LES BIAIS EN ÉVALUATION

ADMEE-Canada

Association pour le développement des méthodologies d'évaluation en éducation

29e session d'étude

Les 18 et 19 octobre 2007

Hôtel Gouverneur Trois-Rivières

ADMEE-2007

Mot de bienvenue

C'est avec un vif plaisir que je vous souhaite la bienvenue à la 29e session d'étude de l'ADMEE Canada. Compte tenu du succès de l'année dernière, le comité exécutif a décidé de tenir la session d'étude à Trois-Rivières pour une deuxième année consécutive. J'espère que vous apprécierez tout autant le programme que nous vous présenterons cette année de même que l'accueil chaleureux des Trifluviens.

Cette année, l'ADMEE propose aux participantes et aux participants de porter un regard sur les biais en évaluation. La thématique du congrès veut vous engager dans une réflexion sur les types de biais qui peuvent fausser l'évaluation. Dans cette perspective, deux ateliers et 24 communications seront présentés. Les ateliers porteront sur l'entrevue de sélection structurée et sur l'évaluation de projet, programme. Les 24 communications se répartissent selon les thèmes suivants : évaluation des compétences, évaluation des apprentissages dans le secteur médical, évaluation de systèmes éducatifs, ainsi que le développement de logiciels dont l'utilisation pourrait aider les intervenants en éducation.

Nous osons espérer que ce programme vous fournira l'occasion d'échanges fructueux et enrichissants tout en vous permettant d'explorer de nouveaux horizons de coopération scientifique dans l'univers francophone.

Bonne session d'étude!

Pierre Valois
Président de l'ADMEE Canada

Jeudi 18 octobre 2007

9 h	***–***	***10 h***	***Accueil et inscription***
10 h 30	***–***	***11 h 45***	***Ateliers 1 et 2***
12 h 00	***–***	***13 h 30***	***Dîner libre***
14 h	***–***	***14 h 30***	***Communications Bloc A***
14 h 45	***–***	***15 h 15***	***Communications Bloc B***
15 h 30	***–***	***16 h***	***Communications Bloc C***
16 h 15	***–***	***16 h 45***	***Communications Bloc D***
16 h45	***–***	***17 h00***	***Lancement du livre de Louise Bélair et al.***
17 h	***–***	***18 h 30***	***Cocktail***
19 h 30	***–***	***21 h 30***	***Assemblée générale de l'ADMEE***

ATELIERS - (JEUDI de 10 h 30 à 11 h 45)

Atelier 1 **L'entrevue de sélection structurée**
Personnes-ressources : Normand Pettersen (UQTR) et André Durivage (UQO)

Atelier 2 **L'évaluation de projet, programme: une question d'indicateurs et de cibles!**
Personne-ressource : Marthe Hurteau (UQAM)

BLOC A - (JEUDI de 14 h à 14 h 30)

A1 **Module pour la Qualité en Évaluation (MQE) : une macro Excel pour l'évaluation des apprentissages en classe**
Conférenciers et conférencière : Pierre Valois, Éric Frenette, Stéphane Germain (Université Laval) et Christina St-Onge (Le Conseil médical du Canada)

A2 **Les biais en évaluation : au-delà de l'acte d'évaluer. Quelques considérations sociologiques**
Conférencier : Patrick Charles (Université de Montréal)

A3 **Les biais de l'évaluateur dans le contexte de l'évaluation des compétences professionnelles en enseignement**
Conférencière : Louise Bélair (UQTR)

BLOC B - (JEUDI de 14 h 45 à 15 h 15)

B1 **Expérimentation d'un modèle multi niveau pour l'évaluation de compétences : résultats d'une recherche menée au premier cycle du secondaire en science et technologie**
Conférenciers et conférencière : Éric Dionne, Michel D. Laurier et Jesus Vazquez-Abad (Université de Montréal)

B2 **Tenir compte des biais, faute de les contrôler ? Un exemple dans le choix des indicateurs d'exclusion éducative dans une population urbaine**
Conférencier : Gérard Figari (Université des sciences sociales – Grenoble)

B3 **Une alternative à l'étude de biais : analyse de profils types pour l'évaluation diagnostique des systèmes éducatifs**
Conférencier : Dany Laveault (Université d'Ottawa)

BLOC C - (JEUDI de 15 h 30 à 16 h)

C1 **Utilisation du progiciel macromedia flash dans un contexte d'évaluation adaptative des apprentissages par ordinateur**
Conférenciers et conférencière : Martin Lesage, Gilles Raîche, Martin Riopel et Komi Sodoke (UQAM)

C2 **Le renouvellement de l'encadrement local en évaluation à l'heure du renouveau pédagogique**
Conférenciers : Claude Robillard et Charles Fournier (C.S. des Affluents)

C3 **Analyse des biais de méthode associés à l'utilisation des échelles d'évaluation du comportement des enfants d'âge primaire**
Conférencière et conférencier : Nathalie Parent et Pierre Valois (Université Laval)

BLOC D - (JEUDI de 16 h 15 à 16 h 45)

D1 **La production automatisée de tâches d'évaluation en mathématiques**
Conférencier et conférencières : Martin Riopel, Fadia Sakr et Mélissa Pilote (UQAM)

D2 **Le facteur « temps limite » dans les tests de performance : condition d'équité et/ou source de biais de méthode**
Conférencier : Pascal Ndinga (UQAM)

D3 **Évaluation subjective et rétroaction fiable : est-ce possible?**
Conférenciers : Serge Sévigny et Julien D'amours-Raymond (Université Laval)

Vendredi 19 octobre 2007

8 h 30	*–*	*9 h*	*Accueil et inscription*
9 h	*–*	*10 h 15*	*Conférence d'ouverture*
10 h 15	*–*	*10 h 45*	*Pause*
10 h 45	*–*	*11 h 15*	*Communications Bloc E*
11 h 30	*–*	*12 h*	*Communications Bloc F*
12 h	*–*	*13 h 30*	*Dîner de l'ADMEE*
13 h 45	*–*	*14 h 15*	*Communications Bloc G*
14 h 30	*–*	*15 h*	*Communications Bloc H*

CONFÉRENCE D'OUVERTURE (VENDREDI de 9 h à 10 h 15)

Titre : **Les biais culturels des tests de QI**
Conférencier : Serge Larivée (Université de Montréal)

BLOC E - (VENDREDI de 10 h 45 à 11 h 15)

E1 **Au-delà des biais méthodologiques en évaluation de programme**
Conférencières : Marthe Hurteau et Stéphanie Mongiat (UQAM)

