



Conseil ontarien  
de la qualité de  
l'enseignement  
supérieur

Un organisme du gouvernement de l'Ontario



## Accroître la valeur : réseau de refonte de l'évaluation des habiletés cognitives pour le développement d'une réflexion de niveau supérieur

Natalie Simper, Brian Frank, Jill Scott et Wanda  
Beyer

Publié par le

## Conseil ontarien de la qualité de l'enseignement supérieur

1 1, rue Yonge, bureau 2402  
Toronto (Ontario) M5E 1E5 Canada

Téléphone : (416) 212-3893  
Télécopie : (416) 212-3899  
Site Web : [www.heqco.ca](http://www.heqco.ca)  
Courriel : [info@heqco.ca](mailto:info@heqco.ca)

### Se référer au présent document comme suit :

Simper, N., Frank, B., Scott, J. et Beyer, W. (2019). *Accroître la valeur : réseau de refonte de l'évaluation des habiletés cognitives pour le développement d'une réflexion de niveau supérieur*. Toronto, Conseil ontarien de la qualité de l'enseignement supérieur.



Les opinions exprimées dans le présent rapport de recherche sont celles des auteurs et ne reflètent pas nécessairement le point de vue ni les politiques officielles du Conseil ontarien de la qualité de l'enseignement supérieur ou des autres organismes ou organisations ayant offert leur soutien, financier ou autre, dans le cadre de ce projet. Imprimeur de la Reine pour l'Ontario, © 2019

## Table des matières

Sommaire .....	4
Introduction .....	6
Réforme de l'évaluation .....	6
Engagement du corps professoral .....	7
Réseau de soutien enchâssé.....	7
Tâches authentiques, apprentissage significatif.....	8
Fiabilité .....	8
Communautés de la pratique .....	10
Questions de recherche.....	10
Méthodes.....	11
Participants .....	12
<i>Instructeurs</i> .....	12
<i>Étudiants</i> .....	13
Mesures pour évaluer l'apprentissage des étudiants .....	13
Test normalisé .....	14
Évaluation fondée sur les cours.....	15
Évaluation valide des grilles d'apprentissage VALUE au premier cycle.....	16
Collecte et notation des réponses des étudiants .....	18
Résultats.....	20
Résultats VALUE.....	20
Résultats HEIghten.....	22
Corrélations entre les mesures.....	25
Sondage auprès des instructeurs .....	26
Rapports narratifs sur les cours.....	27
Discussion.....	28
Modèle pour aider les instructeurs à mettre au point leurs évaluations axées sur le cours .....	28
Liens entre les objectifs du cours et les résultats VALUE et HEIghten .....	30

Valeur ajoutée entre l'acquisition des habiletés cognitives entre la première année et la quatrième année .....	31
Renforcement des capacités des instructeurs.....	31
Propagation de l'initiative .....	33
Conclusions .....	34
Références .....	36

## Liste des tableaux

Tableau 1 : Échantillons de grilles VALUE .....	19
Tableau 2 : Grosseur des échantillons pour les résultats VALUE.....	20
Tableau 3 : Statistiques descriptives HEIghten .....	23
Tableau 4 : Corrélations entre les mesures d'évaluation .....	26
Tableau 5 : Indicateurs de croissance du renforcement des capacités .....	28

## Liste des figures

Figure 1 : Structure du réseau à l'appui de l'évaluation des habiletés cognitives.....	12
Figure 2 : Mesures de l'apprentissage des étudiants .....	14
Figure 3 : Alignement du cours et de la grille VALUE.....	16
Figure 4 : Résultats des habiletés cognitives opérationnalisés à l'aide des critères de la grille VALUE.....	17
Figure 5 : Résultats VALUE pour la pensée critique.....	21
Figure 6 : Résultats VALUE pour la résolution de problèmes .....	21
Figure 7 : Amélioration de la performance HEIghten entre la première année et la dernière année .....	24
Figure 8 : Note HEIghten selon le niveau d'effort par groupe d'année (affichage de l'erreur moyenne) .....	24
Figure 9 : Résultats quantitatifs du sondage sur l'enseignement.....	27

## Sommaire

Le projet de refonte des méthodes d'évaluation cognitive (RMEC) est une étude fondée sur un réseau à l'échelle de l'établissement qui met l'accent sur le développement et l'évaluation d'habiletés cognitives (pensée critique, résolution de problème et, à un moindre degré, réflexion créative) au niveau de l'enseignement de premier cycle. Cette recherche visait à examiner la relation entre l'évaluation de cours, les grilles normalisées et un test normalisé; à déterminer la valeur ajoutée entre l'atteinte des habiletés cognitives la première année et la dernière année; et à examiner le recrutement avec incitatifs d'instructeurs de cours de première année et de quatrième année désireux d'aligner expressément le développement des habiletés à la conception sur mesure d'évaluations de cours et à l'acquisition d'habiletés cognitives. L'apprentissage des étudiants était évalué à l'aide des mesures suivantes : grilles d'évaluation de cours, grilles normalisées (grilles VALUE de l'Association of American Colleges & Universities) et un test normalisé (HEIghten de l'Educational Testing System). Les instructeurs participaient à des sondages avant et après la mise en œuvre et des rapports narratifs détaillés étaient compilés pour suivre qualitativement le processus de changement.

Voici un aperçu des résultats de haut niveau du projet :

- L'analyse des données d'évaluation des étudiants a révélé une forte corrélation entre les notes fondées sur le cours et les notes des grilles VALUE, et les données ont été utilisées pour définir les prochaines étapes pour une partie des cours en cause. La faible corrélation entre les notes HEIghten et l'évaluation fondée sur le cours porte à croire que les évaluations de cours évaluaient plus que la seule pensée critique.
- Il y avait une grande amélioration (la moitié d'un écart-type) au niveau des habiletés en pensée critique entre les données transversales des étudiants de première année et de dernière année comme démontré par les notes de test HEIghten. Le résultat était reflété dans des gains à deux niveaux du niveau médian de rendement en pensée critique sur la grille VALUE avec des gains similaires au niveau de la grille de résolution de problème.
- Le projet de RMEC appuyait 24 instructeurs de 15 départements (voir le Tableau 1) et leur permettait d'aligner minutieusement l'évaluation des habiletés cognitives en utilisant efficacement des mesures de qualité où 40 % des instructeurs se servaient de grilles pour la première fois dans le cadre du cours.
- L'analyse qualitative des rapports de cours narratifs illustre une croissance et des résultats prometteurs au niveau de plusieurs des indicateurs de renforcement des capacités. Une éventuelle

amélioration concerne l'adoption de plus en plus grande de procédés formels pour la notation cohésive des travaux de cours.

- Les données des sondages auprès des instructeurs ont démontré les défis inhérents liés à l'instauration d'un changement de culture au niveau des pratiques d'évaluation, mais aussi le degré de renforcement des capacités atteint. Il est recommandé que ces mesures fassent l'objet d'un suivi à l'avenir pour déterminer les effets à long terme sur le projet.

## Introduction

On a démontré que des pratiques exemplaires pour modifier les protocoles d'évaluation entraînent d'importantes améliorations sur les plans de l'enseignement et de l'apprentissage au niveau des études supérieures (Simper, Frank, Scott et Kaupp, 2018). Le projet de refonte des méthodes d'évaluation cognitive (RMEC) a été conçu à partir des résultats d'une étude du CERA (Consortium sur l'évaluation des résultats (Simper, Frank, Scott et Kaupp, 2018). Le projet du CERA 1 de Queen's était une étude longitudinale de quatre ans qui examinait une évaluation normalisée et fondée sur les grilles de programme pour l'évaluation des résultats d'apprentissage associés à la pensée critique, à la résolution de problème, à la communication écrite et à l'apprentissage continu. Dans le projet du CERA 1, les résultats de deux tests normalisés (le Collegiate Learning Assessment Plus et le Test d'évaluation de la pensée critique) indiquaient que les habiletés des étudiants en matière de pensée critique, de résolution de problème, de communication écrite et d'apprentissage continu s'amélioraient au fil des quatre ans de leur programme, mais ces tests coûtaient cher à mener et étaient assujettis aux problèmes de motivation des étudiants. Les grilles VALUE normalisées (Valid Assessment of Learning in Undergraduate Education) se sont révélées plus économiques et ont fourni une méthode utile pour informer les instructeurs des améliorations à l'enseignement et à l'apprentissage, quoique non sans problèmes liés à la variance entre les tâches. Compte tenu des résultats du projet du CERA 1, le projet de recherche de RMEC met l'accent sur le développement et l'évaluation d'habiletés cognitives en instaurant les grilles VALUE et en orchestrant une approche de réseau pour évaluer le re-développement. Le projet de RMEC permet le regroupement de données fiables et valides (Mathers, Finney et Hathcoat, 2018) en créant un réseau d'instructeurs de diverses disciplines et facultés pour participer à la refonte des cours, un processus qui comprend l'alignement des résultats d'apprentissage, l'élaboration de tâches authentiques et l'ajustement sur mesure des grilles propres aux travaux (Kezar, 2013). Le projet est fondé sur un modèle de mobilisation axé sur le perfectionnement scolaire avec un soutien enchâssé pour mettre au point des tâches authentiques et garantir une évaluation fiable; il s'aligne sur les objectifs d'apprentissage des étudiants tout en étant soutenu par une communauté de la pratique. Suit une courte fondation théorique pour ces principes.

