

Direction des bibliothèques

AVIS

Ce document a été numérisé par la Division de la gestion des documents et des archives de l'Université de Montréal.

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

NOTICE

This document was digitized by the Records Management & Archives Division of Université de Montréal.

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal

Développement de tests mesurant les habiletés de perception et d'interprétation des
radiographies pulmonaires

Par

Salma Nadia Mhiri

Directeur de recherche : Bernard Charlin

Centre de pédagogie appliquée aux sciences de la santé (CPASS)

Faculté de médecine

Département de psychopédagogie et d'andragogie

Faculté des sciences de l'éducation

Mémoire présenté à la Faculté des études supérieures

en vue de l'obtention du grade de

M.A. en Pédagogie universitaire des sciences médicales

Décembre 2008

© Salma Nadia Mhiri



Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé :
Développement de tests mesurant les habiletés de perception et d'interprétation des
radiographies pulmonaires

présenté par
Salma Nadia Mhiri

a été évalué par un jury composé des personnes suivantes :

Jean-Guy Blais
Président rapporteur

Bernard Charlin
Directeur de recherche

Nathalie Loye
Membre du jury

RÉSUMÉ

CONTEXTE : Comme dans les autres disciplines médicales, parfois les radiologues dont la spécialité est à dominance visuelle, font face à des situations d'incertitude et empruntent des voies différentes pour résoudre un problème diagnostic. Dans un tel contexte, l'évaluation des compétences diagnostiques, se prête mal aux tests classiques qui nécessitent un consensus entre les spécialistes. Le test de concordance de scripts (TCS) mesure la concordance des décisions entre un candidat et un panel de spécialistes dans des situations d'incertitude.

BUT : (1) Décrire les composantes perceptuelles et cognitives du raisonnement en radiologie diagnostique; (2) Dégager les concepts à la base de nouveaux tests d'évaluation des compétences diagnostiques en radiologie; (3) Construire deux tests en ligne, basés sur le principe du TCS et mesurant les habiletés de détection, d'identification et d'interprétation des anomalies radiographiques pulmonaires (RxP) en contexte d'incertitude; (4) Décrire les étapes de validation des deux instruments de mesure.

MÉTHODOLOGIE : Description des étapes de construction et des composantes (1) du test de perception, qui mesure les habiletés de détection et d'identification des anomalies RxP; (2) du test d'interprétation qui mesure les habiletés à interpréter des anomalies RxP.

RÉSULTATS et CONCLUSION : Deux TCS en ligne dont le processus de validation est en cours et tenant compte des spécificités visuelles de la radiologie, ont été construits et sont disponibles au site :

<http://www.cme2.umontreal.ca:8080/sct-images-pilote3/>

MOTS CLÉS : radiographies pulmonaires, détection, identification, interprétation, évaluation des résidents, raisonnement clinique en contexte d'incertitude, test de concordance de scripts, tests d'évaluation en ligne.

ABSTRACT

RATIONALE: Radiologists have to face situations of image perception and analysis uncertainty, and do not follow the same paths of thought to solve a radiological problem. Conventional tests do not assess radiological diagnostic skills to solve ill-defined problems. They measure the ability to apply solutions to well-defined problems and need expert consensus. The script concordance test (SCT) measures the consistency of decisions between a candidate and a panel of experts at solving a clinical problem in context of uncertainty.

AIM: (1) To describe the perceptual and cognitive approaches in diagnostic radiology, (2) To develop the basis of new assessment tools to evaluate professional skills in diagnostic radiology, (3) To build two Web-Based tests, based on the principle of SCT, measuring resident's skills to detect, identify (perception test) and interpret (interpretation test) chest X rays abnormalities (RxP) in the context of uncertainty; (4) to describe the validation steps of both instruments.

RESULTS and CONCLUSION: Two Web-Based SCT evaluating perception and interpretation skills, and taking account of visual specificity of radiology, have been built. Their validation process is ongoing. They are available at: <http://www.cme2.umontreal.ca:8080/sct-images-pilote3/>

KEY WORDS: chest X rays, abnormalities perception, detection, identification, abnormalities interpretation, clinical reasoning, residents assessment, ill-defined problems, script concordance test, Web-based assessment.

TABLE DES MATIÈRES

RÉSUMÉ	I
ABSTRACT	II
TABLE DES MATIÈRES	III
LISTE DES TABLEAUX	V
LISTE DES FIGURES	V
LISTE DES ANNEXES	VI
REMERCIEMENTS	VIII
PRÉAMBULE	IX
INTRODUCTION	1
CADRE THÉORIQUE	6
1 LES GRANDES LIGNES DU RAISONNEMENT CLINIQUE	6
1-1 <i>Mécanismes du raisonnement clinique</i>	6
1-2 <i>Les scripts de maladie</i>	7
1-3 <i>Activation des scripts de maladie et raisonnement clinique</i>	12
2 LA DÉMARCHE DIAGNOSTIQUE EN RADIOLOGIE	14
2-1 <i>L'approche perceptuelle de l'expertise radiologique</i>	14
2-1-1 La perception chez les experts.....	15
2-1-2 La perception chez les novices et les résidents.....	15
2-1-3 Évolution de l'apprentissage perceptuel	16
2-2 <i>L'approche cognitive de l'expertise radiologique</i>	17
2-2-1 Organisation cognitive chez les experts	17
2-2-2 Organisation cognitive chez les novices et les résidents.....	18
2-3 <i>Liens entre la perception et la cognition en radiologie</i>	18
2-4 <i>Raisonnement radiologique et scripts radiologiques</i>	19
3 LA VARIABILITÉ DU DIAGNOSTIC EN RADIOLOGIE	20
3-1 <i>Erreurs de perception et d'interprétation</i>	21
3-1-1 Paramètres physiques et techniques.....	21
3-1-2 Facteurs physiologiques	21
3-1-3 Conditions psychophysiques	21
3-1-4 Niveau d'expérience.....	23
3-2 <i>Ambiguïté des images</i>	23
4- CAS PARTICULIER DES RADIOGRAPHIES PULMONAIRES	24
5- ÉVALUATION DU RAISONNEMENT RADIOLOGIQUE EN CONTEXTE D'INCERTITUDE	32
5-1 <i>Les tests classiques</i>	32
5-1-1 QCM.....	32
5-1-2 ECOS	32
5-1-3 Épreuve orale	33
5-2 <i>Le test de concordance des scripts (TCS)</i>	33
5-2-1 Description du TCS.....	34
5-2-2 Études de validation du TCS	35
5-3 <i>Utilisation du TCS en radiologie</i>	37
5-3 <i>Développement du TCS en ligne</i>	39
OBJECTIFS DE LA RECHERCHE	41
MÉTHODOLOGIE	42
1- CONSTRUCTION ET DÉROULEMENT DU TEST DE PERCEPTION (TP)	42
1-1 <i>Critères de sélection des radiographies pulmonaires</i>	43
1-2 <i>Élaboration des vignettes du TP</i>	44

1-3 Panel de référence du TP.....	44
1-4 Calcul des scores du TP.....	45
2- CONSTRUCTION DU TEST D'INTERPRÉTATION (TI)	47
2-1 Sélection des RxP du test d'interprétation.....	51
2-2 Panel de référence du TI.....	51
2-3 Calcul des scores du TI.....	51
3- INTERFACE INFORMATIQUE ET DÉROULEMENT DES TESTS	52
4- PLAN DE VALIDATION DES TESTS	53
4-1 Sujets participant à l'étude.....	53
4-2 Consentement éclairé et comité d'éthique	54
4-3 Validité de contenu	54
4-4 Analyse d'items.....	55
4-5 Mesures de fidélité.....	55
4-6 Pouvoir discriminant des tests.....	57
4-7 Mesure de la validité prédictive des tests.....	58
4-8 Profil des réponses des participants.....	58
4-9 Logiciels statistiques utilisés.....	58
CONCLUSIONS ET BÉNÉFICES ESCOMPTÉS.....	59
RÉFÉRENCES	61
ANNEXES	70

LISTE DES TABLEAUX

Tableau I.	35
Tableau II.	43
Tableau III.	51
Tableau IV.	57

LISTE DES FIGURES

<i>Figure 1.</i> Cadre de compétences CanMEDS	1
<i>Figure 2.</i> Exemples d'activation de scripts de maladie	11
<i>Figure 3.</i> Intégration des scripts dans le modèle du raisonnement clinique	13
<i>Figure 4.</i> Hiérarchie inverse de l'apprentissage perceptuel.....	18
<i>Figure 5.</i> Niveaux d'évidence et reconnaissance d'anomalies radiologiques	24
<i>Figure 6.</i> Schéma du processus diagnostique d'une radiographie pulmonaire	26
<i>Figure 7.</i> Exemple d'une image typique et évidente	28
<i>Figure 8.</i> Exemple d'une anomalie subtile	29
<i>Figure 9.</i> Exemple d'une image compatible avec une image construite	30
<i>Figure 10.</i> Exemple d'une image ambiguë.....	31
<i>Figure 11.</i> Exemple d'une vignette d'interprétation probabiliste.....	49
<i>Figure 12.</i> Exemple d'une vignette d'interprétation analytique	50

LISTE DES ANNEXES

<i>Annexe 1.</i> Test de perception, démonstration en ligne.....	x
<i>Annexe 2.</i> Test de perception, anomalies à détecter et à identifier	xvi
<i>Annexe 3.</i> Contenu des vignettes de perception	xix
<i>Annexe 4.</i> Grille de contrôle de qualité du test de perception.....	xx
<i>Annexe 5.</i> Test d'interprétation : Instructions aux participants.....	xxi
<i>Annexe 6.</i> Contenu des vignettes d'interprétation.....	xxviii
<i>Annexe 7.</i> Grille de contrôle de qualité du test d'interprétation	xxix
<i>Annexe 8.</i> Grille de contrôle de qualité du panel de référence.....	xxx
<i>Annexe 9.</i> Lettres d'information et formulaires de consentement	xxxix
<i>Annexe 10.</i> Questionnaires.....	xxxiii
<i>Annexe 11.</i> Grille d'évaluation des compétences des stages de résidence	xxxvi

LISTE DES SIGLES ET ABRÉVIATIONS

CHUM	Centre Hospitalier de l'Université de Montréal
CPASS	Centre de Pédagogie Appliquée aux Sciences de la Santé
CRMCC	Collège Royal des Médecins et des Chirurgiens du Canada
ECOS	Examen Clinique Objectif Standardisé
JPEG	Joint Photographic Experts Group
PACS	Picture Archiving and Communication System
QCM	Questions à choix multiples
R1,2,3,4,5,6	Résidents de 1 ^{ère} , 2 ^{ème} , 3 ^{ème} , 4 ^{ème} , 5 ^{ème} , 6 ^{ème} année de résidence
Rgen	Radiologues généralistes
Rthx	Radiologie experts en radiologie thoracique
RxP	Radiographie Pulmonaire
SDe	Score de Détection
SDeV	Score de Détection par vignette
SDeT	Score de Détection Total
SId	Score d'Identification
SIdV	Score d'Identification par vignette
SIdT	Score d'Identification Total
SIP	Score d'Interprétation Probabiliste
SIPV	Score d'Interprétation probabiliste par vignette
SIPT	Score d'Interprétation probabiliste Total
SIA	Score d'Interprétation Analytique
SIAV	Score d'Interprétation Analytique par vignette
SIAT	Score d'Interprétation Analytique Total
SD	Déviation Standard
TCS	Test de Concordance des Scripts
TP	Test de Perception
TI	Test d'interprétation

*À ma famille en Tunisie et au Québec
pour leur amour et leur soutien inconditionnel*

À la mémoire de Didi

REMERCIEMENTS

Un Grand Merci

Au Docteur Bernard Charlin, pour ses encouragements tout au long de ce projet qu'il a si patiemment supervisé. Grâce à lui, certains concepts « complexes » du raisonnement clinique me paraissent plus simples.

À Robert Gagnon, pour sa disponibilité, ses conseils judicieux et nos longues discussions autour des processus de validation et des méthodes de mesure.

Au Docteur Louise Samson, qui s'est ralliée au projet avec enthousiasme. Elle a facilité la sélection et l'accès aux radiographies pulmonaires et largement contribué à sa construction.

Au Docteur Driss Kazi-Tani, qui avec talent et patience, a construit les logiciels nécessaires au développement des tests.

Au Docteur Josée Dubois, qui a rendu possible l'accès aux résidents du programme de radiologie diagnostique.

Au Docteur Carole Lambert, une formidable directrice de programme, pour tous ses encouragements.

Aux résidents du programme de radiologie diagnostique et aux radiologues qui ont bien voulu se prêter au jeu et tester la version *Bêta* de ces outils de mesure.

PRÉAMBULE

Ce travail est né d'un heureux concours de circonstances. Au départ, je suis pneumologue. Après un doctorat en physiologie respiratoire, j'ai enseigné la physiologie aux étudiants de la faculté de médecine de Tunis et exercé dans une unité d'explorations fonctionnelles pulmonaires à l'hôpital La Rabta. En 2002, grâce à une bourse des instituts de recherche en santé du Canada, j'ai eu l'occasion d'effectuer un stage de recherche post-doctorale à Montréal, dans mon domaine. J'y ai découvert l'existence du programme maîtrise en pédagogie universitaire des sciences médicales et je me suis empressée de m'y inscrire. Cela venait combler les interrogations suscitées par mon expérience d'enseignement à Tunis.

Pendant les cours de maîtrise, j'ai découvert les publications du Docteur Charlin, relatives à l'évaluation des compétences diagnostiques en radiologie. La possibilité de disposer d'outils quantifiant les situations d'incertitude et de mieux comprendre le raisonnement radiologique m'ont rappelé les discussions diagnostiques passionnantes que j'avais avec les radiologues thoraciques à Tunis. J'ai donc choisi ce projet comme sujet de mémoire. Il s'agissait de développer les outils de mesure les plus réalistes et les plus fidèles possibles, mesurant les compétences diagnostiques radiologiques en situation d'incertitude et en ligne.

La première étape du projet, la construction des instruments de mesure, est achevée. Sa réalisation m'a permis de saisir les subtilités du raisonnement clinique en radiologie, que ce mémoire essaye de conceptualiser et de découvrir des collègues et amis extraordinaires, sans lesquels les tests n'auraient pas vu le jour. Les essais préliminaires de ces tests sont en cours, mais ne seront pas exposés dans ce mémoire. Ils feront l'objet de publications, une fois le processus de validation terminé. Entre-temps, j'ai modifié le cap de ma carrière et choisi une deuxième spécialité, la radio-oncologie où je suis actuellement résidente. Cela me donnera sûrement l'occasion de progresser dans le domaine de l'évaluation en imagerie.

INTRODUCTION

L'opinion des radiologues est souvent décisive dans la prise en charge des patients. Elle détermine le choix d'approfondir ou non les investigations, l'orientation du diagnostic et la prise de décision thérapeutique par le clinicien. L'évaluation des compétences diagnostiques des futurs radiologues est donc cruciale.

Epstein (2002) définit la compétence professionnelle dans le domaine médical, comme étant l'utilisation judicieuse, au service de l'individu et de la communauté, de plusieurs habiletés associant les connaissances scientifiques et techniques, au raisonnement clinique et aux qualités personnelles, dont la communication et la réflexion. Cette définition est reprise dans le cadre des compétences CanMEDS (cf. figure 1, page 1) du Collège Royal des Médecins et des Chirurgiens du Canada (CRMCC)¹, pour qui un médecin compétent est un **expert médical** qui maîtrise les six compétences suivantes : la communication, la collaboration, la gestion, la promotion de la santé, l'érudition et le professionnalisme (Frank, 2005).

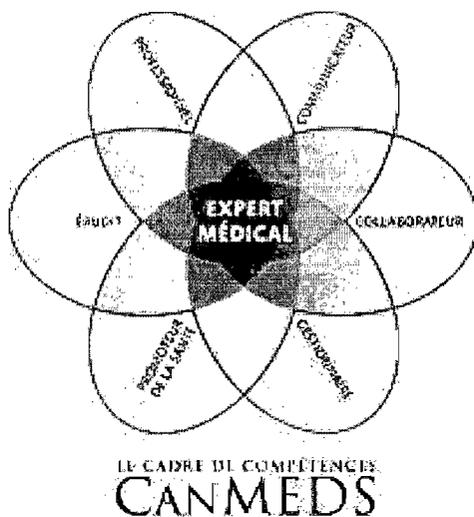


Figure 1. Cadre de compétences CanMEDS

Toujours selon le CRMCC, la **compétence en radiologie diagnostique** inclut

¹ Le Collège Royal des Médecins et des Chirurgiens du Canada a pour mission d'établir des normes de pratique élevées et d'assurer la qualité des soins de santé.

l'expertise médicale et les 6 compétences énumérées ci-dessus. L'expertise médicale radiologique vise les capacités à « 1) *Démontrer les aptitudes diagnostiques et thérapeutiques voulues pour assurer le traitement éthique et efficace du patient*, 2) *Consulter et appliquer l'information pertinente à la pratique clinique*. 3) *Offrir des services de consultation efficaces en matière de soins des patients, séances de sensibilisation et avis juridiques* » (CRMCC, 2007b). Parmi les objectifs spécifiques de l'expertise médicale radiologique, on précise qu'à la fin du programme de résidence en radiologie, le candidat doit « *Faire preuve de dextérité manuelle et d'habiletés techniques ainsi que de compétences en diagnostic et en interprétation* ».

La compétence diagnostique est effectivement fondamentale dans cette spécialité où la prépondérance des habiletés visuelles détermine la précision du diagnostic. Mais les habiletés qui la composent, la **détection**, l'**identification** et l'**interprétation** des images anormales ne sont pas clairement énoncées par le CRMCC. Ces habiletés nécessitent l'intégration de l'information visuelle aux connaissances fondamentales, sémiologiques et au jugement clinique (Rubin, 1989). Elles se développent tout au long de la formation des résidents en radiologie et il est essentiel de s'en assurer par une évaluation adéquate des résidents. Celle-ci est importante à plusieurs titres. 1) Elle permet de suivre leur évolution au cours des cinq années de résidence; 2) elle est primordiale dans l'orientation des apprentissages; 3) elle fournit une rétroaction permettant d'évaluer et d'améliorer le contenu des programmes de formation post-doctorale; 4) elle garantit, par le biais des examens de certification, un niveau de compétence adéquat pour exercer la spécialité.

Il existe plusieurs formats d'évaluation des compétences diagnostiques en radiologie. Le CRMCC en reconnaît trois qu'il utilise dans les examens de certification (CRMCC, 2007a). Une épreuve écrite comprenant des questions à choix multiples (QCM), permet d'évaluer les connaissances factuelles d'ordre médical, chirurgical et fondamental se rapportant à la radiologie diagnostique, ainsi que les aspects théoriques et pratiques liés à l'exercice de la spécialité. La détection des anomalies et leur interprétation est évaluée par des examens cliniques objectifs structurés (ECOS) et des épreuves orales qui se rapprochent de la vie réelle en radiologie diagnostique et offrent des questions relevant de la pratique courante. Ces trois formats d'évaluation ont permis

la certification de nombreuses promotions de radiologues. Ils exigent cependant que les réponses aux questions fassent consensus entre les spécialistes du domaine. Or, dans des situations d'incertitude, les médecins n'empruntent pas toujours le même cheminement pour résoudre des problèmes cliniques (A. S. Elstein, Shulman, & Sprafka, 1978) ou radiologiques (Gunderman, 2005; Gunderman & Nyce, 2002; Robinson, 1997).

Le niveau d'incertitude d'une situation clinique donnée est relié à la disponibilité des informations et détermine la manière dont le problème est résolu. Les situations cliniques où toutes les informations pertinentes sont présentes, définissent les problèmes simples ou bien structurés, résolus avec un niveau de certitude élevé (Schön, 1983). Les situations cliniques incertaines se présentant avec des données manquantes ou ambiguës car atypiques, définissent les problèmes complexes également appelés problèmes mal définis qui n'ont pas de solution univoque. Les spécialistes du même domaine traitent souvent le problème différemment et varient dans leur manière d'utiliser les informations disponibles. Sa résolution n'atteint jamais des niveaux d'accord très élevés. D'où la limite d'une méthode d'évaluation basée sur le consensus.

Le *test de concordance des scripts (TCS)* a été développé dans la dernière décennie. Il essaie de recréer le contexte d'incertitude dans la résolution des problèmes cliniques. Il a pour but d'évaluer les microdécisions que comporte le raisonnement clinique en situation d'incertitude (B. Charlin, Brailovsky, Leduc, & Blouin, 1998; B. Charlin, Roy, Brailovsky, Goulet, & van der Vleuten, 2000; B. Charlin & van der Vleuten, 2004). Son principe assume que chaque expert² est autorisé à jauger différemment les informations présentes et à attribuer un poids relatif aux microdécisions qui en découlent. Il mesure la concordance entre les décisions du candidat et celles d'un panel de références composé d'experts dont la façon de résoudre le problème peut diverger. Les scores obtenus reflèteraient l'organisation des connaissances ou « scripts » du candidat. Un cas clinique comportant de l'incertitude est présenté suivi d'hypothèses pertinentes à la situation. Des informations cliniques sont ensuite données et le candidat doit décider de la direction et de l'effet (positif, négatif ou neutre) de cette information sur le statut de l'hypothèse proposée. Le test mesure la

² Dans le cadre de ce travail, le mot expert désigne un spécialiste dans un domaine donné, un radiologue par exemple.

concordance entre la réponse du candidat et celle d'un panel de référence composé d'experts. Les scores ont montré de bonnes qualités psychométriques (fiabilité et discrimination selon le niveau d'expertise) dans de nombreux domaines de la médecine (Caire, Sol, Moreau, Isidori, & Charlin, 2004; Fournier et al., 2006; Lambert, 2005; Marie et al., 2005; Meterissian, Zabolotny, Gagnon, & Charlin, 2007; Sibert et al., 2006), dont la radiologie diagnostique (Brazeau-Lamontagne, Charlin, Gagnon, Samson, & van der Vleuten, 2004).

Une étude pilote a comparé la progression des habiletés de perception et d'interprétation d'anomalies radiologiques pulmonaires par des résidents de radiologie, en utilisant des QCM pour évaluer la composante perception et un TCS pour la composante interprétation (Brazeau-Lamontagne et al., 2004). Elle a démontré une progression des scores des deux compétences avec l'expérience. Cette progression était plus rapide pour les scores de perception des anomalies que pour ceux de l'interprétation, indiquant l'importance de bien détecter et identifier des images avant de pouvoir les interpréter. La variabilité dans la perception (détection et identification) des anomalies entre les radiologues, n'était toutefois pas prise en compte dans cette étude. L'utilisation de QCM exigeait qu'une réponse au test de perception n'était correcte que si elle faisait consensus entre la plupart (80%) des membres du panel de référence composé de radiologues. Pourtant, la perception de certaines images subtiles ne fait pas toujours l'unanimité des radiologues et la règle du consensus à 80% diminuerait la discrimination du test. Un candidat qui perçoit une anomalie observée par une minorité de radiologues serait donc sanctionné. L'utilisation d'un autre format que des QCM serait donc plus approprié pour évaluer la dimension perception. Par ailleurs, le raisonnement clinique des radiologues comporte des spécificités relatives au caractère visuel de la spécialité. Contrairement au clinicien qui dispose rarement de toutes les informations pertinentes pour élaborer son diagnostic devant un tableau clinique donné, la plupart des éléments visuels nécessaires au diagnostic, aussi ambigus soient-ils, sont souvent accessibles au radiologue sur la radiographie. Ce qui impose une adaptation du format des questions évaluant l'interprétation. Enfin, l'avènement de la technologie numérique dans le stockage des images radiographiques, ou PACS (Picture Archiving and Communication System) a facilité l'accès aux radiographies archivées et leur

utilisation à des fins d'enseignement. Ceci ouvre la voie au développement de nouvelles possibilités d'évaluation en radiologie, prenant en compte les spécificités visuelles de la spécialité et incluant les nouvelles technologies.

Ce travail vise à développer des outils en ligne basés sur le principe du TCS et permettant d'évaluer les différents aspects du raisonnement clinique des résidents en radiologie. Dans le cadre théorique, on décrira brièvement les fondements du raisonnement clinique, le concept de scripts de maladie, les théories du raisonnement radiologique et les possibilités d'application des scripts en radiologie. En dehors de certains concepts, les détails des théories psychologiques et physiologiques de la perception ne feront pas partie de cette présentation. À partir du cadre théorique, on dégagera les concepts à la base de la construction des tests développés dans cette recherche. La section méthodologie comprendra la description des deux tests et des étapes de construction ayant conduit à leur version finale, ainsi que le plan envisagé pour leur validation psychométrique.