E2 **Est-ce qu'il est possible d'obtenir le résultat véritable d'un élève même si celui tente de tricher?**
Conférencières : Gilles Raîche (UQAM), Jean-Guy Blais (Université de Montréal) et David Magis (Université de Liège)

E3 **L'évaluation dans le contexte de la reconnaissance des acquis et des compétences : principes et instrumentation**
Conférencière : Sonia Fradette (MELS)

BLOC F - (VENDREDI de 11 h 30 à 12 h)

F1 **Les biais des pratiques d'évaluation de la formation d'entreprises québécoises performantes**
Conférencier : Alain Dunberry (UQAM)

F2 **Introduction au logiciel Winsteps**
Conférenciers et conférencière : Éric Dionne, Jean-Guy Blais et Julie Grondin (Université de Montréal)

F3 **La certification de médecins spécialistes : défis et enjeux!**
Conférencier : Gary Cole (Collège royal des médecins et chirurgiens du Canada)

BLOC G - (VENDREDI de 13 h 45 à 14 h 15)

G1 **L'éducation ou la formation à l'entrepreneuriat : quelle(s) évaluation(s) possible(s)?**
Conférencier : Narjisse Lassas-Clerc (EM Lyon)

G2 **Contrôle des biais d'évaluation : des méthodes utilisées dans les évaluations à grande échelle et en salle de classe**
Conférencier et conférencière : Pierre Brochu et Mélanie Labrecque (Conseil des ministres de l'éducation, Canada)

G3 **Peut-on déceler le phénomène des « ghost bank » à l'aide d'indices d'ajustement des scores individuels?**
Conférencière : Christina St-Onge (Le Conseil médical du Canada)

BLOC H - (VENDREDI de 14 h 30 à 15 h)

H1 **La définition du concept de compétence et son impact sur les pratiques évaluatives**
Conférencière et conférencier : Marie-Hélène Hébert et Pierre Valois (Université Laval)

H2 **Modèle cognitif : le diagnostic est-il biaisé par les experts?**
Conférencière : Nathalie Loye (Université de Montréal)

H3 **Sources de biais dans l'évaluation de stages cliniques**
Conférencier : Diem-Quyen Nguyen (Université de Montréal – Faculté de médecine)

Jeudi 18 octobre 2007

9 h	***–***	***10 h***	***Accueil et inscription***
10 h 30	***–***	***11 h 45***	***Ateliers 1 et 2***
12 h 00	***–***	***13 h 30***	***Dîner libre***
14 h	***–***	***14 h 30***	***Communications Bloc A***
14 h 45	***–***	***15 h 15***	***Communications Bloc B***
15 h 30	***–***	***16 h***	***Communications Bloc C***
16 h 15	***–***	***16 h 45***	***Communications Bloc D***
16 h 45	***–***	***17 h 00***	***Lancement du livre de Louise Bélair et al.***
17 h	***–***	***18 h 30***	***Cocktail***
19 h 30	***–***	***21 h 30***	***Assemblée générale de l'ADMEE***

ATELIERS - (JEUDI de 10 h 30 à 11 h 45)

Atelier 1 **L'entrevue de sélection structurée**

Personnes-ressources : Normand Pettersen (UQTR) et André Durivage (UQO)

Les recherches récentes indiquent clairement que l'entrevue de sélection structurée est nettement plus efficace que l'entrevue traditionnelle qu'on retrouve encore dans bien des organisations. L'entrevue structurée permet des décisions d'embauche qui sont plus valides, plus équitables et mieux acceptées par les candidats et les tribunaux.

Cet atelier porte en bonne partie sur le contenu du livre *L'entrevue structurée - Pour améliorer la sélection du personnel*, de Normand Pettersen et André Durivage (paru en 2006 aux Presses de l'Université du Québec). On y présentera les connaissances et les techniques les plus à jour en matière d'entrevue de sélection. Basé sur une vision réaliste des nombreuses contraintes du monde organisationnel, on y trouvera diverses façons de conduire une entrevue structurée, avec leurs avantages et leurs limites, de manière à choisir l'entrevue qui convient à chaque situation.

On y expliquera comment préparer une entrevue structurée, comment élaborer les meilleures questions qui soient, comment conduire l'entrevue et comment évaluer les candidats. Des pistes de solutions aux questions et aux problèmes les plus fréquents seront discutées.

Atelier 2 **L'évaluation de projet, programme : une question d'indicateurs et de cibles!**

Personne-ressource : Marthe Hurteau (UQAM)

Que ce soit dans le cadre de programmes, de projets ou de plans de réussite, les établissements sont tenus à en effectuer l'évaluation sur une base régulière afin de souscrire aux exigences de reddition de compte. Dans tous les cas, cette démarche fait état des stratégies mises en place et des effets observés.

Le présent atelier vise à familiariser les participants aux concepts sous-jacents à cette démarche, soit : programme - projet, type d'évaluation, indicateurs et cibles ainsi que les liens logiques qui les unissent, et ce, en illustrant le propos au moyen d'exemples. Il est important de souligner que la démarche proposée s'avère pertinente tant au moment de la conception d'un projet qu'au moment de son évaluation.

Plus précisément, nous verrons comment :

- ✓ Décrire un projet (programme) au moyen du modèle logique
- ✓ Distinguer les différents types d'évaluation
- ✓ Établir les indicateurs pertinents qui prennent en considération les composantes du modèle logique ainsi que les intentions de l'évaluation
- ✓ Établir des cibles.

En tout temps, les participants auront le loisir d'intervenir.

BLOC A - (JEUDI de 14 h à 14 h 30)

A1 **Module pour la Qualité en Évaluation (MQE) : une macro Excel pour l'évaluation des apprentissages en classe**

Conférenciers et conférencière : Pierre Valois, Éric Frenette, Stéphane Germain (Université Laval) et Christina St-Onge (Le Conseil médical du Canada)

L'avènement de l'approche par compétences au sein du système scolaire québécois a suscité un renouveau dans les pratiques d'évaluation des apprentissages. Le succès de ce renouveau en évaluation des apprentissages dans une approche par compétence repose, notamment, sur la conception d'un dispositif d'évaluation de qualité. À cette fin, les enseignants élaborent un dispositif d'évaluation (examens, travaux, portfolio, tâches complexes, etc.) propre à une matière afin de juger les apprentissages réalisés par les élèves. Ce dispositif vise à évaluer les apprentissages des élèves à deux fins : (1) réguler les apprentissages en cours d'année scolaire et (2) établir un bilan du degré de compétence atteint à la fin d'un cycle ou d'une année scolaire. Une quantité importante d'information peut être tirée d'un tel dispositif. Le Module pour la Qualité en Évaluation (MQE) a été élaboré afin d'aider les enseignants à obtenir le maximum d'information sur ce dispositif d'évaluation. Le MQE est une macrocommande intégrée au logiciel Excel. Ce module permet d'obtenir des données tant sur le profil des élèves que sur la qualité du dispositif d'évaluation. Sur le plan des élèves, le MQE permet de poser un diagnostic sur les forces et les défis à relever des élèves. Ce diagnostic peut être ventilé selon les différentes composantes de la compétence. Sur le plan des outils d'évaluation, le MQE permet d'obtenir de l'information sur la fiabilité du dispositif et des différents outils d'évaluation utilisés. Dans le cadre de cet atelier, une présentation détaillée du MQE et des différentes possibilités qu'il offre sera réalisée à l'aide d'un exemple tiré d'une situation réelle d'évaluation en classe.