### Réforme de l'évaluation

Depuis plus de 20 ans, les instances dirigeantes de l'enseignement supérieur collaborent avec des universités et collèges pour élaborer des politiques et des normes qui favorisent des pratiques d'évaluation de haute qualité. En 2008, l'Organisation de coopération et de développement économiques (OCDE) lançait une enquête mondiale axée sur la faisabilité de l'évaluation avec une reconnaissance croissante de l'importance des résultats d'apprentissage (Tremblay, Lalancette et Roseveare, 2012), ce qui a donné lieu au projet d'Évaluation des résultats de l'enseignement supérieur (ERES) (Tremblay, 2013).

Des membres participants de partout dans le monde ont essayé diverses méthodes d'évaluation avec des résultats mixtes. Dans le cadre du processus de « Tuning » de Bologne, en Europe (González et Wagenaar, 2003), plus de 40 pays ont collaboré pour améliorer la qualité de l'éducation en déterminant des résultats d'apprentissage et d'autres procédés pour promouvoir la transparence, la mobilité et l'employabilité. Les efforts de « Tuning USA » ont mené à l'élaboration du profil des grades obtenus (PGO)), un cadre de compétences conçu pour aligner les programmes-cadres et la pédagogie (Jankowski, Hutchings, Ewell, Kinzie et Kuh, 2013).

Plus récemment, l'Association of American Colleges and Universities (AAC&U) a coordonné un projet d'évaluation des grilles dans le cadre d'une évaluation valide de l'apprentissage au niveau des études de premier cycle (VALUE) à grande échelle aux États-Unis (Drezek McConnell et Rhodes, 2017). Beaucoup de raisons expliquent pourquoi le changement donnant suite aux recommandations est lent, mais on peut dire que le changement au niveau de l'enseignement supérieur est difficile à réaliser (Henderson, Beach et Finkelstein, 2011).

### Engagement du corps professoral

Boud et Dochy (2010) présentent sept propositions pour la réforme de l'évaluation et recommandent que l'évaluation de l'apprentissage « soit au cœur de la conception des sujets et des programmes et constitue un point de mire pour le personnel et le développement institutionnel » (p. 3). L'AAC&U recommande que les membres du corps professoral soient directement engagés dans la qualité de l'apprentissage des étudiants (Rhodes, 2011). Un changement réussi exige une participation, une collaboration et un appui soutenus de la part du corps professoral (Bernstein et Greenhoot, 2014), mais un sondage mené par le National Institute of Learning Outcomes Assessment (NILOA) a révélé que les doyens de collèges et d'universités sont d'avis que « l'engagement de plus de membres du corps professoral constitue la principale embûche qui nuit à l'évolution de l'évaluation » (Kinzie, 2010, p. 14). Une stratégie d'engagement consiste à offrir de mini subventions; les fonds peuvent constituer un stimulus pour les activités de réforme et « peuvent avoir de puissantes incidences symboliques... et les personnes qui dirigent ces efforts peuvent avoir un levier utile pour obtenir d'autres ressources institutionnelles » (Miller, Fairweather, Slakey, Smith et King, 2017, p. 40). Le recrutement d'instructeurs du corps professoral dans la recherche permet une approche enchâssée envers la collecte de preuves d'apprentissage au niveau des travaux de cours.

### Réseau de soutien enchâssé

L'engagement du corps professoral constitue un principe central du guide détaillé sur le changement pédagogique de Chasteen et Code (2018), qui fait ressortir le rôle d'experts enchâssés. Le modèle de leadership organisationnel de Hannah et Lester (2009) porte à croire en une stratégie de gestion du changement à multiples niveaux instaurée pour appuyer un réseau de connaissances fondé sur le renforcement de capacités individuelles (micro), sur l'instauration d'un réseau qui facilite le changement

(méso) et, dans une moindre mesure, sur des systèmes de sanction institutionnelle (niveau macro). Quand nous pensons au niveau méso, les microcultures universitaires peuvent « fournir un engagement collégial positif avec les nouveaux professeurs » (Roxa et Martensson, 2015, p. 195). Il a été démontré que l'expertise enchâssée réussit à la fois au niveau d'initiatives systémiques à grande échelle (Chasteen et Code, 2018; Wieman, 2007) et au niveau d'un système viable pour des améliorations de qualité (Wieman, Deslauriers et Gilley, 2013). L'alignement des travaux de cours et des critères d'évaluation désirés exige souvent le développement de tâches qui prennent du temps et de l'expertise à mettre au point et exige aussi un haut degré de dévouement de la part de ceux qui enseignent aux étudiants (Purinton et Alexander, 2013). La recherche indique qu'avec une généreuse infusion de temps et d'expertise, la masse critique nécessaire peut être réalisée pour créer une nouvelle norme (Corbo, Reinholz, Dancy, Deetz et Finkelstein, 2014).

### **Tâches authentiques, apprentissage significatif**

Les principes de l'alignement constructif portent à croire que les travaux de cours doivent être conçus avec soin pour s'aligner avec les résultats attendus (Biggs, 2014) : « Tout comme nous comprenons que ce qui s'enseigne n'est pas la même chose que ce qui s'apprend, nous savons aussi que le but consiste à garantir que les étudiants ont acquis les connaissances et les habiletés particulières à un cours individuel » (Chun, 2010, p. 23). En outre, pour paraphraser Wiggins et McTighe (2005), les buts de l'apprentissage à long terme devraient inclure des expériences de réflexion sur la signification et permettre aux étudiants de transférer des compétences et des habiletés d'un contexte à l'autre. Les tâches authentiques « présentent aux étudiants un défi complexe du monde réel où le scénario, le rôle, le processus et le produit sont tous authentiques; ils doivent ensuite démontrer qu'ils ont les compétences et les connaissances requises pour accomplir la tâche » (Chun, 2010, p. 24). En d'autres mots, le but, c'est que les étudiants appliquent significativement des concepts pour résoudre des problèmes, utilisent des données probantes pour recommander des solutions ou élaborent de façon créative de nouvelles approches ou méthodes.

La nature du problème et du domaine disciplinaire doit être prise en compte. Comme l'argue Jonassen, (1997), les choix de concepts d'évaluation authentique sont tributaires d'un problème bien structuré.

De façon spécifique, les tâches authentiques peuvent ressembler à des projets d'analyse et de recherche, à des projets de conception, à des investigations ou à des enquêtes structurées (Ashford-Rowe, Herrington et Brown, 2013). Les différences au niveau de ces types d'évaluation dans diverses disciplines peuvent avoir des incidences sur l'uniformité quand vient le temps d'évaluer l'apprentissage des étudiants.

### **Fiabilité**

Les évaluations axées sur le cours sont généralement jugées valides (Tremblay, 2013), mais des différences, quand elles ne sont pas correctement alignées, peuvent donner lieu à des erreurs de variance allant jusqu'à 77 % (Hathcoat et Penn, 2012), ce qui a des effets sur la fiabilité des résultats. La recherche précédente du CERA (Frank, Simper et Kaupp, 2016) a observé un désalignement entre les travaux du cours et les critères

d'évaluation résultant de la conception de l'évaluation sans format cohésif pour opérationnaliser les concepts de la pensée critique ou de la résolution de problèmes. Si les étudiants ne sont pas guidés à fournir des réponses qui démontrent la pensée critique et la résolution de problèmes, alors la validité de l'évaluation de tels concepts est remise en cause. On considère généralement que plus on porte attention à l'enseignement et à l'évaluation de ces habiletés, plus les étudiants ont un bon rendement. Le rapport de l'AAC&U affirme que « la concentration du curriculum sur le développement d'habiletés liées à la pensée critique chez les étudiants au niveau des travaux de cours dans les divisions supérieures des majeures » (Drezek McConnell et Rhodes, 2017, p. 4). La fiabilité au niveau des évaluations est nécessaire pour estimer les gains d'apprentissage au fil du programme.