CADRE THÉORIQUE

1 Les grandes lignes du raisonnement clinique

1-1 Mécanismes du raisonnement clinique

Le raisonnement clinique désigne l'ensemble des processus de pensée et de prises de décisions diagnostiques et thérapeutiques, permettant au médecin d'agir de la manière la plus adéquate dans un contexte particulier (Harris, 1993; Higgs & Jones, 2000). Classiquement, la résolution de problèmes cliniques implique des stratégies analytiques et non analytiques. Face à un patient, le clinicien se fait une **représentation mentale** initiale du problème et génère des hypothèses catégorisées et hiérarchisées, dont la vérification systématique permet de réajuster graduellement la perception du problème jusqu'à sa solution. Il s'agit là d'une **stratégie analytique** qui suit une analyse systématique et organisée des liens existant entre les différents symptômes, les signes et les hypothèses diagnostiques possibles (Eva, 2005; G. Norman, 2005). Il est conscient et principalement *hypothético-déductif* selon un mode de *chaînage arrière, ou descendant* (backward reasoning). À partir de données clés rapidement identifiées, des hypothèses diagnostiques pertinentes (4 ou 5 au maximum) sont générées précocement et déclenchent la recherche de nouvelles informations, dont la présence ou l'absence influencera la décision diagnostique. La probabilité de chaque hypothèse résulte d'un faisceau d'arguments permettant de l'étayer ou de l'écarter (A. Elstein, Shulman, & Sprafka, 1990; A. S. Elstein et al., 1978; G. Norman, L. Brooks, C. Colle, & R. Hatala, 1999). Le raisonnement analytique est également possible par *chaînage avant* ou ascendant (forward reasoning). Celui-ci consiste à raisonner inversement, à partir de liens de causalités entre les données cliniques collectées, vers le diagnostic (Patel, Arocha, & Kaufman, 1994; Patel & Goren, 1986). Il serait davantage utilisé devant un tableau clinique complexe, lorsque le clinicien ne perçoit pas d'hypothèses cliniques pertinentes au départ (Eva, 2005). La stratégie **non analytique** est un processus automatique, sans effort conscient, de reconnaissance soit de configurations ou modèles (pattern recognition), soit de similitudes avec des cas auxquels on a déjà été confronté. Dans ce cas, l'esprit reconnaît instantanément une situation généralement simple, par une partie ou l'ensemble des signes caractéristiques qui la composent. À force de

s'exposer à des cas cliniques, on finit par se créer une banque de configurations typiques ou cas concrets, qui serviront de référence aux éléments du problème à résoudre (G. Norman, Young, & Brooks, 2007; G. R. Norman & Brooks, 1997; Schmidt, Norman, & Boshuizen, 1990). Cette stratégie, qui est plus importante et plus efficace chez les experts, est également utilisable par les novices (Coderre, Mandin, Harasym, & Fick, 2003). Elle est particulièrement présente dans des domaines impliquant des habiletés visuelles comme la dermatologie, la lecture d'électrocardiogrammes (Hatala, Norman, & Brooks, 1996; G. R. Norman, Rosenthal, Brooks, Allen, & Muzzin, 1989), ou la radiologie (G. R. Norman, Brooks, Coblentz, & Babcook, 1992; G. R. Norman, Coblentz, Brooks, & Babcook, 1992).

Dans les faits, le raisonnement clinique est probablement mixte, analytique et non analytique. Un diagnostic reposant exclusivement sur une approche de reconnaissance de modèles, risque d'engendrer des hypothèses diagnostiques erronées et une fermeture prématurée du raisonnement, si les hypothèses ne sont pas validées par une approche analytique (Coderre et al., 2003). Alors qu'un diagnostic exclusivement analytique serait laborieux et le raisonnement submergé par une pléthore d'informations à intégrer (Ark, Brooks, & Eva, 2007; GR. Norman, LR. Brooks, CaL. Colle, & RM Hatala, 1999). Eva (2005) suggère un modèle bidirectionnel mobilisant simultanément des stratégies interactives analytiques et non analytiques. Dans ce modèle, des hypothèses sont induites par la représentation mentale initiale, que se fait le médecin devant un patient. Elles sont ensuite vérifiées par un va-et-vient entre les deux processus analytique et non analytique, qui modulent la représentation mentale jusqu'au diagnostic final. La complémentarité de ces deux stratégies améliore la précision du diagnostic même chez les novices (Ark et al., 2007; Eva, Hatala, Leblanc, & Brooks, 2007; GR. Norman et al., 1999).

1-2 Les scripts de maladie

Aboutir à une décision clinique adéquate ne se limite pas qu'au raisonnement, mais requiert aussi une organisation appropriée des connaissances. Parmi les différents modèles d'organisation des connaissances proposés (M. Nendaz, Charlin, Leblanc, & Bordage, 2005), la théorie des scripts offre une stratégie efficace pour mieux

comprendre la dynamique de la résolution des problèmes cliniques. En psychologie cognitive, les schémas correspondant à un cadre facilitant l'organisation et l'interprétation des éléments qu'il contient. Les *scripts* sont des séquences d'actions stéréotypées, associées à des situations familières de la vie courante (Anderson & Conway, 1993; Schank & Abelson, 1977). Ils contiennent des éléments ayant des liens temporels, spatiaux ou de causalité et forment des réseaux de connaissances bien structurés et mobilisables au besoin, pour résoudre spécifiquement différents problèmes et leurs variantes. À titre d'exemple, dans le script « aller voir un spectacle », l'individu s'attend à effectuer des actions stéréotypées, comme aller sur place, faire la file, acheter son billet, le présenter à l'ouvreuse, rentrer dans la salle et voir le spectacle. Ce script est générique, car selon le type de spectacle à voir, la nature des actions va légèrement changer, mais l'ensemble du script aura la même forme. S'il s'agit d'un film par exemple, les billets seront achetés sur place, le film sera précédé d'une séance de publicité, puis visualisé en entier. Pour une pièce de théâtre, par contre, les billets sont réservés à l'avance, le spectacle commencera sans capsules publicitaires préalables et il sera interrompu par un entracte en principe. Les scripts se distinguent des schémas par leur caractère dynamique et par la possibilité d'y inclure des éléments inattendus. Par exemple, le script « d'aller à la dernière minute au théâtre » alors qu'il n'y a plus de billets disponibles, peut inclure la solution inhabituelle d'en obtenir par des personnes qui se désistent au dernier moment.

En médecine, le concept des scripts a été adapté pour donner les *scripts de maladie*, qui sont des entités de connaissances organisées, permettant de faire face aux exigences des tâches cliniques et de les accomplir (Feltovich, Barrows, & 1984;128)42., 1984). Il peut s'agir de scripts diagnostiques, d'investigations ou thérapeutiques. Un script contient des connaissances peu nombreuses mais essentielles pour résoudre un problème clinique. Ces connaissances sont groupées en rubriques comme les facteurs prédisposant l'apparition d'une maladie, ses modes de présentation, ses conséquences biologiques, son imagerie ou ses modalités thérapeutiques. La représentation mentale des symptômes d'un patient, active des scripts appropriés permettant de faire des inférences et d'interpréter les informations collectées et leurs liens (H. S. Barrows & Feltovich, 1987; B. Charlin, Gagnon, Sauve, & Coletti, 2007; B. Charlin, Tardif, &

Boshuizen, 2000; Feltovich et al., 1984; G. Norman, 2005). Les connaissances figurant dans un script contiennent des données dont la présence ou l'absence ou l'association à d'autres éléments va agir sur la probabilité de l'hypothèse évoquée (Schmidt et al., 1990).

La figure 2 (cf. figure 2, page 11) montre un exemple de script de maladie illustré par la situation clinique suivante : « *Un homme de 38 ans, fumeur, **toxicomane, itinérant et séropositif pour le VIH**, se présente à une clinique en se plaignant d'une toux depuis 3 mois. Il crache du sang (**hémoptysies**) et il a perdu 5 kilogrammes* ». Les éléments clés du tableau clinique (en caractères gras) composent la représentation mentale et activent immédiatement le script de la tuberculose chez le clinicien (cf figure 2, page 11). Le patient présente effectivement quelques facteurs favorisants et des symptômes généraux et respiratoires dont l'association est très évocatrice de la maladie. Ce qui va induire la recherche d'autres signes, comme la notion de contact avec des personnes tuberculeuses, une fièvre nocturne, ou la présence d'images pulmonaires apicales mal systématisées, pour renforcer l'hypothèse. Le clinicien envisage déjà d'entreprendre un examen direct et des cultures des expectorations pour la confirmer ou l'éliminer. Parallèlement, l'élément « immunosuppression » active des scripts de maladies possibles dans ce cas, mais moins probables, comme la pneumocystose ou d'autres infections à bactéries opportunistes. Ces scripts peuvent contenir des éléments communs avec celui de la tuberculose. Dans le cas de la pneumocystose par exemple, l'immunosuppression et la toux sont en faveur du diagnostic, mais l'absence de dyspnée (essoufflement) et la présence d'une hémoptysie en diminuent les probabilités. Ainsi, l'émergence d'autres scripts que le principal, va pousser à approfondir le raisonnement et à faire des liens entre les informations, et de hiérarchiser les hypothèses en attendant de les confirmer ou de les réfuter. Les scripts de maladies comportent plusieurs niveaux. Il peut s'agir de génériques représentant des classes de maladies, ou de prototypes de maladies comme l'exemple ci-dessus, voire des cas particuliers de patients vus dans le passé et pouvant servir de référence à des situations nouvelles (B. Charlin, Gagnon et al., 2007; Custers, Boshuizen, & Schmidt, 1998; Kulatunga-Moruzi, Brooks, & Norman, 2001). Leur contenu est spécifique du domaine, mais il varie d'un médecin à l'autre. Il progresse avec l'exposition itérative à une situation clinique donnée, devenant

plus riche, plus varié, mieux organisé et plus facilement mobilisable avec l'expérience. Les experts sélectionnent les informations les plus pertinentes dès les premières minutes de l'entrevue avec le patient. Ils augmentent ainsi les chances d'aboutir au vrai diagnostic, en activant les scripts les plus compatibles (Neufeld, Norman, Feightner, & Barrows, 1981). Contrairement aux débutants, ils utilisent moins souvent leurs connaissances des mécanismes fondamentaux des maladies, sauf si le cas est trop complexe (Schmidt et al., 1990).

<i>Histoire clinique</i>
Un homme de 38 ans, fumeur, toxicomane itinérant et séropositif pour le VIH, se présente à une clinique en se plaignant d'une toux depuis 3 mois. Il crache du sang (hémoptysies) et il a perdu 5 kilogrammes depuis quelques mois.
<i>Représentation mentale</i>
Homme jeune, fumeur, toxicomane, itinérant, VIH Toux, hémoptysies traînantes Amaigrissement

Scripts activés



<i>Script Tuberculose</i>	
Facteurs favorisants	<ul style="list-style-type: none"> ▪Précarité socio-économique ▪Contact tuberculeux ▪Originaire d'une zone endémique ▪Antécédents tuberculeux ▪Immunosuppression ▪Recrudescence récente de la tuberculose
Mode d'apparition	<ul style="list-style-type: none"> ▪Lent ▪Subaigu
Symptômes	<ul style="list-style-type: none"> ▪Altération de l'état général, fatigue ▪Fièvre vespérale ▪Amaigrissement ▪Signes respiratoires : toux, hémoptysies
Imagerie	<ul style="list-style-type: none"> ▪Opacités pulmonaires mal systématisées possiblement apicales, exca ▪Images miliaires ▪Tuberculome ▪Pleurésie...
Biologie	<ul style="list-style-type: none"> ▪Intradermoréaction à la tuberculine positive (négative si immunosuppression) ▪Bacille de Koch dans les expectorations ▪Taux de CD4 bas
Envisager	<ul style="list-style-type: none"> ▪Traitement anti-infectieux approprié ▪Déclaration obligatoire ▪Isolement préventif

<i>Autres Scripts activables mais moins probables</i>
Pneumonies bactériennes usuelles
Pneumonies bactériennes à germes opportunistes
Pneumonie à pneumocystis jirovecii
Cancer du poumon

Figure 2. Exemples d'activation de scripts de maladie

1-3 Activation des scripts de maladie et raisonnement clinique

L'activation des scripts de maladie correspond finalement à l'opérationnalisation du raisonnement clinique. Le script de maladie est une entité activée par voie analytique ou non analytique dès que l'on se fait une représentation mentale d'un cas clinique (B. Charlin, Boshuizen, Custers, & Feltovich, 2007). Son induction se fait probablement le plus souvent sur un mode non analytique par la reconnaissance automatique d'une configuration ou de cas concrets. Dans une situation typique, le script le plus probable est immédiatement activé et retenu comme diagnostic jusqu'à preuve du contraire, même si certains de ses éléments ne sont pas précisés dans la présentation initiale. En fait, les informations collectées sont parfois tellement typiques d'une maladie donnée, qu'on présume que d'autres éléments moins importants le sont également. Il est cependant essentiel de confirmer ensuite le diagnostic par d'autres données cliniques ou para-cliniques, ce qui implique une transition vers un raisonnement plus analytique. Dans le cas de la figure 2 (cf. figure 2, page 11), par exemple, il n'est pas nécessaire de chercher d'autres symptômes comme la fièvre vespérale. L'association de la précarité sociale et de l'immunosuppression aux hémoptysies est tellement en faveur de la tuberculose, que la présence de fièvre est assumée. D'ailleurs, son absence ne diminuera pas la probabilité que le patient souffre de tuberculose. Par contre, d'autres rubriques du script, comme l'aspect radiologique et la présence du bacille de Koch ou des tests plus sophistiqués, doivent être vérifiées pour retenir le diagnostic. Lorsque la situation est moins usuelle ou atypique, l'activation des scripts ne sera pas automatique et se fera par une induction délibérée (B. Charlin, Boshuizen et al., 2007). Si dans l'exemple de la figure 2 (cf. figure 2, page 11), la radiologie n'évoque pas une tuberculose ou que la biologie est négative, le raisonnement analytique sera prépondérant et générera d'autres scripts d'infections pulmonaires non tuberculeuses ou de cancer. L'élément tabagisme, par exemple, prendrait toute son ampleur et activerait le script cancer pulmonaire. La figure 3 (cf. figure 3, page 13) propose un schéma d'intégration des scripts de maladie dans le processus de raisonnement clinique. Une validation des hypothèses selon le modèle bidirectionnel proposé par Eva (Eva, 2005), fait suite à l'activation non analytique ou analytique de scripts de maladie. Ce qui permet de faire évoluer la représentation mentale jusqu'au diagnostic.

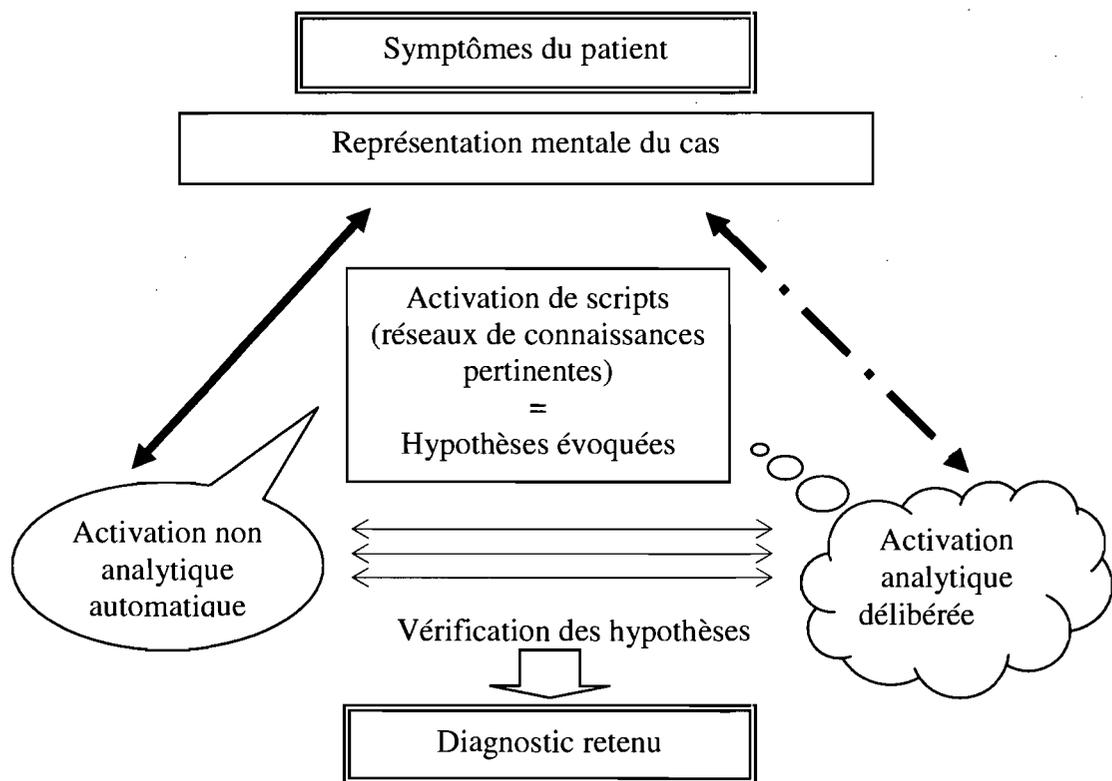


Figure 3. Intégration des scripts dans le modèle du raisonnement clinique

2 La démarche diagnostique en radiologie

La radiologie est une spécialité à dominance visuelle, où les informations, relatives à l'état clinique du patient, sont recueillies à partir de l'observation et de l'interprétation d'images. Le raisonnement radiologique va donc dépendre de la dimension visuelle et de l'activation des « scripts radiologiques » qui seront définis plus loin. Schématiquement, l'approche d'une radiographie comporte une **étape perceptuelle**, qui commence par la *recherche* et la *détection* d'images suspectes, suivie par la *prise de décision* de normalité ou d'anormalité et *identification* des éléments observés (A. M. Lesgold et al., 1988; G. R. Norman, Coblenz et al., 1992). Cette étape déclenche **l'étape d'interprétation**, qui met en relation les différents éléments observés, induit un processus conscient de recherche de nouveaux indices visuels et/ou cliniques et consolide les hypothèses diagnostiques plausibles conduisant au diagnostic final (A. M. Lesgold et al., 1988). Les phases de perception et d'interprétation des images radiologiques sont étroitement liées et souvent simultanées. Leur progression avec l'expertise a été étudiée par deux approches complémentaires, l'approche perceptuelle et l'approche cognitive. Dans les sections suivantes, le mot expert désignera les médecins radiologues et le mot novices, les résidents en cours de spécialisation ou les étudiants en médecine.

2-1 L'approche perceptuelle de l'expertise radiologique

Selon Kundel et coll. (H.L. Kundel, 2000), la perception visuelle d'anomalies radiologiques implique un processus de recherche ciblé et attentif, que l'on pourrait comparer à la recherche d'un objet du quotidien. Pour retrouver un livre sur les étagères d'une bibliothèque, par exemple, cela nécessite : 1) une connaissance préalable d'une ou de plusieurs de ses caractéristiques comme le titre, l'auteur, le sujet, le format et d'avoir une idée sur son emplacement; 2) d'entreprendre une recherche attentive et sélective de l'objet. De ce fait, la tâche de perception visuelle ne se limite pas à un processus exclusivement sensoriel de traitement du signal visuel. Elle fait intervenir une recherche consciente et orientée, qui induit la détection, la localisation, la comparaison, la catégorisation et l'identification des anomalies radiologiques (H.L. Kundel, 2000). En

étudiant l'analyse de la position des yeux, lors du balayage visuel des radiographies pulmonaires, Kundell et coll. (H. L. Kundel & Nodine, 1983; H. L. Kundel, Nodine, & Krupinski, 1989) ont conceptualisé la tâche perceptuelle en 4 étapes comportant : 1) une impression globale initiale; 2) une phase de recherche et de découverte; 3) une phase de recherche réflexive et 4) une phase de vérification post-recherche.

2-1-1 La perception chez les experts

Les experts sont *plus systématiques* et *plus prompts* que les novices à découvrir les lésions radiologiques. Ils adoptent des *stratégies efficaces et spécifiques de recherche organisée* des anomalies. Ces stratégies sont guidées par les lésions détectées et leurs connaissances préalables en anatomie, en pathologie, en clinique ainsi que leur expérience radiologique (H. L. Kundel & Wright, 1969; Raufaste & Eyrolle, 1998). Sur une radiographie pulmonaire par exemple, **1)** Ils se font une **représentation mentale** des images perçues par l'ensemble de la rétine grâce à un bref balayage visuel délibéré et circonférentiel du thorax, centré sur le poumon. **2)** Ce qui leur permet de **repérer rapidement** les zones potentiellement anormales (H. L. Kundel & LaFollette, 1972; Llewellyn Thomas & Lansdown, 1963). Cet échantillonnage visuel est influencé par la tâche à accomplir et le contenu de l'image. La distribution des points fixés est spécifique à chaque individu. **3)** Ensuite, **ils scrutent** délibérément et successivement les détails observés, ils **sélectionnent** les zones initialement identifiées comme suspectes et les caractérisent. Simultanément, ils tentent de découvrir d'autres anomalies dans des zones à haut risque de lésions. Ce qui conduit à une **prise de décision perceptuelle** d'anomalie et à l'élaboration d'une liste d'hypothèses diagnostiques initiales. **4)** En même temps, ils continuent à scruter et à définir les lésions identifiées, pour **valider et nuancer leurs hypothèses**, de la plus probable à la moins probable (H.L. Kundel, 2000; Nodine & Mello-Thoms, 2000).

2-1-2 La perception chez les novices et les résidents

Les connaissances anatomiques et cliniques des novices et des résidents sont en construction et leur exploration visuelle est large et aléatoire. Sur une radiographie pulmonaire, par exemple, leur regard s'attarde sur les contours importants, comme la silhouette cardiaque ou la paroi thoracique. Alors que les radiologues expérimentés

identifient la majorité des lésions présentes (vrais positifs) du premier coup d'œil, en 200 ms ou moins, (Mello-Thoms et al., 2005; Mugglestone, 1995), les résidents sont plus lents et en découvrent moins. Plus ils mettent de temps à observer la radiographie, plus ils rapportent de faux positifs (lésions n'existant pas) (Christensen et al., 1981). Contrairement aux experts, la plupart des faux négatifs (vraies lésions non identifiées) sont dus à une erreur de détection et non à une erreur d'identification (Mello-Thoms, 2003). Leurs capacités de détection s'améliorent avec l'entraînement et l'exposition aux images radiologiques. Ce qui induit un processus d'apprentissage perceptuel automatique, guidé par les besoins de la tâche visée (Ahissar & Hochstein, 2004).

2-1-3 Évolution de l'apprentissage perceptuel

Le traitement de l'information visuelle est hiérarchisé. Il comprend des niveaux perceptuels inférieurs de détection de composantes élémentaires, comme des lignes ou des contrastes, et des niveaux perceptuels supérieurs. Ceux-ci correspondent à la représentation et à la catégorisation d'objets complexes que l'individu confronte aux références emmagasinées dans sa mémoire visuelle. Selon Ahissar et Hochstein (2004), l'amélioration de la perception suit une hiérarchie inverse ou descendante. L'apprentissage perceptuel à des niveaux inférieurs serait dicté par le besoin et ne surviendrait qu'après un apprentissage perceptuel de haut niveau (cf. figure 4, page 18). Autrement dit, l'individu développe une expérience perceptuelle graduelle. Il commencerait par assimiler la reconnaissance d'une image dans son ensemble et ne s'initierait à l'identification des images élémentaires, que si la tâche à accomplir est difficile et nécessite des éléments de différenciation. Par exemple, un enfant reconnaît une voiture dans la rue par sa forme générale. Par contre, s'il aperçoit un engin inhabituel et difficile à identifier, il analysera ses différentes composantes et affirmera qu'il s'agit d'une voiture s'il reconnaît les éléments carrosserie, volant et 4 roues. En fait, lorsque la distinction d'une image est ardue, en raison d'un important bruit de fond comme en radiologie, l'identification des stimulus élémentaires favorise la perception et la classification de l'information (Garrigan & Kellman, 2008). Cette théorie de hiérarchie inverse de l'apprentissage perceptuel est appuyée par des données expérimentales en physiologie. Sowden et coll. (2000) ont démontré que l'expérience en radiologie améliore la perception visuelle d'anomalies difficiles à détecter, faisant

intervenir des zones cérébrales associées à la mémoire visuelle et au processus de remémoration (Haller & Radue, 2005).

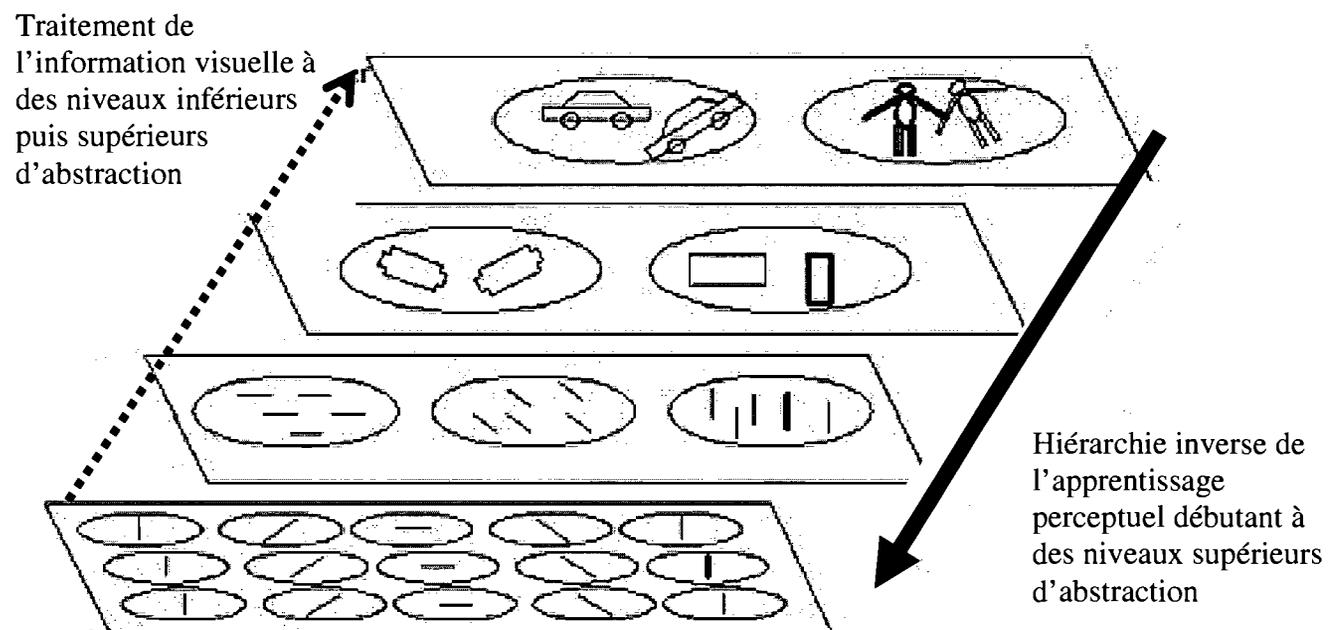


Figure 4. Hiérarchie inverse de l'apprentissage perceptuel

2-2 L'approche cognitive de l'expertise radiologique

Cette approche découle de travaux sur l'organisation des connaissances des experts. Elle met en relief les liens entre la représentation mentale du cas radiologique et les schémas anatomiques, radiologiques et cliniques activés. Lesgold et coll. (1988) ont analysé les compte rendus verbaux d'interprétations de radiographies pulmonaire, provenant de radiologues ayant dix ans d'expérience au minimum et de résidents ayant différents niveaux de formation. À l'instar des études basées sur les stratégies de recherche visuelle, ils retrouvent une étape initiale de *prise de décision perceptuelle*, découlant d'une représentation préalable de l'anatomie thoracique normale, stockée dans la mémoire sous forme de schémas. Ceux-ci correspondent à des cadres mentaux de représentations anatomiques normales permettant de les discerner des anomalies. Elle est suivie d'une étape de *prise de décision analytique* dans un cadre cognitif de résolution de problèmes diagnostiques, qui induit la transformation des images perçues en constatations sémantiques radiologiques à la base du raisonnement diagnostique.