A2 **Les biais en évaluation : au-delà de l'acte d'évaluer. Quelques considérations sociologiques**

Conférencier : Patrick Charles (Université de Montréal)

Le courant philosophique postmodernisme véhicule les concepts de culture, de diversité et valorise la différence. La venue d'immigrants de cultures et de religions différentes transforme le portrait ethnique du Québec. Ainsi, la problématique de la production des instruments d'évaluation dans le système scolaire québécois se pose.

En d'autres termes, doit-on construire des instruments d'évaluation pour chaque groupe ethnique ou un instrument unique sur la base de ce qui est commun à toutes les cultures? Au MELS, la pratique actuelle de production des instruments d'évaluation se fait à partir de la définition du domaine. Cela oblige les spécialistes à distinguer les biais éventuels en évaluation afin de répondre aux principes de validité, de fidélité, d'équité et de justice.

Premièrement, cette présentation veut montrer que, quoique le discours scientifique en évaluation considère la question de biais à toutes les étapes de l'acte d'évaluer, pour leur part les documents officiels (les Principes d'équité relatifs aux pratiques d'évaluation des apprentissages scolaires au Canada; le Cadre de référence de l'évaluation des apprentissages au préscolaire et au primaire; la Politique d'évaluation des apprentissages) ignorent pratiquement cette problématique, à part la subjectivité inhérente au jugement. Deuxièmement, d'un point de vue sociologique, elle veut aussi montrer que ces instruments d'évaluation reflètent d'abord l'idéologie de l'élite qui les produit. Par conséquent, la pratique actuelle de l'évaluation et la valorisation de la différence culturelle, telle que prônée par le postmodernisme, sont incompatibles.

A3 **Les biais de l'évaluateur dans le contexte de l'évaluation des compétences professionnelles en enseignement**

Conférencière : Louise Bélair (UQTR)

Cette communication se veut une réflexion sur la place de la fidélité et de la validité dans l'évaluation des compétences professionnelles en enseignement. Les critères de la scientificité (fiabilité, faisabilité, crédibilité, fidélité et, par conséquent, validité) seront explorés par une mise à plat des divers biais à prendre en compte à la lumière des quatre temps de la démarche d'évaluation.

Comme la compétence est par définition complexe, intégrative et évolutive, il s'avère nécessaire de comprendre les enjeux relatifs à l'intention de départ, à la collecte de données, à l'interprétation et au jugement pour être en mesure de s'assurer d'une décision d'évaluation valide et contrôlée du point de vue des biais relatifs à l'évaluateur.

Le contexte de l'évaluation des futurs enseignants sera pris en exemple pour illustrer diverses questions à se poser avant de prendre une décision sur le niveau de compétence en enseignement.

Un tableau synthèse sera remis aux participantes et participants.

BLOC B - (JEUDI de 14 h 45 à 15 h 15)

B1 **Expérimentation d'un modèle multiniveau pour l'évaluation de compétences : résultats d'une recherche menée au premier cycle du secondaire en science et technologie**

Conférenciers et conférencière : Éric Dionne, Michel D. Laurier et Jesus Vazquez-Abad (Université de Montréal)

L'évaluation des compétences pose de nombreux défis au secondaire. Certaines disciplines, comme celle de science et technologie, doivent composer avec de nombreux changements pédagogiques et didactiques qui ont des impacts majeurs sur l'évaluation. Qui plus est, le contexte particulier de l'évaluation certificative – qui est par définition à enjeux critiques – possède en lui-même des contraintes qui peuvent affecter les modalités pratiques liées à la démarche évaluative. Compte tenu de cette situation, nous avons conçu et expérimenté un modèle d'évaluation multiniveau, inspiré des travaux de Rey et ses collaborateurs (Rey, Carette, Defrance & Kahn, 2003) visant à mieux cerner une des compétences à développer en science et technologie. Dans la présentation, nous expliquerons les dimensions à prendre en compte dans l'élaboration de tâches dont le niveau de complexité varie. Nous avons expérimenté le modèle auprès de 560 élèves du premier cycle du secondaire issus de cinq commissions scolaires et de 14 écoles au terme de l'année scolaire 2005-2006. Nous présenterons les résultats que nous avons obtenus concernant l'effet d'une approche multiniveau afin de juger du développement d'une compétence chez des élèves du premier cycle. Les analyses démontrent que le modèle est utile afin de cibler le niveau de développement de la compétence pour chacun des élèves en catégorisant chacun de ces derniers en trois catégories : en réussite, en difficulté passagère et en difficulté majeure. Nous avons également remarqué que le modèle permet de collecter une information plus variée et plus riche qu'un modèle classique caractérisé par des épreuves uniques.

B2 **Tenir compte des biais, faute de les contrôler? Un exemple dans le choix des indicateurs d'exclusion éducative dans une population urbaine**

Conférencier : Gérard Figari (Université des sciences sociales – Grenoble)

Il s'agit d'une réflexion sur une forme de biais méthodologiques relatifs au choix d'indicateurs dans le cadre d'une étude destinée à construire une mutualisation des données en éducation afin de déterminer les risques d'exclusion éducative d'une population d'enfants et de jeunes d'une collectivité urbaine afin d'accompagner une politique locale d'intervention en éducation.

- Contexte : une communauté de communes voulant mutualiser des données permettant la prédiction de risques d'exclusion éducative (scolaire, péri-scolaire, extra-scolaire).
- Commande : construire une batterie d'indicateurs permettant de collecter des données comparables d'une commune à l'autre et utilisables pour ajuster une politique éducative locale de lutte contre l'exclusion de nature éducative.
- Méthode : 1- étude des conceptions de l'exclusion éducative des acteurs de la politique éducative dans les sept communes concernées; 2- formulation d'un référentiel commun d'évaluation de l'« exclusion éducative » débouchant sur 150 indicateurs potentiels; 3- enquête auprès des acteurs de la politique éducative pour déterminer le choix de 15 indicateurs communs; 4- analyse des biais méthodologiques de l'opération : en particulier, le problème de la représentativité des indicateurs choisis et de la pertinence qu'il y a à les utiliser pour déterminer des risques d'exclusion.

