardiaque modérée	
Rythme scapulo-huméral inversé lors de l'élévation	
Légère atrophie deltoïde droit	
Patron de limitations des mouvements actifs et passifs	
Les sensations de fin de mouvement (SFM)	
Patron de réponses aux mouvements isométriques résistés	
Glissement inférieur soulage légèrement la douleur	
Douleur à droite à la palpation au pourtour de l'acromion	

Quel serait le 2^e diagnostic (structure et processus pathologique) le plus probable dans ce cas?

Réponse : capsulite G-H

En réutilisant l'échelle ci-dessous, déterminez le poids relatif pour chacune des caractéristiques cliniques qui ont influencé votre jugement quant au 2^e diagnostic le plus probable.

+3 = supporte fortement +2 = supporte modérément +1 = supporte légèrement 0 = n'influence pas
 -3 = questionne fortement -2 = questionne modérément -1 = questionne légèrement

Caractéristiques cliniques

Poids relatifs

Âge de la patiente	
Intensité de la douleur	
Apparition à la suite du grand ménage	
Antécédent de déchirure partielle de la coiffe des rotateurs	
Patiente droitrière	
Diabète de type 1	
Insuffisance cardiaque modérée	
Rythme scapulo-huméral inversé lors de l'élévation	
Légère atrophie deltoïde droit	
Patron de limitations des mouvements actifs et passifs	
Les sensations de fin de mouvement (SFM)	
Patron de réponse aux mouvements isométriques résistés	
Glissement inférieur soulage légèrement la douleur	
Douleur à droite à la palpation au pourtour de l'acromion	

Vignette 3

Votre patient, âgé de 74 ans, droitier, a subi une fracture de la partie distale du radius à droite il y a 7 semaines. Depuis sa fracture, il présente des engourdissements au niveau des 2^e et 3^e doigts de la main droite surtout la nuit. Il y a une semaine, son plâtre a été enlevé et, selon la radiographie, la fracture était bien consolidée. Ce patient souffre d'arthrite rhumatoïde, mais n'est pas dans une phase active. Il consulte pour avoir un programme d'exercices à la maison.

Observations spécifiques :

Cedème résiduel localisé à la base du pouce. Pas de chaleur ni de rougeur au niveau du poignet droit

		BILAN ARTICULAIRE			FORCE	
Articulation poignet	ROM	Droit (A/P)	(SFM)	Gauche (A/P)	Droit	Gauche
	Flexion	50/55 dlr légère	Étirement des tissus mous	60/65	4+/5	5/5
	Extension	40/42 dlr légère	Étirement des tissus mous	60/65	4/5	5/5
	Déviation ulnaire	23/30 dlr légère	Étirement des tissus mous	20/25	4+/5	5/5
	Déviation radiale	5/10 dlr légère	Os à os	10/15	4+/5	5/5
	Pronation	50/55 dlr légère	Étirement des tissus mous	70/75	4+/5	5/5
	Supination	70/75 dlr légère	Étirement des tissus mous	75/80	4+/5	5/5
	Préhension Jamar				12 kg avec dlr	25 kg

Articulations adjacentes :

ROM coude droit : normal et sans douleur

Test	Résultat
Phalen	Positif : provoque engourdissements

Palpation : légère douleur diffuse au niveau de tous les os du carpe

Basé sur les informations du cas, quel serait l'exercice le plus pertinent que vous pourriez donner (nommez l'articulation ou le groupe musculaire ciblé et les paramètres de l'exercice)?

Réponse : étirement des fléchisseurs du poignet, 3*20 sec, intensité faible

En utilisant l'échelle ci-dessous, déterminez le poids relatif pour chacune des caractéristiques cliniques qui ont influencé votre jugement afin de déterminer l'exercice le plus pertinent.

+3 = supporte fortement +2 = supporte modérément +1 = supporte légèrement 0 = n'influence pas
 -3 = questionne fortement -2 = questionne modérément -1 = questionne légèrement

Caractéristiques cliniques

Poids relatifs

Âge du patient	
Plâtre enlevé il y a une semaine	
Diagnostic d'arthrite rhumatoïde	
Engourdissements au niveau des 2 ^e et 3 ^e doigts surtout la nuit	
Cedème résiduel localisé à la base du pouce	
Pas de chaleur ni de rougeur localement	
Force de préhension limitée à droite	
Patron de limitation des mouvements actifs et passifs du poignet	
Les sensations de fin de mouvements (SFM) du poignet D	
La force au poignet droit	
Test de Phalen : positif	

Quel serait le 2^e exercice le plus pertinent que vous pourriez donner (nommez l'articulation ou le groupe musculaire ciblé et les paramètres de l'exercice)?

Réponse : circumduction du poignet, 3*10

En réutilisant l'échelle ci-dessous, déterminez le poids relatif pour chacune des caractéristiques cliniques qui ont influencé votre jugement afin de déterminer le traitement le plus pertinent.