Beaucoup d'autres ont articulé des principes de conception d'évaluation efficace (Popham, 1999; Wiggins et McTighe, 2005, par exemple), mais sans plan d'évaluation délibéré, les grands énoncés de résultats sont difficiles à corroborer. Un plan d'évaluation doit tenir compte de plusieurs facteurs (à titre d'exemple, chaque contexte d'apprentissage est différent, ce qui complique l'application d'une approche unifiée. Les grilles fondées sur des critères sont de plus en plus utilisées pour évaluer l'apprentissage des étudiants (Dawson, 2017). On note toutefois des lacunes dans la conception et l'utilisation de grilles, comme des grilles qui incluent des listes de vérification plutôt que des indicateurs de qualité (Popham, 1997), ou des processus questionnaires pour établir la fiabilité entre les taux (Jonsson et Svingby, 2007). Une méta-analyse menée par Svingby (2007) porte à croire que « la notation fiable d'évaluations de rendement peut être améliorée en utilisant des grilles. En ce qui touche les questions de fiabilité, les grilles devraient être analytiques, spécifiques à un sujet et complémentées d'exemples et/ou de formation du notateur » (p. 141). La formation du notateur garantit que les notateurs comprennent comment appliquer les grilles. Le processus comprend une discussion de la terminologie de la pratique de notation utilisant des exemples de travaux, ce qui donne l'occasion aux participants d'expliquer la raison de la note donnée.

Certains arguent en faveur d'une approche de tests normalisés plutôt que d'une approche fondée sur les grilles pour quantifier le perfectionnement des habiletés (Council for Aid to Education, 2016; Stein et Haynes, 2011). Toutefois, ces tests sont encore perçus par beaucoup d'étudiants et d'instructeurs comme déconnectés du contexte disciplinaire, administrés comme une évaluation à faible enjeu et pas associés à la connaissance du domaine et leurs bienfaits sont donc jugés questionnaires (Madaus et Clarke, 2001). Les tests normalisés qui ont peu ou pas de conséquences pour les personnes qui les prennent sont sujets à des problèmes motivationnels et logistiques (Liu, Bridgeman et Adler, 2012; Simper, Frank, Kaupp, Mulligan et Scott, 2018). Si les étudiants ne sont pas motivés à performer, les résultats ne refléteront pas avec précision leurs apprentissages (Banta et Palomba, 2014). Par contre, les évaluations qui sont considérées à enjeux élevés par les étudiants (pour l'admission) ou par les établissements (pour du financement) peuvent détériorer l'environnement d'apprentissage. Cette détérioration tient à la réorientation du point de mire sur des compétences qui ne sont pas testées ou à l'exclusion des étudiants aux notes faibles, ce qui gonfle par conséquent la performance des étudiants (Koretz, 2008).

## Communautés de la pratique

Wenger (2000, 2011) décrit les communautés de la pratique comme celles qui favorisent une intégration entre des membres autochoisis avec des intérêts communs dans le but de partager des connaissances et de solutionner collectivement des problèmes. La confiance mutuelle existe au sein de tels groupes et plusieurs considèrent qu'ils ont la « structure sociale idéale pour la gérance des connaissances » (Wenger, McDermott et Snyder, 2002, p. 12). Hutchings, Huber et Ciccone, (2006) proposent une stratégie similaire pour l'amélioration collective de l'enseignant supérieure, soit un « lieu communal axé sur l'enseignement », un espace conceptuel où des communautés d'éducateurs engagés envers le questionnement et l'innovation pédagogiques se réunissent pour échanger des idées sur l'enseignement et l'apprentissage et s'en servent pour relever le défi de préparer les étudiants à la vie professionnelle et civique » (p. 26).

## Questions de recherche

Le concept du projet de Refonte des méthodes d'évaluation cognitive (RMEC) a tiré profit des pratiques efficaces, comme susmentionné, reposait sur un appui enchâssé, mettait l'accent sur la participation du corps professoral pour produire des évaluations de l'apprentissage significatives pour les étudiants, et évaluait de multiples méthodes. En outre, le projet de RMEC concordait directement avec le plan d'études de l'Université Queen's et renforçait les procédés et capacités d'évaluation des résultats d'apprentissage. Les questions de recherche spécifiques utilisées pour alimenter le processus et les méthodes analytiques étaient les suivantes :

1. Dans quelle mesure le modèle de refonte cognitif s'avère-t-il utile pour aider les instructeurs à élaborer des évaluations d'habiletés cognitives fondées sur le cours?
2. Quelle est la relation entre l'évaluation valide de l'apprentissage au niveau des études de premier cycle (VALUE) et les résultats de HEIghten?
3. Quelle est la valeur ajoutée entre l'acquisition d'habiletés cognitives en première année et en dernière année à l'échelle de l'Université Queen's?
4. Quels indicateurs démontrent que le projet renforce les capacités des instructeurs?
5. Dans quelle mesure le projet s'est-il avéré efficace en vue de renforcer les capacités et de propager l'initiative?

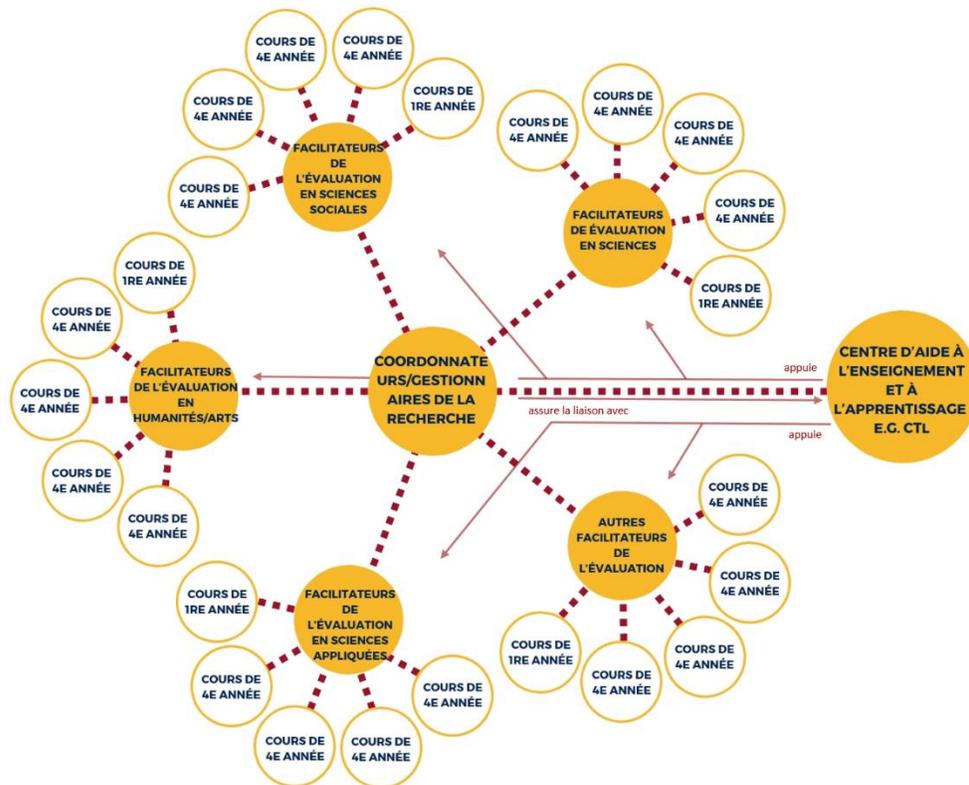
## Méthodes

Le projet de RMEC a été conçu pour renforcer les capacités au niveau des évaluations constructives qui s'alignent avec les dimensions de pensée critique, de résolution de problèmes et de pensée créative. Ce projet a été mis en œuvre en quatre étapes :

- **Prémise en œuvre** : Tenue de conversations et de consultations ouvertes avec des professionnels de l'éducation de divers départements (janvier-avril 2017).
- **Initiation du projet** : Création d'un site Web pour donner un aperçu du projet et pour les appels de propositions (juin juillet 2017).
- **Lancement du projet** : Début de la création du réseau, y compris la promotion du projet auprès des chefs de départements, l'orientation des facilitateurs de l'évaluation et la tenue de consultations initiales avec les instructeurs (août-septembre 2017).
- **Mise en œuvre** : Collaboration entre les facilitateurs de l'évaluation et les facultés participantes, recrutement des étudiants de première année et de quatrième année, administration de tests normalisés, notation VALUE et tenue d'analyses de données (septembre 2017–juillet 2018).