B3 **Une alternative à l'étude de biais : analyse de profils types pour l'évaluation diagnostique des systèmes éducatifs**

Conférencier : Dany Laveault (Université d'Ottawa)

Les différents niveaux du système éducatif sont interpellés différemment par les résultats d'une évaluation à grande échelle. Pour ce qui est du biais et de l'impact, un item qui manifeste un fonctionnement différentiel (FDI) chez les garçons et les filles dans un Conseil peut ne pas en présenter dans un autre. Le même phénomène se produit à l'intérieur d'un même Conseil pour différentes écoles ou différents groupes (linguistiques, ethniques, etc.). Dans de tels contextes, l'étude du FDI dépend du groupe de référence et conduit à des conclusions difficilement généralisables. Il apparaît plus utile de qualifier le type d'interaction entre l'item et le niveau du système éducatif (classe, école, Conseil) que de se prononcer sur la présence d'un biais. Cette approche est d'autant plus nécessaire lorsque les différences ne sont pas uniformes.

La présentation utilisera les résultats d'un Conseil scolaire de l'Ontario pour illustrer une procédure qui consiste à présenter les analyses d'items par domaine (approche critériée) et par niveau de rendement (approche normative) à la fois pour l'école, pour le Conseil et pour la province. L'étude de plusieurs cas d'analyses d'items a permis de distinguer des profils types. L'analyse de profils types semble préférable à la distinction de biais pour l'évaluation diagnostique à tous les niveaux d'un système éducatif. La présentation offrira des exemples de ces profils et portera sur le cas particulier d'une école ayant connu une diminution du rendement de ses élèves.

BLOC C - (JEUDI de 15 h 30 à 16 h)

C1 **Utilisation du progiciel macromedia flash dans un contexte d'évaluation adaptative des apprentissages par ordinateur**
Conférenciers et conférencière : Martin Lesage, Gilles Raîche, Martin Riopel et Komi Sodoke (UQAM)

Un logiciel permettant la création et l'affichage de questions selon la théorie de la réponse à l'item dénommé « PersonFit » a été développé par le CAMRI et sert présentement à modéliser des banques de questions. Le logiciel « PersonFit » est développé avec la technologie Java ayant certaines limites dans la création d'interfaces usager conviviales comportant des images, clips sonores, vidéo et animations graphiques. Le logiciel « PersonFit » code les items dans des fichiers à l'aide de normes développées par des informaticiens s'intéressant aux environnements éducatifs. Afin d'explorer de nouvelles options, le CAMRI étudie présentement la possibilité de développer un logiciel similaire avec le progiciel Macromedia Flash qui utilise le langage de script « ActionScript ». Ce logiciel doit de plus être compatible avec « PersonFit » en utilisant la norme IMS QTI et par l'accès à des bases de données communes à ces deux logiciels. Des résultats préliminaires ont étés obtenus par la création d'une application développée en Macromédia Flash étant capable de créer et d'afficher des questions en plus d'avoir accès aux fonctionnalités et au grand potentiel multimédia de Flash. Cette application crée présentement des fichiers sur un poste de travail client. Le CAMRI est présentement en train d'étendre les fonctionnalités de cette application afin qu'elle puisse saisir des questions sur le poste du client pour ensuite les écrire sur un serveur. Des travaux futurs de ce projet viseront le développement d'une sous-norme de « IMS-QTI » effectuant la modélisation des paramètres d'item. Cette sous-norme sera finalement incluse dans les systèmes d'apprentissage à distance par ordinateur ayant des fonctionnalités de *testing* adaptatif.

C2 **Le renouvellement de l'encadrement local en évaluation à l'heure du renouveau pédagogique**
Conférenciers : Claude Robillard et Charles Fournier (C.S. des Affluents)

L'article 96.15 de la Loi sur l'instruction publique et la Politique d'évaluation des apprentissages obligent les établissements scolaires à renouveler leur encadrement local en évaluation, lequel aura un impact sur les pratiques évaluatives des enseignants. À cet égard, l'objectif premier visé par cette présentation réside dans la volonté d'amener les participants à prendre connaissance d'une démarche vécue par un comité régionale composé de directions d'établissement scolaire et d'enseignants représentant les quatorze écoles secondaires de la commission scolaire des Affluents. Cette démarche était composée de formations sur un référentiel en évaluation au secondaire et sur le Guide concernant le renouvellement de l'encadrement local en évaluation des apprentissages. Elle était aussi composée d'une production de normes et modalités articulée autour des étapes du processus d'évaluation des apprentissages à des fins de consultation par les équipes-école et d'implantation par la suite.

C3 **Analyse des biais de méthode associés à l'utilisation des échelles d'évaluation du comportement des enfants d'âge primaire**
Conférencière et conférencier : Nathalie Parent et Pierre Valois (Université Laval)

Parmi les instruments de mesure disponibles pour évaluer les troubles du comportement (TC), les échelles d'évaluation remplies par les parents et les enseignants sont les plus utilisés. Différents biais de méthode peuvent toutefois altérer la validité de ces échelles. Cette recherche a pour objectif d'étudier les biais de méthode associés à l'utilisation des échelles d'évaluation des TC. Les biais de système d'évaluation et de répondant y sont comparés sous deux angles: la théorie classique des tests et la théorie de la généralisabilité. L'échantillon comprend 396 enfants dont chacun est évalué par un parent et son enseignant. Trois échelles d'évaluation sont remplies par chaque répondant : celles d'Achenbach (Achenbach & Rescorla, 2001), de Conners (1997) et de Bullock et Wilson (Parent, Tremblay & Valois, 2006). Les traits ciblés sont l'agressivité, l'inattention

et le retrait social. Les résultats montrent que les biais de méthode sont responsables de 17 à 41 % de la variance dans les scores. Le fait de se référer au parent ou à l'enseignant explique de 59 à 77 % de l'erreur de mesure alors que l'influence du choix du système d'évaluation est de 9 à 21 %. Par ailleurs, les analyses factorielles confirmatoires réalisées montrent que les échelles d'évaluation du comportement fournissent un portrait valide des TC pour chaque type de répondant et trait mesuré. Ces résultats confirment la pertinence d'utiliser les échelles d'évaluation du comportement et la contribution unique et complémentaire de chaque type de répondants dans l'évaluation. Ainsi, l'utilisation des échelles d'évaluation du comportement est recommandée, mais recourir à une stratégie d'évaluation multirépondant est nécessaire afin d'augmenter la fiabilité de la prise de décision.