+3 = supporte fortement +2 = supporte modérément +1 = supporte légèrement 0 = n'influence pas
 -3 = questionne fortement -2 = questionne modérément -1 = questionne légèrement

Caractéristiques cliniques

Poids relatifs

Âge du patient	
Plâtre enlevé il y a une semaine	
Diagnostic d'arthrite rhumatoïde	
Engourdissements au niveau des 2 ^e et 3 ^e doigts surtout la nuit	
Cedème résiduel localisé à la base du pouce	
Pas de chaleur ni de rougeur localement	
Force de préhension limitée à droite	
Patron de limitation des mouvements actifs et passifs du poignet	
Les sensations de fin de mouvements (SFM) du poignet D	
La force au poignet droit	
Test de Phalen : positif	

Vignette 4

Vous évaluez un policier de 40 ans, à vélo, pour une douleur à la face antérieure de la hanche droite qui est apparue graduellement depuis 1 an et demi. Monsieur est un ancien joueur de football universitaire et a déjà subi plusieurs blessures aux MI's (entorses de chevilles bilatérales et déchirure du LCA droit).

Depuis 2 mois, vers la fin de sa saison de touch-football, la douleur à sa hanche s'est accentuée. Celle-ci est parfois présente au repos et pire en fin de journée ou selon les activités. La tolérance debout et en position assise est diminuée. Monsieur a de plus en plus de difficultés à faire son travail. Il ressent un peu de soulagement avec le vélo avec modération. Il y a présence de raideur matinale pendant 30 minutes qui s'améliore avec le mouvement.

Posture : augmentation lordose (debout), antériorisation de la tête fémorale droite (décubitus dorsal)

Articulation	ROM	AMPLITUDE ARTICULAIRE		SFM	Droite
		Gauche (A/P)	Droite (A/P)		
Hanche	Flexion	110/115	90/95 dlr	Capsulaire raide ++	dlr
	Extension	10/15	0/5	Capsulaire raide+	indolore
	Abduction	30/35	25/30	Capsulaire raide	Indolore
	Adduction	20/25	15/20	Capsulaire raide	Indolore
	Rotation externe	35/40	25/30	Capsulaire raide	Indolore
	Rotation interne	20/25	5/12	Capsulaire raide+	indolore

Mouvements combinés hanche droite :

Flexion/adduction : dlr, raide ++

Ext/add/RI: dlr, raide ++

FABERE : 1/3 dlr

Mouvements accessoires : diminution glissement postérieur coxo-fémorale droite

Autres mouvements : N/E

Scan lombaire : sans particularité

Flexibilité :

→ Test Thomas : manque souplesse fléchisseurs hanche droite > gauche

→ Test 90/90 SLR : manque souplesse ischios droit = gauche (-25 degrés bilat.)

Palpation : tension et inconfort palpation du ventre musculaire et du tendon iliopsoas droit, atrophie grand fessier droit

Basé sur les informations du cas, quel est le diagnostic (structure et processus pathologique) le plus probable?

Réponse : arthrose C-F

En utilisant l'échelle ci-dessous, déterminez le poids relatif pour chacune des caractéristiques cliniques qui ont influencé votre jugement quant au diagnostic.

+3 = supporte fortement +2 = supporte modérément +1 = supporte légèrement 0 = n'influence pas
 -3 = questionne fortement -2 = questionne modérément -1 = questionne légèrement

Caractéristiques cliniques

Poids relatifs

Âge du patient	
Apparition graduelle de la douleur depuis 1 an et demi	
Douleur localisée face antérieure hanche	
Antécédent de multiples blessures aux membres inférieurs	
Raideur matinale x 30 minutes mieux avec mouvement	
Antériorisation tête fémorale	
Diminution de flexion A/P 90°/95° avec douleur hanche droite	
SFM capsulaire raide toutes directions	
Hypomobilité et dlr à tous les mouvements combinés	
Douleur au mouvement résisté de flexion	
Diminution de glissement postérieur coxo-fémoral droite	
Thomas test positif pour fléchisseur hanche D	
Atrophie grand fessier	
Tension et inconfort palpation iliopsoas droit	

Quel serait le 2^e diagnostic (structure et processus pathologique) le plus probable dans ce cas?

Réponse : tendinopathie ilio-psoas

En réutilisant l'échelle ci-dessous, déterminez le poids relatif pour chacune des caractéristiques cliniques qui ont influencé votre jugement quant au 2^e diagnostic le plus probable.