La structure du réseau est chapeauté par le gestionnaire de la recherche qui a collaboré étroitement avec les parties intéressées, qui est sensible aux besoins des hauts dirigeants et qui respecte les normes et pratiques disciplinaires. Le gestionnaire de la recherche a chapeauté le personnel de soutien (facilitateurs de l'évaluation) qui avait l'expertise disciplinaire pour fournir une aide à l'évaluation aux instructeurs participants. Les cinq facilitateurs de l'évaluation embauchés une journée par semaine aidaient les membres du corps professoral à élaborer des évaluations qui s'alignaient avec l'ensemble des mesures. Lors de leurs réunions hebdomadaires, ils recevaient une formation pour les aider à élaborer des tâches authentiques et des processus et procédures d'évaluation. Trois des facilitateurs des évaluations travaillaient en même temps comme adjoints de recherche, un comme concepteur pédagogique et un comme professeur auxiliaire. Dans le cadre de leurs rôles actuels, ils inspiraient confiance et étaient respectés au niveau de leurs disciplines. En outre, ils avaient le capital intellectuel et humain requis pour naviguer et aider à recruter des instructeurs du corps professoral (Hanson, 2001). Chaque facilitateur de l'évaluation travaillait avec un instructeur de première année et avec de multiples instructeurs de quatrième (ou dernière) année en ce qui avait trait à l'enseignement et à l'évaluation.

Figure 1 : Structure du réseau à l'appui de l'évaluation des habiletés cognitives



## Participants

### Instructeurs

Les principaux enquêteurs du CERA ont pris contact avec des doyens adjoints et des chefs de département pour obtenir leur appui en vue de recruter des instructeurs pour le projet. Une page Web a aussi été créée décrivant le projet et annonçant des appels de déclaration d'intérêt. La participation des instructeurs a été formalisée grâce à la soumission de courtes déclarations d'intérêt (DI). Les trois questions réflexives dans la DI étaient : (1) Comment encouragez-vous le perfectionnement des habiletés cognitives dans le cadre de votre cours à l'heure actuelle? (2) Comment évaluez-vous les habiletés cognitives à l'heure actuelle? (3) Comment pourriez-vous améliorer ces évaluations? Il y avait 24 instructeurs participants relevant de cinq groupes de disciplines, avec une représentation des départements de sciences biomédicales, de biologie, de théâtre, de génie (chimique, civil, géologique et multidisciplinaire), d'anglais, de géographie, de sciences de la santé, de médecine, de sciences infirmières, de psychologie et de sociologie. Chaque instructeur participant recevait un petit cachet monétaire à utiliser pour améliorer les cours, comme des conférenciers invités, des recherches propres au cours ou des séances d'orientation à l'intention des étudiants.

## Étudiants

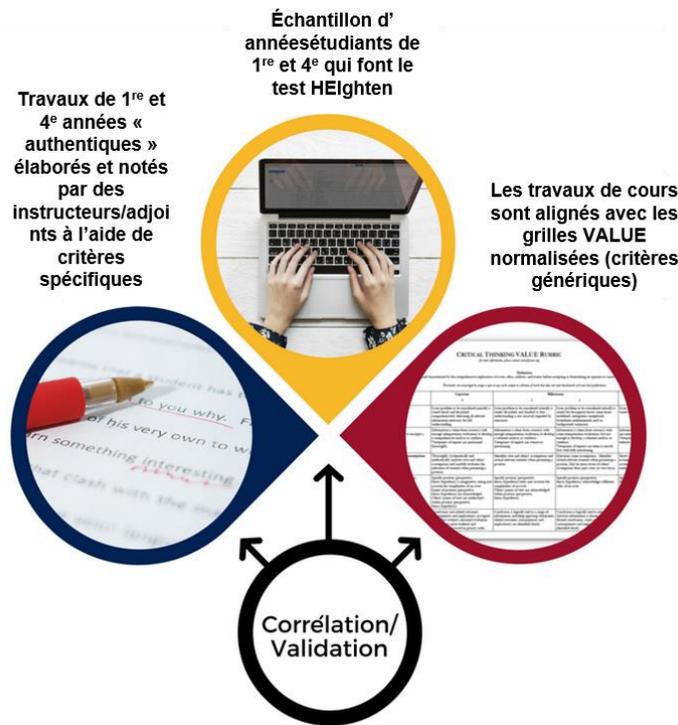
Les étudiants inscrits à des cours de première et quatrième années participants ont été recrutés pendant les heures de cours par le facilitateur de l'évaluation désigné ou par le gestionnaire de la recherche. Dans certains cas (p. ex., cours en ligne), les étudiants ont été recrutés par le facilitateur de l'évaluation ou le gestionnaire de la recherche par le biais du système de gestion de l'apprentissage. Le recrutement des étudiants avait deux objectifs : premièrement, obtenir le consentement des étudiants à participer conformément aux lignes directrices éthiques et, deuxièmement, s'assurer que les étudiants étaient au courant des efforts engagés par leur instructeur de cours et l'Université Queen's pour améliorer et évaluer leurs habiletés cognitives. Un concept transactionnel a été utilisé. Par contre, la grosseur de l'échantillon variait d'un cours à l'autre. Dans certains des cours de dernière année, moins d'étudiants étaient inscrits (n=12), donc l'échantillon de consentement maximum était limité par le nombre d'étudiants inscrits. De plus, il y avait des étudiants dans une gamme d'années de programme inscrits à plusieurs cours participants. Par conséquent, l'échantillon était choisi par leur année au sein du programme (première ou dernière). Ce chevauchement de participants limitait le nombre disponible de réponses d'étudiants. Dans la mesure du possible, l'échantillon était tiré de ceux qui avaient fait le test HEIghten. Dans le cas des cours avec de nombreuses inscriptions avec la majorité d'étudiants consentants, un processus de stratification a été utilisé pour garantir la représentativité. Des étudiants consentants ont été choisis pour représenter ceux avec des résultats élevés, moyens et faibles au niveau du test HEIghten, y compris un équilibre entre les sexes conforme avec la répartition au niveau du programme. Les étudiants d'anglais langue seconde ont été inclus dans la mesure du possible pour garantir une représentation adéquate au niveau du test HEIghten puisqu'une partie du test comprend des réponses écrites. Les titres des travaux du cours et la grosseur de l'échantillon sont énumérés au Tableau 1.

## Mesures pour évaluer l'apprentissage des étudiants

Les trois outils d'évaluation utilisés pour évaluer l'apprentissage des étudiants sont illustrés à la Figure 2. Il s'agit des outils suivants :

- Test normalisé (HEIghten Educational Testing System)
- Évaluation fondée sur le cours (grilles et clés de notation)
- Grilles normalisées (grilles VALUE de l'AAC&U)

Figure 2 : Mesures de l'apprentissage des étudiants



### Test normalisé

Les membres de l'équipe de recherche ont examiné et mis en œuvre le test Collegiate Learning Assessment Plus, le Test d'évaluation de la pensée critique et le test HEIghten dans le cadre d'une étude antérieure du CETA. Ils ont constaté que le test HEIghten était le plus utile, rentable et réalisable pour satisfaire aux objectifs du projet de RMEC (Simper, Frank, Kaupp, Mulligan et Scott, 2018). En outre, le test HEIghten triangulait les données puisque cette évaluation normalisée captait des aspects de la pensée critique différents de ceux des grilles VALUE. Le test HEIghten a été mis au point par l'Educational Testing Service (ETS). Il s'agit d'un test en ligne portant sur 26 éléments de la pensée critique qui comprend un sondage de sortie avec des questions sur la démographie, l'engagement et l'effort. Le test HEIghten a été conçu pour évaluer des habiletés analytiques (analyse de la structure d'argument, pour l'évaluation de la structure d'argument et pour l'évaluation des données probantes et de leur utilisation) et des habiletés de synthèse (formulation d'arguments valides ou solides, choix d'information qui constitue cet argument ou qui y contribue pour une position particulière, reconnaissance de conclusions, extrapolation d'implications, ou reconnaissance ou formulation d'explications pour des phénomènes décrits). Il est administré dans le cadre

d'une séance d'une heure soit surveillée par une personne à travers le portail du test, soit non surveillée par courriel d'invitation. Les résultats sont accessibles immédiatement. Le test offre l'option d'afficher les résultats de l'étudiant à la fin du test. Il affiche leurs résultats pour les concepts d'évaluation comparativement à la moyenne générale. Les résultats sont aussi accessibles en téléchargeant le rapport du cours ou le fichier de données (.csv) dès la fin de la séance. Le test a été enchâssé dans 10 cours et une invitation a été envoyée à tous les étudiants participants inscrits à des cours de quatrième année qui n'avaient pas enchâssé le test. Les incitatifs à la participation au test varient, d'aucun incitatif à une exigence du cours, au tirage d'un prix ou à un pourcentage des notes de cours attribué pour professionnalisme pour avoir participé (voir le Tableau 3 pour les statistiques descriptives).