2-2-1 Organisation cognitive chez les experts

Les travaux de Lesgold et coll (1981; 1988) ont démontré que les experts possèdent des *schémas d'anatomie radiologique normale riches, structurés et discriminants*. Ils détectent plus rapidement et davantage d'anomalies que les résidents en examinant une radiographie et ils réactivent et identifient d'emblée leurs schémas. Ce qui leur permet de se consacrer à la recherche d'indices signalant des anomalies potentielles. La réactivation de schémas normaux sert également de référence pour orienter la recherche et l'identification de lésions. Les experts se distinguent aussi des novices par leur aptitude à *regrouper un grand nombre de données observées* et de meilleures *capacités à établir de multiples liens* entre les éléments constatés et leurs connaissances préalables. Ils intègrent, aisément, leurs découvertes radiologiques *dans de grandes catégories de représentations mentales du problème radiologique* (A. M. Lesgold et al., 1988; Myles-Worsley, Johnston, & Simons, 1988; Raufaste & Eyrolle, 1998). Après une brève exposition, ils génèrent d'emblée, au cours des 2 premières secondes (A. M. Lesgold et al., 1988), des *hypothèses diagnostiques hiérarchisées*. Leurs modes de résolution de problèmes spécifiques restent *flexibles et adaptés à la situation*. Ils ont

plus d'aisance à les rectifier s'ils ne concordent pas avec le contexte, ou si la perception initiale est incomplète ou inadéquate. Ils démontrent également une plus grande aptitude à *généraliser* leurs découvertes radiologiques spécifiques, en puisant à travers l'histoire clinique et les particularités anatomiques du patient.

Raufaste et coll (1998) distinguent deux catégories d'experts, les « experts de base » et les « super experts », dans la résolution de cas radiologiques difficiles. Le raisonnement des experts de base est efficace, mais serait restreint à l'identification des indices et à la vérification des seules hypothèses diagnostiques qui en découlent. Celui des super experts, définis comme des experts dont les *raisonnements complexes comportent de nombreux pas d'inférences successifs*, est plus approfondi (Raufaste et al., 1998). Ils adoptent des stratégies mixtes de recherche topographique associées à un raisonnement exhaustif, qui leur permet de découvrir des anomalies inattendues et de générer davantage d'hypothèses pertinentes. En plus de leur expertise en radiologie, les super experts disposent souvent d'un champ de pratique ciblé à un domaine spécifique de l'imagerie médicale et d'une expérience académique et de recherche, qu'ils mettent à profit en cas de difficulté diagnostique.

2-2-2 Organisation cognitive chez les novices et les résidents

L'approche des novices et des résidents intermédiaires est plus laborieuse que celle des experts (A. M. Lesgold et al., 1988). Les structures qu'ils perçoivent sont plus petites, moins organisées donc plus nombreuses et plus superficielles. Une fois l'anomalie détectée et identifiée, leur raisonnement est moins profond et varié. Souvent, ils ne disposent pas encore des connaissances nécessaires pour résoudre le problème, ni des habiletés leur permettant d'établir des liens entre les observations qu'ils font et leurs connaissances théoriques. Ils se limitent à la description des lésions et émettent plus rarement des hypothèses diagnostiques que les experts (A. Lesgold et al., 1981; A. M. Lesgold et al., 1988; Raufaste & Eyrolle, 1998; Raufaste et al., 1998).

2-3 Liens entre la perception et la cognition en radiologie

Le diagnostic radiologique est donc un processus constructif impliquant la détection d'images anormales et l'activation précoce d'un schéma qui oriente la recherche visuelle et l'élaboration de l'hypothèse diagnostique. Les théories, perceptuelle et

cognitive du diagnostic radiologique sont complémentaires. L'interprétation des images radiologiques n'est pas purement perceptuelle et fait intervenir des habiletés cognitives d'une part. D'autre part, le raisonnement radiologique ne peut se faire adéquatement si la base perceptuelle est déficiente. L'approche perceptuelle s'appuie sur la théorie de détection du signal visuel de l'anomalie radiologique. La perception d'une anomalie est assimilée à la dimension « discrimination visuelle » entre un signal anormal et le bruit de fond composé par les images normales. La prise de décision d'anormalité correspond aux dimensions « seuil d'identification » et interprétation. Dans l'approche cognitive, la composante perceptuelle est assimilée à l'étape initiale de « reconnaissance non-verbale » de l'anomalie et la composante cognitive à son interprétation. Grâce à ces deux approches, on a pu démontrer que l'expertise radiologique est caractérisée par une aisance dans la détection et la reconnaissance de l'image anormale, la sélection d'hypothèses diagnostiques appropriées et une grande souplesse dans l'ajustement de la décision diagnostique si de nouvelles informations pertinentes sont disponibles.

Les progrès récents de l'imagerie cérébrale fonctionnelle et de la neurobiologie permettent une meilleure compréhension de ces deux habiletés perceptuelles et cognitives. La propagation de l'information, par les voies visuelles, active des zones spécifiques du cortex cérébral. Ce qui exercerait un effet de facilitation descendante de la reconnaissance préliminaire d'une image donnée et induit une comparaison avec les représentations mentales de référence (Bar et al., 2006). L'attention sélective oriente la recherche visuelle et le mouvement des yeux dans le recueil précoce des éléments évidents composant les images balayées par la rétine. Avec l'expérience, le système visuel construit une carte de saillance correspondant à une carte topographique, qui combine les éléments visuels saillants et une évaluation cognitive de leur pertinence (Thompson, Bichot, & Sato, 2005; Treue, 2003).

2-4 Raisonnement radiologique et scripts radiologiques

Le raisonnement clinique est-il transposable au raisonnement radiologique ? Peut-on parler de scripts radiologiques? Les radiologues sont des médecins qui doivent raisonner à partir d'informations visuelles. Même si la base de leur raisonnement diffère de celle des cliniciens, ils suivent des étapes similaires. (1) Ils découvrent des anomalies et se

font une représentation mentale de la situation, (2) ils cherchent d'autres anomalies, (3) ils confrontent leurs trouvailles aux données cliniques et posent leur diagnostic. La représentation mentale induirait la recherche de nouvelles anomalies visuelles par l'activation de « scripts radiologiques ». Le raisonnement à partir de scripts radiologiques se distingue de celui des scripts de maladie, par la disponibilité immédiate des informations sur la radiographie. Tandis que le clinicien doit rechercher par d'autres moyens les éléments pertinents du script activé, le radiologue les visualise instantanément. Sa représentation mentale initiale renferme souvent la plupart des éléments du script radiologique dont l'activation se ferait selon le mode non analytique, par la reconnaissance de configuration (H. L. Kundel & Nodine, 1983; G. R. Norman, Coblenz et al., 1992). Si les images sont typiques du script activé, le diagnostic est confirmé, mais le radiologue continuera ses investigations visuelles à la recherche d'éléments en rapport avec la situation clinique et ou d'autres anomalies concomitantes. Dans certains cas, les anomalies sont plus ambiguës et évoquent plusieurs hypothèses et la reconnaissance de configurations n'est plus suffisante pour poser le diagnostic. Le script radiologique serait alors activé par voie analytique. Selon Norman(1992) et Lesgold (1988), le raisonnement analytique en radiologie serait descendant pour résoudre le problème.

3 La variabilité du diagnostic en radiologie

Le cheminement vers le diagnostic n'est pas identique chez les médecins (H. Barrows, Norman, Neufeld, & Feightner, 1982; A. S. Elstein et al., 1978). Leurs stratégies de collecte de données et d'évocation des hypothèses sont différentes (M. R. Nendaz et al., 2004). Malgré cette variabilité, ils aboutissent au bon diagnostic, car (1) ils se font une bonne représentation mentale du problème; (2) leur plan de collecte des informations relatives au problème est cohérent et ciblé; (3) la validation des informations obtenues est efficace; (4) ils sont capables de synthétiser les données recueillies et de redresser leur point de vue au cas où ils s'éloigneraient du bon diagnostic (Hasnain, Bordage, Connell, & Sinacore, 2001; M. R. Nendaz et al., 2004).

En radiologie, le niveau élevé de précision et de certitude visés. la variabilité diagnostique reste toutefois non négligeable. Elle est estimée entre 2 et 40% par

différentes études (Cochrane & Garland, 1952; Elmore, Wells, Lee, Howard, & Feinstein, 1994; Garland & Cochrane, 1952; Gunderman & Nyce, 2002; Hopstaken, Witbraad, van Engelshoven, & Dinant, 2004; Krupinski, 2003; Lilienfeld & Kordan, 1966; Llewellyn Thomas & Lansdown, 1963; G. R. Norman, Coblenz et al., 1992; Robinson, 1997). Si cette variabilité relève parfois d'erreurs de perception ou de prises de décision, la composante inhérente au caractère ambigu de l'image n'est pas négligeable et entraîne une divergence d'opinion et de cheminement diagnostique entre les experts (Gunderman & Nyce, 2002; Robinson, 1997).

3-1 Erreurs de perception et d'interprétation

La perception et l'interprétation des images radiologiques peut être sujette à des erreurs d'origine technique, des erreurs inhérentes aux modalités d'acheminement du stimulus visuel à travers le système neurosensoriel, ou des erreurs d'interprétation par le système psychophysiologique.

3-1-1 Paramètres physiques et techniques

Les paramètres techniques, comme la netteté des contours, la différence de densité, le niveau de luminance, le bruit de fond, la longueur d'ondes de la lumière et la complexité des formes et le positionnement du patient influencent la discrimination des signaux visuels.

3-1-2 Facteurs physiologiques

Les fonctions de la rétine et son architecture cellulaire déterminent le pouvoir de résolution et de contraste. Des images perçues par la fovéa, ou zone de vision centrale, auront une résolution spatiale maximale, alors que celles balayées par la rétine périphérique, dont la discrimination est moindre, seront moins précises.

3-1-3 Conditions psychophysiologiques

- Les facultés d'accommodation visuelle et d'adaptation à la lumière, ou la fatigue visuelle modifient la perception des images.
- *La disponibilité d'une histoire clinique préalable* améliore la précision de la lecture des images radiologiques (Loy & Irwig, 2004). Cela cible l'attention de l'observateur vers la recherche et la détection d'éléments pertinents d'une part et

leur interprétation dans un contexte clinique définit d'autre part (Leblanc, Brooks, & Norman, 2002). Différentes études montrent que la disponibilité d'indices cliniques améliore le taux de vrais diagnostics positifs détectés sur les clichés radiologiques sans pour autant augmenter les faux diagnostics négatifs (Berbaum et al., 1988; Berbaum et al., 1986; G. R. Norman, Coblenz et al., 1992).

- *Le mode de balayage oculaire de l'image radiologique scrutée et la reconnaissance des images* joue un rôle dans les faux diagnostics positifs ou négatifs. Il est spécifique à chaque individu et il n'est jamais uniforme sur le cliché à observer (Llewellyn Thomas & Lansdown, 1963).
 - Les faux diagnostics positifs correspondent généralement à une attraction du regard par un leurre visuel composé de différentes structures anatomiques superposées et induisant plusieurs observateurs en erreur. Plus rarement, il s'agit de faux diagnostics positifs sporadiques, détectés par un seul individu et ne correspondant à aucun substratum anatomique ni pathologique. Ils sont habituellement générés par une plus longue fixation visuelle de la radiographie induisant une fausse impression d'une anomalie (H. L. Kundel et al., 1989).
 - Une fois les facteurs techniques éliminés, les faux diagnostics négatifs sont principalement dus à l'observateur. Leurs conséquences peuvent être graves, comme dans le cas d'un nodule pulmonaire cancéreux qui n'est découvert qu'en rétrospectif. Plusieurs catégories d'erreurs sont décrites (Krupinski, 2003) : 1) L'erreur de recherche visuelle découle de l'absence de fixation de la zone d'intérêt par la rétine fovéale. 2) L'erreur de reconnaissance résulte d'une fixation fovéale de l'anomalie sur une durée insuffisante ne permettant pas de la reconnaître, surtout si elle est discrète. 3) L'erreur de décision ou de jugement relève d'une incapacité à reconnaître consciemment, en tant qu'anomalie, une image pourtant longuement fixée par la fovéa. 4) L'erreur due à la satisfaction de la recherche provient d'un arrêt prématuré de la recherche visuelle dès qu'une anomalie importante est découverte, malgré la présence d'autres

lésions subtiles et souvent pertinentes (Samuel, Kundel, Nodine, & Toto, 1995). À l'inverse, la présence de plusieurs anomalies sur un même cliché radiographique, peut détourner l'attention de l'observateur d'une lésion importante (Fuhrman et al., 2002).

3-1-4 Niveau d'expérience

Les connaissances et l'expérience préalables orientent la détection, la sélection, la combinaison des signaux visuels et la formulation d'hypothèses diagnostiques appropriées. Les experts décident de la probabilité que le signal perçu soit un vrai positif significatif, ce qui réduit incontestablement l'erreur diagnostique.

3-2 Ambiguïté des images

La certitude diagnostique n'est pas toujours possible en radiologie. Dans le cas des anomalies subtiles, difficiles à détecter, complexes, ou atypiques, le diagnostic est émis avec un niveau de probabilité plus ou moins élevé. Une divergence d'opinion entre les experts est donc possible sans refléter forcément une erreur (Hobby, Tom, Todd, Bearcroft, & Dixon; Robinson, 1997). Robinson (1997) a schématisé les concepts d'erreur et de variabilité dans le diagnostic radiologique à l'aide d'un diagramme (cf. figure 5, page 24) décrivant la relation entre le niveau de certitude diagnostique (axe des abscisses) et l'évidence de l'anomalie (axe des ordonnées). Le point (A) désigne les anomalies évidentes et facilement identifiables, où le diagnostic peut être énoncé avec exactitude, une bulle d'emphysème ou un pneumothorax par exemple. Le point (B) représente des anomalies qui, malgré leur subtilité, sont identifiées avec certitude, tel un épanchement pleural liquidien minime. Au point (C), les anomalies sont évidentes, mais leur nature est incertaine, comme dans le cas d'une opacité pulmonaire pouvant évoquer une tumeur ou une image construite. Enfin, d'autres anomalies sont trop discrètes ou trop atypiques et sont identifiées avec un très faible degré de certitude même par les experts dont l'opinion peut diverger (point D). Tout écart de diagnostic engendrerait une erreur diagnostique pour les anomalies situées autour de la zone d'extrême certitude (point A). Il serait considéré comme une divergence dans l'opinion des experts pour celles qui sont situées dans la zone d'extrême ambiguïté (point D). La réponse exacte à ces images équivoques ne peut être connue que par l'évolution du patient, d'autres

méthodes diagnostiques de référence (imagerie plus sophistiquée, exploration chirurgicale, ou résultats de pathologie). La divergence d'opinion entre les experts est alors « permise » et la frontière entre les zones d'erreur et de variabilité d'opinion est déterminée par les limites de l'acceptabilité professionnelle. Dans des situations problématiques, cette divergence est aussi utile que dans les autres spécialités médicales. Elle permet au radiologue de nuancer son opinion selon le contexte clinique et de garder une ouverture sur d'autres possibilités diagnostiques (Gunderman & Nyce, 2002).

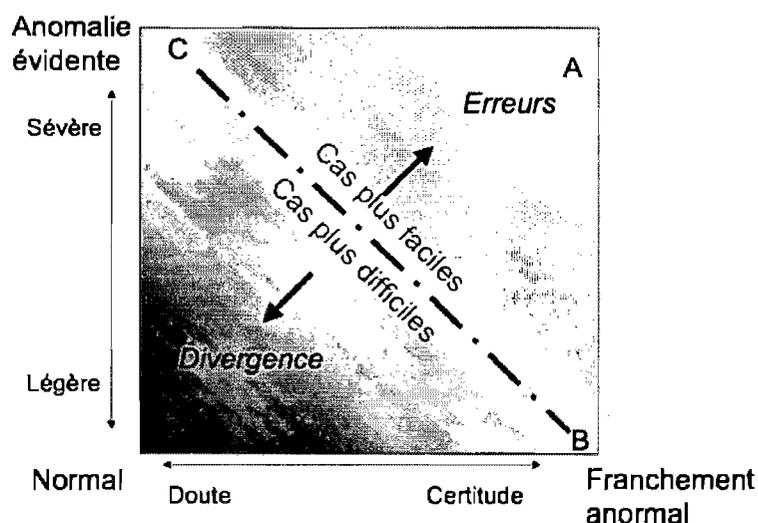


Figure 5. Niveaux d'évidence et reconnaissance d'anomalies radiologiques

4- Cas particulier des radiographies pulmonaires

La radiographie pulmonaire (RxP) est un examen peu coûteux, très utilisé dans la pratique médicale courante. Il oriente le diagnostic, le choix d'investigations plus poussées et les décisions thérapeutiques. On y a recours dans diverses situations, comme la présence de symptômes respiratoires ou non-respiratoires, dans le cadre de bilans préopératoires ou de programmes de dépistage de maladies respiratoires. Tout manque de précision dans la lecture de cet examen entraînerait des erreurs et des retards au diagnostic qui sont possiblement néfastes à la santé du patient. La variabilité dans la

lecture des RxP n'est cependant pas négligeable (Bloomfield, Teele, Voss, Knight, & Harding, 1999; Cherian et al., 2005; Garland & Cochrane, 1952; Hopstaken et al., 2004; Lilienfeld & Kordan, 1966; Sarria et al., 2003). Près de 20% des nodules pulmonaires étaient rétrospectivement présents sur des RxP effectuées plusieurs mois avant le diagnostic d'un cancer des poumons d'après l'étude de Quekel et coll (1999). En dehors des erreurs d'observation détaillées au §3-1 (cf. figure 5, page 24), l'incertitude diagnostique est générée par la complexité et la superposition des structures intrathoraciques. Ce qui rend les anomalies subtiles, atypiques ou ambiguës. Les hypothèses diagnostiques plausibles sont alors émises avec des probabilités dépendant du niveau de confiance et d'expertise de l'observateur. En situation réelle, les radiologues examinent les RxP à la demande des cliniciens et disposent de renseignements cliniques succincts pour orienter leurs recherches. Trois catégories de réponses sont généralement émises 1) absence d'anomalie ou variante anatomique normale; 2) image de composition, ou artéfact dû à des problèmes techniques, ou à la superposition de structures anatomiques; 3) RxP franchement anormale. Une fois l'anomalie identifiée, elle est classée dans une catégorie diagnostique et la formulation du diagnostic variera selon le niveau de certitude. La figure 6 (cf. figure 6, page 26) propose un schéma du processus diagnostique radiologique, où les hypothèses évoquées activeraient des scripts radiologiques par voie analytique ou non analytique selon la situation.

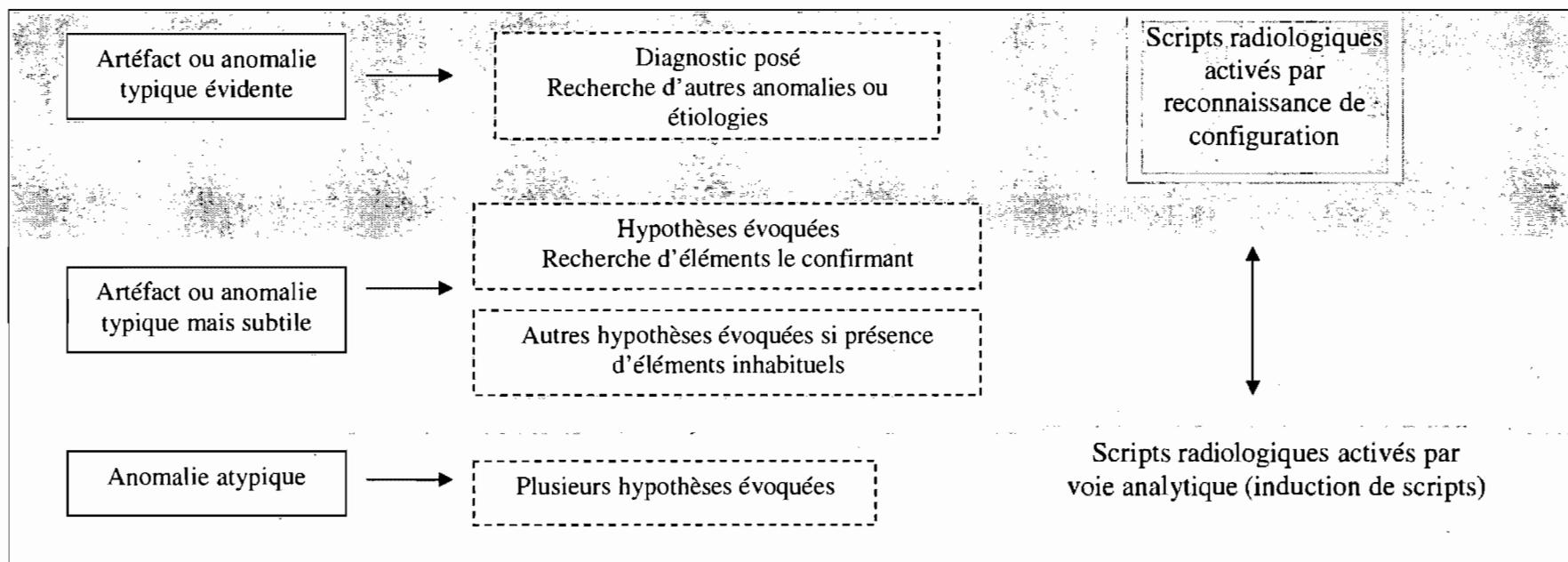
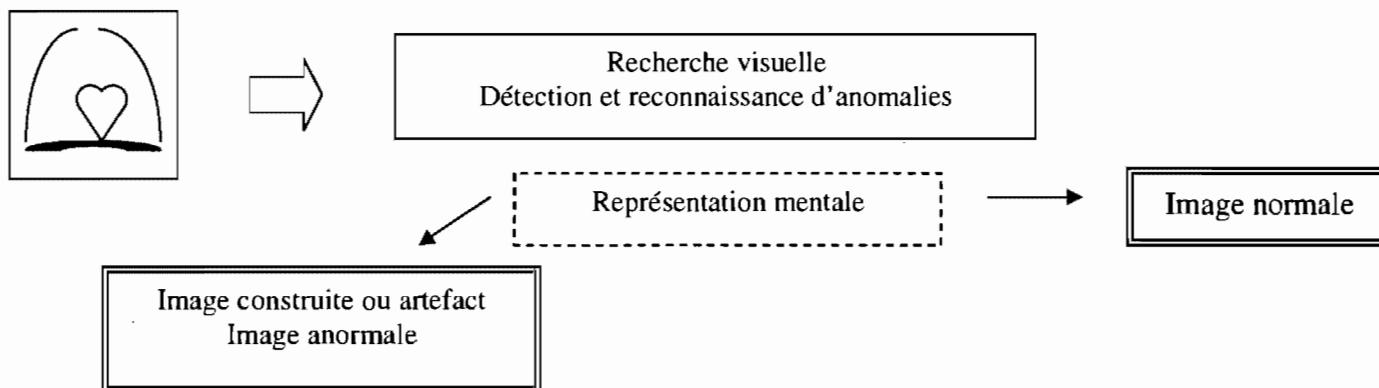


Figure 6. Schéma du processus diagnostique d'une radiographie pulmonaire

1. Les images évidentes et typiques sont d'emblée identifiées par reconnaissance de configuration et le diagnostic est posé avec certitude. On parle d'images pathognomoniques où les possibles diagnostics différentiels sont automatiquement écartés. Le radiologue s'attarde plutôt à rechercher des éléments indiquant une étiologie, le niveau de gravité, ou des diagnostics concomitants. Cette situation se voit par exemple dans le pneumothorax, qui correspond à un épanchement gazeux dû au décollement du poumon de la paroi thoracique et dont les images sont typiques en général (cf. figures 6 et 7, pages 26 et 28).
2. Si l'anomalie est relativement typique mais subtile (cf. figures 6 et 8, pages 26 et 29), ou si elle prête à confusion avec une variante de la normale ou une image construite (cf. figures 6 et 9, pages 26 et 30), le script radiologique de l'hypothèse la plus probable serait activé par reconnaissance de modèles et des éléments visuels pouvant la confirmer sont recherchés. Simultanément, l'induction d'autres scripts se ferait par voie analytique. Une liste de diagnostics différentiels, plus ou moins probables, est alors émise. Le radiologue complètera son diagnostic à l'aide d'autres investigations : des incidences radiographiques ou un scanner thoracique par exemple.
3. Enfin, lorsque l'image est plus ambiguë, évoquant plusieurs hypothèses difficiles à départager, le radiologue recherche des arguments visuels et cliniques pour pondérer et prioriser un nombre restreint d'hypothèses (3 ou 4 en général). Chacun de ces arguments va appuyer ou non l'hypothèse évoquée pour aboutir au diagnostic le plus probable (cf. figures 6 et 10, pages 26 et 31).

En résumé, à l'instar du clinicien, le radiologue est confronté à des problèmes diagnostiques de complexité variable, auxquels il répond par une mobilisation adaptée de ses compétences. Les cas les plus simples et les plus évidents sont résolus sur un mode économique mais efficace de reconnaissance de modèles suivi d'une analyse consciente pour valider l'impression diagnostique. Dans les cas moins typiques et moins évidents, il emprunte une voie davantage analytique afin de confirmer ou de réfuter ses hypothèses. Les diagnostics possibles sont hiérarchisés du plus au moins probable.