+3 = supporte fortement +2 = supporte modérément +1 = supporte légèrement 0 = n'influence pas
 -3 = questionne fortement -2 = questionne modérément -1 = questionne légèrement

Caractéristiques cliniques

Poids relatifs

Âge du patient	
Apparition graduelle de la douleur depuis 1 an et demi	
Douleur localisée face antérieure hanche	
Antécédent de multiples blessures aux membres inférieurs	
Raideur matinale x 30 minutes mieux avec mouvement	
Antériorisation tête fémorale	
Diminution de flexion A/P 90°/95° avec douleur hanche droite	
SFM capsulaire raide toutes directions	
Hypomobilité et dlr à tous les mouvements combinés	
Douleur au mouvement résisté de flexion	
Diminution de glissement postérieur coxo-fémoral droite	
Thomas test positif pour fléchisseur hanche D	
Atrophie grand fessier	
Tension et inconfort palpation iliopsoas droit	

Vignette 5

M. Duguay, 43 ans, a de la douleur diffuse dans le coude droit (**D1**) depuis environ 1 an (début graduel). Ses douleurs ont augmenté depuis 3 mois (a fait beaucoup d'heures supplémentaires en tant qu'électricien). Il mentionne avoir la main droite froide et des engourdissements quand il travaille. Il a aussi de la douleur cervicale droite (**D2**). La douleur ne perturbe pas son sommeil.

Résultats de l'évaluation objective

Observations : port de tête antérieur, enroulement des épaules, proéminence cervicothoracique, respiration apicale

Bilan articulaire cervical		Bilan articulaire coude
Flexion	3/4 tire à droite	Tous les mouvements sont normaux et indolores
Extension	3/4 D2	
Flexion latérale d	3/4 D2	
Flexion latérale g	Complet tire à droite	
Rotation d	3/4 D2	
Rotation g	Complète	
Flexion CV	Tire à droite en sous-occipital	
Extension CV	Complet	

Traction (neutre) : légère diminution de la douleur cervicale (diminue de 2 à 1/10)

Évaluation de dépistage à l'épaule : limite de la normale

Évaluation neurologique :

- Dermatome (C4-T1) : normal
- Réflexes ostéotendineux (C5, C7) : normal et symétrique
- Myotome (C4-T1) : fort et symétrique

Mobilité accessoire : douleur et raideur sur les piliers articulaires de C2-C3 et C5-C6 à droite

Palpation : douleur au niveau des scalènes bilatéralement

Préhension : Jamar – à droite : 20 kg (**D1**), à gauche : 35 kg

Tests spéciaux : extension résistée du poignet à droite reproduit **D1**

ULNT 1(G) : atteint 0 degré extension du coude sans symptômes

ULNT 1(D) : -20 degrés extension coude; tire au niveau antérieur de l'avant-bras

Pouls radial : symétrique

Palpation : dlr (**D1**) a/n de l'épicondyle latéral

Basé sur les informations du cas, quel est le diagnostic (structure et processus pathologique) le plus probable pour **D1** (la douleur au coude)?

Réponse : épicondylite chronique latérale

En utilisant l'échelle ci-dessous, déterminez le poids relatif pour chacune des caractéristiques cliniques qui ont influencé votre jugement quant au diagnostic.

+3 = supporte fortement +2 = supporte modérément +1 = supporte légèrement 0 = n'influence pas
 -3 = questionne fortement -2 = questionne modérément -1 = questionne légèrement

Caractéristiques cliniques

Poids relatifs

Âge du patient	
Main froide	
Engourdissements de la main	
Apparition graduelle	
Type de travail (électricien)	
Posture (port de tête antérieur, enroulement des épaules, proéminence cervicothoracique)	
Respiration apicale	
Effet de la traction cervicale	
Patron de mouvements cervical	
Évaluation neurologique normale	
<u>Mobilité accessoire</u> : (douleur et raideur sur les piliers articulaires de C2-C3 et C5-C6)	
Faiblesse de la préhension à droite	
Pouls radial symétrique	
ULNT 1 : limité à droite	
Douleur à la palpation de l'épicondyle latéral	

Quel serait le 2^e diagnostic (structure et processus pathologique) le plus probable dans ce cas pour **D1** (la douleur au coude)?

Réponse : syndrome du tunnel radial

En réutilisant l'échelle ci-dessous, déterminez le poids relatif pour chacune des caractéristiques cliniques qui ont influencé votre jugement quant au 2^e diagnostic le plus probable.

+3 = supporte fortement +2 = supporte modérément +1 = supporte légèrement 0 = n'influence pas
 -3 = questionne fortement -2 = questionne modérément -1 = questionne légèrement

Caractéristiques cliniques

Poids relatifs

Âge du patient	
Main froide	
Engourdissements de la main	
Apparition graduelle	
Type de travail (électricien)	
Posture (Port de tête antérieur, enroulement des épaules, proéminence cervicothoracique)	
Respiration apicale	
Effet de la traction cervicale	
Patron de mouvements cervical	
Évaluation neurologique normale	
<u>Mobilité accessoire</u> : (douleur et raideur sur les piliers articulaires de C2-C3 et C5-C6)	
Faiblesse de la préhension à droite	
Pouls radial symétrique	
ULNT 1 : limité à droite	
Douleur à la palpation de l'épicondyle latéral	

Vignette 6

Vous évaluez un professeur d'éducation physique de 31 ans. Il a subi une blessure au genou gauche, il y a 3 jours, lors d'un match amical de hockey cosom.