BLOC D - (JEUDI de 16 h 15 à 16 h 45)

D1 **La production automatisée de tâches d'évaluation en mathématiques**

Conférencier et conférencières : Martin Riopel, Fadia Sakr et Mélissa Pilote (UQAM)

Les recherches sur la production automatisée de tâches d'évaluation s'intéressent à la création d'un très grand nombre de nouvelles tâches dont les caractéristiques seraient prédites à l'aide d'un modèle formel. Dans ce contexte, le processus de validation ne ferait plus intervenir individuellement chacune des tâches créées mais plutôt le modèle dans son ensemble. Une fois le modèle validé, de nouvelles tâches adaptées pourraient être créées automatiquement au besoin lors de l'évaluation des sujets. Cette façon de procéder présenterait plusieurs avantages sur le plan du coût de développement de nouvelles tâches (qui pourraient être produites en grand nombre rapidement) et de la sécurité des tests produits (un grand nombre de tâches diminue significativement la possibilité de répétition de la même tâche). Bien qu'elle soit soutenue par les modélisations issues de la théorie de la réponse à l'item et qu'elle convienne bien aux évaluations adaptatives et informatisées, la production automatisée de tâches d'évaluation peut aussi être appliquée lors d'évaluations classiques en salle de classe. Nous présenterons les travaux qui ont jeté les bases de la production automatisée de tâches d'évaluation ainsi que quelques exemples d'application convenant au contexte de l'évaluation informatisée.

D2 **Le facteur « temps limite » dans les tests de performance : condition d'équité et/ou source de biais de méthode**

Conférencier : Pascal Ndinga (UQAM)

La standardisation des tests constitue une réponse appropriée à la préoccupation manifeste d'équité à l'égard des examinés. Il s'agit d'uniformiser les conditions de passation du test, en particulier dans des cas d'évaluation de masse ou de groupe. Tous les sujets sont ainsi, objectivement, soumis au même dispositif du *testing*. C'est une condition essentielle à toute comparaison des performances des participants ou des groupes, un des critères inhérents à la crédibilité de toute étude scientifique. Un des aspects fondamentaux de l'uniformité d'un test est sa durée (limite de temps – temps alloué) de passation. La plupart des tests standardisés sont en effet administrés à l'intérieur d'un temps prescrit (imposé).
Or, cette condition peut s'avérer pour certains sujets, en l'occurrence les plus lents, similaire à celle d'un test de vitesse. La limite de temps aux tests d'accomplissement tend à être imposée principalement pour une convenance administrative et ces tests sont destinés à permettre de 80 à 90 % des examinés de terminer le test à l'intérieur du temps prescrit (Munger et Loyd, 1991). Ainsi, même dans ce contexte, fort louable, des tests standardisés, 10 à 20 % des examinés ne peuvent pas terminer le test à l'intérieur du temps alloué. Pour eux, les items non répondus peuvent aussi signifier items non atteints, faute de temps suffisant. Dans ce contexte, que faire afin que limite de temps du test implique à la fois différenciation des candidats et pleine expression des connaissances et habiletés de tous?
Le présent exposé vise à faire le point de la recherche sur cette question et à discuter des pistes de solution proposées, dans le contexte de la réforme scolaire actuellement en cours au Québec.

D3 **Évaluation subjective et rétroaction fiable : est-ce possible?**

Conférenciers : Serge Sévigny et Julien D'amours-Raymond (Université Laval)

Les scores holistiques utilisés lors de l'évaluation de l'écriture ou lors de l'évaluation des compétences ne semblent pas fournir les informations qui permettraient d'accroître, par le biais de l'enseignement, les compétences en écriture des élèves canadiens (Ercikan, 2006). Les enseignants se plaignent de cette situation car ils estiment que les fonds canadiens pourraient être mieux investis dans le système d'éducation (CTF, 1999). La meilleure façon d'améliorer le rendement de l'élève serait-elle de connaître

leurs faiblesses? Le présent projet met à l'épreuve deux types d'évaluation par les juges susceptibles (1) de nous informer sur les faiblesses des élèves et (2) de favoriser l'accord interjuges et, par conséquent, la fidélité de la mesure ainsi que la validité de l'interprétation des résultats. La collaboration de professeurs de français expérimentés a permis de tester deux types d'évaluation :

Type 1 : Lire les productions écrites des élèves en examinant en une seule lecture les six composantes spécifiques de l'écriture selon un critère précis : *déterminer la principale composante à améliorer en priorité.*

Type 2 : Lire les productions écrites des élèves en examinant en une seule lecture les six mêmes composantes que celles du type 1 selon le critère suivant : *déterminer les deux composantes à améliorer en priorité.*

Un échantillon de 160 textes a permis d'évaluer l'accord interjuges pour chaque type d'évaluation. Les taux d'accord interjuges de chaque paire de juges par type d'évaluation seront comparés, d'abord entre eux, et ensuite avec les taux actuels associés aux évaluations à grande échelle (fidélité). Les résultats permettront de connaître les principales faiblesses des élèves en écriture et d'évaluer les taux d'accord interjuges pour chaque type d'évaluation. La discussion examinera l'apport des résultats quant à l'évaluation subjective, la formation des juges, la fidélité des scores, la validité des interprétations et la rétroaction aux gens du milieu de l'éducation.

Vendredi 19 octobre 2007

8 h 30	*–*	*9 h*	*Accueil et inscription*
9 h	*–*	*10 h 15*	*Conférence d'ouverture*
10 h 15	*–*	*10 h 45*	*Pause*
10 h 45	*–*	*11 h 15*	*Communications Bloc E*
11 h 30	*–*	*12 h*	*Communications Bloc F*
12 h	*–*	*13 h 30*	*Dîner de l'ADMEE*
13 h 45	*–*	*14 h 15*	*Communications Bloc G*
14 h 30	*–*	*15h*	*Communications Bloc H*

CONFÉRENCE D'OUVERTURE (VENDREDI de 9 h à 10 h 15)

Titre : **Les biais culturels des tests de QI**
Conférencier : Serge Larivée (Université de Montréal)

BLOC E - (VENDREDI de 10 h 45 à 11 h 15)

E1 **Au-delà des biais méthodologiques en évaluation de programme**
Conférencières : Marthe Hurteau et Stéphanie Mongiat (UQAM)