Représentation mentale

Pneumothorax sous tension

Script radiologique activé pneumothorax sous tension

- Visualisation de la ligne pleurale viscérale
- Collapsus complet (ou partiel) du poumon
- Hyperclarté de l'hémi thorax droit
- Distension de l'hémi thorax droit et aplatissement de la coupole diaphragmatique
- Refoulement du médiastin et de la trachée
- Emphysème sous-cutané plus accessoirement

Diagnostic

Pneumothorax sous tension

Figure 7. Exemple d'une image typique et évidente

L'expert reconnaît immédiatement le script dans son ensemble et s'attache à rechercher d'éventuels signes de gravité ou étiologiques, comme une bulle d'emphysème ou une fracture de côtes. Il identifiera également des détails plus secondaires comme la présence de tubes endotrachéal et naso-gastrique indiquant que le patient est dans un état critique et probablement hospitalisé aux soins intensifs.



Représentation mentale

Opacité sphérique unique lobe supérieur gauche = nodule unique du lobe supérieur du poumon gauche

Principal script radiologique activé Nodule pulmonaire malin

- Localisation : lobes **supérieurs**
- Diamètre de l'ordre de **2 cm**
- Aspect du contour et homogénéité ?
- Présence d'autres nodules ?
- Âge du patient ? Tabagisme ? Toux ?
hémoptysie ? histoire de cancer ?

Hypothèses diagnostiques

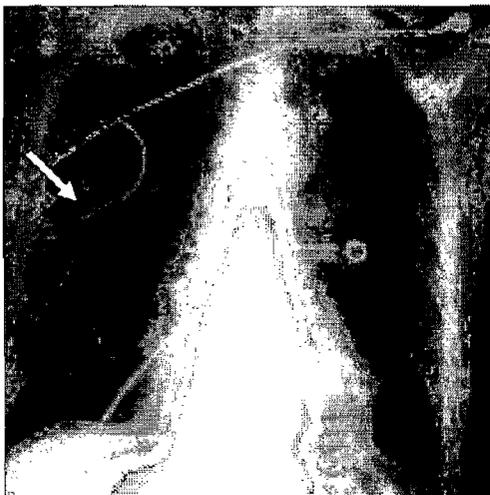
Cancer pulmonaire
Métastase d'un autre cancer
Tumeur bénigne
Granulome infectieux

Classification des diagnostics possibles

Probables : cancer pulmonaire vs métastase d'un autre cancer
Possibles mais moins probables : tumeur bénigne, granulome infectieux

Figure 8. Exemple d'une anomalie subtile

La représentation mentale induit plusieurs hypothèses, mais active principalement le script des nodules d'origine maligne. Le radiologue va rechercher des éléments visuels et cliniques en faveur de ce diagnostic. Simultanément, il active d'autres scripts d'hypothèses possibles. Un scanner thoracique sera nécessaire pour faire avancer le diagnostic.



Représentation mentale

Opacité pulmonaire à bord linéaire gauche avec visualisation du parenchyme pulmonaire au-delà du trait linéaire

Hypothèses diagnostiques

Pli cutané ou empreinte de draps
Pneumothorax partiel

Principal script radiologique activé : Pneumothorax partiel

- Visualisation de la ligne pleurale viscérale ?
- Collapsus partiel du poumon ?
- Hyperclarté de l'hémi-thorax droit ?
- Absence de parenchyme pulmonaire au-delà de la ligne pleurale ?

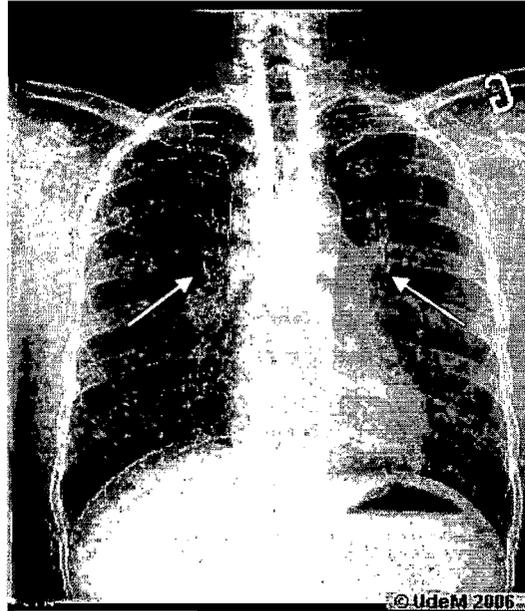
Classification des diagnostics possibles

Probable : pli cutané dû à chez un patient alité (empreinte de drap)

Très peu probable : pneumothorax partiel

Figure 9. Exemple d'une image compatible avec une image construite

La représentation mentale génère les hypothèses d'image construite par la superposition de replis cutanés, chez un patient alité vs celle d'un pneumothorax. Et active le script de pneumothorax linéaire. Dans ce cas, on s'attachera à confirmer l'absence de tous les éléments de ce script avant d'affirmer qu'il s'agit d'une image de composition. Au moindre doute, d'autres incidences radiographiques voire un scanner seront effectuées pour départager les hypothèses.



Représentation mentale

Adénopathies (ganglions) hilaires bilatérales

Principales hypothèses diagnostiques

- Sarcoïdose
- Lymphome
- Métastases ganglionnaires
- Tuberculose

Classification des hypothèses possibles

Le plus probable : Sarcoïdose
*Possibles mais moins probables :
 tuberculose, métastases, lymphome*

Scripts radiologiques activés

- **Sarcoïdose**
 Adénopathies hilaires symétriques
 Bien définies
 Possibles images réticulo-nodulaires
 prédominant aux sommets
- **Métastases ganglionnaires**
 Adénopathies asymétriques en général
 Parfois associées à des métastases
 intrabronchiques
 Contexte néoplasique (tête et cou,
 thyroïde, génito-urinaire, mélanome...)
- **Tuberculose**
 Adénopathies asymétriques en général
 Rares chez les adultes
 Épanchement pleural associé possible
- **Lymphoma**
 Adénopathies asymétriques en général
 Prédominant au médiastin supérieur

Figure 10. Exemple d'une image ambiguë

La représentation mentale génère les hypothèses les plus pertinentes et active les scripts radiologiques ayant pour point commun l'élément « adénopathies hilaires ». Les hypothèses sont hiérarchisées selon la présence ou l'absence d'éléments propres à chaque script et le contexte clinique du patient. Le diagnostic de sarcoïdose paraît le plus probable car les adénopathies sont symétriques et on devine la présence d'images réticulo-nodulaires. Les hypothèses de tuberculose, lymphome ou métastases sont possibles, mais paraissent moins probables car les adénopathies y sont généralement plus asymétriques. Dans tous les scripts activés, l'âge et l'histoire du patient ainsi que d'autres éléments cliniques ou paracliniques peuvent fournir des informations précieuses. Le diagnostic final ne sera confirmé qu'avec des investigations plus poussées.

5- Évaluation du raisonnement radiologique en contexte d'incertitude

Plusieurs tests à visée formative ou sommative ont fait leurs preuves, depuis de nombreuses années, dans l'évaluation des compétences des résidents en radiologie, pendant et à la fin de leur formation. Les QCM, les ECOS et les épreuves orales figurent parmi les trois formats d'évaluation des examens de certification du collège royal.

5-1 Les tests classiques

5-1-1 QCM

Dans le cadre des examens de certification, l'épreuve des QCM comprend 180 questions à choix multiple, d'une durée de trois heures. Elle évalue surtout les connaissances factuelles médicales chirurgicales et fondamentales (CRMCC, 2007a). Dans d'autres situations, les QCM comprennent des questions d'ordre diagnostique radiologique (McCoubrie & McKnight, 2008). Le candidat doit alors choisir, parmi 4 ou 5 propositions, le diagnostic le plus probable après avoir observé une radiographie. Ce type d'épreuve est objectif et mesure avec précision les connaissances du candidat. Il reflète sa compétence à reconnaître des anomalies dont le diagnostic est évident et il possède un bon pouvoir prédictif, même s'il ne mesure que des connaissances factuelles (B. Charlin, Bordage, & Van der Vleuten, 2003). Les QCM ne reflètent néanmoins pas l'opinion du candidat sur de potentiels diagnostics alternatifs dont le niveau de certitude est moindre. Ils ont également l'inconvénient d'orienter et de suggérer la réponse parmi les propositions énoncées.

5-1-2 ECOS

Dans l'examen du CRMCC, les ECOS sont composés de stations d'examen d'une durée totale de 2 heures et 45 minutes. Le candidat observe des radiographies et doit reporter et identifier, sur un schéma, les anomalies détectées. La grille de correction comporte des critères standards relatifs aux éléments présents sur le cliché radiologique et assurant l'objectivité des scores. Ils testent aussi bien les capacités de détection et d'interprétation des anomalies que les connaissances anatomiques. Les ECOS sont objectifs, standardisés et génèrent moins de stress qu'une épreuve orale (R. Peterson, 2000). Ils permettent de déceler les défaillances dans la détection et l'interprétation (R. Peterson, 2000), ainsi que le manque de connaissances cliniques (Morag et al., 2001).

Elle ne fournit toutefois qu'une idée ponctuelle sur l'attitude du candidat devant les images observées. Elle ne reflète pas spécifiquement l'organisation des connaissances du candidat, ni son raisonnement radiologique.

5-1-3 Épreuve orale

Pendant la partie orale des examens du CRMCC, les candidats discutent des cas présentés, devant un jury pendant 3 heures. Le tiers du temps est alloué à l'imagerie thoracique et mammaire. Elle porte principalement sur les capacités de détection et d'interprétation des images. Cette épreuve interactive évalue l'attitude du candidat et son raisonnement radiologique devant une situation ambiguë. L'épreuve orale peut constituer un facteur de stress pour le candidat qui doit « performer » devant un examinateur. Cette part de subjectivité dans l'interaction entre le candidat et les examinateurs, constitue un biais potentiel de l'évaluation. De plus, l'aspect variabilité entre les experts y est probablement estompé, puisque les questions choisies font généralement consensus entre les examinateurs.

En somme, l'évaluation spécifique et objective du raisonnement radiologique en contexte d'incertitude, à l'aide de QCM, d'ECOS ou d'épreuves orales, paraît peu appropriée. Les grilles de corrections de ces tests sont basées sur le principe de consensus des examinateurs et les habiletés d'un candidat à détecter et interpréter des images pour lesquels les experts divergent, risquent de ne pas être évalués par de tels tests.

5-2 Le test de concordance des scripts (TCS)

Le test de concordance de scripts s'inspire de la théorie des scripts en psychologie cognitive (B. Charlin, Tardif et al., 2000). Il vise à mesurer le raisonnement utilisé pour résoudre des problèmes cliniques en contexte d'incertitude, également qualifiés par les termes « mal définis », ou « complexes ». Il permet d'évaluer le cheminement emprunté par le candidat pour confirmer ou éliminer différentes hypothèses diagnostiques, ou options thérapeutiques dans un contexte dynamique d'informations cliniques. De plus, il tient compte de la divergence d'opinion entre les experts du domaine étudié (B. Charlin, Brailovsky, Leduc et al., 1998; B. Charlin, Roy et al., 2000; B. Charlin & van der Vleuten, 2004).

5-2-1 Description du TCS

Le TCS est composé de cas cliniques complexes présentés dans une vignette. Plusieurs hypothèses diagnostiques, d'investigations, ou thérapeutiques sont possibles pour chaque scénario. Ces hypothèses sont présentées au candidat à la suite de la vignette initiale, sous forme de questions successives. Celui-ci doit se prononcer sur l'effet d'une nouvelle information quant à la pertinence de l'hypothèse en question. L'information est fournie juste après la formulation de l'hypothèse et porte sur des détails aussi bien cliniques que para-cliniques relatifs au cas présenté (cf. figure 10, page 31). Chaque réponse est nuancée selon une échelle de type Likert à 5 points, allant de l'effet fortement positif à l'effet fortement négatif, en passant par la position neutre (-2, -1, 0, +1, +2). La réponse à chaque question reflète une partie du script de maladie tel qu'il est perçu par le candidat. Les réponses sont comparées à celle d'un panel de référence composé d'experts et le score obtenu pour chaque question reflète la concordance du candidats à ceux des spécialistes du domaine, sur la tâche cognitive spécifiquement demandée.

Les scores sont calculés par la méthode des scores combinés, qui mesurent le degré de similitude entre les réponses données par les candidats et celles données par un panel de référence composé de spécialistes du domaine, tout en tenant compte de leur divergence d'opinion. Le score obtenu à chaque réponse est attribué selon la proportion d'experts ayant répondu comme le candidat. Il correspond au rapport entre le nombre d'experts ayant répondu à cette question et la valeur modale qui exprime le nombre maximum d'experts ayant donné une même réponse à la question. L'exemple suivant illustre le calcul des scores obtenus à une question d'une épreuve où le panel d'experts est composé de 15 membres, dont 8 ont répondu (+2), 5 autres ont choisi (+1) et 2 sont restés neutres (0). Le candidat A dont la réponse est (+2) aura répondu comme 8 experts et obtiendra un score de 1 (8/8). Le candidat B qui a répondu +1 comme 5 experts, obtiendra le score 0,625 (5/8). Le score du candidat C qui aura répondu 0 sera de 0,25 (2/8). Les candidats D et E auront un score de 0 car aucun expert n'aura répondu ni (-1), ni (-2) à cette question (cf. tableau I, page 35).

Tableau I.
Mode d'attribution des scores à une question

Échelle de réponse	-2	-1	0	+1	+2
Réponse des 15 experts	0	0	2	5	8
Score initial	0	0	2/15	5/15	8/15
Score transformé	0	0	2/8	5/8	8/8
Crédit du candidat	0	0	0,25	0,625	1
Candidats	E	D	C	B	A

5-2-2 Études de validation du TCS

Le TCS est un outil d'évaluation relativement jeune, qui a évolué pendant les dix dernières années. Depuis les premières versions, développées par Charlin (1998; 1998), le test a été appliqué dans divers domaines de la santé médicaux et non-médicaux (Cohen, Fitzgerald, Lane, & Boninger, 2005; Demeester, 2004; Khonputsas, Besinque, Fisher, & Gong, 2006). Ce qui a permis d'analyser ses qualités psychométriques à travers différentes études et de démontrer ses validités de construit et prédictive (Bland, Kreiter, & Gordon, 2005; Brailovsky, Charlin, Beausoleil, Cote, & Van der Vleuten, 2001; Brazeau-Lamontagne et al., 2004; Carrière, Gagnon, Charlin, Downing, & Bordage, 2008; B. Charlin, Desaulniers, Gagnon, Blouin, & van der Vleuten, 2002; B. Charlin et al., 2006; B. Charlin, Gagnon et al., 2007; Cohen et al., 2005; Demeester, 2004; Fournier et al., 2006; Gagnon et al., 2006; Khonputsas et al., 2006; Labelle et al., 2004; Lambert, 2005; Marie et al., 2005; Meterissian et al., 2007; Sibert, Charlin, Corcos, Gagnon, Grise et al., 2002; Sibert, Charlin, Corcos, Gagnon, Lechevallier et al., 2002; Sibert et al., 2001; Sibert et al., 2006).

Le test démontre généralement une bonne fidélité. Le coefficient de Cronbach varie de 0,70 à 0,90, et ce en une heure de temps de passage dans la plupart des tests développés, notamment en gynécologie (B. Charlin, Brailovsky, Leduc et al., 1998), en urologie (Sibert, Charlin, Corcos, Gagnon, Lechevallier et al., 2002; Sibert et al., 2006), en chirurgie générale (Meterissian et al., 2007), en médecine générale (Gagnon, Charlin, Coletti, Sauve, & van der Vleuten, 2005) et interne (Marie et al., 2005), en radio-

oncologie (Lambert, 2005), en médecine d'urgence (Carrière et al., 2008; Fournier et al., 2006) et en soins intensifs (Gibot & Bollaert, 2008). Les études de généralisabilité de Gagnon et coll. (2008) montrent que, pour un temps de passation d'environ 1 heure, la fidélité serait optimale pour un test comportant 3 à 5 questions par cas et un ensemble de 20 à 30 cas par test, lequel devrait comporter un total d'au moins 54 questions. Le degré de précision de l'estimation de la fidélité du test est également meilleure lorsque le panel d'experts comporte au moins 10 membres (15 à 20 si l'enjeu est important) (Gagnon et al., 2005).

Le pouvoir discriminant du TCS est illustré dans les différentes études par la amélioration des scores des candidats avec le niveau d'expérience. Cette amélioration est le plus souvent significative entre les groupes de résidents juniors, seniors et les experts (Lambert, 2005; Meterissian et al., 2007) mais elle paraît moins nette entre les différentes années de résidence. L'effet intermédiaire, qui représente un défaut des tests traditionnels où les scores des experts étant très proches voire inférieurs à ceux des résidents finissants, n'est pas observé (B. Charlin, Roy et al., 2000). Cet effet est sans doute dû à ce que les tests conventionnels mesurent davantage l'interprétation de connaissances factuelles que le processus d'interprétation de données et de prise de décision de la pratique quotidienne. La stabilité interculturelle du test a été démontrée par une étude en urologie, qui a comparé les scores d'étudiants et de résidents appartenant à des universités francophones (canadienne et française) et à une université canadienne anglophone (Sibert, Charlin, Corcos, Gagnon, Grise et al., 2002). Ils ont montré que les scores étaient stables malgré les différences de milieux et linguistiques. Ils ont aussi démontré que les scores étaient plus élevés lorsque la grille de correction était élaborée à partir de réponses d'une panel provenant du pays d'origine. Enfin, le test semble bien mesurer les étapes successives du raisonnement clinique. Selon l'étude de Gagnon et coll. (2006), la vitesse de la réponse aux questions dépend du degré de typicité de l'information fournie. Plus l'information est atypique, plus l'activation du script est lente. Ce qui appuie l'hypothèse à la base de la construction du TCS postulant que l'activation des scripts se ferait principalement par voie non analytique dans un contexte d'incertitude (B. Charlin, Boshuizen et al., 2007).

La validité prédictive a été initialement documentée par Brailovsky et coll. (2001). Ils ont mis en évidence une corrélation entre les résultats obtenus au TCS par des étudiants de médecine et ceux obtenus à la résolution de problèmes par des réponses ouvertes et courtes ($R : 0,45$) et à des tests oraux ($R : 0,45$). Ils n'ont cependant pas observé de corrélation avec les ECOS, qui semblaient tester davantage un savoir faire factuel que pratique et décisionnel. La valeur prédictive du TCS est également confirmée par des travaux plus récents utilisant des examens de fin de stages (Carrière et al., 2008) ou des épreuves d'examens classants nationaux (Gibot & Bollaert, 2008) comme outils de référence. Les résultats des études de corrélation avec des tests de référence ne sont cependant pas toujours significatifs (Demeester, 2004). La nature du test utilisé comme référence serait probablement le principal facteur limitant, étant donné qu'aucune méthode d'évaluation conventionnelle ne mesure, de manière objective et spécifique, le raisonnement clinique jusqu'à présent.

5-3 Utilisation du TCS en radiologie

En radiologie diagnostique, le principe du TCS semble un outil de choix pour évaluer le raisonnement clinique, surtout dans les situations où le diagnostic visuel est ambigu et ne fait pas consensus entre les experts. Il a été appliqué pour mesurer la capacité des étudiants en médecine et des résidents en radiologie à interpréter des radiographies pulmonaires dans une étude exploratoire publiée en 1998 (B. Charlin, Brailovsky, Brazeau-Lamontagne et al., 1998). Les 30 candidats dont 9 experts n'avaient pas accès aux images. Ils devaient répondre à une série de questions relatives à l'effet possible d'un détail observé sur une anomalie radiologique pulmonaire. Chaque cas comportait environ 10 questions et l'échelle de type Likert était de 7 points, allant de l'hypothèse diagnostique complètement exclue, très peu probable, peu probable, un peu plus probable, très probable, ou certaine, avec un point neutre où l'information n'avait aucune influence sur l'hypothèse. L'étude a démontré une bonne fiabilité du test d'interprétation (TCS) avec un coefficient alpha de Cronbach de 0,83 et aucun effet intermédiaire n'a été noté. Les résultats encourageants de cette étude ont conduit au développement d'un TCS radiologique portant sur l'interprétation de vrais films de radiographies pulmonaires (Brazeau-Lamontagne et al., 2004). Les scores du nouveau TCS d'interprétation (TI) ont été comparés aux scores d'un test de perception (TP) dans

un groupe de 60 candidats composé d'externes, résidents juniors (R1 et R2) et résidents seniors (R3, R4, R5). Le panel d'experts était formé de 11 radiologues. Les deux tests portaient sur des clichés différents. Ils ont démontré un bon pouvoir discriminant et le coefficient α de Cronbach était de 0,79 pour le TP et de 0,81 pour le TCS. Les habiletés de perception progressaient plus rapidement que les habiletés d'interprétation, car les scores de perception s'amélioraient avant les scores d'interprétation. Bien que le TP évaluait principalement des habiletés perceptuelles, cette différence de progression et la corrélation entre les deux tests pourraient témoigner indirectement de la valeur prédictive du TCS en radiologie. Les habiletés d'interprétation ne pouvant évoluer que sur une bonne base perceptuelle.

Le format du test d'interprétation utilisé dans cette étude pouvait cependant poser des problèmes de validité apparente, dans le contexte particulier de la radiologie. Les questions successives étaient orientées sur présence éventuelle de détails élémentaires de l'anomalie à diagnostiquer. Or, la disponibilité simultanée de toutes les images élémentaires d'une anomalie radiologique caractérise les spécialités visuelles. Décortiquer les caractéristiques élémentaires de chaque anomalie testée permettait probablement de sonder l'attitude diagnostique de l'observateur, mais ce mode de questionnement est perçu comme artificiel par les radiologues. Il s'apparente davantage à un contexte d'enseignement qu'à leur fonctionnement dans la réalité quotidienne. Ce qui peut mettre mal à l'aise l'observateur, surtout quand l'élément dont il est question est absent de l'anomalie observée. On pourrait même se demander si certaines questions ne testeraient pas plus des compétences factuelles que décisionnelles. Des modifications du format des questions paraissaient donc nécessaires pour faciliter l'utilisation du TCS, le rendre plus réaliste et se rapprocher le plus possible de la dimension résolution de problèmes qu'il vise à évaluer. Le test de perception consistait à répondre à des QCM interrogeant le candidat sur la possible présence d'une anomalie sur une RxP qu'il est en train d'observer. La réponse était donnée parmi 4 choix de réponses. Ce format de test permettait de mesurer jusqu'à une certaine mesure la variabilité des experts en matière de perception. En proposant un choix limité de réponses, l'utilisation de QCM peut avoir cependant orienter la réponse des participants en raison de leur effet suggestif. Enfin, l'utilisation de films comme support des images était contraignante et limitait la diffusion du test et son organisation logistique. Cela exigeait que chaque série d'images

soit examinée séparément par chaque participant, puis de les faire circuler d'un participant à l'autre. La disponibilité de systèmes de stockages numériques permet de contourner cet écueil et de développer des tests en ligne.

5-3 Développement du TCS en ligne

Les technologies de l'information et de la communication sont de plus en plus utilisées dans l'enseignement médical (Lee, 2006; Triola et al., 2006) et surtout dans son évaluation (Frohna, Gruppen, Fliegel, & Mangrulkar, 2006; Hammoud & Barclay, 2002; Hawkins et al., 2004; Lieberman, Frye, Litwins, Rasmusson, & Boulet, 2003; M. W. Peterson, Gordon, Elliott, & Kreiter, 2004). L'informatisation des tests d'évaluation rend plus souple leur organisation et donne un accès immédiat aux scores et à un feedback utile aux étudiants lors des évaluations formatives. Elle permet de stocker les scores et les informations sur les comportements des candidats, qui seront analysables ultérieurement par les examinateurs. Elle rend disponibles des banques de questions facilement accessibles pour la construction de nouveaux tests. La possibilité d'incorporer différents types de supports multimédias comme des images radiologiques ou histologiques, des films vidéo et des sons, multiplie les possibilités d'évaluation et les formats de tests. Enfin, l'utilisation de l'internet offre des opportunités de diffusion des tests et permet une plus grande participation dans différents centres. L'introduction des tests en ligne et de supports multimédias, sont particulièrement intéressants dans l'évaluation du raisonnement clinique par le TCS. Celui-ci a été initialement conçu sous forme d'une version papier et adapté à un format électronique en ligne depuis quelques années (Caire et al., 2004; B. Charlin, Kazi-Tani, Gagnon, & Thivierge, 2005; Sibert et al., 2006; Sibert, Darmoni, Dahamna, Weber, & Charlin, 2005). Sibert et coll. (2006; 2005) ont développé un TCS en urologie et démontré sa faisabilité à travers le Web et à grande échelle. Un format adapté à l'autoévaluation existe en neurochirurgie en France (Caire et al., 2004; Kalamarides, Caire, Dauger, Brassier, & Moreau, 2008), tandis que les neurologues de l'Université McGill à Montréal ont mis au point un TCS en ligne incluant des fichiers d'imagerie cérébrale (Lubarsky).

L'inclusion d'images radiologiques numérique au TCS est rendue possible grâce à la disponibilité du système de stockage numérique. La large diffusion du *PACS* (Picture Archiving and Communication System), un système de stockage numérique des

images, a facilité l'accès et la visualisation des images radiologiques dans les hôpitaux. Grâce à ce concept, l'utilisation de films classiques comme supports d'images n'est plus nécessaire. L'environnement PACS comporte un système numérique d'acquisition, de lecture, de communication et d'archivage des images (Bick & Lenzen, 1999). Les radiologues y ont accès, directement sur leur poste de travail, à travers un logiciel convivial spécialement dédié à la visualisation des images. L'observateur peut intervenir entre autres sur les conditions de visualisation (luminosité et l'agrandissement des images observées), marquer et mesurer les dimensions d'une anomalie notée. Les résidents de radiologie s'entraînent de plus en plus à lire les radiographies sur le PACS et sont devenus moins familiers avec la visualisation des films classiques. Il était donc logique de développer des tests d'évaluation se rapprochant des conditions de pratique quotidienne. Par ailleurs, les images étant stockées numériquement, l'utilisation du PACS évite toute la procédure de déplacer les films vers les lieux de l'examen et de devoir les ramener à l'hôpital après l'examen. Dans un test où les radiographies sont importées du PACS sous un fichier JPEG (Joint Photographic Experts Group), les images seront toujours disponibles dans le dossier informatique du patient, tout en constituant une banque de questions permanentes accessible à partir du lieu de l'examen. Enfin, l'évolution du TCS en radiologie vers un format électronique comporte l'avantage de pouvoir le faire passer à plusieurs participants simultanément et indépendamment de leur localisation géographique. Ce qui facilite sa faisabilité, tenant compte de leur agenda chargé. Toutefois, en dehors des nombreux sites Web de formation continue, peu de tests en ligne ont été développés pour évaluer les compétences diagnostiques en radiologie. Finlay et coll. (2006) ont récemment construit, à Toronto, un test en ligne visant à évaluer les connaissances des résidents en radiologie diagnostique. Leur outil présentait des cas, relativement faciles, de radiographies thoraciques, digestives ou musculo-squelettiques. Les candidats devaient sélectionner une réponse parmi un menu exhaustif de diagnostics. Ils ont démontré la faisabilité du test en ligne et le pouvoir discriminant des différents niveaux d'expérience était acceptable. Ce format de test n'est toutefois pas destiné à comparer les niveaux de compétence des résidents avec celui des radiologues et ne mesurait pas spécifiquement les dimensions perception et interprétation. De plus, les cas présentés étaient choisis de manière à faire consensus entre les experts.