Alors qu'il voulait effectuer un pivot, son pied a glissé et il a senti une torsion ainsi qu'une vive douleur au genou. Il est retourné au banc en boitant, n'a pas terminé la partie et a appliqué de la glace une fois à la maison. Son genou a enflé graduellement au fil de la soirée. Il ne peut marcher sans boiterie à cause de la douleur, mais n'utilise pas d'aide technique. La douleur, d'une intensité moyenne de 5/10, est localisée à la face interne du genou; elle est pire avec la MEC et en fin d'amplitude articulaire. Il a de la difficulté à déplier le genou complètement. Il a un inconfort constant au genou qui augmente au fil de la journée. Le sommeil est peu affecté, mais le genou est plus raide au réveil. Monsieur n'est pas allé travailler depuis sa blessure et est limité dans ses déplacements.

Observation : genou positionné à environ 15° de flexion, œdème léger à modéré

Bilan articulaire genou gauche (A/P) :

Flexion : 120° raideur et légère dlr/125° SFM rebond avec douleur

Extension : -10° douleur/-8° SFM vide

Résisté isométrique sous maximal (effectués à 30 degrés de flexion) :

Extension : difficulté à initier contraction du quad

Flexion : contraction douloureuse

Tests ligamentaires genou gauche :

Lachman : négatif

Tiroir postérieur et sag test : négatif

Stress test valgus à 10 degrés et 30 degrés flexion : pas de laxité notée, mais douleur avec les deux positions

Stress test varus à 10 et 30 degrés : négatif

Tests spécifiques genou gauche :

McMurray : douleur lors du test en RE du tibia vers 115°

Test de vague : positif (petite vague)

Palpation genou gauche : douleur à la palpation interligne articulaire médiale, pas de douleur à la palpation du condyle fémoral interne et portion médiale de la tubérosité du tibia.

Basé sur les informations du cas, quel est le diagnostic (structure et processus pathologique) le plus probable?

Réponse : déchirure méniscale médiale

En utilisant l'échelle ci-dessous, déterminez le poids relatif pour chacune des caractéristiques cliniques qui ont influencé votre jugement quant au diagnostic.

+3 = supporte fortement +2 = supporte modérément +1 = supporte légèrement 0 = n'influence pas
 -3 = questionne fortement -2 = questionne modérément -1 = questionne légèrement

Caractéristiques cliniques

Poids relatifs

Âge du patient	
Mécanisme de blessure en torsion	
Enflure graduelle en soirée	
Localisation de la douleur face interne du genou	
Douleur augmente à la MEC	
Cedème léger à modéré	
Flexion A/P : 120/125 dlr avec SFM rebond	
Extension A/P : -10/-8 avec FSM vide	
Douleur avec contraction ischio-jambiers	
Douleur sans laxité aux stress tests en valgus à -10° et 30° flexion	
McMurray douloureux en RE à 115° flexion	
Test de la vague positif	
Douleur à la palpation interligne médiale	
Pas de douleur palpation condyle fémoral interne et portion interne de la tubérosité tibiale	

Quel serait le 2^e diagnostic (structure et processus pathologique) le plus probable dans ce cas?

Réponse : entorse grade 1 LCM

En réutilisant l'échelle ci-dessous, déterminez le poids relatif pour chacune des caractéristiques cliniques qui ont influencé votre jugement quant au 2^e diagnostic le plus probable.

+3 = supporte fortement +2 = supporte modérément +1 = supporte légèrement 0 = n'influence pas
 -3 = questionne fortement -2 = questionne modérément -1 = questionne légèrement

Caractéristiques cliniques

Poids relatifs

Âge du patient	
Mécanisme de blessure en torsion	
Enflure graduelle en soirée	
Localisation de la douleur face interne du genou	
Douleur augmente à la MEC	
Cedème léger à modéré	
Flexion A/P : 120/125 dlr avec SFM rebond	
Extension A/P : -10/-8 avec FSM vide	
Douleur avec contraction ischio-jambiers	
Douleur sans laxité aux stress tests en valgus à -10° et 30° flexion	
McMurray douloureux en RE à 115° flexion	
Test de la vague positif	
Douleur à la palpation interligne médiale	
Pas de douleur palpation condyle fémoral interne et portion interne de la tubérosité tibiale	

Vignette 7

Votre patiente de 56 ans a des douleurs diffuses au niveau postérieur du cou à gauche en fin de journée de travail (secrétariat) à l'ordinateur. Ces douleurs sont apparues de façon graduelle, depuis 3 mois, sans circonstances particulières. La patiente est gauchère et a une bonne santé générale.

Observations spécifiques :

Posture : port de tête antérieur, proéminence cervicothoracique. Épaule droite plus basse que la gauche.

Mobilité articulaire

Mouvements	Bilan articulaire actif cervical
Flexion	3/4 tire à gauche
Extension	1/2 D1
Flexion latérale d	3/4 D1
Flexion latérale g	Complète
Rotation d	3/4 D1
Rotation g	Complet D1
Flexion CV	Tire bilat. en sous-occipital 1\2 AA
Extension CV	Complet

Articulations adjacentes : ROM épaules présente -10° en flexion antérieure et abduction.