Même si l'évaluation de programme a bénéficié de nombreux développements (méthodes, modèles, etc.) au cours des trente dernières années, il n'en demeure pas moins qu'elle fait l'objet de constantes et nombreuses remises en question. Plusieurs d'entre elles portent sur la légitimité des jugements générés ainsi que sur la pertinence des recommandations qui en découlent.
La présente étude s'inscrit dans le cadre d'un programme de recherche qui porte sur les pratiques en évaluation de programme et qui vise à mieux les comprendre afin de les améliorer. Après avoir développé et validé une modélisation de la pratique, cette phase s'intéresse plus particulièrement à décrire la nature des jugements générés par la démarche évaluative ainsi que leurs assises.
Dans cette perspective, nous établirons dès le point de départ ce que nous entendons par processus spécifique à l'évaluation de programme qui permet d'émettre un jugement et nous effectuerons certaines distinctions telles : déclaration/jugement, jugement légitime/argumenté. Par la suite, nous présenterons la perspective de l'étude, ses fondements théoriques et la méthodologie afin de s'attarder davantage aux résultats obtenus ainsi que leurs conséquences sur la pratique en évaluation de programme. Notons brièvement que sur les quarante rapports provenant de la banque ERIC qui ont été analysés, seulement 50 % d'entre eux ont généré un jugement et seulement 15 % de ces jugements peuvent être considérés comme légitimes. À eux seuls, ces seuls résultats permettent d'induire les difficultés auxquelles se confronte la pratique évaluative ainsi que les biais qui peuvent en émaner.

E2 **Est-ce qu'il est possible d'obtenir le résultat véritable d'un élève même si celui tente de tricher?**
Conférenciers : Gilles Raîche (UQAM), Jean-Guy Blais (Université de Montréal) et David Magis (Université de Liège)

Le résultat obtenu par un élève lorsqu'on lui administre une épreuve d'évaluation de ses apprentissages n'est pas toujours représentatif de son niveau de compétence. À titre illustratif, celui-ci peut avoir copié ses réponses sur celles d'un autre : son résultat est alors surestimé. Dans

d'autres situations, des élèves sous-performent intentionnellement à des tests de classement en langue seconde pour pouvoir, par la suite, se la couler douce à l'intérieur de leurs cours. Dans tous ces cas, cela se traduit par la production d'un patron de réponses inapproprié qui ne permet pas de connaître le niveau de compétence véritable de l'élève. Récemment, des stratégies édumétriques ont été expérimentées pour tenter d'obtenir le résultat véritable d'un élève même si celui-ci tente de tricher. Elles sont toutes basées sur des modifications apportées aux modélisations issues de la théorie de la réponse à l'item et misent sur trois aspects spécifiques de ces modélisations. Ces aspects se traduisent en un paramètre de discrimination, de pseudo-chance et d'inattention propres à chacun des élèves à qui on administre une épreuve d'évaluation. Dans les modélisations usuelles issues de la théorie de la réponse à l'item, ces mêmes paramètres sont plutôt spécifiques aux items et non pas aux élèves. Ces différentes stratégies pour tenter d'obtenir le résultat véritable d'un élève seront décrites et leurs forces et faiblesses seront abordées.

E3 **L'évaluation dans le contexte de la reconnaissance des acquis et des compétences : principes et instrumentation**
Conférencière : Sonia Fradette (MELS)

L'instrumentation de reconnaissance des acquis et des compétences (RAC) développée dans le contexte de l'approche harmonisée en formation professionnelle et technique repose sur la notion du « cœur de la compétence ». En effet, le mandat de la RAC l'oblige à se placer à la charnière de deux univers : celui du travail et des entreprises, d'une part, et celui de l'éducation et des établissements scolaires, d'autre part. Dans chacun de ces univers, la mise en perspective des compétences relève d'une logique bien différente : dans le monde de l'éducation, les compétences sont conçues et définies d'abord à partir d'une logique de développement tandis que dans le monde du travail et de l'emploi, la façon d'appréhender les compétences relèverait surtout d'une logique de mobilisation.
En RAC, étant placée à la charnière entre ces deux univers, force nous est de constater que cette position est déterminante en qui a trait à l'élaboration de l'instrumentation requise au regard d'un programme d'études. Cependant, comme il s'agit d'élaborer une instrumentation adaptée au contexte spécifique de la RAC, laquelle fait appel essentiellement à des pratiques d'évaluation, il faut nécessairement libérer le matériel d'évaluation requis à cette fin des contraintes et des particularités propres au matériel utilisé à des fins de formation.
L'objectif de cet atelier est de présenter les principes et méthodes qui guident le développement de l'instrumentation requise: fiche descriptive, conditions de reconnaissance, fiches d'évaluation et guide d'accompagnement.

BLOC F - (VENDREDI de 11 h 30 à 12 h)

F1 **Les biais des pratiques d'évaluation de la formation d'entreprises québécoises performantes**
Conférencier : Alain Dunberry (UQAM)

Les entreprises évaluent peu leur formation. Cette situation est d'autant déplorable que bon nombre d'entre elles hésitent à investir en formation, mettant en doute les effets positifs qu'elles peuvent en retirer. La recherche visait à identifier les pratiques d'entreprises jugées performantes afin d'inciter les entreprises à évaluer leur formation et ainsi mieux juger les bénéfices qu'elles en retirent. Une recension des pratiques d'évaluation formelles et informelles de la formation a été effectuée auprès de 12 entreprises québécoises considérées comme performantes en formation. Ces 12 entreprises étaient issues de trois secteurs différents. Les pratiques ont été recensées et analysées selon qu'il s'agisse d'évaluation de la satisfaction, des apprentissages, des comportements en poste de travail ou des résultats sur l'entreprise. À ces quatre niveaux, proposés par le modèle de Kirkpatrick (1998), a été ajouté un cinquième niveau, celui de l'évaluation du rendement financier (Phillips, 1997). Les résultats font ressortir que ces entreprises évaluent principalement les apprentissages et les comportements, surtout de formations reliées de près à l'entraînement à la tâche. L'évaluation de la satisfaction vient en troisième lieu. Il existe très peu d'évaluations des impacts de la formation sur l'organisation ou de son rendement financier. Deux biais principaux sont identifiés : (1) on procède très rarement à une intégration formelle des données d'évaluation permettant une vision critique de l'ingénierie de la formation dans son ensemble; (2) les pratiques formelles d'évaluation privilégient une approche de contrôle de conformité qui valorise peu la réflexion critique de l'apprenant et de là, le développement de pratiques autonomes.