En résumé, la lecture des images radiologiques fait intervenir des habiletés perceptuelles et cognitives fortement intriquées, faisant la particularité du raisonnement radiologique. Il a cependant plusieurs points communs avec le raisonnement clinique, impliquant l'utilisation de la reconnaissance des modèles pour les cas typiques et le raisonnement analytique dans des situations plus élaborées. Son évaluation par des outils de mesure classiques a jusqu'à présent permis la certification de plusieurs promotions de radiologues. Toutefois, l'aspect raisonnement radiologique en contexte d'incertitude, qui s'acquiert avec l'expérience, reste difficile à mesurer avec les moyens existants.

OBJECTIFS DE LA RECHERCHE

Les principaux objectifs de ce projet sont de :

- 1. Construire deux instruments de mesure, réalistes et accessibles **en ligne**, basés sur le concept du TCS. Chaque outil évaluera respectivement les **habiletés à percevoir et à interpréter des anomalies radiographiques thoraciques en contexte d'incertitude**, par les **résidents de radiologie en cinquième année de résidence**. La dimension perception explorera les sous-dimensions détection et identification. La dimension interprétation analysera le processus diagnostique en s'inspirant des modèles de reconnaissance de configuration et d'activation des scripts.*
- 2. Mettre en place un protocole de validation de ces instruments. Des stratégies seront proposées afin :*
 - d'optimiser les scores de perception, notamment les scores de détection en fonction de la possible variabilité entre les experts.*
 - d'évaluer le niveau d'expertise et le nombre d'experts nécessaires dans le panel d'experts de TCS à contenu visuel.*
 - de mesurer la variabilité entre les experts et quantifier son importance dans l'évaluation en contexte d'incertitude, en radiologie diagnostique.*

MÉTHODOLOGIE

Cette section décrit la construction d'un test de perception (TP) et d'un test d'interprétation (TI) des anomalies RxP, élaborés durant l'année universitaire 2006-2007. Ils ont été construits par deux médecins, une radiologue spécialisée en imagerie thoracique (LS) et une pneumologue (SM). Les radiographies pulmonaires (RxP) étaient choisies parmi les dossiers des archives numériques du département de radiologie du Centre Hospitalier de l'Université de Montréal (CHUM). Une première sélection de radiographies, représentative de l'ensemble des pathologies couramment observées, compose une banque de 400 RxP. Pour chacun des tests, des clichés différents répondant au programme de formation post-doctorale en radiologie diagnostique de l'université de Montréal, ainsi qu'aux objectifs d'évaluation du CRMCC et à des critères spécifiques à chaque test ont été sélectionnés. Les questions étaient élaborées, puis révisées conjointement par les deux médecins.

1- Construction et déroulement du test de Perception (TP)

Le test de perception³ vise la composante « perceptuelle » de la lecture d'une RxP. Son objectif est d'évaluer séparément les capacités du résident de *détecter* et *d'identifier* des *anomalies* radiologiques thoraciques *subtiles ou ambiguës*. Il ne s'agit pas d'un TCS classique, mais sa construction s'inspire du concept du TCS. On tiendra compte simultanément des impératifs de précision et de la possibilité de divergence entre les experts dans la perception des images radiologiques. Il comporte plusieurs vignettes radiologiques composées chacune d'une radiographie pulmonaire, contenant une ou plusieurs anomalies. En observant la RxP sur un écran d'ordinateur, le participant doit (1) détecter chaque anomalie présente en la pointant et en cliquant dessus avec le curseur. (2) Ensuite, il doit l'identifier parmi les catégories d'anomalies proposées sur un menu déroulant qui s'affiche à l'écran dès qu'une anomalie a été détectée (annexe 1, pages x-xv). La tâche d'identification est considérée comme une phase intermédiaire initiatrice du processus d'interprétation, correspondant à une analyse sémiologique des

³ Une version de démonstration du test de perception est disponible à l'adresse suivante : <http://www.fpcmed2.umontreal.ca:8080/sctimagesdemo3/>

éléments morphologiques de l'anomalie et permettant de l'identifier. Ce qui correspond à la prise de décision perceptuelle décrite plus haut (cf. Cadre théorique §2-1 et §2-3).

1-1 Critères de sélection des radiographies pulmonaires

Les RxP contenues dans chaque vignette du TP sont choisies parmi la banque des 400 clichés initialement sélectionnés. En plus des objectifs du programme de formation post-doctorale et d'évaluation du CRMCC, elles devaient répondre aux critères suivants :

1. Présence d'une ou de plusieurs anomalies subtiles potentiellement difficiles à détecter par un résident finissant en difficulté (R5). Le nombre d'anomalies à détecter et à identifier varie de 1 à 6. Le nombre de clics est limité à 6 par vignette et précisé dans les instructions précédant le texte. Si plus de 6 clics sont possibles, le candidat cliquera sur les anomalies les plus pertinentes. Une vignette de RxP normale a été incluse dans le test. Le niveau de difficulté des anomalies à détecter est établi conjointement, par les constructrices du test. Il est inspiré d'une grille échelonnée de 1 à 3 (cf. tableau II, page 43) tirée des recommandations de la Société de Radiologie d'Amérique du Nord (RSNA), dans cadre du projet RadLex⁴. Celui-ci vise à unifier la terminologie en radiologie diagnostique (Langlotz, 2006). Le présent projet visant à détecter les difficultés des résidents finissants, la plupart des anomalies à détecter sur les RxP du TP sont représentatives des niveaux de difficulté 2 et 3. Afin de ne pas décourager les externes et les résidents débutants, des anomalies de niveau 1 sont toutefois présentes dans plusieurs vignettes. Le niveau de difficulté de chaque vignette représente la moyenne de la difficulté des questions (annexe 3, page xix). La grille fera l'objet d'un processus de validation par le panel de référence et devrait refléter les compétences de détection supposées à chaque niveau d'expérience.

Tableau II.

Grille de difficulté des images à détecter

Niveau de difficulté	Détection
1	Image de base dont la détection est facile
2	Image dont la détection est moyennement difficile
3	Image dont la détection est difficile

⁴ <http://radlex.org/viewer>

2. Une anomalie est présente sur une radiographie lorsqu'elle est détectée par LS qui est experte en imagerie thoracique et confirmée obligatoirement par une autre imagerie (tomodensitométrie, une résonance magnétique) ou des examens de laboratoire (pathologie, bactériologie...).
3. La concordance entre le domaine d'évaluation et le domaine de performance, qui représente les anomalies radiologiques que tout futur radiologiste doit savoir détecter et identifier. La part des anomalies présentes tient compte de la prévalence et / ou de la gravité de la pathologie à détecter. L'identification des anomalies se fait à partir du menu déroulant qui comporte 11 catégories d'anomalies regroupées selon leurs sites anatomiques et décrites selon les termes préconisés dans le projet RadLex (annexe 2, pages xvi-xviii).
4. Chaque vignette radiologique et chaque anomalie à détecter doivent répondre aux critères de la grille de contrôle de qualité figurant à l'annexe 4 (page xx).

1-2 Élaboration des vignettes du TP

Chaque vignette ou cas radiologique correspond à une radiographie pulmonaire qui peut être normale ou présenter jusqu'à 6 anomalies dont la subtilité est variable. Chaque anomalie correspond à une question (de même que la RxP normale). Pour ne pas influencer la recherche visuelle, aucun énoncé clinique préalable n'introduit le cas (Leblanc et al., 2002; G. R. Norman, Brooks et al., 1992). Le TP a été limité à 20 vignettes après un essai préliminaire d'une version de 30 cas, jugée trop longue et risquant de décourager les participants. Les 20 vignettes contiennent un total de 51 questions, correspondant à différents syndromes thoraciques, résumés dans le tableau de l'annexe 3 (page xix). Elles sont présentées successivement et de manière aléatoire.

1-3 Panel de référence du TP

Tel que suggéré par Charlin et coll (2005), 20 radiologues seront soumis au test et leurs réponses serviront à établir les scores des candidats. Parmi eux, 10 seront sur-spécialisés en radiologie thoracique et 10 auront une pratique de radiologie générale. Un radiologue sera considéré sur-spécialiste s'il consacre au moins 50% de ses activités professionnelles à l'imagerie thoracique. La composition du panel permettra d'évaluer si la présence de sur-spécialistes dans le panel de référence, modifie le pouvoir

discriminant du test. On s'attend à ce qu'au moins un membre du panel détecte les anomalies présentes sur les RxP. Celles-ci ayant déjà été détectées par l'une des constructrices du test, sur-spécialisée en imagerie thoracique (LS) et leur présence confirmée par des examens de référence. La composition du panel de référence sera évaluée à l'aide d'une grille de contrôle de qualité (annexe 8, page xxx). Elle s'assurera que le nombre d'experts répond aux critères de Gagnon et al (2005), qui démontrent qu'un panel comportant 15 à 20 experts permet d'obtenir une fidélité acceptable et une bonne corrélation des scores entre l'échantillon étudié et le panel. Dans le cadre particulier de ce projet qui vise à déceler les difficultés des R5 finissants, les experts auront des fonctions académiques et certains d'entre seront surspécialisés en imagerie thoracique. Enfin, on s'assurera que les conditions de réalisation des tests, dont les instructions fournies, la résolution de l'écran et l'environnement sont comparables entre tous les participants.

1-4 Calcul des scores du TP

Deux scores mesurant respectivement la détection (SDe) et l'identification des anomalies (SIId) seront calculés pour chaque anomalie détectée.

Calcul du score de détection : chaque vignette aura un score (SDeV) représentant la moyenne des SDe obtenus pour chaque anomalie détectée. Le SDe total (SDeT) correspondra à la somme des SDeV ramenée à 100.

Le SDe tiendra compte de la particularité visuelle du test. Malgré sa subtilité, une anomalie réellement présente⁵ est considérée détectable même si elle n'est décelée que par un seul spécialiste. De ce fait, le calcul le SDe par la technique classique des scores combinés, avantagerait les candidats faibles en détection, si la plupart des membres du panel ne la détectent pas. Dans le cadre de cette recherche, on comparera trois méthodes de calcul des scores de détection pour déterminer et valider celle qui fournit le meilleur pouvoir discriminant. Un score de détection dichotomique sera attribué à la base. Il sera enrichi par un système de pondération ou de boni sensés augmenter le pouvoir discriminant du test. À terme, la méthode de calcul des scores la plus valide et la plus discriminante sera retenue.

- **Système de scores dichotomiques**

⁵ La présence de toutes les anomalies subtiles ou ambiguës étant confirmée lors de la construction du test par d'autres examens de référence.

- Un score = 0 sera accordé à un candidat qui ne détecte pas l'anomalie présente et confirmée par au moins un expert.
- Un score = 1 sera accordé à un candidat qui aura détecté une lésion présente, indépendamment du nombre de radiologues qui l'auront détectée.

- **Système de pondération :**

On pourrait se baser sur les niveaux de difficultés de chaque anomalie à détecter. On utiliserait une échelle de pondération selon la difficulté à percevoir les anomalies. Le candidat obtiendrait un score = 0 s'il ne détecte pas l'anomalie présente, un score = 1 s'il détecte une lésion facile, 1,25 pour une lésion moyenne et 1,5 pour une lésion difficile. Cette méthode de calcul des scores devrait refléter la capacité du candidat à détecter des anomalies difficiles et de plus en plus subtiles ou atypiques. En théorie, les résidents seniors auraient des scores de détection comparables à ceux des experts. Elle ne tiendra toutefois pas compte de la variabilité entre les experts, tout comme le système de scores dichotomiques.

- **Système de bonification :**

Après avoir attribué un score dichotomique, ce système accorderait un bonus constant de 0,25 au candidat qui aura détecté une anomalie que seule une minorité de radiologues auront repérée. Cette minorité serait arbitrairement fixée à 25% des radiologues. Par exemple, si une lésion présente n'est vue que par 25% des radiologues ou moins, le candidat l'ayant décelée se verra attribuer le score de 1 + 0,25. L'utilisation de boni aurait pour but d'augmenter la variabilité de l'échelle des scores de détection. Jusqu'à présent, les scores de détection dans les autres formats de tests ne tenaient compte que du consensus d'experts, souvent fixé à 80% d'accord (Brazeau-Lamontagne et al., 2004). Or, la détection de certaines anomalies réelles, mais subtiles qui ne sont détectées que par une minorité de radiologues parfois, peut changer les décisions thérapeutiques et le pronostic d'un patient. Dans le cadre de cette recherche, nous faisons l'hypothèse que le système de bonification des scores permettrait de distinguer les candidats détectant les anomalies comme une minorité d'experts. Évidemment, cette différence ne serait visible que si leurs scores de détection aux autres questions sont

meilleurs ou comparables au reste de l'échantillon. La validité du système de bonification sera comparée aux systèmes dichotomiques et pondérés.

Le score d'identification, correspondra à un score d'analyse sémiologique. Dans le cadre particulier de cette recherche et afin de rendre le test le plus réaliste possible, tel que mentionné dans les objectifs initiaux, il sera relié à la tâche de détection. Seules les anomalies détectées seront identifiées, mais le score sera calculé indépendamment du SDe, car une anomalie peut être perçue mais incorrectement identifiée. Étant donné qu'une même anomalie peut être décrite par des termes équivalents, on ne tiendra pas compte des subtilités de langage et on regroupera les termes identifiant la même entité, lors de la correction du test. Par exemple, certains spécialistes sont plus à l'aise pour identifier une anomalie hilare dans la rubrique « masse / nodule », alors que d'autres parleront de « hiles dense », ce qui désigne la même entité sémiologique. Les différentes réponses seront regroupées, dans ce cas, dans une catégorie « anomalies hilaires ». De manière plus générale, si des termes différents mais acceptables sont utilisés pour désigner une même anomalie, les réponses seront rassemblées dans une même rubrique. Le SId sera déterminé par la méthode des scores combinés (c.f. cadre théorique §5-2-1). Chaque vignette aura un score (SIdV) représentant la moyenne des SId obtenus pour chaque anomalie identifiée. Le SId total (SIdT) correspondra à la somme des SIdV ramenée à 100.

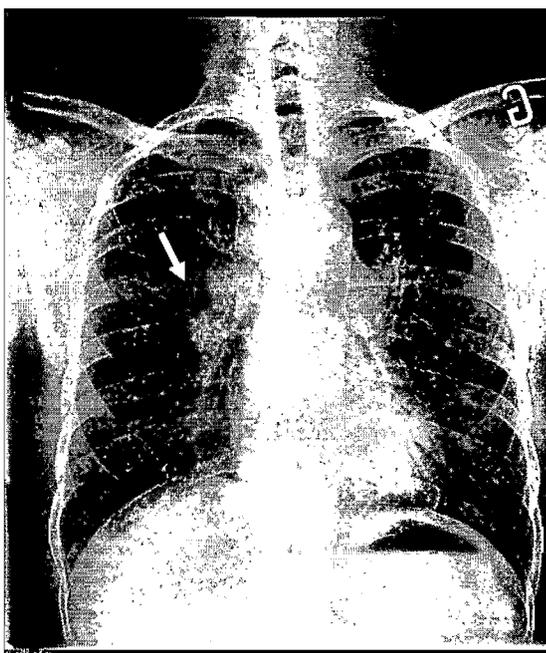
2- Construction du Test d'interprétation (TI)

Alors que le test de perception mesure les habiletés de détection et de description des anomalies observées sur une RxP, le test d'interprétation a pour but de mesurer le jugement diagnostique radiologique. Chaque vignette du TI comporte de brefs renseignements cliniques introduisant une radiographie pulmonaire, contenant des anomalies présentant des degrés variables d'ambiguïté. Trois à cinq hypothèses diagnostiques sont successivement proposées et le candidat doit **juger de la plausibilité** de chacune d'entre elles⁶. Afin de se rapprocher de la réalité, le test comporte deux formats de vignettes évaluant le jugement radiologique devant deux types de situations incertaines :

⁶ Une version de démonstration du test d'interprétation est disponible à l'adresse suivante : <http://www.fpcmed2.umontreal.ca:8080/sctimagesdemo3/>

1. Dans le cas d'anomalies suffisamment typiques, l'activation des scripts (probablement par voie non analytique chez les spécialistes), aboutit rapidement au diagnostic le plus probable, mais sans éliminer formellement d'autres diagnostics (exemple de la figure 9, page 30). Dans ce cas, on utilisera des *vignettes d'interprétation probabiliste* (IP), où le candidat devra se prononcer directement sur la plausibilité de plusieurs hypothèses avancées. Celles-ci sont sélectionnées de manière à ne pas comporter d'hypothèse neutre pour éviter les réponses standards, notamment chez les novices ne désirant pas se prononcer. Bien que l'usage d'échelles à 5 ou 7 points soit généralement conseillé (Carifio & Perla, 2007; Jamieson, 2004), on utilisera une échelle de type Likert à 4 points, dont la fiabilité est démontrée (Chang, 1994), pour inciter la personne interrogée à se prononcer. L'échelle comprendra les options : très peu probable, peu probable, probable, très probable (cf. figure 11, page 49 et annexe 5, pages xxi-xxvii). Cette terminologie a été retenue afin de se rapprocher de la réalité quotidienne en radiologie, où les possibilités diagnostiques sont émises, par abus de langage, en termes de probabilités (Langlotz, 2006), sans toutefois désigner un pourcentage précis. En pratique, les différentes hypothèses de chaque vignette sont présentées successivement. Les faire figurer simultanément à l'écran pourrait avoir des effets suggestifs et inciter certains candidats à procéder par élimination. Ceci induira cependant un biais de dépendance entre les différentes hypothèses, dont il faudra tenir compte lors de l'analyse des données.
2. Les situations où les anomalies présentes nécessitent un raisonnement analytique plus poussé, pour étayer ou éliminer les hypothèses diagnostiques (cf. exemple de la figure 12, page 51), seront évaluées par des *vignettes d'interprétation analytique* (IA). De nouvelles informations sont alors fournies après chaque hypothèse émise, pour moduler la probabilité de l'hypothèse avancée. Elles comportent des données cliniques, évolutives, ou radiologiques : images de nouvelles RxP ou de tomodensitométrie (cf. figure 12, page 50 et annexe 5, pages xxi- xxvii). L'effet de chaque information sur la probabilité de l'hypothèse avancée est coté sur une échelle de type Likert à 5 points, allant de -2 à +2 (- 2 = exclue ou quasi improbable, - 1 = moins probable, 0 = n'a aucun effet sur l'hypothèse, + 1 = plus probable, + 2 = quasi probable).

Mme L, 29 ans, consulte pour dyspnée, toux et vomissements. On observe les anomalies suivantes sur la radiographie pulmonaire.



Indiquez la vraisemblance de chacun des quatre diagnostics proposés

Sarcoidose	Très peu probable <input type="checkbox"/>	Peu probable <input type="checkbox"/>	Probable <input type="checkbox"/>	Très probable <input type="checkbox"/>
Lymphome	Très peu probable <input type="checkbox"/>	Peu probable <input type="checkbox"/>	Probable <input type="checkbox"/>	Très probable <input type="checkbox"/>
Métastases d'un cancer	Très peu probable <input type="checkbox"/>	Peu probable <input type="checkbox"/>	Probable <input type="checkbox"/>	Très probable <input type="checkbox"/>
Tuberculose	Très peu probable <input type="checkbox"/>	Peu probable <input type="checkbox"/>	Probable <input type="checkbox"/>	Très probable <input type="checkbox"/>

Figure 11. Exemple d'une vignette d'interprétation probabiliste

Mme B, 40 ans, consulte pour fièvre et dyspnée. On obtient la radiographie pulmonaire suivante



Question 1 :

Si on vous suggère...	et que...	...vous en concluez que
un œdème pulmonaire	la patiente n'a pas d'antécédents cardio-vasculaires	- 2 = l'hypothèse est très peu probable
		- 1 = l'hypothèse est moins probable
		0 = cela n'a aucun effet sur l'hypothèse
		+ 1 = l'hypothèse est plus probable
		+ 2 = l'hypothèse est très probable

Question 2 :

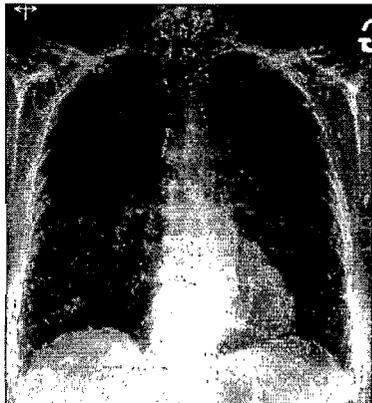
Si on vous suggère...	et que vous observez la radiographie suivante après 3 semaines d'évolution	...vous en concluez que
une alvéolite allergique extrinsèque		- 2 = l'hypothèse est très peu probable
		- 1 = l'hypothèse est moins probable
		0 = cela n'a aucun effet sur l'hypothèse
		+ 1 = l'hypothèse est plus probable
		+ 2 = l'hypothèse est très probable

Figure 12. Exemple d'une vignette d'interprétation analytique

Le TI comporte 30 vignettes radiologiques et un total de 92 questions, également réparties entre les deux types de questions (annexe 6, page xxviii).

2-1 Sélection des RxP du test d'interprétation

Les RxP figurant dans le TI sont différentes de celles qui sont présentées dans le test de perception. Elles sont sélectionnées selon :

1. L'ambiguïté des anomalies, qui seraient potentiellement difficiles à interpréter par un résident finissant (R5) en difficulté. Elle est jugée sur 3 niveaux de difficultés (cf. tableau III, page 50), conformément aux recommandations RadLex⁷
2. La concordance entre le domaine d'évaluation et le domaine de performance, déterminée par les objectifs de formation post-doctorale en imagerie thoracique, qui sont en concordance avec les objectifs de certification du CRMCC, tout en tenant compte de la prévalence et/ou de la gravité de la pathologie à interpréter.

Tableau III.

Échelle d'ambiguïté des images à interpréter

Niveau de difficulté	Interprétation	
1	Image de base dont l'interprétation est facile	Typique
2	Image dont l'interprétation est moyennement difficile	Atypique
3	Image dont l'interprétation est difficile	Mime une autre entité

3. Chaque vignette radiologique du TI doit répondre aux critères de la grille de contrôle de qualité figurant à l'annexe 7 (page xxix).

2-2 Panel de référence du TI

Les mêmes 20 radiologues, 10 en radiologie générale et 10 en radiologie thoracique seront soumis au test. Ils détermineront les scores des candidats en répondant aux questions dans les mêmes conditions.

2-3 Calcul des scores du TI

Les questions des vignettes IP auront chacune un score SIP et chaque vignette IP aura un score SIPV, représentant la moyenne des SIP. Le SIP total (SIPT) correspondra à la somme des SIPV ramenée à 100. On procédera de la même manière pour le calcul des

⁷ <http://radlex.org/viewer>

scores des vignettes d'interprétation analytiques (SIAV), qui représenteront la moyenne des scores à chaque hypothèse (ou question). La somme des SIAV, rapportée à 100, donnera le score total des vignettes IA (SIAT).

3- Interface informatique et déroulement des tests

Les tests sont programmés et mis en ligne sur le serveur du Centre de Pédagogie Appliquée aux Sciences de la Santé (CPASS) à l'université de Montréal, par les informaticiens du centre. La programmation est effectuée sur une interface utilisant le langage JAVA associé à un système de gestion de base de données du type *MySQL*. L'exécution du programme, via Internet, est assurée par le serveur Web *APACHE TOMCAT*, un serveur d'applications Java. L'accès en ligne est sécurisé et chaque candidat dispose d'un identifiant et d'un mot de passe uniques, pour accéder aux tests à travers des navigateurs de type Windows Explorer ou Firefox. L'anonymat des participants est assuré pendant toutes les étapes du test par un système de codage des coordonnées personnelles. Seules les informations concernant le niveau et le domaine d'expertise, ainsi que les stages effectués seront accessibles à des fins d'analyses. Les images radiographiques sont également anonymisées. Les images sont importées du PACS dans un format JPEG haute résolution et insérées dans chaque vignette. Dans un premier temps, seule l'option d'agrandissement sera disponible. La possibilité de modification de la luminosité et des contrastes sera disponible ultérieurement.

Afin d'optimiser et d'unifier les conditions du test, les candidats seront convoqués en groupes à l'Université de Montréal, pour exécuter le test dans une salle dédiée aux examens en ligne (dont la première partie de l'examen d'aptitude du Conseil Médical du Canada). La salle est équipée d'ordinateurs connectés à l'Internet et munis d'écrans de 17 pouces. Les radiologues, dont la disponibilité est plus limitée, auront le choix de faire le test à partir de l'hôpital sur des ordinateurs possédant les mêmes caractéristiques que ceux qui sont utilisés par les candidats. La durée estimée de chaque test est en moyenne de 45 à 60 minutes. Cependant et dans le cadre de cette recherche, le temps alloué à leur exécution est illimité pour éviter de décourager les candidats peu expérimentés et pour laisser assez de flexibilité aux radiologues, qui risquent d'être régulièrement interrompus. En prévision de possibles interruptions, une option « pause » est prévue, permettant au participant de reprendre le test à la vignette où il

s'est arrêté. Un système de sauvegarde automatique des données est prévu pour parer à d'éventuelles pannes informatiques. Les réponses des candidats et celles des radiologues aux deux tests sont colligées et importées sur le correcteur en ligne du TCS⁸. Les scores des candidats seront calculés lorsque la totalité des experts auront effectué le test.