### Évaluation fondée sur les cours

Le facilitateur de l'évaluation et l'instructeur du cours ont passé à travers un processus itératif pour concevoir ou reconcevoir une tâche de cours. L'apprentissage des étudiants a été évalué à l'aide d'un type de cadre de qualité, surtout sous forme de grille. Par contre, il y a eu quelques cas où la notation fondée sur le cours a été menée de façon holistique en utilisant des descripteurs de notes. Ce travail visait à aligner les tâches du cours avec une grille d'évaluation qui, elle, s'alignait avec les résultats de la grille VALUE choisie (voir la Figure 3). Les grilles VALUE ont été utilisées pour opérationnaliser les concepts de pensée critique, de résolution de problème et de pensée créative. Le langage des grilles VALUE du cours appuyaient la détermination des grilles du cours et fournissaient un point de discussion pour le facilitateur de l'évaluation et l'instructeur. Une fois l'évaluation fondée sur le cours et la grille définie, les instructeurs examinaient le travail et la grille avec les étudiants de la même façon qu'ils le feraient pour n'importe quel autre travail de cours. La Figure 3 illustre l'alignement des cours et des grilles VALUE en relation avec les données probantes exhibées dans un artefact d'étudiant.

Figure 3 : Alignement du cours et de la grille VALUE



### Évaluation valide des grilles d'apprentissage VALUE au premier cycle

Il existe 16 grilles VALUE générales mises au point par l'AAC&U (Rhodes et Finley, 2013). La première page de chaque grille fournit une description et une définition des termes. Les critères sont à l'arrière et la colonne de gauche énumère les dimensions tandis que la rangée du haut indique les niveaux de réalisation (Capstone 4, Milestones 2 et 3 et Benchmark 1). Il y a des indicateurs de qualité pour chaque niveau et dimension. Les résultats attendus de ces trois grilles sont résumés à la Figure 4.

Figure 4 : Résultats des habiletés cognitives opérationnalisés à l'aide des critères de la grille VALUE



Les habiletés cognitives s'expriment différemment selon les normes, les pratiques, les contextes et les hypothèses courants propres à une discipline. Ces différences d'opinions concernant les concepts d'évaluation peuvent constituer un obstacle au travail. À titre d'exemple, on pourrait passer beaucoup de temps à discuter à savoir en quoi la pensée critique au niveau des humanités diffère de la pensée critique au niveau des sciences. Par contre, un tel débat mine le but, qui consiste à évaluer le concept. Par conséquent, la tenue pour acquis d'un cadre prédéfini comme la grille VALUE s'avère bénéfique pour opérationnaliser les habiletés cognitives.

## Collecte et notation des réponses des étudiants

Si possible, les travaux des étudiants consentants ont été recueillis à l'aide du système de gestion de l'apprentissage, mais dans le cas des cours où les travaux étaient soumis en version papier, nous avons recueilli et numérisé les documents. La réponse de chaque étudiant de l'échantillon était notée indépendamment par deux personnes. Une était le facilitateur de l'évaluation engagé dans le développement des travaux du cours et l'autre était un expert disciplinaire. Une procédure cohésive était suivie pour noter à l'aide des grilles VALUE. La procédure supposait une compréhension commune en lisant les instructions des travaux et un exemple de réponse, en déterminant comment les critères de dimension avaient été démontrés et en parvenant à une entente concernant le niveau du travail. Les notateurs notaient ensuite individuellement cinq à dix échantillons de travaux à la fois, compilant une liste annotée pour justifier la décision pour chaque critère et assignant un niveau de performance pour chaque dimension. L'étape finale était celle de la calibration, où deux notateurs utilisaient leurs notations pour discuter de toute différence entre les niveaux assignés et, si possible, en venir à une entente quant au niveau final. Les niveaux avant et après la calibration étaient enregistrés pour faire le suivi de la fiabilité de la notation. L'Université Queen's a recours à la notation à l'aide de grilles VALUE depuis cinq ans. En ce sens, elle a accès à des notateurs d'expérience et a établi des procédés pour initier de nouveaux notateurs. La fiabilité pré-calibration entre les deux notateurs était élevée (entre 68 % et 78 %) et le processus de calibration assurait un niveau d'entente de 99 %.

**Tableau 1 : Échantillons de grilles VALUE**

Faculté	Cours	Reconception d'évaluation	Échantillon
<b>Sciences sociales</b>	PSYC 100	Laboratoire de pensée critique en trois phases	41
	PSYC 453	Réponse critique et proposition de recherches	8
	PSYC 450	Réponse multipartite en pensée critique	16
	GPHY 401	Projet d'analyse des systèmes alimentaires	10
	SOCY 424	Pouvoir, inégalités et justice sociale - affiches/présentations	15
<b>Humanités</b>	ENGL 100	Essai final : analyse d'un poème	50
	ENGL442	Réflexion es essai critique	10
	ENGL 487	Réponse de lecture attentive	24
	ENGL 489	Utilisation de la discussion pour éclairer un papier de sources secondaires	26
	DRAM 439	Réflexion sur le processus de performance et rétroactions des pairs	9
<b>Sciences appliquées</b>	APSC 100	Rapport de faisabilité sur les colonies sur Mars	191
	MECH 495	Conception d'une station de travail	56
	CIVL 500	Thèse de recherche et séance d'affiches	16
	GEOE 447	Rapport de recherches	10
	CHEE 470	Rapport sur les dangers environnementaux	64
<b>Sciences</b>	BIOL 103	Laboratoire de pensée critique et rapport de recherches	42
	BIOL 402	Conception de recherches et rapport de laboratoire	3
<b>Sciences de la santé</b>	NURS 101	Réflexion critiques : santé autochtone	42
	NURS 401	Réflexion critique : enjeux liés à la santé	36
	HLTH 102	Évaluation des sources de données probantes	44
	ANAT 100	Étude de cas : application des connaissances anatomiques	14
	BMED 173	Série de blogues sur la pensée critique (expérimentation animale)	24
	MEDS 112	Révision WIKI Med et projet de référencement	17
	OT 871	Réflexions sur les consultations de patients novices-experts	20
	TOTAL		788

## Résultats

### Résultats VALUE

Les notes médianes de la première année et de la quatrième année sur les dimensions de la pensée critique sont affichées au Tableau 2. Comme indiqué, la note médiane pour la première année pour toutes les dimensions était Benchmark 1 (n= 391), la note médiane pour la dernière année pour « explication des enjeux » était Capstone 4 et les autres dimensions étaient Milestone 3 (n= 268). Le pourcentage d'étudiants notés à chaque niveau des dimensions de la pensée critique est affiché à la Figure 5. Les résultats au niveau du cours sont affichés sous forme de graphiques à barres et joints à l'Annexe 1.

La note médiane de la première année pour les dimensions de résolution de problèmes « définition du problème », « identification du problème » et « solution proposée » était Milestone 2. La note médiane de première année pour « évaluation de la solution » était Benchmark 1, « évaluation des résultats » était plus faible que Benchmark 1 (notée avec un niveau zéro) (n= 203). Les notes médianes pour toutes les dimensions étaient Milestone 3 (les travaux n'ont pas été notés pour chaque dimension n=168 à n=178). Le pourcentage d'étudiants notés à chaque niveau des dimensions de résolution de problèmes est affiché à la Figure 6.

**Tableau 2 : Grosseur des échantillons pour les résultats VALUE**

		Première année	Médiane	Dernière année	Médiane
Pensée critique	Exploration d'enjeux	391	1	268	4
	Utilisation de données probantes	357	1	268	3
	Contexte et hypothèses	391	1	268	3
	Position/Hypothèse	391	1	212	3
	Conclusions et résultats	391	1	268	3
Résolution de problèmes	Définition du problème	203	2	178	3
	Détermination des stratégies	203	2	174	3
	Proposition de solutions	203	2	177	3
	Évaluation de la solution	203	1	168	3
	Évaluation des résultats	191	0	169	3