F2 **Introduction au logiciel Winsteps**

Conférenciers et conférencière : Éric Dionne, Jean-Guy Blais et Julie Grondin (Université de Montréal)

Le logiciel Winsteps® (Linacre, 2006) est un outil conçu pour modéliser des données brutes au moyen de différents modèles de mesure. À titre d'exemple, le logiciel permet de modéliser les données qui sont sous formes dichotomiques ou polychotomiques comme le modèle *rating scale* (Andrich, 1978, 1988) ou le modèle *partial credit* (Masters, 1982) qui sont des modèles fréquemment utilisés en éducation et en science sociale compte tenu de la nature des données qui y sont traitées. Le recours au logiciel Winsteps permet, notamment, de situer les données sur une échelle d'intervalle, ce qui est un avantage indéniable. Le but de la présentation est de réaliser une initiation au logiciel afin que les participants puissent apprivoiser les possibilités qu'il offre. Pour ce faire, nous réaliserons quelques analyses complètes en partant des données brutes jusqu'à la lecture et à l'interprétation des sorties statistiques que fournit le logiciel. Nous montrerons comment il est possible d'obtenir rapidement et facilement des données fort pertinentes telles l'erreur de mesure, les statistiques d'ajustement (infit et outfit), l'estimation de la difficulté d'un item ou l'estimation du niveau d'habileté d'un sujet. Également, nous indiquerons comment il est possible de transférer les données vers des logiciels de traitement de données quantitatives comme Excel® ou SPSS® afin de réaliser des analyses plus avancées.

F3 **La certification de médecins spécialistes : défis et enjeux!**

Conférencier : Gary Cole (Collège royal des médecins et chirurgiens du Canada)

Le Collège royal des médecins et chirurgiens du Canada est responsable de la certification de tous les médecins spécialistes au Canada, ce qui englobe 49 différentes spécialités. Ce processus inclut l'accréditation des programmes des 17 universités qui offrent la formation de médecins spécialistes en plus de la certification des résidents.
La formation et l'évaluation des spécialistes s'appuient sur le modèle CanMEDS qui est un cadre de références des sept compétences que devraient développer tous les médecins spécialistes. Ce modèle développé sur la base des besoins de notre société est maintenant adopté par plusieurs autres pays.
L'évaluation est composée de deux aspects : l'évaluation en formation et l'évaluation finale. L'évaluation pendant la formation peut être à caractère formatif ou sommatif. Bien que chaque programme soit différent, ils connaissent des similarités quant aux défis et enjeux liés à l'évaluation en formation, notamment le besoin d'avoir de multiples sources d'information, l'évaluation objective à partir de standard reconnu et le besoin de rétroaction qui doit accompagner l'évaluation formative.
Les examens sont à caractère sommatif et critérié, et plusieurs défis et enjeux y sont associés, notamment les petits échantillons, l'élaboration de tableau de spécification (plan directeur), la standardisation, le point de passation et la fidélité de la mesure. Chacun des 49 examens est unique et peut combiner une ou plusieurs composantes. La fidélité au cœur de nos préoccupations, tout au long du processus et particulièrement lorsque des patients standardisés et des médecins adoptent le rôle d'examinateur. Depuis 2005, un simulateur cardio-pulmonaire est utilisé dans le cadre de certains examens. Ce simulateur très réaliste peut être programmé pour manifester différents symptômes. Bien que ce type de simulateur soit utilisé depuis longtemps en formation, son utilisation lors d'évaluation à enjeux élevés est récente et a donc besoin d'être validée.

BLOC G - (VENDREDI de 13 h 45 à 14 h 15)

G1 **L'éducation ou la formation à l'entrepreneuriat : quelle(s) évaluation(s) possible(s)?**

Conférencier : Narjisse Lassas-Clerc (EM Lyon)

L'entrepreneuriat est reconnu aujourd'hui comme un des facteurs socioéconomiques les plus efficaces pour la croissance des sociétés. L'expansion de ce phénomène a été suivie par la multiplication des initiatives éducatives et de formation spécialisées. Notre recherche se penche sur la question de l'évaluation de ces initiatives. De nombreux auteurs soulignent la nécessité de recourir à des méthodes scientifiques rigoureuses. Pour répondre à ce besoin, nous nous proposons de faire appel aux sciences de l'éducation en nous appuyant sur des construits issus de la psychologie et de l'entrepreneuriat. À partir des travaux dominants publiés dans les trois corpus concernés, cette communication présente un modèle qui nous permettrait de penser et de pratiquer l'évaluation des initiatives en question. Après avoir défini notre objet de recherche et discuté la problématique d'évaluation de programme et l'impact sur des variables socio-affectives, nous évoquerons les différentes déclinaisons possibles de

notre modèle. C'est notamment le modèle de Kirkpatrick, discuté et adapté, qui est utilisé ici comme cadre général dans lequel les différents niveaux de transfert sont représentés par les objectifs éducatifs de Bloom et opérationnalisés à l'aide de construits de psychologie cognitive adaptés à l'entrepreneuriat, à savoir les attitudes, les connaissances et les compétences entrepreneuriales (une revue de documentation reprendra, dans la communication, les travaux dominants dans le champ de la cognition entrepreneuriale). Cette communication a un double objectif : il s'agit pour nous, en confrontant nos travaux aux experts de l'évaluation éducative, d'échanger sur notre approche théorique et de recueillir leurs points de vue sur notre conception des aspects méthodologiques.

G2 **Contrôle des biais d'évaluation : des méthodes utilisées dans les évaluations à grande échelle et en salle de classe**

Conférencier et conférencière : Pierre Brochu et Mélanie Labrecque (Conseil des ministres de l'éducation, Canada)

Le Programme pancanadien d'évaluation 2007 (PPCE) a permis d'évaluer la performance des élèves de 13 ans dans toutes les provinces et un territoire dans le domaine de la lecture (composante majeure) ainsi que les mathématiques et les sciences (composantes mineures). Que ce soit sur le plan de l'échantillonnage, de l'instrument d'évaluation, de son administration, ou de sa correction, il existe différents facteurs extérieurs qui peuvent avoir un impact considérable sur le rendement des élèves au test. Il est bien important de tenir compte de ces facteurs puisqu'ils peuvent affecter la qualité des items et du test et ainsi fausser les résultats obtenus en sous-évaluant ou en surévaluant le rendement des élèves. Dans un projet d'une grande envergure tel que le PPCE, il était nécessaire de s'assurer que le test administré aux élèves de 13 ans évalue réellement leurs habiletés et que les résultats obtenus soient représentatifs.
Cette présentation consiste à présenter les procédures suivies dans le cadre du PPCE pour le contrôle des différents types de biais dans cette évaluation à grande échelle. Parmi ces procédures, nous retrouvons :
• Définition des caractéristiques mesurées
• Les techniques d'échantillonnage
• Révision linguistique des items
• Règles d'administration standardisées
• Correction à l'aveugle
• Correction en chaîne
• Double correction
• Tests de fidélité
• Analyse des biais, etc.