4- Plan de validation des tests

Le canevas de validation proposé ci-dessous vise à explorer les différents aspects de la validité de construit suggérés pour les tests psychométriques (Cook & Thomas, 2006; Downing, 2003; Messick, 1995).

4-1 Sujets participant à l'étude

Un échantillon composé de 4 groupes de sujets sera étudié :

1. Un panel référence qui comporte 20 radiologues, parmi lesquels 10 exercent dans des domaines variés, dont la radiologie thoracique (Rgen) et 10 autres sont sur-spécialisés en radiologie thoracique (Rthx). Un radiologue est considéré sur-spécialisé, si plus de 50% de son temps est alloué à l'imagerie thoracique. Dans le cadre de ce projet, tous les radiologues auront une pratique hospitalo-universitaire.
2. Un groupe de résidents seniors (S) composé de résidents en 4^{ème} et 5^{ème} année de résidence et « fellows » qui sont des résidents de 6^{ème} année, ayant réussi leur examen de certification du CRMCC et débutant une formation en sur-spécialité en radiologie. Ce regroupement des R4, R5, R6 est préalablement établi par les programmes de formation post-doctorale.
3. Un groupe de résidents juniors (J) en 1^{ère}, 2^{ème} et 3^{ème} année de résidence en radiologie diagnostique (R1, R2, R3), tel que défini dans les programmes de formation post-doctorale.
4. Des externes en médecine (Ext) à la fin de leur stage de radiologie.

La taille de l'échantillon est calculée de manière à détecter une différence équivalente à une déviation standard entre les moyennes des scores totaux des différents sous-groupes, pour une puissance prédéfinie $1 - \beta = 80\%$ et un risque de première espèce $\alpha <$

⁸ <http://www.fpcmed.umontreal.ca/tcs/fr/correction.html>

0,05. Le nombre de sujets nécessaires dans chaque sous-groupe est déterminé à la lumière des différences de scores estimées d'après les précédentes études portant sur le TCS (Brazeau-Lamontagne et al., 2004; Carrière et al., 2008; Gibot & Bollaert, 2008; Lambert, 2005). La différence attendue entre les score moyens des radiologues et des résidents seniors est de 6%, 7% entre les résidents seniors et les juniors et 10% entre les juniors et les externes. Le nombre de sujets, calculé à l'aide du logiciel Power on X (Power And Precision, Biostat, Engelwood, New Jersey, USA), sera de 18, 22 et 30 respectivement dans les sous-groupes de résidents seniors, juniors et des externes. Le test s'adressera aux résidents et externes des programmes de radiologie des trois facultés de médecine francophones québécoises.

4-2 Consentement éclairé et comité d'éthique

Avant d'exécuter les tests, chaque participant signera le formulaire de consentement qui fait suite à la lettre d'information (Annexe 9, pages xxxi-xxxii). Après avoir passé le test, chaque participant remplira individuellement, le questionnaire démographique destiné à chaque groupe de participants (Annexe 10, pages xxxiii-xxxiv). Ce document recueille les informations pertinentes concernant le niveau d'expérience et l'existence de possibles facteurs confondants, ainsi que leurs commentaires. Le projet est déjà approuvé par les comités d'éthique du CHUM pour les droits d'utilisation des radiographies pulmonaires et par celui de l'université de Montréal.

4-3 Validité de contenu

Un comité formé de 2 experts en imagerie thoracique, indépendants de la construction du test, jugeront de la validité de contenu. Une grille de contrôle de qualité, adaptée de celle que Carrière et Caire proposent sur le site Web du TCS ⁹ et ayant déjà servi à sa construction (annexe 4, page xx et annexe 7, page xxix) sera utilisée à cet effet. Pour chaque test, elle évaluera la composition du panel de référence et chaque vignette radiologique, notamment sa pertinence, son mode de présentation, les niveaux de difficulté et d'incertitude des anomalies à détecter ou à interpréter.

- Pour le test de détection, elle évaluera selon les critères RadLex¹⁰, la pertinence des différentes catégories d'identification et des termes sémiologiques utilisés.

⁹ <http://www.fpcmed.umontreal.ca/tcs/fr/validate.html>

¹⁰ <http://radlex.org/viewer>

- Pour le test d'interprétation, elle évaluera en plus du niveau de difficulté, les hypothèses diagnostiques et les nouvelles informations radiologiques ou cliniques fournies.

4-4 Analyse d'items

En théorie, l'analyse d'items devrait se faire uniquement sur les cas (ou vignettes) car l'indépendance des questions nichées dans chaque vignette n'est pas garantie. Cependant, étant donné le caractère visuel des questions et afin de vérifier leurs qualités psychométriques, on procèdera à une double analyse d'items évaluant chaque question dans un premier temps, puis chaque vignette dans un second temps. À partir des scores SDe et SDeV du TP, on mesurera respectivement les *indices de difficulté* de chaque anomalie à détecter. Pour chacun des tests TP et TI, la *discrimination des scores de chaque question SDe, SId, SIP et SIA*, et des scores de chaque cas *SDeV, SIdV, SIPV et SIAV*, sera mesurée à l'aide du coefficient de corrélation item-total corrigé. On calculera également les corrélations inter-items afin d'évaluer leur homogénéité dans chaque test. Au terme de cette double analyse, les items dont le coefficient de corrélation item-total sera bas ($< 0,1$) seront éliminés. On éliminera soit les cas soit les questions de moindre qualité, selon les résultats de l'analyse d'items. Pour le TP, cela consisterait à ne pas tenir compte de certaines anomalies sur un cliché RxP donné, au lieu d'éliminer le cas en entier.

4-5 Mesures de fidélité

Là fidélité des instruments de mesure sera abordée par une étude de cohérence interne et de stabilité temporelle, ainsi que par une étude de généralisabilité. Le *coefficient de cohérence interne alpha de Cronbach* sera calculé à partir des scores SDe, SId, SIP, SIA et des sommes (SDe + SId) et (SIP + SIA) de chaque sous-groupe. Il est à noter que les scores obtenus par chaque radiologue seront calculés à l'aide des réponses du panel, excluant sa propre réponse afin d'éviter tout biais dans les scores des radiologues. La *stabilité temporelle* de chaque test sera évaluée à l'aide d'une corrélation de Pearson entre les scores initiaux des résidents et des externes à chaque test et ceux obtenus deux mois plus tard. Ce délai de deux mois est un compromis pour minimiser « l'effet mémorisation ». On s'attend à une stabilité voire une amélioration des scores des résidents avec le temps due à « l'effet de pratique ».

Une *étude de généralisabilité* permettra d'apprécier la précision de chaque test et de vérifier dans quelle mesure les données recueillies sont généralisables (Bertrand & Blais, 2004; Crocker & Algina, 1986). Pour chacun des scores totaux calculés dans les deux tests, elle permettra d'estimer la fidélité en prenant en compte l'influence du nombre de cas, du nombre de questions par cas et de la composition du panel de radiologues. On pourra ainsi, vérifier si les résultats des études précédentes concernant le nombre de questions par cas (Gagnon et al., 2008) et l'influence de la composition du panel de référence sur les scores (B. Charlin, Gagnon et al., 2007), sont applicables aux TCS destinés aux spécialités visuelles. Les facettes étudiées, leur nichage et leurs croisements sont détaillés dans le tableau IV (cf. tableau IV, page 57).

Toutes les facettes seront considérées aléatoires infinies. Les facettes candidats constitueront la facette de différenciation et les facettes cas, questions, questions par cas et radiologues seront les facettes d'instrumentation. Les radiologues seront classés en deux facettes : radiologues généraux et thoraciques à deux niveaux chacune. Ce qui permettra d'évaluer l'effet de la composition du panel sur la variabilité des scores de chaque test. Pour chaque score, on mesurera un coefficient de généralisabilité relatif G estimant les différentes sources de variance en rapport avec les facettes énumérées. On procédera ensuite à une étude D permettant d'optimiser le coefficient de généralisabilité absolu, en agissant sur les éléments générant les composantes de variance les plus élevées. Elle permettra entre autres de déterminer, si dans ce genre de tests, il est nécessaire d'avoir des radiologues sur-spécialisés dans le panel de référence. Le nombre de radiologues figurant dans le panel d'experts pourrait limiter la faisabilité de l'étude de généralisabilité. On pourrait l'étendre aux experts des universités Laval, Shebrooke et McGill (dont la plupart des radiologues sont bilingues).

Tableau IV.
Étude de généralisabilité, plan d'observation

Test	Facettes à étudier	Composantes des variances
TP	1. Candidats	Candidats
	2. Radiologues généraux (10)	Radiologues généraux
	• <i>SDeT</i> N=5	Radiologues thoraciques
	▪ <i>SIdT</i> N=5	Cas
	▪ <i>S.Total</i> (SDeT+SIdT) N=5	Anomalies
	N=5	Anomalies : Cas
TI	3. Radiologues thoraciques (10)	Candidats x Radiologues généraux
	4. Cas	Candidats x Radiologues thoraciques
	5. Anomalies	Candidats x Anomalies
	6. Anomalies par cas	Candidats x Anomalies : Cas
	1. Candidats	Candidats
	2. Radiologues généraux (10)	Radiologues généraux
▪ <i>SIPT</i> N= 5	• N= 5	Radiologues thoraciques
	• N=5	Cas
	▪ <i>SIAT</i> N=5	Questions
	▪ <i>S.Total</i> (SIPT+SIAT) N=5	Questions : Cas
	• N=5	Candidats x Radiologues généraux
	• N=5	Candidats x Radiologues thoraciques
4. Cas	Candidats x Cas	
5. Questions	Candidats x Questions	
6. Questions par cas	Candidats x Questions : Cas	

4-6 Pouvoir discriminant des tests

Le pouvoir discriminant de chaque instrument sera mesuré après le contrôle de l'homogénéité de variance entre les différents groupes, par un test de Levene. On comparera les moyennes des scores obtenus par chacun des 4 groupes, à l'aide d'une analyse de variance de type ANOVA, vérifiée par un test *post-hoc* de Newman-Keuls. Si l'homogénéité de variance n'est pas confirmée, la comparaison sera effectuée à l'aide de tests non paramétriques de Kruskal-Wallis et de Mann-Whitney pour comparer les groupe deux par deux. On mesurera également l'impact du degré de variabilité entre les radiologues, sur l'amplitude de la différence des scores entre les différents groupes. Dans ce dessein, on étudiera la taille d'effet en fonction du niveau de variabilité des réponses des radiologues. Celle-ci sera classée en trois niveaux déterminés par le niveau de l'erreur-type (SEM) des scores des membres du panel. Un SEM détermine une faible variabilité, s'il se situe entre 0,00 et 0,49, une variabilité modérée entre 0.50 et 0.99 et une forte variabilité s'il dépasse 1 (B. Charlin et al., 2006).

4-7 Mesure de la validité prédictive des tests

La mesure de la validité prédictive vise à vérifier si les tests mis au point sont représentatifs : 1) à court terme, des compétences cliniques évaluées pendant les stages de résidence ; 2) à plus long terme, des compétences mesurées par les examens de certification du CRMCC. Elle sera étudiée en mesurant le coefficient de corrélation de Spearman entre la moyenne des scores de chaque test et la moyenne annuelle obtenue aux points 6-8, 10-13, 15, 17 et 25, de la grille d'évaluation de la maîtrise des compétences des stages de résidence de l'année en cours (Annexe 11, page xxxvii). Les points de comparaison sélectionnés sur la grille devraient refléter certaines facettes du raisonnement clinique. On mesurera également les corrélations entre les moyennes des scores avec ceux des examens de certification du CRMCC ou du Collège des Médecins de Famille du Canada pour les candidats (résidents et externes).

4-8 Profil des réponses des participants

La mesure du délai de réponse des participants à chaque question et la durée totale nécessaire pour passer chaque test, seront mesurables. On pourra donc comparer les délais moyens et les temps moyens entre les 4 groupes à l'aide d'un test t de Student. Pour le test de perception, on comparera également le nombre d'anomalies pointées à l'écran mais non choisie, qui reflètera le degré d'hésitation des participants. On mesurera également le degré de corrélation entre les scores de détection, d'identification et d'interprétation, à l'aide de corrélation de Pearson. Toutes ces informations permettront de mieux comprendre la progression du raisonnement radiologique avec l'expérience.

4-9 Logiciels statistiques utilisés

Le calcul du nombre de candidats nécessaires à l'étude sera effectué avec le logiciel Power And Precision, Biostat, Engelwood, New Jersey, USA. Les analyses statistiques seront effectuées à l'aide du logiciel SPSS (Statistical Package for the Social Sciences), version 13.0. L'étude de généralisabilité sera réalisée grâce au logiciel EduG, version 3.04.

CONCLUSIONS ET BÉNÉFICES ESCOMPTÉS

Ce travail aura été l'occasion d'une réflexion sur le raisonnement radiologique, ses différentes modalités et les moyens d'évaluer chacune de ses facettes. Il a permis la construction de deux outils de mesure destinés à détecter les résidents finissants en radiologie diagnostique, éprouvant des difficultés à la perception et à l'interprétation des anomalies radiographiques thoraciques. Une fois le processus de validation achevé, ces tests compléteront l'arsenal des instruments évaluant les compétences radiologiques. Ils auront l'avantage de simuler les tâches quotidiennes du radiologue, et de :

1. mesurer pas à pas le jugement clinique radiologique, de la détection des anomalies au raisonnement complexe, afin d'identifier les résidents en difficulté.
2. fournir les premiers outils permettant de quantifier la compétence diagnostique en radiologie en contexte d'incertitude et tenant compte de la variabilité des radiologues dans de telles situations.
3. proposer un format convivial, informatisé, familial et accessible en ligne.

À plus long terme, ces tests devraient être adaptables à divers usages formatifs destinés aux externes ou aux résidents de radiologie ou d'autres spécialités telles que la médecine familiale, la pneumologie ou la chirurgie thoracique. Ce qui permettra de diagnostiquer les difficultés rencontrées pendant la formation et de les corriger précocement. Ces tests constitueront un grand atout dans les examens de certification du CRMCC et dans la formation continue des spécialistes. Enfin, d'autres spécialités où les informations visuelles sont importantes pourront utiliser la plateforme informatique de ces instruments et développer des tests similaires, comme c'est déjà le cas en hématologie et en neurochirurgie.

Des questions, auxquelles le processus de validation devraient répondre, restent pour l'instant posées :

- Concernant le test de perception : 1) Existe-t-il un système de scores optimal permettant de tenir compte de la variabilité des experts ? 2) La composition du panel d'expert, en terme de nombre et d'expérience, doit-elle être comparable à

celle des TCS classiques ? 3) Le TP visant à simuler la réalité, comment contourner le biais introduit par l'utilisation des mêmes vignettes pour déterminer les scores de détection et d'identifications ?

- Pour le test d'interprétation : 1) Quel panel de référence utiliser ? 2) Les vignettes probabilistes sont-elles adéquates pour être administrées dans un format du type TCS ? reflètent-elles vraiment le raisonnement par reconnaissance de modèles ? la réponse à cette question pourrait être approfondie dans des recherches ultérieures. 3) Les vignettes analytiques permettront-elles de détecter les résidents en difficulté ?

Enfin, les données concernant le comportement des participants, renseignera sur la dynamique et les interactions des habiletés de perception et d'interprétation radiologique et leur progression avec l'expérience.

RÉFÉRENCES

- Ahissar, M., & Hoshstein, S. (2004). The reverse hierarchy theory of visual perceptual training. *Trends Cogn Sci*, 8(10), 457-464.
- Anderson, S., & Conway, M. (1993). Investigating the structure of autobiographical memories. *Journal of Experimental Psychology*, 19, 1-19.
- Ark, T. K., Brooks, L. R., & Eva, K. W. (2007). The benefits of flexibility: the pedagogical value of instructions to adopt multifaceted diagnostic reasoning strategies. *Med Educ*, 41(3), 281-287.
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proc Natl Acad Sci U S A*, 103(2), 449-454.
- Barrows, H., Norman, G., Neufeld, V., & Feightner, J. (1982). The clinical reasoning of randomly selected physicians in general medical practice. *Clin Invest Med*, 5, 49-55.
- Barrows, H. S., & Feltovich, P. J. (1987). The clinical reasoning process. *Med Educ*, 21(2), 86-91.
- Berbaum, K. S., el-Khoury, G. Y., Franken, E. A., Jr., Kathol, M., Montgomery, W. J., & Hesson, W. (1988). Impact of clinical history on fracture detection with radiography. *Radiology*, 168(2), 507-511.
- Berbaum, K. S., Franken, E. A., Jr., Dorfman, D. D., Barloon, T., Ell, S. R., Lu, C. H., et al. (1986). Tentative diagnoses facilitate the detection of diverse lesions in chest radiographs. *Invest Radiol*, 21(7), 532-539.
- Bertrand, R., & Blais, J.-G. (2004). *Les modèles de mesure dans le cadre de la théorie de généralisabilité*. Québec: Presses de l'université du Québec.
- Bick, U., & Lenzen, H. (1999). PACS: the silent revolution. *Eur Radiol*, 9(6), 1152-1160.
- Bland, A. C., Kreiter, C. D., & Gordon, J. A. (2005). The psychometric properties of five scoring methods applied to the script concordance test. *Acad Med*, 80(4), 395-399.
- Bloomfield, F. H., Teele, R. L., Voss, M., Knight, D. B., & Harding, J. E. (1999). Inter- and intra-observer variability in the assessment of atelectasis and consolidation in neonatal chest radiographs. *Pediatr Radiol*, 29(6), 459-462.
- 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.
- Caire, F., Sol, J. C., Moreau, J. J., Isidori, P., & Charlin, B. (2004). [Self-assessment for neurosurgery residents by script concordance test (SCT). The process of test elaboration]. *Neurochirurgie*, 50(1), 66-72.

- Carifio, J., & Perla, R. (2007). Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes. *Journal of Social Sciences*, 3(3), 106-116.
- Carrière, B., Gagnon, R., Charlin, B., Downing, S., & Bordage, G. (2008). Assessing Clinical Reasoning in Pediatric Emergency Medicine: Validity Evidence for a Script Concordance Test. *Ann Emerg Med*.
- Chang, L. (1994). A psychometric evaluation of 4-point and 6-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement*, 18(3), 205-215.
- Charlin, B., Bordage, G., & Van der Vleuten, C. (2003). L'évaluation du raisonnement clinique. *Pédagogie Médicale*, 4, 42-51.
- 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. *Adv Health Sci Educ Theory Pract*, 3(1), 51-58.
- Charlin, B., Brailovsky, C. A., Brazeau-Lamontagne, L., Samson, L., Leduc, C., & Vleuten, C. (1998). Script questionnaires: their use for assessment of diagnostic knowledge in radiology. *Med Teach*, 20(6), 567-571.
- 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., et al. (2006). Assessment of clinical reasoning in the context of uncertainty: the effect of variability within the reference panel. *Med Educ*, 40(9), 848-854.
- Charlin, B., Gagnon, R., Sauve, 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., Kazi-Tani, D., Gagnon, R., & Thivierge, R. (2005). Le test de concordance comme outil d'évaluation en ligne du raisonnement des professionnels en situation d'incertitude. *Revue internationale des technologies en pédagogie universitaire*, 2(2), 22-27.
- 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.
- Charlin, B., Tardif, J., & Boshuizen, H. P. (2000). Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Acad Med*, 75(2), 182-190.
- Charlin, B., & van der Vleuten, C. (2004). Standardized assessment of reasoning in contexts of uncertainty: the script concordance approach. *Eval Health Prof*, 27(3), 304-319.
- Cherian, T., Mulholland, E. K., Carlin, J. B., Ostensen, H., Amin, R., de Campo, M., et al. (2005). Standardized interpretation of paediatric chest radiographs for the diagnosis of pneumonia in epidemiological studies. *Bull World Health Organ*, 83(5), 353-359.

- Christensen, E. E., Murry, R. C., Holland, K., Reynolds, J., Landay, M. J., & Moore, J. G. (1981). The effect of search time on perception. *Radiology*, *138*(2), 361-365.
- Cochrane, A. L., & Garland, L. H. (1952). Observer error in the interpretation of chest films; an international investigation. *Lancet*, *2*(11), 505-509.
- Coderre, S., Mandin, H., Harasym, P. H., & Fick, G. H. (2003). Diagnostic reasoning strategies and diagnostic success. *Med Educ*, *37*(8), 695-703.
- 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.
- Cook, D., & Thomas, J. (2006). Current Concepts in Validity and Reliability for Psychometric Instruments: Theory and Application. *Am J Med*. 2006 Feb;*119*(2):166.e7-16., *119*(2), 166.e167-e116.
- CRMCC. (2007a). Forme de l'examen objectif intégré en Radiologie diagnostique. Retrieved 8 march 2008, 2008, from http://crmcc.medical.org/residency/certification/examformats/117_f.php
- (2007b). Objectifs et exigences de la formation spécialisée en radiologie diagnostique (Publication., from Collège Royal des Médecins et Chirurgiens du Canada: <http://crmcc.medical.org/information/index.php>
- Crocker, L., & Algina, J. (1986). Introduction to generalizability theory. In *Introduction to classical and modern test theory* (pp. 157-191). Belmont: Wadsworth, Thompson Learning.
- Custers, E., Boshuizen, H., & Schmidt, H. (1998). The role of illness scripts in the development of medical diagnostic expertise: results from an interview study. *Cognit Instr*, *16*(4), 367-398.
- Demeester, A. (2004). *Evaluation du raisonnement clinique des étudiants sages-femmes par le test de concordance de script*. Université Paris 13, Paris.
- Downing, S. (2003). Validity: on the meaningful interpretation of assessment data. *Medical Education*, *37*, 830-837.
- Elmore, J. G., Wells, C. K., Lee, C. H., Howard, D. H., & Feinstein, A. R. (1994). Variability in radiologists' interpretations of mammograms. *N Engl J Med*, *331*(22), 1493-1499.
- Elstein, A., Shulman, L., & Sprafka, S. (1990). Medical Problem Solving, a Ten-Year Retrospective. *Evaluation & The Health Profession*, *13*, 5-36.
- Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical Problem Solving: An Analysis of Clinical Reasoning*. Cambridge.
- Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. *Jama*, *287*(2), 226-235.
- Eva, K. W. (2005). Ce que tout enseignant devrait savoir concernant le raisonnement clinique. *Pédagogie Médicale*, (6), 225-234.
- Eva, K. W., Hatala, R. M., Leblanc, V. R., & Brooks, L. R. (2007). Teaching from the clinical reasoning literature: combined reasoning strategies help novice diagnosticians overcome misleading information. *Med Educ*, *41*(12), 1152-1158.
- Feltovich, P., Barrows, H., & 1984;128)42. (1984). Issues of generality in medical problem solving. In H. Schmidt & M. De Volder (Eds.), (pp. 128-142). Assen, Holland: Van Gorcum: xccv zcx.

- Finlay, K., Norman, G. R., Keane, D. R., & Stolberg, H. (2006). A web-based test of residents' skills in diagnostic radiology. *Can Assoc Radiol J*, 57(2), 106-116.
- Fournier, J. P., Thiercelin, D., Pulcini, C., Alunni-Perret, V., Gilbert, E., Minguet, J. M., & Bertrand, F. (2006). Évaluation du raisonnement clinique en médecine d'urgence : les tests de concordance des scripts décèlent mieux l'expérience clinique que les questions à choix multiples à contexte riche. *Pédagogie Médicale*, 7(1), 20-30.
- Frank, J. (2005). *Le Cadre de compétences CanMEDS 2005. L'excellence des normes, des médecins et des soins*. Ottawa: Le Collège royal des médecins et chirurgiens du Canada
- Frohna, J. G., Gruppen, L. D., Fliegel, J. E., & Mangrulkar, R. S. (2006). Development of an evaluation of medical student competence in evidence-based medicine using a computer-based OSCE station. *Teach Learn Med*, 18(3), 267-272.
- Fuhrman, C. R., Britton, C. A., Bender, T., Sumkin, J. H., Brown, M. L., Holbert, J. M., et al. (2002). Observer performance studies: detection of single versus multiple abnormalities of the chest. *AJR Am J Roentgenol*, 179(6), 1551-1553.
- 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.
- Gagnon, R., Charlin, B., Lambert, C., Carriere, B., & Van der Vleuten, C. (2008). Script concordance testing: more cases or more questions? *Adv Health Sci Educ Theory Pract*.
- Gagnon, R., Charlin, B., Roy, L., St-Martin, M., Sauve, E., Boshuizen, H. P., et al. (2006). The cognitive validity of the Script Concordance Test: a processing time study. *Teach Learn Med*, 18(1), 22-27.
- Garland, L. H., & Cochrane, A. L. (1952). Results of an international test in chest roentgenogram interpretation. *J Am Med Assoc*, 149(7), 631-634.
- Garrigan, P., & Kellman, P. J. (2008). Perceptual learning depends on perceptual constancy. *Proc Natl Acad Sci U S A*, 105(6), 2248-2253.
- Gibot, S., & Bollaert, P. (2008). Le test de concordance de script comme outil d'évaluation formative en réanimation médicale. *Pédagogie Médicale*, 9, 7-18.
- Gunderman, R. B. (2005). Education and the Art of Uncertainty. *Radiology*, 237(3), 801-802.
- Gunderman, R. B., & Nyce, J. M. (2002). The Tyranny of Accuracy in Radiologic Education. *Radiology*, 222(2), 297-300.
- Haller, S., & Radue, E. W. (2005). What is different about a radiologist's brain? *Radiology*, 236(3), 983-989.
- Hammoud, M. M., & Barclay, M. L. (2002). Development of a Web-based question database for students' self-assessment. *Acad Med*, 77(9), 925.
- Harris, I. (1993). New Expectations for Professional Competence. In W. Curry L, J.F (Ed.), *Educating Professionals. Responding to new expectations for competence and accountability* (pp. 17-52). San Francisco: Jossey-Bath.
- Hasnain, M., Bordage, G., Connell, K. J., & Sinacore, J. M. (2001). History-taking behaviors associated with diagnostic competence of clerks: an exploratory study. *Acad Med*, 76(10 Suppl), S14-17.