Figure 5 : Résultats VALUE pour la pensée critique

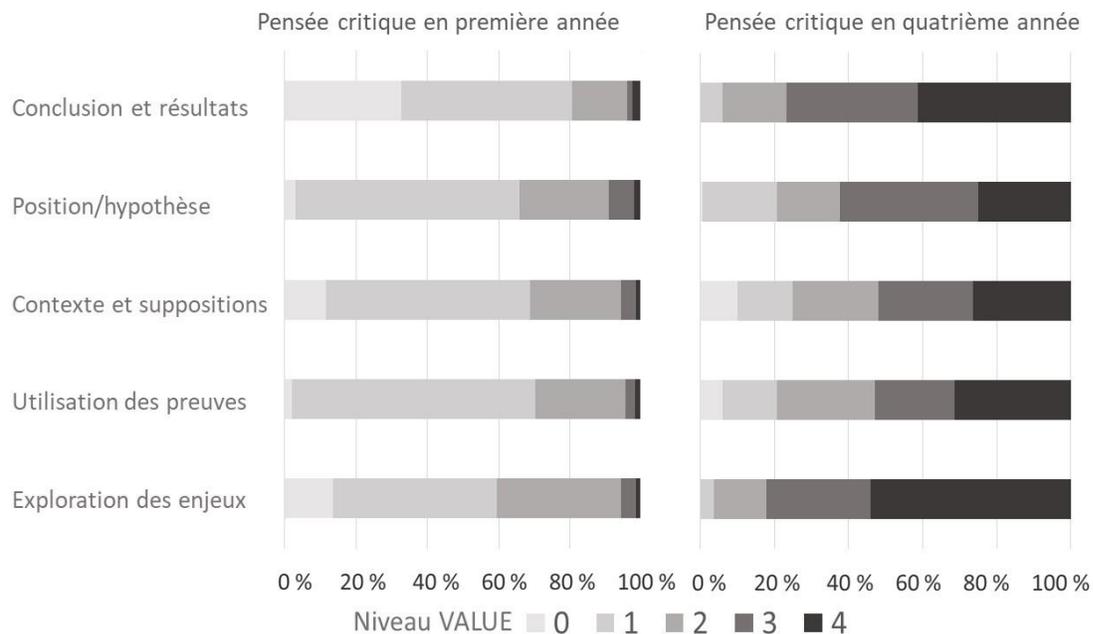
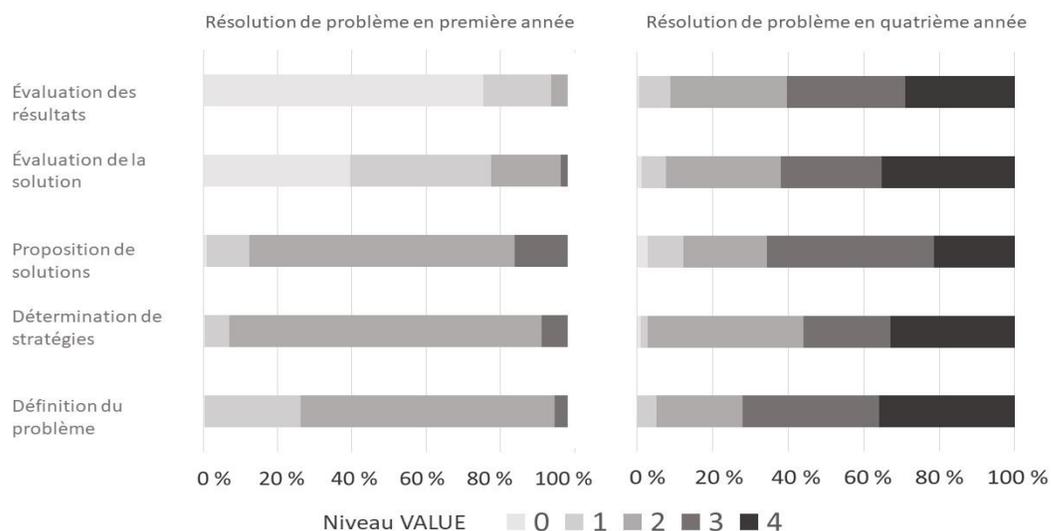


Figure 6 : Résultats VALUE pour la résolution de problèmes



**Remarque :** Il n'y avait pas assez d'échantillons des travaux évalués au niveau « Appliquer la solution » pour être représentatifs à l'échelle de l'établissement. Par conséquent, ils ne sont pas inclus à la Figure 6.

## Résultats HEIghten

Les statistiques descriptives pour les échantillons du test HEIghten sont affichées au Tableau 3. Il est à noter que les résultats de cours individuels ont été tenus anonymes pour les instructeurs participants. Le plus grand pourcentage de participation et d'achèvement de test a été observé par le biais d'un test surveillé sur les lieux, conformément aux exigences du cours, ou avec un pourcentage de la note de cours attribuée pour professionnalisme au niveau de la participation au test. Le résultat du test HEIghten n'est pas jugé valide si les étudiants ont terminé moins de 75 % du test. Seuls les tests terminés ont été inclus dans l'analyse.

Le portail de surveillance HEIghten permet la production de rapports, fournit des comparaisons de notes et des seuils de définition. Un exemple de rapport HEIghten produit pour tous les étudiants de cours de première année ayant participé au test est joint à l'Annexe 3. Le système de test permet aussi de télécharger des données immédiatement après la fin du test. Des analyses distinctes ont été menées en fonction des échantillons d'étudiants en première année et en dernière année de leur programme qui avaient terminé au moins 75 % du test et indiqué avoir investi au moins un certain effort dans le test. L'analyse de variance a démontré d'importantes différences entre les performances en première année ( $n=1523$ ) et en dernière année ( $n=149$ ) au niveau du test HEIghten  $F(1,1683)=30.82$   $p<.001$  (voir la Figure 7). La grosseur de l'effet Cohen était  $d= .497$  (une amélioration d'une moitié d'écart-type de la première année à la dernière année). Les répartitions par niveau de programme figurent sous forme de diagrammes en boîte joints à l'Annexe 2.

Les efforts des étudiants étaient autodéclarés sur un barème de 5 points. L'effort constitue un important facteur pour les résultats du test puisque l'examen de la régression linéaire a révélé que les efforts en première année représentaient 8 % de la variance de points totale ( $R^2=.08$ ). On notait des résultats semblables chez les étudiants de dernière année alors que les efforts représentaient 7 % de la variance de points totale ( $R^2=.07$ ). La Figure 8 affiche le pointage moyen pour chacun des niveaux de barèmes d'efforts. Fait à noter, même s'il y avait de grandes différences entre les performances en première année et en dernière année, les étudiants de première année qui avaient investi beaucoup d'efforts réussissaient mieux que ceux de dernière année qui avaient investi peu d'efforts.

**Tableau 3 : Statistiques descriptives HEIghten**

Cours	Contexte du test et incitatif	Échantillon total pour le test	Test terminé	Consentement total	% de consentement et terminé	Sexe			> 75 % terminé	Note globale	
						% femmes	% Hommes	% Autre		Moyenne	ET
# 1	Sans surveillance- 0,5 %	289	174	73	18%	69 %	28 %	3 %	72 %	165,4	7,21
# 2	Surveillance – 2 %	915	882	709	74 %	29 %	70 %	1 %	95 %	168	5,74
# 3	Sans surveillance – 5 %	934	898	791	82 %	74 %	25 %	1 %	96 %	164,4	7,19
# 15	Sans surveillance – exigence de cours	87	71	83	78 %	90 %	9 %	1 %	83 %	159,4	6,58
# 18	Surveillance – exigence de cours	136	134	124	90 %	94 %	6 %	<1 %	98 %	165,4	6,4
# 21	Surveillance – 5 %	58	56	57	95 %	91 %	9 %	<1 %	96 %	162,4	6,92
Première année - TOTAL		2 419	2 215	1 837	76 %	56 %	41%	4 %	90 %	164,2	6,67
# 7	Surveillance – 5 % et nourriture	18	17	17	94 %	44 %	51 %	5 %	94 %	172,9	3,59
# 16	Surveillance-nourriture	56	54	45	80 %	47 %	53 %	0 %	98 %	170,1	5,92
# 19	Surveillance – exigence de cours	78	77	58	74 %	90 %	9 %	1 %	98 %	166,9	5,36
Groupe de 4 <sup>e</sup> année	Sans surveillance-tirage d'un prix	38	30	38	79 %	58 %	42 %	1 %	96 %	165,9	4,06
Quatrième année - TOTAL		190	178	158	94 %	66 %	33 %	1 %	93 %	169,0	5,22
# 20 (Cours de 2 <sup>e</sup> cycle)	Surveillance-exigence de cours	70	59	52	74 %	90 %	6 %	4 %	79 %	168,1	6,11

Figure 7 : Amélioration de la performance HEIghten entre la première année et la dernière année