Flexibilité musculaire : rétraction au niveau des sous-occipitaux bilatéralement

Mouvements accessoires : douleur (D1) sur les piliers articulaires de C2-C3, C3-C4 bilatéralement

Évaluation neurologique : normal

Tests spéciaux :

Upper limb neurodynamic test (1) : (-) bilatéralement

Palpation : myalgie bilatérale trapèze supérieur; induration à gauche (point douloureux, pas de douleur irradiée) également tissus musculaires indurés en sous-occipital bilatéralement.

Basé sur les informations du cas, quel serait l'exercice le plus pertinent que vous pourriez donner (nommez l'articulation ou le groupe musculaire ciblé et les paramètres de l'exercice)?

Réponse : extension axiale (rentrez le menton), 10*5 sec, 6 fois par jour

En utilisant l'échelle ci-dessous, déterminez le poids relatif pour chacune des caractéristiques cliniques qui ont influencé votre jugement afin de déterminer l'exercice le plus pertinent.

+3 = supporte fortement +2 = supporte modérément +1 = supporte légèrement 0 = n'influence pas
 -3 = questionne fortement -2 = questionne modérément -1 = questionne légèrement

Caractéristiques cliniques

Poids relatifs

Patiente gauchère	
Type de travail (secrétariat à l'ordinateur)	
Posture : port de tête antérieur	
Posture : proéminence cervicothoracique	
Posture : épaule droite plus basse que gauche	
ROM épaules présente -10° en flexion antérieure et abduction	
Flexibilité musculaire : rétraction au niveau des sous-occipitaux bilatéralement	
Mouvements accessoires : douleur (D1) sur les piliers articulaires de C2-C3, C3-C4 bilatéralement	
Upper limb neurodynamic test : (-) bilatéralement	
Palpation : myalgie trapèze bilatérale; induration à gauche	
Palpation : tissus musculaires indurés en sous-occipital bilatéralement	

Quel serait le 2^e exercice le plus pertinent que vous pourriez donner (nommez l'articulation ou le groupe musculaire ciblé et les paramètres de l'exercice)?

Réponse : étirement trapèze sup., 3*30 sec

En réutilisant l'échelle ci-dessous, déterminez le poids relatif pour chacune des caractéristiques cliniques qui ont influencé votre jugement afin de déterminer le traitement le plus pertinent.

+3 = supporte fortement +2 = supporte modérément +1 = supporte légèrement 0 = n'influence pas
 -3 = questionne fortement -2 = questionne modérément -1 = questionne légèrement

Caractéristiques cliniques

Poids relatifs

Patiente gauchère	
Type de travail (secrétariat à l'ordinateur)	
Posture : port de tête antérieur	
Posture : proéminence cervicothoracique	
Posture : épaule droite plus basse que gauche	
ROM épaules présente 10° en flexion antérieure et abduction	
Flexibilité musculaire : rétraction au niveau des sous-occipitaux bilatéralement	
Mouvements accessoires : douleur (D1) sur les piliers articulaires de C2-C3, C3-C4 bilatéralement	
Upper limb neurodynamic test : (-) bilatéralement	
Palpation : myalgie trapèze bilatérale; induration à gauche	
Palpation : tissus musculaires indurés en sous-occipital bilatéralement	

Vignette 8

Vous évaluez Mme Desjardins, 36 ans, coiffeuse, pour une douleur à la hanche droite débutant graduellement depuis le début de sa participation à des cours de Zumba il y a 4 semaines.

La douleur est localisée à la face latérale de la hanche et elle est de type élancement. L'intensité varie entre 0/10 (au repos) et 4/10. Elle est augmentée par l'effort physique comme le Zumba, la marche ou la station debout prolongée.

Elle continue d'aller à ses cours de Zumba, mais commence à être limitée dans son intensité.
Médication : Tyféno

Observation : augmentation de l'angle Q droit > gauche avec genu valgus bilatéral, léger surpoids

Articulation	ROM	AMPLITUDE ARTICULAIRE		Force	
		Gauche (A/P)	Droite (A/P)	Gauche	Droite
Hanche	Flexion	115°/120°	115°/120° dlr	5	5
	Extension	15°/20°	15°/20°	4+	4+
	Abduction	40°/45°	40°/45°	4+	4
	Adduction	25°/30°	25°/30° étirement sensible	5	5
	Rotation externe	40°/45°	40°/45°	5	5
	Rotation interne	35°/40°	35°/40°	5	5
Genou	Flexion	Complète	Complète	5	5
	Extension	0°/+10°	0°/+10°	5	5

Scan lombaire : sans particularité

Souplesse :

Ober : diminution de souplesse de la bandelette ilio-tibial droite > gauche

Tests fonctionnels :

Step-down : D : ↑ modérée add. et RI fémur, ↑dlr+; G : légère augmentation d'add. et RI fémur

Palpation : légère sensibilité dans la région du grand trochanter droit, pas d'œdème noté.