- Hatala, R., Norman, G., & Brooks, L. (1996). The effect of clinical history on physicians' ECG interpretation skills. *Acad Med*, 71(10 Suppl), S68-S70.
- Hawkins, R., MacKrell Gaglione, M., LaDuca, T., Leung, C., Sample, L., Gliva-McConvey, G., et al. (2004). Assessment of patient management skills and clinical skills of practising doctors using computer-based case simulations and standardised patients. *Med Educ*, 38(9), 958-968.
- Higgs, J., & Jones, M. (2000). Clinical reasoning in the health professions. In J. Higgs & M. Jones (Eds.), *Clinical Reasoning in the Health Professions* (2 ed., pp. 3-14): Oxford: Butterworth-Heinemann Ltd.
- Hobby, J. L., Tom, B. D., Todd, C., Bearcroft, P. W., & Dixon, A. K. (2000). Communication of doubt and certainty in radiological reports. *The British Journal of Radiology*, 73(873), 999-1001.
- Hopstaken, R. M., Witbraad, T., van Engelshoven, J. M., & Dinant, G. J. (2004). Inter-observer variation in the interpretation of chest radiographs for pneumonia in community-acquired lower respiratory tract infections. *Clin Radiol*, 59(8), 743-752.
- Jamieson, S. (2004). Likert scales: how to (ab)use them. *Med Educ*, 38(12), 1212-1218.
- Kalamarides, M., Caire, F., Dager, F., Brassier, G., & Moreau, J. J. (2008). Creation of self-assessment tools for on-line continuing medical education. Modelization of a training session. *Neurochirurgie*, 54(1), 21-27.
- Khonputsa, P., Besinque, K., Fisher, D., & Gong, W. C. (2006). Use of script concordance test to assess pharmaceutical diabetic care: a pilot study in Thailand. *Med Teach*, 28(6), 570-573.
- Krupinski, E. A. (2003). Medical image perception issues for PACS deployment. *Semin Roentgenol*, 38(3), 231-243.
- Kulatunga-Moruzi, C., Brooks, L. R., & Norman, G. R. (2001). Coordination of analytic and similarity-based processing strategies and expertise in dermatological diagnosis. *Teach Learn Med*, 13(2), 110-116.
- Kundel, H. L. (2000). Visual search in medical images. In J. Beutel, H. L. Kundel & V. M. R.L. (Eds.), *Handbook of medical imaging* (Vol. 1, pp. 837-858). Bellingham, Washington: SPIE Press.
- Kundel, H. L., & LaFollette, P. S. (1972). Visual search patterns and experience with radiological images. *Radiology*, 103, 523-528.
- Kundel, H. L., & Nodine, C. F. (1983). A visual concept shapes image perception. *Radiology*, 146(2), 363-368.
- Kundel, H. L., Nodine, C. F., & Krupinski, E. A. (1989). Searching for lung nodules. Visual dwell indicates locations of false-positive and false-negative decisions. *Invest Radiol*, 24(6), 472-478.
- Kundel, H. L., & Wright, D. J. (1969). The influence of prior knowledge on visual search strategies during the viewing of chest radiographs. *Radiology*, 93(2), 315-320.
- Labelle, M., Beaulieu, M., Paquette, D., Fournier, C., Bessette, L., Choquette, D., et al. (2004). An integrated approach to improving appropriate use of anti-inflammatory medication in the treatment of osteoarthritis in Quebec (Canada): the CURATA model. *Med Teach*, 26(5), 463-470.
- Lambert, C. (2005). *Le test de concordance de scripts : un outil pour évaluer le raisonnement clinique des résidents en radio-oncologie*. Université de Montréal, Montréal.

- Langlotz, C. P. (2006). RadLex: a new method for indexing online educational materials. *Radiographics*, 26(6), 1595-1597.
- Leblanc, V. R., Brooks, L. R., & Norman, G. R. (2002). Believing is seeing: the influence of a diagnostic hypothesis on the interpretation of clinical features. *Acad Med*, 77(10 Suppl), S67-69.
- Lee, W. R. (2006). Computer-based learning in medical education: a critical view. *J Am Coll Radiol*, 3(10), 793-798.
- Lesgold, A., Feltovich, P., Glaser, R., & Wang, Y. (1981). *The Acquisition of Perceptual Diagnostic Skill in Radiology*. Pittsburgh: Pittsburgh Univ PA Learning Research And Development Center
- Lesgold, A. M., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: diagnosing x-ray pictures. In G. R. ICHI MTH, Farr MJ (Ed.), (pp. 311-342): Hillsdale, NJ, Lawrence Erlbaum. Associates.
- Lieberman, S. A., Frye, A. W., Litwins, S. D., Rasmusson, K. A., & Boulet, J. R. (2003). Introduction of patient video clips into computer-based testing: effects on item statistics and reliability estimates. *Acad Med*, 78(10 Suppl), S48-51.
- Lilienfeld, A. M., & Kordan, B. (1966). A study of variability in the interpretation of chest x-rays in the detection of lung cancer. *Cancer Res*, 26(10), 2145-2147.
- Llewellyn Thomas, E., & Lansdown, E. L. (1963). Visual search patterns of radiologists in training. *RADIOLOGY*, 81, 288-291.
- Loy, C. T., & Irwig, L. (2004). Accuracy of diagnostic tests read with and without clinical information: a systematic review. *Jama*, 292(13), 1602-1609.
- Lubarsky, S. (2007). The Neurology Script Concordance Test. from <http://www.fpcmed2.umontreal.ca/neurology.mcgill.ca/>
- Marie, I., Sibert, L., Roussel, F., Hellot, M. F., Lechevallier, J., & Weber, J. (2005). [The script concordance test: a new evaluation method of both clinical reasoning and skills in internal medicine]. *Rev Med Interne*, 26(6), 501-507.
- McCoubrie, P., & McKnight, L. (2008). Single best answer MCQs: a new format for the FRCR part 2a exam. *Clin Radiol*, 63(5), 506-510.
- Mello-Thoms, C. (2003). Perception of breast cancer: eye-position analysis of mammogram interpretation. *Acad Radiol*, 10(1), 4-12.
- Mello-Thoms, C., Hardesty, L., Sumkin, J., Ganott, M., Hakim, C., Britton, C., et al. (2005). Effects of lesion conspicuity on visual search in mammogram reading. *Acad Radiol*, 12(7), 830-840.
- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Am Psychologist*, 50, 741-749.
- Meterissian, S., Zabolotny, B., Gagnon, R., & Charlin, B. (2007). Is the script concordance test a valid instrument for assessment of intraoperative decision-making skills? *Am J Surg*, 193(2), 248-251.
- Morag, E., Lieberman, G., Volkan, K., Shaffer, K., Novelline, R., & Lang, E. V. (2001). Clinical competence assessment in radiology: introduction of an objective structured clinical examination in the medical school curriculum. *Acad Radiol*, 8(1), 74-81.

- Mugglestone, C. J. (1995). Breast cancer and hormonal supplements in postmenopausal women. Increase in risk of breast cancer is small. *Bmj*, *311*(7018), 1505.
- Myles-Worsley, M., Johnston, W. A., & Simons, M. A. (1988). The influence of expertise on X-ray image processing. *J Exp Psychol Learn Mem Cogn*, *14*(3), 553-557.
- 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*, 235-254.
- Nendaz, M. R., Gut, A. M., Perrier, A., Reuille, O., Louis-Simonet, M., Junod, A. F., et al. (2004). Degree of concurrency among experts in data collection and diagnostic hypothesis generation during clinical encounters. *Med Educ*, *38*(1), 25-31.
- Neufeld, V. R., Norman, G. R., Feightner, J. W., & Barrows, H. S. (1981). Clinical problem-solving by medical students: a cross-sectional and longitudinal analysis. *Med Educ*, *15*(5), 315-322.
- Nodine, C. F., & Mello-Thoms, C. (2000). The nature of expertise in radiology. In J. Beutel, H. L. Kundel & V. M. R.L. (Eds.), *Handbook of medical imaging* (Vol. 1). Bellingham, Wash.: SPIE Press.
- Norman, G. (2005). Research in clinical reasoning: past history and current trends. *Med Educ*, *39*(4), 418-427.
- Norman, G., Brooks, L., Colle, C., & Hatala, R. (1999). The Benefit of Diagnostic Hypotheses in Clinical Reasoning: Experimental Study of an Instructional Intervention for Forward and Backward Reasoning. *Cognition and Instruction*, *17*(4), 433 — 448.
- (1999). The benefit of diagnostic hypotheses in clinical reasoning: Experimental study of an instructional intervention for forward and backward reasoning. *Cognition and instruction* *17*, 433-448.
- Norman, G., Young, M., & Brooks, L. (2007). Non-analytical models of clinical reasoning: the role of experience. *Med Educ*, *41*(12), 1140-1145.
- Norman, G. R., & Brooks, L. R. (1997). The Non-Analytical Basis of Clinical Reasoning. *Adv Health Sci Educ Theory Pract*, *2*(2), 173-184.
- Norman, G. R., Brooks, L. R., Coblenz, C. L., & Babcock, C. J. (1992). The correlation of feature identification and category judgments in diagnostic radiology. *Mem Cognit*, *20*(4), 344-355.
- Norman, G. R., Coblenz, C. L., Brooks, L. R., & Babcock, C. J. (1992). Expertise in visual diagnosis: a review of the literature. *Acad Med*, *67*(10 Suppl), S78-83.
- Norman, G. R., Rosenthal, D., Brooks, L. R., Allen, S. W., & Muzzin, L. J. (1989). The development of expertise in dermatology. *Arch Dermatol*, *125*(8), 1063-1068.
- Patel, V., Arocha, J., & Kaufman, D. (1994). Diagnostic reasoning and medical expertise. *Psychol Reason Motiv*, *31*, 187-252.
- Patel, V., & Goren, G. (1986). Knowledge-based solution strategies in medical reasoning. *Cognit Sci*, *10*, 91-116.
- Peterson, M. W., Gordon, J., Elliott, S., & Kreiter, C. (2004). Computer-based testing: initial report of extensive use in a medical school curriculum. *Teach Learn Med*, *16*(1), 51-59.
- Peterson, R. (2000). Examining the examination: Canadian versus US certification exam. *Can Assoc Radiol*, *51*(3), 209.

- Quekel, L. G., Kessels, A. G., Goei, R., & van Engelshoven, J. M. (1999). Miss rate of lung cancer on the chest radiograph in clinical practice. *Chest*, *115*(3), 720-724.
- Raufaste, E., & Eyrolle, H. (1998). [Radiological expertise and diagnosis. I. Theoretical advances]. *J Radiol*, *79*(3), 227-234.
- Raufaste, E., Verderi-Raufaste, D., & Eyrolle, H. (1998). [Radiological expertise and diagnosis. II. Empirical study]. *J Radiol*, *79*(3), 235-240.
- Robinson, P. J. (1997). Radiology's Achilles' heel: error and variation in the interpretation of the Rontgen image. *Br J Radiol*, *70*(839), 1085-1098.
- Rubin, A. (1989). Contributions of cognitive science and educational technology to training in radiology. *Invest Radiol*, *24*(9), 729-732.
- Samuel, S., Kundel, H. L., Nodine, C. F., & Toto, L. C. (1995). Mechanism of satisfaction of search: eye position recordings in the reading of chest radiographs. *Radiology*, *194*(3), 895-902.
- Sarria, E., Fischer, G. B., Lima, J. A., Menna Barreto, S. S., Flores, J. A., & Sukiennik, R. (2003). [Interobserver agreement in the radiological diagnosis of lower respiratory tract infections in children]. *J Pediatr (Rio J)*, *79*(6), 497-503.
- Schank, R., & Abelson, R. (1977). *Scripts, Plans, Goals, and Understanding: An Inquiry Into Human Knowledge Structures*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- 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.
- Schön, D. (1983). *The reflective practitioner : How professionals think in action*. New York.
- 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.
- 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.
- Sibert, L., Charlin, B., Gagnon, R., Corcos, J., Khalaf, A., & Grise, P. (2001). [Evaluation of clinical reasoning in urology: contribution of the Script Concordance Test]. *Prog Urol*, *11*(6), 1213-1219.
- Sibert, L., Darmoni, S. J., Dahamna, B., Hellot, M. F., Weber, J., & Charlin, B. (2006). On line clinical reasoning assessment with Script Concordance test in urology: results of a French pilot study. *BMC Med Educ*, *6*, 45.
- Sibert, L., Darmoni, S. J., Dahamna, B., Weber, J., & Charlin, B. (2005). Online clinical reasoning assessment with the Script Concordance test: a feasibility study. *BMC Med Inform Decis Mak*, *5*, 18.
- Sowden, P. T., Davies, I. R., & Roling, P. (2000). Perceptual learning of the detection of features in X-ray images: a functional role for improvements in adults' visual sensitivity? *J Exp Psychol Hum Percept Perform*, *26*(1), 379-390.
- Thompson, K. G., Bichot, N. P., & Sato, T. R. (2005). Frontal eye field activity before visual search errors reveals the integration of bottom-up and top-down salience. *J Neurophysiol*, *93*(1), 337-351.

- Treue, S. (2003). Visual attention: the where, what, how and why of saliency. *Curr Opin Neurobiol*, 13(4), 428-432.
- Triola, M., Feldman, H., Kalet, A. L., Zabar, S., Kachur, E. K., Gillespie, C., et al. (2006). A randomized trial of teaching clinical skills using virtual and live standardized patients. *J Gen Intern Med*, 21(5), 424-429.

ANNEXES



Index of the whole available SCTs : 2 test(s) found

[Administration Accueil](#)

Click on the title to watch the demo.

1. **DÉMO: DESCRIPTION** (SCT) de 10m (10 minutes)
 - Author : Selma Mahiri ([selmahiri@umontreal.com](mailto:selmamahiri@umontreal.com))
 - Status : Ready (Participants can request their answers)
 - Summary : Ce test consiste à détecter et à identifier des anomalies observées sur des radiographies pulmonaires. Il s'agit de la version démo et ne comporte qu'un seul choix.
2. **DÉMO: INTERPRÉTATION** (SCT) de 10m (10 minutes)
 - Author : Selma Mahiri ([selmahiri@umontreal.com](mailto:selmamahiri@umontreal.com))
 - Status : Ready (Participants can request their answers)
 - Summary : Le test d'interprétation consiste à interpréter des radiographies de deux scanners différents, en indiquant la vraisemblance des hypothèses diagnostiques possibles. Cette version démo comporte deux choix et six réponses.



Annexe 1. Test de perception, démonstration en ligne

Page d'accueil

<http://www.fpcmed2.umontreal.ca:8080/sctimagesdemo3/>



Test de concordance de script

appliqué à l'imagerie

Dr. demoPrenom demoNom connecté au test
"DEMO-1-PERCEPTION"

➤ Ce test contient 1 cas (Total des questions : 1) :

Description :

INSTRUCTIONS

Les radiographies pulmonaires qui vous seront présentées peuvent être normales, comporter une ou plusieurs images pathologiques, des images de composition, ou des artefacts.

1. Si le cliché est normal, vous cliquez sur la case : "Je pense qu'il s'agit d'un cliché normal" et vous passez au cliché suivant.
2. Si vous détectez des anomalies localisées (images pathologiques, de composition ou artefact), vous cliquez séparément sur chaque anomalie observée et vous l'identifiez à partir du menu proposé qui s'ouvre automatiquement. Ensuite, vous passez à la détection et à l'identification de l'anomalie suivante. Une fois toutes les anomalies identifiées, vous passez au cliché suivant. Vous avez droit à 7 clics au maximum par cliché.
3. Si l'anomalie est diffuse (i.e. hyperinflation pulmonaire ou infiltrat interstitiel diffus), vous cliquez sur n'importe quelle zone du cliché pour l'identifier à partir du menu proposé.

Test de perception : Instructions aux participants

Dr. demoPrenom demoNom connecté au test
"DÉMO-1-PERCEPTION"

[Recharger cette page](#) [Mettre en pause ce test](#)

[Terminer ce test maintenant](#)

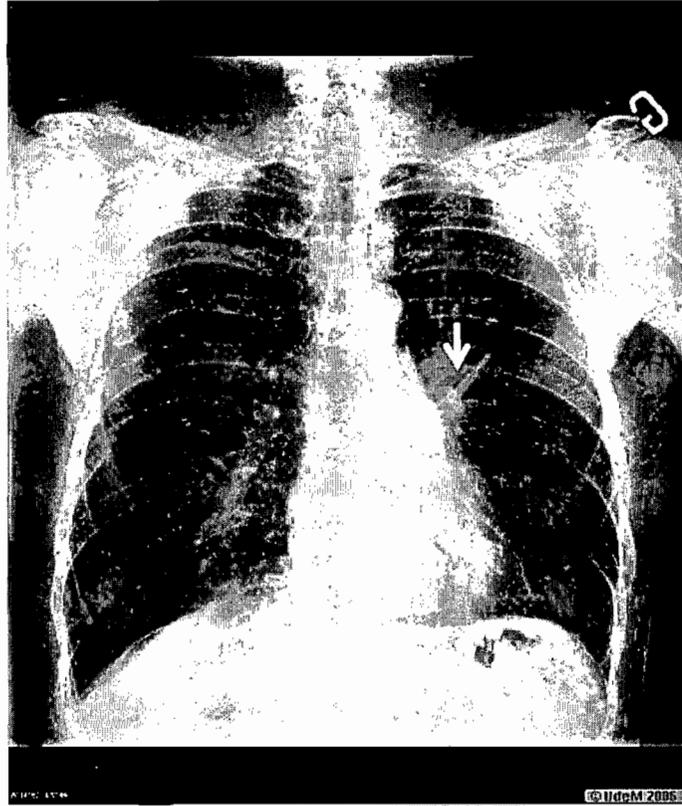
Votre décision vis à vis du cliché ci-dessous (cliché 1 de 1)

Vous pouvez :

- effectuer jusqu'à 3 **selections** sur ce cliché
- OU classer ce cliché comme étant **normal**.

(Veuillez attendre que l'image ci-dessous soit complètement chargée avant d'agir.)

[Taille initiale](#) [Agrandir l'image](#) [Agrandir 2 fois l'image](#)



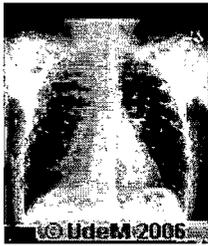
[Taille initiale](#) [Agrandir l'image](#) [Agrandir 2 fois l'image](#)

Test de perception : Étape d'identification d'une anomalie (1)

Votre décision vis à vis du cliché ci-dessous (cliché 1 de 1)

Vos sélection(s) :

Selection n°1: (...interprétation en cours...)



Associez votre sélection à l'une des sémiologies suivantes : (cliquez sur pour ouvrir une catégorie)

Image de composition ou artéfact

Anomalies

Vous pouvez :

- effectuer jusqu'à **3 selections** sur ce cliché

(Veuillez attendre que l'image ci-bas soit complètement chargée avant d'agir.)

Test de perception : Étape d'identification d'une
anomalie (2)

Votre décision vis à vis du cliché ci-dessous (cliché 1 de 1)

Vos sélection(s) : [Effacer tout]

Selection n°1: (...interprétation en cours...) [Effacer]

Associez votre sélection à l'une des sémiologies suivantes : (cliquez sur [] pour ouvrir une catégorie)



Image de composition ou artéfact

- Anomalies
 - Anomalies parenchymateuses
 - Alvéoles
 - Atélectasies
 - Images cicatricielles
 - Images interstitielles
 - Nodules / Masses pulmonaires
 - nodule / masse
 - nodule / masse calcifié(e)
 - nodule / masse cavitaire
 - cavitation avec niveau hydro-aérique
 - croissant aérien
 - Radiotransparences
 - Volumes pulmonaires
- Aorte Thoracique
- Bronches
- Hiles / Artères pulmonaires
- Médiastin
- Trachée
- Plèvre
- Paroi / Diaphragme
- Tubes / Canules / Cathéters / appareillages
- Cou
- Abdomen

Vous pouvez :

- effectuer jusqu'à 3 selections sur ce cliché

(Veuillez attendre que l'image ci-bas soit complètement chargée avant d'agir.)

[Taille initiale] [Agrandir l'image] [Agrandir 2 fois l'image]

Test de perception : Étape d'identification d'une anomalie (3)

Votre décision vis à vis du cliché ci-dessous (cliché 1 de 1)

Vos sélection(s) :

Selection n°1: nodule / masse

Vous pouvez :

- effectuer jusqu'à 3 selections sur ce cliché

Cochez cette case si vous avez déjà vu ce cas dans le passé

Vous avez terminé, passez au cliché suivant

(Veuillez attendre que l'image ci-bas soit complètement chargée avant d'agir.)

Test de perception : Étape d'identification d'une anomalie (4)

Annexe 2. Test de perception, anomalies à détecter et à identifier

Terminologie utilisée

1- Anomalies parenchymateuses	
Alvéoles	<ul style="list-style-type: none"> ○ Opacités floues ○ Consolidation ○ Bronchogramme aérique ○ Opacités en verre dépoli ○ Pneumonie ronde
Atélectasies	<ul style="list-style-type: none"> ○ Specifications du lobe concerné
Images cicatricielles	<ul style="list-style-type: none"> ○ Remaniement cicatriciel ○ Rétraction cicatricielle
Images interstitielles	<ul style="list-style-type: none"> ○ Reticulations ○ Opacités miliaires ○ Opacités micro-nodulaires ○ Lignes septales de kerley b ○ Opacités micro-nodulaires ○ Images en rayon de miel ○ Images kystique
Nodules / masses pulmonaires	<ul style="list-style-type: none"> ○ Nodule / masse ○ Nodule / masse calcifié(e) ○ Nodule / masse cavaire ○ Cavitation avec niveau hydro-aérique ○ Croissant aérien
Radiotransparances	<ul style="list-style-type: none"> ○ Emphysème ○ Bulle d'emphysème
Volumes pulmonaires	<ul style="list-style-type: none"> ○ Hyperinflation ○ Volumes réduits
2. Aorte thoracique	
<ul style="list-style-type: none"> ○ Anévrisme sacculaire ○ Dilatée ○ Dissection ○ Calcifications 	

3. Bronches
<ul style="list-style-type: none"> ○ Épaississement de la paroi bronchique ○ Bronchiectasies
4. Hiles / artères pulmonaires
<ul style="list-style-type: none"> ○ Adénopathie ○ Adénopathie calcifiée ○ Hile dense ○ Artères pulmonaires dilatées
5. Médiastin
<ul style="list-style-type: none"> ○ Adénopathie ○ Adénopathie calcifiée ○ Élargissement médiastinal ○ Épanchement péricardique ○ pneumomédiastin ○ Cardiomégalie ○ Masse médiastinale supérieure ○ Masse médiastinale antérieure ○ Masse médiastinale moyenne ○ Achalasie ○ Hernie hiatale ○ Niveau hydro-aérique dans l'oesophage
6. Trachée
<ul style="list-style-type: none"> ○ Comblement rétrotrachéal ○ Déviation trachéale ○ Masse trachéale ○ Sténose trachéale
7. Plèvre
<ul style="list-style-type: none"> ○ Épaississement pleural ○ Épanchement pleural ○ Épanchement pleural cloisonné ○ Épanchement intra-scissural ○ Scissure accessoire ○ Fibrothorax ○ Masse(s) pleurale(s) ○ Plaques pleurales ○ Pneumothorax ○ Pneumothorax cloisonné ○ Hydro-pneumothorax

8. Paroi thoracique	
Diaphragme	<ul style="list-style-type: none"> ○ Hernie de bochdalek ○ Surélévation
Côtes	<ul style="list-style-type: none"> ○ Fracture récente ○ Fracture ancienne
Sternum	<ul style="list-style-type: none"> ○ Lésion lytique ○ Lésion blastique ○ Ilôt d'ostéocondensation ○ Pectus excavatum ○ Dysplasie fibreuse ○ Enchondrome
Rachis	<ul style="list-style-type: none"> ○ Arthrose ○ Écrasement vertébral ○ Rugger jersey spine ○ Vertèbres en H ○ Lésion lytique ○ Lésion blastique ○ Ilôt d'ostéocondensatio ○ Dysplasie fibreuse
Épaule (clavicule, omoplate, humérus)	<ul style="list-style-type: none"> ○ Fracture ○ Luxation ○ Lésion lytique ○ Lésion blastique ilôt d'ostéocondensation ○ dysplasie fibreuse ○ Dysplasie fibreuse
Tissus mous	<ul style="list-style-type: none"> ○ Add new ○ Emphysème sous-cutané ○ Masse pariétale ○ Mastectomie
9. Cou	
<ul style="list-style-type: none"> ○ Calcifications carotidiennes ○ Larynx ○ Deviation ○ Masse 	
10. Abdomen	
<ul style="list-style-type: none"> ○ Agraffes chirurgicales ○ Calcifications ○ Pneumopéritoine ○ Splénomégalie 	

11. Tubes / canules / cathéters /
<ul style="list-style-type: none"> ○ Agraffes chirurgicales ○ Catheters ○ Jugulaire ○ Sous-clavier ○ Canule de trachéostomie ○ Pace-maker transveineux ○ Tube endo-trachéal ○ Tube naso-gastrique

Test de perception : catégories et anomalies à détecter

Catégories d'anomalies	Anomalies à détecter	Nombre d'anomalies
Médiastin	Adénopathies hilaires	1
	Adénopathies médiastinales	3
	Masse médiastinale	1
	Élargissement médiastinal	1
	Pneumomédiastin	1
	Hernie hiatale	1
Cardio-vasculaires	Cardiomégalie	1
	Anomalies de l'aorte thoracique	2
	Anomalies des artères pulmonaires et hiles	1
	Cardiomégalie	1
Parenchyme pulmonaire	Images alvéolaires	2
	Atélectasies	3
	Nodules et masses pulmonaires	12
	Remaniements cicatriciels	1
	Images interstitielles	2
	Hyperinflation pulmonaire	1
Voies respiratoires	Déviations trachéales	1
Plèvre	Épaississement scissural	1
	Pneumothorax	1
	Plaques pleurales	1
	Épanchement pleural	1
	Épanchement pleural	1
	Épanchement pleural	1
	Plaques pleurales	1
Paroi	Cal osseux	1
	Lésion lytique costale	1
	Fracture costale	1
	Lésion lytique	1
Divers	Tube naso-gastrique	1
	Emphysème sous-cutané	1
	Agrafes chirurgicales	1
	Agrafes chirurgicales	1
	Image de composition ou artéfact	1
Total		51

Vignettes TP	Nombre d'anomalies	Types d'anomalies	Difficulté
1	2	Fracture costale, anomalies alvéolaires	2,5
2	3	Vaisseaux pulmonaires, nodules et masses pulmonaires, atélectasie	2,0
3	3	Adénopathies hilaires, anomalies alvéolaires, épaissement scissural	2,3
4	1	Nodules et masses pulmonaires	3,0
5	2	Pneumothorax, pneumomédiastin	2,0
6	2	Images interstitielles, plaques pleurales	1,5
7	4	Adénopathies médiastinales, plaques pleurales	2,0
8	3	Plèvre, nodules et masses pulmonaires, atélectasie	2,0
9	4	Nodules et masses pulmonaires, hyperinflation, divers (tubulure)	1,8
10	1	Normal	1,0
11	2	Nodules et masses pulmonaires, lésion lytique	2,5
12	1	Plaque pleurale	2,0
13	1	Masse médiastinale	3,0
14	3	Pneumomédiastin, images interstitielles, divers (emphysème sous-cutané)	2,0
15	1	Anomalie de la trachée	3,0
16	6	Nodules et masses pulmonaires, hernie hiatale, divers (agrafes chirurgicales), images de composition ou artéfact, anomalie de la paroi	2,0
17	5	Nodules et masses pulmonaires, anomalie vasculaire (aorte thoracique)	2,8
18	1	Élargissement médiastinal	2,0
19	4	Nodules et masses pulmonaires, atélectasie, cardiomégalie, anomalie vasculaire (aorte thoracique)	1,8
20	2	Nodules et masses pulmonaires, lésion lytique costale	2,3

Annexe 3. Contenu des vignettes de perception : Liste des anomalies par vignette et niveaux de difficulté

Le niveau de difficulté de chaque question est établi par les constructeurs du test, selon l'échelle de difficulté RadLex (Langlotz, 2006). La difficulté de chaque vignette correspond à la moyenne du niveau de difficulté des questions.