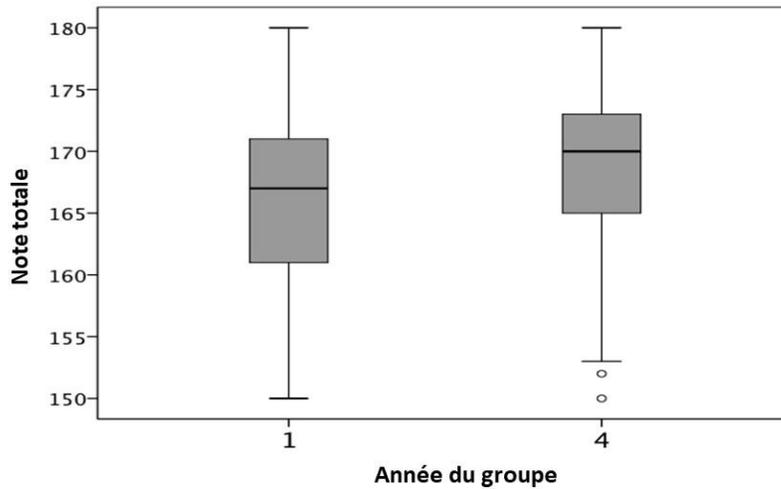
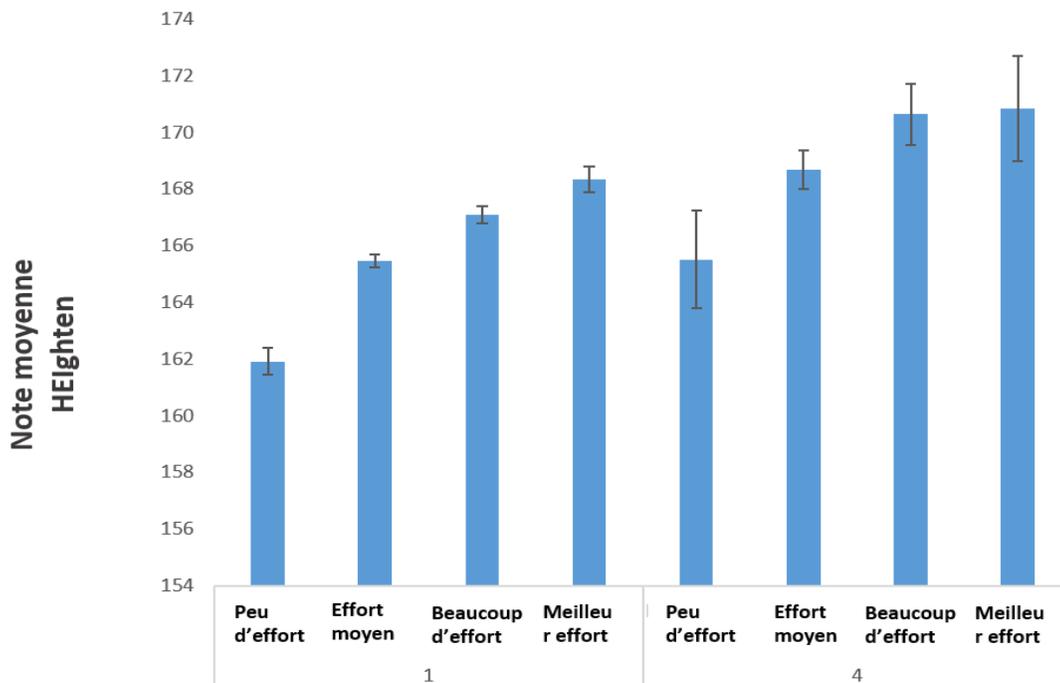


Figure 8 : Note HEIghten selon le niveau d'effort par groupe d'année (affichage de l'erreur moyenne)



## Corrélations entre les mesures

Compte tenu que les données VALUE étaient normalement réparties sans schématisateur et qu'il y avait un assortiment de données ordinales et continues, les corrélations de Pearson ont été calculées entre les notes des grilles VALUE par dimension, par évaluation de cours, par pourcentage de cours, par évaluation de cours, par pourcentage de cours, par moyenne pondérée cumulative (MPC) par session, par MPVC cumulative ainsi que par l'évaluation HEIghten (voir le Tableau 4). Il y avait des corrélations significatives entre toutes les notes des dimensions VALUE et leurs notes de travaux respectives. La plus grande corrélation au niveau du cours était entre la dimension de résolution de problème « Appliquer la solution » et l'année des travaux ( $r(70) = ,880$   $p < ,001$ ) et le plus faible était la dimension de pensée critique « Conclusions et résultats » ( $r(340) = ,221$   $p < ,001$ ). Puisque la note de travail constitue une composante des dimensions d'évaluation, les corrélations ont été calculées à partir d'une moyenne VALUE de la pensée critique (PC) et une moyenne pour la résolution de problème (RP). Les corrélations entre la note du travail de cours et la PC étaient  $r(400) = ,303$   $p < ,00$  et entre la note de travail et de cours et la RP étaient  $r(355) = ,319$   $p < ,001$ . Les corrélations au pourcentage du cours étaient plus élevées que la note de travail, avec la PC elles étaient  $r(518) = ,355$   $p < ,001$ , et avec la RP, elles étaient  $r(248) = ,603$   $p < ,001$ . Les corrélations avec la MPC de session chutaient au point où la corrélation à la MPC cumulative n'était pas significative pour la PC. Les corrélations entre les dimensions VALUE et les notes HEIghten étaient similaires aux corrélations du travail de cours VALUE. Les plus fortes corrélations à HEIghten étaient à la MPC cumulative ( $r(2011) = ,318$   $p < ,001$ ). Le tableau de corrélations au niveau des dimensions est joint en tant qu'Annexe 4.

**Tableau 4 : Corrélations entre les mesures d'évaluation**

	Année du travail	Pourcentage du cours	MPC de la session	MPC cumulative	Moyenne de la PC VALUE	Moyenne de la RP VALUE
Pourcentage du cours	,236** 358					
MPC de la session	,130** 457	.633** 1437				
MPC cumulative	,108* 457	,534** 1437	,932** 2606			
Moyenne de la PC VALUE	,303** 400	,355** 518	,140** 740	0,002 740		
Moyenne de la RP VALUE	,319** 355	,603** 284	,425** 394	,253** 394	,899** 337	
Note HEIghten totale	0,086 302	,198** 1 045	,279** 2 011	,318** 2 011	,193** 491	,241** 274

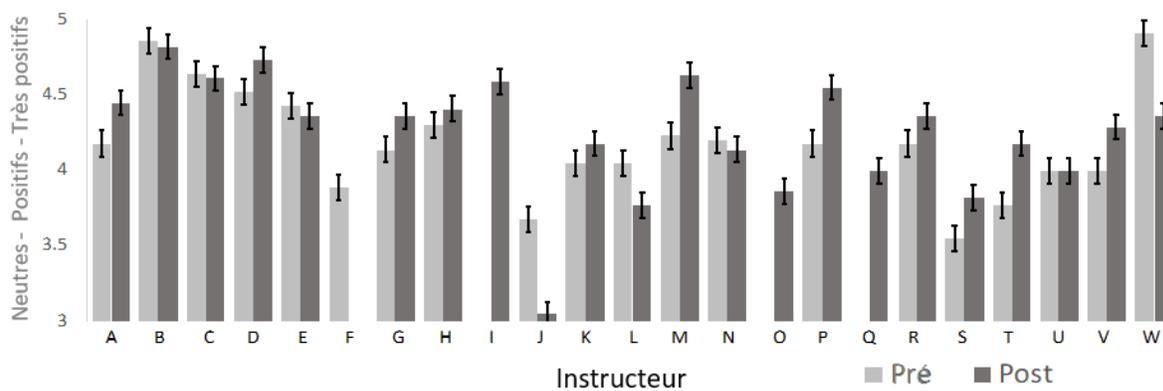
\*\* La corrélation est significative au niveau 0,01 (2- queue).

\* La corrélation est significative au niveau 0,05 (2-queue).

### Sondage auprès des instructeurs

Avant le début de la reconception de l'évaluation, les instructeurs en cause ont participé à un sondage sur les attitudes face à l'enseignement et les pratiques d'enseignement et d'évaluation. Le sondage incluait des questions ayant trait à l'intervention d'évaluation spécifique et une section portait sur les attitudes face à l'enseignement (15 points déclarés sur une échelle d'entente de 5 points et 5 points déclarés sur une échelle d'importance de 5 points). La fiabilité de l'échelle pour ces 18 points était calculée à l'aide de l'alpha = ,86 de Cronbach. Un exemple de question associée à l'échelle d'entente est « Pour enseigner efficacement, il faut savoir comment les étudiants apprennent un sujet et pas seulement connaître le sujet ». Un exemple de questions de l'échelle d'importance est « Comprendre de quelle façon les étudiants apprennent un sujet particulier ». Vingt-trois instructeurs ont participé au sondage dont 19 qui ont participé à la fois au présondage et au post-sondage. Les notes d'échelle pour le présondage et le post-sondage pour chacun des instructeurs de cours sont affichées à la Figure 9. Les données du sondage portent à croire que, parmi les 19 réponses jumelées, il y avait un changement positif significatif dans le cas de huit instructeurs, aucun changement significatif dans le cas de huit instructeurs et un changement négatif dans le cas de trois instructeurs.