Patron de marche : léger Trendelenburg

Basé sur les informations du cas, quel serait l'exercice le plus pertinent que vous pourriez donner (nommez l'articulation ou le groupe musculaire ciblé et les paramètres de l'exercice)?

Réponse : squat unipodal (correction alignement)

En utilisant l'échelle ci-dessous, déterminez le poids relatif pour chacune des caractéristiques cliniques qui ont influencé votre jugement afin de déterminer le traitement le plus pertinent.

+3 = supporte fortement +2 = supporte modérément +1 = supporte légèrement 0 = n'influence pas
 -3 = questionne fortement -2 = questionne modérément -1 = questionne légèrement

Caractéristiques cliniques

Poids relatifs

Apparition de la douleur coïncide avec début cours de Zumba	
Résultat du bilan articulaire	
Augmentation de l'angle Q (D > G)	
Présence de genu valgus bilatéral	
Résultat du bilan musculaire	
Diminution de souplesse bandelette ilio-tibial droite > gauche	
Résultat du step down test	
Léger Trendelenburg à la marche	

Quel serait le deuxième exercice le plus pertinent que vous pourriez donner (nommez l'articulation ou le groupe musculaire ciblé et les paramètres de l'exercice)?

Réponse : étirement BIT 3*30 sec

En réutilisant l'échelle ci-dessous, déterminez le poids relatif pour chacune des caractéristiques cliniques qui ont influencé votre jugement afin déterminer le traitement le plus pertinent.

+3 = supporte fortement +2 = supporte modérément +1 = supporte légèrement 0 = n'influence pas
-3 = questionne fortement -2 = questionne modérément -1 = questionne légèrement

Caractéristiques cliniques

Poids relatifs

Apparition de la douleur coïncide avec début cours de Zumba	
Résultat du bilan articulaire	
Augmentation de l'angle Q (D>G)	
Présence de genu valgus bilatéral	
Résultat du bilan musculaire	
Diminution de souplesse bandelette ilio-tibial droite > gauche	
Résultat du step down test	
Léger Trendelenburg à la marche	

FIN DES PROBLÈMES DE RAISONNEMENT CLINIQUE

ANNEXE 6 TABLEAU DE SPÉCIFICATIONS TCS ET PRC

Thèmes	items	Évaluation	Intervention
Vignette 1 Région : poignet/main Condition : fracture	Item 1	x	
	Item 2	x	
	Item 3	x	
Vignette 2 Région : coude Condition : épicondylalgie	Item 4	x	
	Item 5	x	
	Item 6	x	
Vignette 3 Région lombaire Condition : protrusion discale	Item 7	x	
	Item 8	x	
	Item 9	x	
	Item 10	x	
Vignette 4 Région lombaire Condition : sténose spinale	Item 11	x	
	Item 12	x	
	Item 13	x	
	Item 14	x	
Vignette 5 Région cervicale Condition : cervicalgie vs douleur systémique	Item 15	x	
	Item 16	x	
	Item 17	x	
Vignette 6 Région main/poignet Condition : blessure de surutilisation	Item 18		x
	Item 19		x
	Item 20		x
Vignette 7 Région main/poignet Condition : post fracture	Item 21		x
	Item 22		x
	Item 23		x
Vignette 8 Région lombaire Condition : lombalgie chronique	Item 24		x
	Item 25		x
	Item 26		x
Vignette 9 Région lombaire Condition : lombalgie mécanique	Item 27		x
	Item 28		x
	Item 29		x
Vignette 10 Région : coude Condition : tendinopathie	Item 30		x
	Item 31		x
	Item 32		x
Vignette 11 Région cervicale Condition : cervicalgie	Item 33		x
	Item 34		x
	Item 35		x
Vignette 12 Région : mollet Condition : déchirure musculaire	Item 36	x	
	Item 37	x	
	Item 38	x	
Vignette 13 Région : épaule Condition : blessure de surutilisation	Item 39	x	
	Item 40	x	
	Item 41	x	

Thèmes	items	Évaluation	Intervention
Vignette 14 Région : genou Condition : blessure traumatique	Item 42	x	
	Item 43	x	
	Item 44	x	
	Item 45	x	
Vignette 15 Région : genou Condition : blessure traumatique	Item 46	x	
	Item 47	x	
	Item 48	x	
Vignette 16 Région genou Condition : condition post opératoire	Item 49		x
	Item 50		x
	Item 51		x
Vignette 17 Région genou Condition : condition traumatique	Item 52		x
	Item 53		x
	Item 54		x
Vignette 18 Région épaule Condition : condition traumatique	Item 55	x	
	Item 56	x	
	Item 57	x	
Vignette 19 Région hanche Condition : blessure de surutilisation	Item 58		x
	Item 59		x
	Item 60		x
	Item 61		x
Vignette 20 Région genou Condition : électrothérapie/exercices	Item 62		x
	Item 63		x
	Item 64		x
Vignette 21 Région épaule Condition : exercices	Item 65		x
	Item 66		x
	Item 67		x
Vignette 22 Région hanche Condition : blessure de surutilisation	Item 68	x	
	Item 69	x	
	Item 70	x	
Vignette 23 Région : hanche Condition : tendinopathie	Item 71	x	
	Item 72	x	
	Item 73	x	
Vignette 24 Région épaule Condition : déficience principale	Item 74	x	
	Item 75	x	
	Item 76	x	
Vignette 25 Région : hanche Condition : arthrose	Item 77	x	
	Item 78	x	
	Item 79	x	
Vignette 26 Région : genou Condition : blessure de surutilisation	Item 80	x	
	Item 81	x	
	Item 82	x	
Total		48/82	34/82
% des items		58,5 %	41,5 %
Après optimisation (Items en rouge enlevés)		37/67	30/67
% des items		55,2 %	44,8 %