Vignette de Détection (TP)		
1. Le cliché RxP ne présente pas de défauts techniques (faible ou forte pénétration, images amputées...)	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
2. La radiographie montre une ou des anomalies possiblement difficiles à détecter, même par un radiologue (anomalies subtiles pouvant passer inaperçues)	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
3. Identification et catégorie de chaque anomalie	Niveau de difficulté	
• Anomalie 1	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/>	
• Anomalie 2	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/>	
• Anomalie 3	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/>	
• Anomalie 4	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/>	
• Anomalie 5	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/>	
• Anomalie 6	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/>	
4. L'ensemble des vignettes RxP présentées couvrent des situations fréquemment rencontrées dans la pratique courante	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
5. L'ensemble des vignettes RxP présentées sont représentatives des objectifs de formation post-doctorale en imagerie thoracique	Oui <input type="checkbox"/>	Non <input type="checkbox"/>

Annexe 4. Grille de contrôle de qualité du test de perception

Test de concordance de script

appliqué à l'imagerie

Dr. demoPrenom demoNom connecté au test
"DEMO-2-INTERPRETATION"

Ce test contient 2 cas (Total des questions : 6) :

Description :

Chaque cas se présente sous forme d'une vignette radiologique. Les questions ont deux formats possibles:
Format 1: La vignette radiologique est suivie d'une hypothèse diagnostique. Ensuite, une nouvelle information, clinique ou radiologique, vous est fournie. Selon cette information, vous devez juger de la vraisemblance de l'hypothèse proposée, en pondérant votre réponse sur une échelle allant de -2 à +2 correspondant aux options suivantes :

- 2 = exclue ou pratiquement exclue
- 1 = moins probable
- 0 = n'a aucun effet sur l'hypothèse
- + 1 = plus probable
- + 2 = pratiquement probable

Format 2: La vignette radiologique est suivie d'une série de diagnostics différentiels possibles. Votre tâche consiste à indiquer la probabilité de chacun de ces diagnostics sur une échelle de 4 points:

Très peu probable
Peu probable
Probable
Très probable

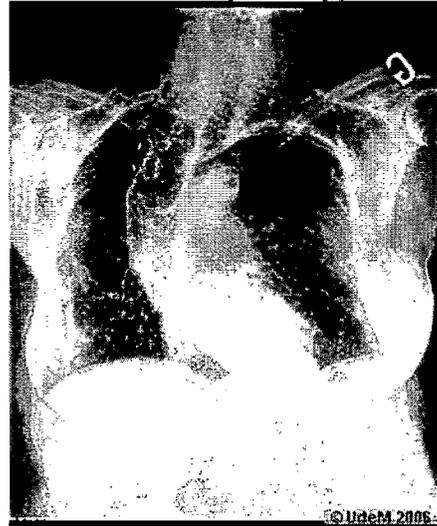
Annexe 5. Test d'interprétation : Instructions aux participants

<http://www.fpcmed2.umontreal.ca:8080/sctimagesdemo3/>

Clinical Vignette :

Mme B, 58 ans, astmatique, présente la radiographie pulmonaire suivante

(Click to enlarge this image)



Parmi les 3 diagnostics possibles, indiquez la vraisemblance de chacun

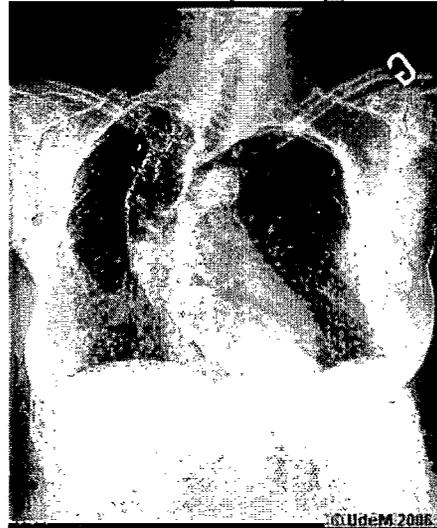
Hypothèse N°1	
Tumeur fibreuse bénigne de la plèvre	<input type="radio"/> Très peu probable <input type="radio"/> Peu probable <input type="radio"/> Probable <input type="radio"/> Très probable
<input type="checkbox"/> Cochez cette case si vous avez déjà vu ce cas dans le passé	
<input type="button" value="Valider"/>	

Test d'interprétation : 1^{ère} question d'une vignette d'interprétation probabiliste

Clinical Vignette :

Mme B, 58 ans, astmatique, présente la radiographie pulmonaire suivante

(Click to enlarge this image)



Parmi les 3 diagnostics possibles, indiquez la vraisemblance de chacun

Hypothèse N°2

Mésothéliome malin

- Très peu probable
- Peu probable
- Probable
- Très probable

Cochez cette case si vous avez déjà vu ce cas dans le passé

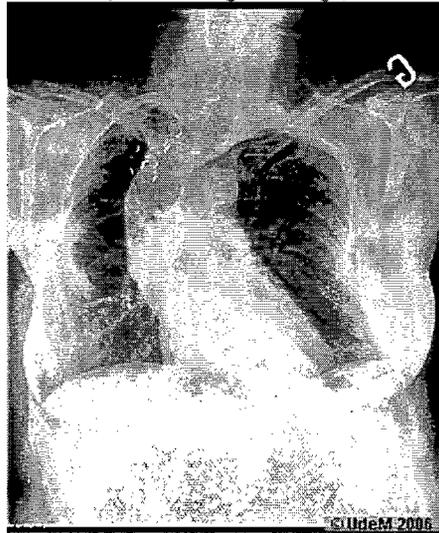
Valider

Test d'interprétation : 2^{ème} question d'une vignette d'interprétation probabiliste

Clinical Vignette :

Mme B. 58 ans, astmatique, présente la radiographie pulmonaire suivante

(Click to enlarge this image)



Parmi les 3 diagnostics possibles, indiquez la vraisemblance de chacun

Hypothèse N°3

Néoplasie pulmonaire

- Très peu probable
- Peu probable
- Probable
- Très probable

Cochez cette case si vous avez déjà vu ce cas dans le passé

Valider

Test d'interprétation : 3^{ème} question d'une vignette d'interprétation probabiliste

Cas clinique N°27 (of 30), question 1 of 3 :

Clinical Vignette :

M. D, 62 ans retraité, se plaint de douleurs thoraciques et présente la radiographie pulmonaire suivante

(Cliquez pour agrandir)



Si on vous suggère

et que vous apprenez que le patient

vous en concluez

une métastase pleurale	a des antécédents de cancer du rein	<p><input type="radio"/> -2 : que l'hypothèse est très peu probable ou exclue</p> <p><input type="radio"/> -1 : que l'hypothèse est moins probable</p> <p><input type="radio"/> 0 : que cela n'a aucun effet sur l'hypothèse</p> <p><input type="radio"/> +1 : que l'hypothèse est plus probable</p> <p><input type="radio"/> +2 : que l'hypothèse est quasi probable</p> <p><input type="checkbox"/> Cochez cette case si vous avez déjà vu ce cas dans le passé</p> <p><input type="button" value="valider"/></p>
------------------------	-------------------------------------	---

Test d'interprétation : 1^{ère} question d'une vignette d'interprétation analytique

Cas clinique N°29 (of 30), question 2 of 3 :

Clinical Vignette :

M. D., 62 ans retraité, se plaint de douleurs thoraciques et présente la radiographie pulmonaire suivante

(Cliquez pour agrandir)



Si on vous suggère

et que le patient

vous en concluez

un myélome

se plaint de douleurs thoraciques

- 2 : que l'hypothèse est très peu probable ou exclue
- 1 : que l'hypothèse est moins probable
- 0 : que cela n'a aucun effet sur l'hypothèse
- +1 : que l'hypothèse est plus probable
- +2 : que l'hypothèse est quasi probable

Cochez cette case si vous avez déjà vu ce cas dans le passé

Valider

Test d'interprétation : 2^{ème} question d'une vignette d'interprétation analytique

Cas clinique N°29 (of 30), question 3 of 3 :

Clinical Vignette :

M. D, 62 ans retraité, se plaint de douleurs thoraciques et présente la radiographie pulmonaire suivante

(Cliquez pour agrandir)



Si on vous suggère	et que le patient	vous en concluez
tumeur fibreuse bénigne de la plèvre	présente des hypoglycémies à répétition	<p><input type="radio"/> -2 : que l'hypothèse est très peu probable ou exclue</p> <p><input type="radio"/> -1 : que l'hypothèse est moins probable</p> <p><input type="radio"/> 0 : que cela n'a aucun effet sur l'hypothèse</p> <p><input type="radio"/> +1 : que l'hypothèse est plus probable</p> <p><input type="radio"/> +2 : que l'hypothèse est quasi probable</p> <p><input type="checkbox"/> Cochez cette case si vous avez déjà vu ce cas dans le passé</p> <p><input type="button" value="Valider"/></p>

Test d'interprétation : 3^{ème} question d'une vignette d'interprétation analytique

Vignette	Catégorie d'anomalie	Diagnostic	Difficulté	Type de vignette
1	Pleurale	Plaques pleurales	2	IP
2	Pleurale	Hémothorax	1	IP
3	Pleurale	Tumeur fibreuse bénigne de la plèvre	3	IP
4	Pleurale	Mésothéliome malin	2	IP
5	Pleurale, parenchymateuse	Épanchement pleural néoplasique	2	IP
6	Parenchymateuse	Silicose	2	IP
7	Parenchymateuse	Réaction médicamenteuse	3	IP
8	Parenchymateuse	Silicose	2	IP
9	Parenchymateuse	Amyloïdose diffuse	3	IP
10	Parenchymateuse	Infection à <i>Pneumocystis jiroveci</i>	3	IP
11	Parenchymateuse	Pneumonie lipidique	3	IP
12	Parenchymateuse	Tuberculose	1	IP
13	Parenchymateuse	Miliaire tuberculeuse		IP
14	Parenchymateuse	Histoplasmose	3	IP
15	Médiastin, adénomégalies	Sarcoïdose	2	IP
16	Paroi	Myélome	3	IA
17	Parenchymateuse	Nodule métastatique	2	IA
18	Parenchymateuse	Pneumonite radique	2	IA
19	Parenchymateuse	Alvéolite allergique extrinsèque	2	IA
20	Parenchymateuse	Infection à <i>Pneumocystis jiroveci</i>	2	IA
21	Médiastin, adénomégalies	Lymphome	2	IA
22	Paroi	Thalassémie	2	IA
23	Parenchymateuse	Séquestration	3	IA
24	Parenchymateuse	Vasculite	3	IA
25	Parenchymateuse	Métastases, embolie pulmonaire	3	IA
26	Parenchymateuse	Fibrose pulmonaire idiopathique	1	IA
27	Parenchymateuse	Lymphangite carcinomateuse	2	IA
28	Parenchymateuse	Sarcoïdose	2	IA
29	Parenchymateuse	Lymphangite carcinomateuse	2	IA
30	Parenchymateuse	Fibrose pulmonaires	2	IA

Annexe 6. Contenu des vignettes d'interprétation
Type d'anomalies, diagnostics et niveau de difficulté

Vignette d'interprétation (TI)		
1. Le cliché RxP ne présente pas de défauts techniques (faible ou forte pénétration,...)	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
2. Présente une RxP possiblement problématique pour un radiologue (images atypiques ou ambiguës évoquant plusieurs hypothèses)	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
3. La RxP montre des images appropriées au niveau de formation des personnes évaluées	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
4. Catégorie de la principale anomalie à interpréter	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
5. Niveau de difficulté de la RxP à interpréter	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/>	
6. L'énoncé clinique précédant la RxP est clair et il reproduit une situation réaliste	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
7. L'ensemble des vignettes RxP présentées couvrent des situations fréquemment rencontrées dans la pratique courante	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
8. L'ensemble des vignettes RxP présentées sont représentatives des objectifs de formation post-doctorale en imagerie thoracique	Oui <input type="checkbox"/>	Non <input type="checkbox"/>

Hypothèses diagnostiques		
1. Pertinentes à la situation clinique, du point de vue d'un expert	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
2. La même option ne se retrouve pas dans deux questions consécutives	Oui <input type="checkbox"/>	Non <input type="checkbox"/>

Informations complémentaires fournies		
1. L'information permet de tester la force de l'association entre l'hypothèse et l'information nouvelle, dans le contexte de la vignette radiologique	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
2. Les questions construites correspondent à des questions que l'on se pose réellement en situation clinique	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
3. La forme de la question est compatible avec le raisonnement dans la discipline	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
4. Les questions sont construites de façon à répartir les réponses des experts de façon équilibrée entre: -2, -1, 0, +1, +2 ; ou très peu, peu, probable, ou très probable	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
5. La vignette RxP est plus appropriée à un format IP ou IA	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
6. La répartition des vignettes RxP d'IA et d'IP est équilibrée	Oui <input type="checkbox"/>	Non <input type="checkbox"/>

Annexe 7. Grille de contrôle de qualité du test d'interprétation



Composition du panel de référence pour chaque test		
1. Le nombre de radiologues est suffisamment important	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
2. Les membres du panel peuvent légitimement faire partie d'un panel de référence, compte tenu des caractéristiques des personnes examinées	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
3. La composition en spécialistes et en sur-spécialistes paraît adéquate	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
4. Chaque membre du panel a passé le test individuellement, dans des conditions similaires à celles des candidats, spécifier si ce n'est pas le cas	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
5. Chaque membre du panel a reçu l'instruction de répondre de son mieux, même s'il ne se sent pas particulièrement compétent pour répondre à une question donnée	Oui <input type="checkbox"/>	Non <input type="checkbox"/>

Annexe 8. Grille de contrôle de qualité du panel de référence

Annexe 9. Lettres d'information et formulaires de consentement

Lettre d'information et formulaire de consentement destinés aux résidents et aux externes

Bonjour,

Dans le cadre du projet recherche en pédagogie médicale "*Measurement of perception and interpretation skills along radiology training: Validity and reliability of the script concordance approach*" financé par le Conseil médical du Canada, nous développons deux tests d'évaluation informatisés en imagerie thoracique. Ils visent à mesurer : 1) les aptitudes à détecter et à identifier les images subtiles sur des radiographies thoraciques ; 2) les capacités de raisonnement en contexte d'incertitude, dans le domaine de l'imagerie thoracique.

Ce projet, conçu à l'université de Montréal et approuvé par les comités d'éthique, comporte deux tests utilisant le concept de concordance des scripts (TCS.) Le TCS est un outil de mesure réaliste et précis pour les situations frustrantes où le consensus est difficile. Il compare la concordance des réponses des candidats à celles d'un groupe d'experts tout en tenant compte de leurs divergences d'opinion. Après leur validation, les tests pourraient servir dans des examens de certification, dans l'évaluation formative des résidents et des externes, ainsi qu'en formation médicale continue.

Les tests sont en cours de validation et nous avons besoin de votre participation à titre de résident ou d'externe. En plus de contribuer à l'amélioration des tests d'évaluation, votre collaboration vous donnera l'occasion de vous familiariser avec un nouveau format de test développé dans le but d'être utilisé dans les examens de certification du collège royal du Canada.

Ce travail, réalisable en ligne, devrait prendre une heure à une heure et demie de votre temps pour chacun des tests. Les questions portent sur les connaissances de pratique courante. Vos fichiers de réponses seront immédiatement codés puis traités de façon absolument anonyme.

Une démonstration des tests est disponible à l'adresse suivante :

<http://www.fpcmed2.umontreal.ca:8080/sctimagesdemo3/>

Enfin, comme il s'agit d'une version initiale, une fois les tests complétés, vos suggestions et commentaires, seront les bienvenus pour nous aider à l'améliorer.

Pour confirmer votre participation, veuillez compléter et signer le formulaire ci-dessous après avoir lu le document de consentement et me le retourner par courriel.

Consentement :

Je, soussigné, _____, déclare avoir lu et compris les termes du formulaire de consentement et accepte, par la présente, à participer au projet de recherche suivant, dans les conditions décrites dans le formulaire de consentement.

Merci infiniment pour votre attention et votre aide.

Selma Mehiri

Lettre d'information et formulaire de consentement destinés aux radiologues

Cher Docteur,

Dans le cadre de mon projet de maîtrise en pédagogie médicale, je développe **deux tests d'évaluation informatisés en imagerie thoracique**, avec les Docteurs Louise Samson et Bernard Charlin. Il s'agit d'une recherche financée par le Conseil médical du Canada sous l'intitulé "*Measurement of perception and interpretation skills along radiology training: Validity and reliability of the script concordance approach*". À terme, ces tests serviront aux examens de certification, à l'évaluation formative des résidents et à la formation médicale continue.

Il s'agit de tests conviviaux, qui ont pour **but** d'évaluer des résidents finissants (R5) en radiologie en mesurant : 1) leurs aptitudes à détecter et à identifier les images subtiles ou ambiguës sur des radiographies thoraciques ; 2) leurs capacités de raisonnement en contexte d'incertitude, dans le domaine de l'imagerie thoracique. Ils ont la particularité de tenir compte à la fois de la précision nécessaire pour établir un diagnostic radiologique et de la divergence qui peut exister entre les experts.

Le projet est approuvé par les comités d'éthique de l'université de Montréal et du CHUM. Il comporte deux tests en ligne (sécurisés) qui reproduiront des conditions proches de la pratique courante. Le premier test consiste à détecter puis à identifier des anomalies subtiles sur des clichés de radiographies pulmonaires. Le second test comprend des clichés de radiographies pulmonaires pouvant comporter une ambiguïté diagnostique. La tâche consiste à déterminer la vraisemblance d'hypothèses diagnostiques en fonction des informations radiologiques et cliniques proposées. Le système de notation utilisé est basé sur le principe du test de concordance des scripts, qui tient compte de la divergence d'opinion entre les experts et s'appuie sur la théorie hypothético-déductive du raisonnement clinique. Les réponses des candidats seront comparées à celles de médecins experts radiologistes regroupés en un panel de référence.

Présentement, les tests sont en cours de validation. Votre participation au projet, en tant qu'expert, est cruciale. Vous ferez partie du panel d'experts dont les réponses permettront d'établir les scores des candidats. Ce travail est réalisable **en ligne** et devrait prendre **trois quarts d'heure à une heure pour chacun des tests**. Les questions portent sur les connaissances utilisées dans la pratique clinique courante. Les fichiers de réponses seront immédiatement codés puis traités de façon absolument **anonyme**.

Une **démonstration en ligne est disponible** à l'adresse suivante :

<http://www.fpcmed2.umontreal.ca:8080/sctimagesdemo3/>. Des informations supplémentaires sur le TCS sont également disponibles : <http://sedmc03.meddir.umontreal.ca/tcs/>

Pour confirmer votre participation, je vous demanderai de me retourner ce document encompétant le formulaire ci-dessous après avoir lu le document de consentement. Ensuite, je vous communiquerai, par courriel, vos codes d'accès pour réaliser les tests.

Consentement :

Je, soussigné, _____, déclare avoir lu et compris les termes du formulaire de consentement et accepte, par la présente, à participer au projet de recherche suivant, dans les conditions décrites dans le formulaire de consentement.

Merci infiniment pour votre attention et votre aide.

Selma Mehiri MD

Annexe 10. Questionnaires

Questionnaire destiné aux externes

1. Indiquez la période d'externat actuelle
Période 1 à 20
2. Indiquez la période d'externat durant laquelle vous avez effectué votre stage de radiologie
Période 1 à 20
3. Indiquez si vous avez déjà effectué votre stage en imagerie thoracique
 - a. Oui Si oui, indiquez-en la durée :
 1. 1 jour
 2. 2 jours
 - b. Non
4. Avez-vous effectué un stage en :

a. Pneumologie	Si oui, indiquez la période	Période (1 à 20)
b. Cardiologie	Si oui, indiquez la période	Période (1 à 20)
c. Médecine interne	Si oui, indiquez la période	Période (1 à 20)
d. Soins intensifs	Si oui, indiquez la période	Période (1 à 20)
e. Chirurgie thoracique	Si oui, indiquez la période	Période (1 à 20)
f. Urgences	Si oui, indiquez la période	Période (1 à 20)

Sur une échelle de 1 à 10, indiquer :

1. Le niveau de difficulté perçue pour le test de perception :
2. Le niveau de difficulté perçue pour le test d'interprétation :

Commentaires :

Merci de votre collaboration !

Questionnaire destiné aux résidents

Nom et Prénom :

Code réservé à l'administration

Pour des fins d'analyses, veuillez répondre aux questions suivantes :

1. Indiquez votre niveau de résidence
2. Indiquez la période de stage actuelle
3. Si vous êtes Résident 1, indiquez si vous avez déjà complété ou êtes en train de compléter :
 - a) votre stage en pneumologie
 - b) vos stages d'introduction à la radiologie Si oui, indiquez lequel :
4. Si vous êtes Résident 2, indiquez si vous avez complété, ou êtes en train de compléter vos stages d'imagerie thoracique Si oui, indiquez lequel :
5. Si vous êtes Résident 3, 4 ou 5, indiquez si vous avez déjà complété, ou êtes en train de compléter vos stages obligatoires en imagerie thoracique Si oui, indiquez le nombre total de stages obligatoires déjà accomplis :
6. Si vous êtes Fellow (R6), indiquez si vous avez complété tous vos stages obligatoires en imagerie thoracique
7. Quelque soit votre niveau de résidence, indiquez si vous avez déjà complété, ou êtes en train de compléter des stages optionnels en imagerie thoracique Si oui, indiquez le nombre de stages optionnels déjà accomplis :

Sur une échelle de 1 à 10, indiquer :

- Le niveau de difficulté perçue pour le test de perception :
- Le niveau de difficulté perçue pour le test d'interprétation :

Commentaires :

Merci de votre collaboration !

Questionnaire destiné aux radiologues

- 1) **Indiquez le nombre d'années de pratique en radiologie :**
 - a) 0 à 5 ans
 - b) 5 à 10 ans
 - c) 10 à 15 ans
 - d) 15 à 20 ans
 - e) Supérieur à 20 ans
- 2) **Indiquez le pourcentage de temps consacré à l'imagerie thoracique dans votre pratique :**
 - a) < 50%
 - b) ≥ 50%
- 3) **Indiquez depuis combien d'années, vous consacrez au moins 75% de votre temps à l'imagerie thoracique :**
 - a) < 5 ans
 - b) 5 à 10 ans
 - c) 10 à 15 ans
 - d) 15 à 20 ans
 - e) Supérieur à 20 ans

Sur une échelle de 1 à 10, indiquer :

- Le niveau de difficulté perçue pour le test de perception :
- Le niveau de difficulté perçue pour le test d'interprétation :

Commentaires :

Merci de votre collaboration !

Insuffisante
Score = 1
Inférieure
Score = 2
Conforme
Score = 3
Dépasse
Score = 4

Professionalisme

1. Sens des responsabilités
2. Respect et honnêteté dans ses rapports professionnels
3. Empathie envers ses patients
4. Capacité d'appliquer les principes éthiques en clinique
5. Capacité d'autocritique

Expertise

6. Connaissances cliniques
7. Connaissances fondamentales
8. Anamnèse organisée, dirigée et pertinente
9. Maîtrise des gestes d'examen physique appropriés
10. Formulation du problème à partir des données cliniques et paracliniques
11. Formulation et justification des conduites à tenir
12. Élaboration d'un plan de suivi approprié
13. Identification et prise en charge des urgences
14. Habiletés techniques
15. Élaboration d'hypothèses diagnostiques appropriées

Communication

16. Habiletés de communication : patients, familles, professionnels de la santé.
17. Documentation écrite utile et pertinente (rapports de procédures)

Collaboration

18. Consultation efficace et collaboration avec les autres équipes médicales
19. Collaborations constructives lors du travail interdisciplinaire
20. **Gestion**
21. Organisation du travail clinique
22. Utilisation judicieuse des ressources dans les soins de ses patients

Promotion de la santé

23. Intégration des concepts de médecine préventive aux conduites à tenir
24. Capacité d'enseignement au patient des principes de promotion de la santé

Érudition

25. Capacité d'évaluation critique de la littérature et des enseignements
26. Contribution académique et à l'enseignement
27. Motivation et autonomie à lire, questionner et apprendre