De plus, les enseignantes ou enseignants ainsi que les responsables des programmes d'évaluation peuvent se baser sur ces différentes approches pour limiter la présence de biais dans leurs propres évaluations. Nous allons aussi discuter de la manière dont certaines de ces pratiques pourraient être applicables pour les évaluations en salle de classe.

G3 **Peut-on déceler le phénomène des « ghost bank » à l'aide d'indices d'ajustement des scores individuels?**

Conférencière : Christina St-Onge (Le Conseil médical du Canada)

L'évaluation de type certificative se doit d'être aussi fidèle et valide que possible, car les décisions qui en découlent ont d'importantes répercussions. Lors d'évaluations où les enjeux sont élevés, par exemple lors d'épreuves de certification, il arrive que les instruments de mesure soient composés d'items ayant déjà été utilisés en plus de nouveaux items. On insère souvent des items déjà utilisés, car on peut les choisir selon leurs qualités psychométriques et ainsi contribuer à la validité et à la fidélité de l'instrument de mesure.
Il arrive, lors d'évaluations de type certificative, que les candidats s'entraident d'une année à l'autre en discutant des examens passés. Ce phénomène peut avoir des répercussions sur la performance des candidats et sur les qualités psychométriques des examens. Toutefois, ce phénomène est difficile à repérer sur le plan des données et passe souvent inaperçu lors des analyses statistiques traditionnelles de fidélité et de validité.
Toutefois, un groupe d'indices, les indices d'ajustement des scores individuels, pourrait peut-être faciliter la reconnaissance de ce phénomène lors d'évaluation où les enjeux sont élevés. Une étude Monte Carlo sera réalisée pour vérifier si l'utilisation de l'indice d'ajustement des scores individuels, peut nous permettre de repérer le phénomène communément connu sous le nom de « ghost bank » dans le cadre d'examen à choix multiples.

BLOC H - (VENDREDI de 14 h 30 à 15 h)

H1 **La définition du concept de compétence et son impact sur les pratiques évaluatives**
Conférencière et conférencier : Marie-Hélène Hébert et Pierre Valois (Université Laval)

L'avènement de l'approche par compétences a introduit un nouveau jargon dans les milieux scolaires québécois et étrangers. La notion de compétence est dorénavant sur toutes les lèvres. Mais sa signification n'est toujours pas univoque, avec pour conséquences des pratiques évaluatives disparates d'un ordre d'enseignement à l'autre, d'un pays à l'autre, etc. Pour les uns, la compétence désigne une procédure automatisable (p.ex. additionner deux fractions). Pour les autres, elle renvoie à un savoir-agir fondé sur la mobilisation d'un ensemble de ressources (p.ex. résoudre une situation-problème mathématique). Proposant un exercice de clarification autour du concept de compétence et prenant appui sur le *Programme de formation de l'école québécoise*, la communication mettra en lumière différentes définitions du concept de compétence et leurs impacts sur les pratiques évaluatives.

H2 **Modèle cognitif : le diagnostic est-il biaisé par les experts?**
Conférencière : Nathalie Loye (Université de Montréal)

Les modèles cognitifs offrent aujourd'hui de nombreuses perspectives pour réaliser une évaluation diagnostique des élèves. Ils proposent aussi bien d'extraire une information diagnostique à partir des données d'un test existant que de présider à la création d'un test diagnostique. Ils peuvent s'appliquer aux données d'une évaluation à grande échelle ou d'une évaluation élaborée par des enseignants pour leur propre usage. Dans tous les cas, le diagnostic prend la forme d'une liste d'attributs cognitifs dans laquelle il est mentionné si l'élève maîtrise ou non chacun des attributs. Préalablement à la modélisation des données à l'aide d'un modèle cognitif, des experts doivent élaborer une matrice qui contient la liste des attributs cognitifs visés par le test, ainsi que les liens qui existent entre chaque attribut et chaque item du test. Le modèle permet ensuite d'inférer la maîtrise ou non de chaque attribut à partir du schéma de réponse de l'élève. Ces experts doivent être familiers avec les élèves concernés et posséder les connaissances et compétences liées au contenu du test. Une bonne conception de ce qui est attendu d'eux, tant sur le plan du contenu que du mécanisme de collecte de l'information, est également indispensable. Cette présentation vise à mettre en évidence les biais liés à la création d'une telle matrice à l'aide des résultats d'une étude empirique menée sur les données du test diagnostique en mathématique de l'école polytechnique de Montréal. Elle aborde, dans un premier temps, l'impact que peuvent avoir les consignes sur le travail des experts. Dans un deuxième temps, elle illustre les différences qui existent entre les matrices créées par différents experts. Pour finir, elle montre de quelle manière le diagnostic des étudiants est influencé par les choix de chaque expert dans la conception de sa matrice.

H3 **Sources de biais dans l'évaluation de stages cliniques**
Conférencier : Diem-Quyen Nguyen (Université de Montréal – Faculté de médecine)

La formation médicale au stade postdoctoral a pour but de développer les compétences cliniques de futurs médecins généralistes et spécialistes. Pour cela, elle est principalement composée de stages dans différents milieux cliniques afin de mieux les préparer à leur futur milieu de travail. Malgré plusieurs avantages qu'on leur reconnaît, les difficultés de planification des activités évaluatives ont toujours été un défi. Ces difficultés découlent des problèmes inhérents à la nature même de ce type de formation: un cheminement variable de chaque étudiant, l'incertitude des problèmes cliniques rencontrés au cours des stages et le nombre élevé des professeurs engagés dans l'enseignement clinique.
Traditionnellement, et peut-être aussi pour contourner ces problèmes, l'évaluation des étudiants se fait principalement par une fiche d'évaluation globale. Ces fiches ont l'avantage d'être souples et faciles à utiliser, mais elles sont aussi souvent décriées pour leur manque de fiabilité. Les biais qui les affectent sont l'effet de complaisance, de halo et la tendance à une déviation vers les scores élevés.
Les recherches en mesure et évaluation en pédagogie médicale se multiplient depuis les dernières années afin de mieux contrôler ces biais, mais le problème reste encore important.
Nos réflexions, à ce stade-ci, visent alors à explorer les liens entre les contraintes rencontrées dans une formation par stage et les différents types de biais afin de proposer des pistes de solutions. Ces dernières tiendraient alors compte de la faisabilité et aussi des données de recherche.