Tableau de spécifications PRC

Thèmes	Items	Évaluation	Intervention
Vignette 1 Région : région lombaire Condition : sténose spinale	Caractéristiques : Item 1 à 28 +diagnostic : 2 items	x	
Vignette 2 Région : épaule Condition : blessure de surutilisation	Caractéristiques : Item 29 à 56 +diagnostic : 2 items	x	
Vignette 3 Région main/ poignet Condition : post fracture	Caractéristiques : Item 57 à 78 +intervention : 2 items		x
Vignette 4 Région hanche Condition : blessure de surutilisation	Caractéristiques : Item 79 à 106 +diagnostic : 2 items	x	
Vignette 5 Région cervicale et coude Condition : cervicalgie vs épicondylalgie	Caractéristiques : Item 107 à 136 +diagnostic : 2 items	x	
Vignette 6 Région genou Condition : blessure traumatique	Caractéristiques : Item 137 à 164 +diagnostic : 2 items	x	
Vignette 7 Région cervicalgie Condition : blessure de surutilisation	Caractéristiques : Item 165 à 186 +intervention : 2 items		x
Vignette 8 Région hanche Condition : blessure de surutilisation	Caractéristiques : Item 187 à 202 +intervention : 2 items		x
Total	202 items de caractéristiques et 16 items d'hypothèse de diagnostics/intervention		

ANNEXE 7 SONDAGE ÉTUDIANTS



Faculté des sciences de l'éducation

Département d'administration et fondements de l'éducation

Sondage

Le présent sondage vise à recueillir vos commentaires sur le format des questions d'évaluation du raisonnement clinique que vous avez expérimenté.

Les tests de concordance de script

Svp entourez la réponse qui correspond à votre opinion

1. Est-ce que les scénarios cliniques étaient en lien avec les connaissances apprises dans le curriculum?

Pas du tout en lien
Un peu en lien
Moyennement en lien
Très en lien
Complètement en lien

Commentaires : _____

2. Est-ce que vous avez trouvé difficile de vous adapter à ce type de questions?

Pas du tout difficile
Un peu difficile
Moyennement difficile
Très difficile
Extrêmement difficile

Commentaires : _____

3. Est-ce que vous trouvez ce type d'évaluation pertinent pour évaluer le raisonnement clinique?

Pas du tout pertinent
Un peu pertinent
Moyennement pertinent
Très pertinent
Extrêmement pertinent

Commentaires : _____

Les problèmes de raisonnement clinique

SVP entourez la réponse qui correspond à votre opinion

4. Est-ce que les scénarios cliniques étaient en lien avec les connaissances apprises dans le curriculum?

Pas du tout en lien
Un peu en lien
Moyennement en lien
Très en lien
Complètement en lien

Commentaires : _____

5. Est-ce que vous avez trouvé difficile de vous adapter à ce type de questions?

Pas du tout difficile
Un peu difficile
Moyennement difficile
Très difficile
Extrêmement difficile

Commentaires : _____

6. Est-ce que vous trouvez ce type d'évaluation pertinent pour évaluer le raisonnement clinique?

Pas du tout pertinent
Un peu pertinent
Moyennement pertinent
Très pertinent
Extrêmement pertinent

Commentaires : _____

Merci de votre participation !

ANNEXE 8 SONDAGE EXPERTS



Faculté des sciences de l'éducation

Département d'administration et fondements de l'éducation

Questionnaire membre du panel d'expert

Pour des fins d'analyses, SVP, veuillez répondre aux quelques questions suivantes :

1. Nombre d'années de pratique :

- a. 0 à 5
- b. 5 à 10
- c. 10 à 15
- d. 15 à 20
- e. Plus de 20

2. Type de pratique

- a. Exclusivement en musculosquelettique en pratique privée
- b. Exclusivement en musculosquelettique en milieu public
- c. Principalement en musculosquelettique mais touchant aussi à d'autres domaine (neurologie, cardiorespiratoire, etc.)

3. Implication en enseignement de la physiothérapie

- a. Cours universitaire de base
- b. Formation continue
- c. a et b
- d. Non impliqué en enseignement

Sondage

Ce sondage vise à recueillir vos commentaires sur le format des questions d'évaluation du raisonnement clinique que vous avez expérimenté.

Les tests de concordance de script

Svp entourez la réponse qui correspond à votre opinion

1. Est-ce que vous avez trouvé difficile de vous adapter à ce type de questions?

Pas du tout difficile
Un peu difficile
Moyennement difficile
Très difficile
Extrêmement difficile

Commentaires :

2. Est-ce que vous trouvez ce type d'évaluation pertinent pour évaluer le raisonnement clinique?

Pas du tout pertinent
Un peu pertinent
Moyennement pertinent
Très pertinent
Extrêmement pertinent

Commentaires :

Les problèmes de raisonnement clinique

SVP entourez la réponse qui correspond à votre opinion

3. Est-ce que vous avez trouvé difficile de vous adapter à ce type de questions?

- Pas du tout difficile**
- Un peu difficile**
- Moyennement difficile**
- Très difficile**
- Extrêmement difficile**

Commentaires :

4. Est-ce que vous trouvez ce type d'évaluation pertinent pour évaluer le raisonnement clinique?

- Pas du tout pertinent**
- Un peu pertinent**
- Moyennement pertinent**
- Très pertinent**
- Extrêmement pertinent**

Commentaires :

Merci de votre participation !

ANNEXE 9 GRILLE D'ANALYSE DES FORMULAIRES DE STAGE

Étudiant	Énoncés en lien avec le RC	Types de difficulté* si présente	Degré mineur/majeur	Cotation au-delà des attentes **	Cotation sous les attentes **
Stage 2					
Stage 3					
Stage 4					
Stage 5					
Stage 6					
Catégorisation pour l'ensemble des stages (1-4) :					

*** Étapes principales du raisonnement clinique**

Génération d'hypothèses et récolte de données

Raffinement et traitement des hypothèses

Diagnostic final et élaboration d'un plan d'intervention

**** Pour les rôles suivants**

Rôle expert (évaluation)

Rôle expert (intervention)

Rôle praticien érudit

ANNEXE 10 CITATIONS DANS LA LANGUE D'ORIGINE

« Simply, clinical reasoning is the sum of the thinking and decision making processes associated with clinical practice; it is a critical skill in the health professions, central to the practice of professional autonomy, and it enables practitioners to take wise action taking the best judged action in a specific context » (Higgs & Jones, 2008, p.4)

« Validation of an instrument calls for an integration of many types of evidence. The varieties of investigation are not alternatives, any one of which would be adequate. The investigations supplement one another. The person validating a test should give thought to all questions suggested under the three rubrics listed by the committee, though the relative importance of the questions varies from test to test. » (Cronbach & Thorndike, 1971, p. 445)

« If all data are potentially relevant to validity, where should one start, and how much evidence is needed to adequately support a proposed interpretation or use? » (Kane, 2013, p.8.).

« On the one hand, validity theory was becoming so big that it was increasingly unclear how to translate it into validation practice, particularly if it was supposed to embrace both scientific and ethical concerns. » (Newton & Shaw, 2014, p. 135)

ANNEXE 11 EXEMPLE DE TYPES DE QUESTIONS DE RAISONNEMENT CLINIQUE APPLIQUÉS À LA PHYSIOTHÉRAPIE

Éléments clés (incluant une question à choix multiples et une réponse construite)

Votre patient de 52 ans est diabétique et présente une faiblesse importante lors de l'extension du gros orteil droit, ainsi qu'une diminution de sensation sur la partie médiale et dorsale du pied. Le tout a débuté de façon insidieuse.

1. Laquelle des conditions suivantes a **la plus forte** probabilité de causer cette déficience?
 - a. Radiculopathie L5
 - b. Neuropathie du nerf saphène
 - c. Tumeur de la moelle épinière
 - d. Un hallus valgus
 - e. Tendinite du l'extenseur du gros orteil

Nommez un (1) test clinique qui pourrait vous aider à confirmer votre impression.

Questions à appariement étendu

- a. Fracture de la grosse tubérosité de l'humérus
- b. Capsulite de l'articulation gléno-humérale
- c. Tendinite du supraépineux
- d. Déchirure partielle de la coiffe des rotateurs
- e. Syndrome d'accrochage
- f. Tendinose de l'infra épineux
- g. Radiculopathie de C5 avec déficit moteur
- h. Point gâchette du deltoïde moyen

Pour chaque patient, sélectionner le diagnostic le plus probable.

1. Un homme de 32 ans, droitier, se présente à votre clinique en première ligne pour des douleurs constantes à l'épaule gauche. Ces douleurs sont apparues, il y a 3 jours à la suite d'une chute sur l'épaule. À l'examen, il présente un arc douloureux entre 80 et 100 degrés, ainsi qu'une faiblesse importante et une douleur à 3/10 à l'épreuve isométrique résisté en abduction. Le mouvement passif d'abduction est presque complet, mais que le mouvement actif est de 60 degrés.

2. Une femme de 55 ans présente des douleurs à l'épaule droite depuis 6 mois, une limitation importante de tous ces mouvements passifs et actifs et peu de douleur au niveau des mouvements résistés.