Figure 9 : Résultats quantitatifs du sondage sur l'enseignement



### Rapports narratifs sur les cours

Les efforts de reconception de l'évaluation ont été déclarés par les facilitateurs de l'évaluation à l'aide d'un gabarit de rapport narratif. Les questions dans le gabarit ont été mises au point par le groupe de recherche pour répondre aux objectifs de recherche et pour surveiller le processus, les problèmes et les résultats qualitatifs. Il y a eu 25 rapports narratifs (un pour chaque cours) de cinq pages chacun environ. Les faits saillants les plus notoires ont été résumés dans un tableau de sept pages avec les activités regroupées dans la colonne descriptive et les résultats dans une autre colonne (le tableau est joint en tant qu'Annexe 5). Un processus d'analyse inductive a suivi la réflexion sur les points communs, en généralisant sur les comportements et les résultats démontrant une croissance professionnelle.

La prochaine étape consistait à regrouper les comportements communs à la majorité des cas et inclus en tant qu'indicateurs de croissance. Les fréquences de ces actions de la part des instructeurs de cours participants sont indiquées au Tableau 5. Des exemples de l'indicateur « découverte » étaient des choses comme modifier considérablement sa perspective et son style d'enseignement suivant des conversations avec les étudiants et le facilitateur de l'évaluation, ou la réalisation, par l'instructeur du cours, de l'apprentissage fortuit enchâssé dans l'activité de pensée critique, avait pour effet d'améliorer la persistance et le professionnalisme. Ces découvertes n'étaient pas toujours positives. À titre d'exemple, quand, dans le cadre d'un rigoureux processus de pensée critique, on réalise que le point de mire du contenu n'est pas aligné avec les critères des dimensions de la pensée critique de la grille VALUE.

**Tableau 5 : Indicateurs de croissance du renforcement des capacités**

	Nombre d'instructeurs	% d'instructeurs participants
Utilisation d'une grille fondée sur un critère pour la première fois	10	40 %
Découverte/révélation évaluation pédagogique/cognitive de la discipline	5	20 %
Adoption d'un processus de notation officiel et cohésif (AE) pour les travaux de cours	6	24 %
Recours à une stratégie pour promouvoir l'effort des étudiants dans le cadre d'une évaluation normalisée	8	32 %
Présentation donnée par l'instructeur lors de conférence/symposium	6	24 %
Instructeur partage son expertise de façon formelle (entraîneur en évaluation)	3	12 %
Travail se poursuit (instructeur/cours participe de nouveau)	12	48 %

## Discussion

### Modèle pour aider les instructeurs à mettre au point leurs évaluations axées sur le cours

Une approche de réseau donne la chance de négocier et de mieux interpréter les pratiques d'apprentissage, d'enseignement et d'évaluation (Roxå, Mårtensson et Alveteg, 2011). La RMEC a permis la création d'un réseau de soutien pour engager la participation du corps professoral et pour promouvoir un changement au niveau des pratiques d'évaluation. Travaillant au « niveau méso » pour favoriser ce changement, les facilitateurs de l'évaluation ont appuyé avec succès 24 instructeurs pour les aider à reconcevoir leurs évaluations pour cibler de façon particulière la pensée critique, la résolution de problème et, dans un cas, la pensée créative. Ces travaux étaient diversifiés (voir le Tableau 1) et exigeaient un processus itératif de planification, de conception et de peaufinage. Dans le cas de certains des instructeurs participants, ce processus constituait une courbe d'apprentissage difficile; un instructeur a dit « nous avons essayé un format radicalement différent et avons eu des difficultés au niveau de la mise en œuvre » (instructeur Q). Un autre ajoutait « un cours intensif sur le vocabulaire de la pédagogie serait excellent — peut-être une

réunion additionnelle de 90 minutes où nous pourrions nous familiariser avec la signification des termes » (instructeur H). Outre l'élaboration de l'évaluation, les facilitateurs de l'évaluation ont aidé les instructeurs à articuler la terminologie spécifique des grilles VALUE. À titre d'exemple, « pour décomposer les grandes habiletés cognitives en composantes, puis tenter de voir comment/quand les étudiants s'en servent » (instructeur J). Ce dialogue essentiel a fait ressortir les nuances propres à la discipline de la terminologie des grilles VALUE. Ce degré de soutien au niveau « micro » a aidé les instructeurs à aligner leurs travaux de cours aux grilles VALUE. À titre d'exemple, ils « doivent devenir familiers avec la grille VALUE avant de concevoir l'évaluation pour avoir une idée de comment les habiletés cognitives peuvent être évaluées, puis travailler à reculons au moment de concevoir le « concept à l'envers » de l'évaluation » (instructeur O).

Compte tenu du fait que le changement au niveau de l'enseignement et de l'apprentissage exige une participation soutenue et un soutien (Bernstein et Greenhoot, 2014), certains instructeurs n'étaient pas en mesure d'instaurer entièrement le travail qu'ils avaient modifié pour l'itération 2018 du cours. Dans le cas des cours en ligne, les travaux sont déterminés d'avance et font partie du processus de développement de l'assurance de la qualité du cours en ligne, ce qui signifie que peu de révisions sont permises après le début de la session. Certains instructeurs ont déclaré que les travaux du cours prenaient beaucoup de temps à noter, spécialement en ce qui touchait les rétroactions. L'adoption d'une approche axée sur l'amélioration continue a permis à un certain nombre d'instructeurs de réfléchir de façon critique à leur engagement et à adapter leurs travaux en réponse à ces préoccupations. On a sollicité les rétroactions des étudiants concernant le processus d'évaluation dans certains cours, ce qui s'est avéré une précieuse source de données probantes à utiliser pour améliorer le cours. Un instructeur a affirmé :

« Je vais recommander que nous donnions suite aux recommandations des étudiants! Cela signifie leur fournir des exemples et leur permettre de s'exercer à passer à travers le processus de tenue d'un projet de recherche indépendant avant de s'attendre à ce qu'ils soient capables de leur faire tout seuls... la chose la plus importante, c'est de définir ce que nous voulons dire quand nous parlons de pensée critique à nos étudiants et de leur donner le temps de s'exercer à passer à travers le processus de pensée critique (avec beaucoup d'appui) avant de tenir pour acquis qu'ils seront capables de le faire par eux-mêmes » (instructeur M)

Les instructeurs de cours avaient aussi d'excellents conseils à prodiguer à d'autres qui se lancent dans des travaux semblables. Voici quelques exemples de conseils formulés :

Pensez d'abord à ce que voulez que vos étudiants soient en mesure de démontrer à la fin du cours, puis définissez des stratégies pour enseigner ces habiletés cognitives. Ne tenez pas pour acquis que vos façons de faire dans le passé évaluent avec exactitude ce que vous avez l'intention d'évaluer dans votre cours (instructeur K).

Je dirais essayez d'avoir un aperçu avant et après des habiletés pour donner à vos étudiants amplement de rétroactions et pour écouter ce que les étudiants ont à dire, leurs réflexions et leurs buts dans leur propre langage, et de demander conseil (instructeur H).

Commencez par un remue-méninges (avec des collègues, des concepteurs de programme, etc. de confiance) pour déterminer les grands objectifs de la pensée critique pour le cours et comment les atteindre avec le perfectionnement d'habiletés spécifiques (encore une fois, quelles habiletés, à quel niveau, dans quel ordre). Idéalement, ceci devrait exiger de porter attention à la place du cours dans le programme de l'étudiant (la mesure de sa pertinence). Et il faut encourager l'examen de la littérature sur la pensée critique dans le contexte d'une discipline particulière ainsi que ce que font les collègues (instructeur D).

Clarifier quelle habileté cognitive ils aimeraient que leurs étudiants perfectionnent davantage et créer/adopter un cadre de réflexion pour cette habileté particulière pour aider à guider les grilles, pur justifier le travail, les instructions aux étudiants, la notation et les rétroactions (instructeur A).

Ces commentaires démontrent la compréhension des instructeurs de cours du processus d'évaluation des habiletés cognitives et des bienfaits d'un alignement constructif des objectifs d'apprentissage.

### Liens entre les objectifs du cours et les résultats VALUE et HEIghten

Comme susmentionné, les définitions d'habiletés cognitives sont variées et chacune est probablement influencée par les contextes disciplinaires exprimés différemment selon les normes, les pratiques, les contextes et les hypothèses au sein d'un domaine. Des efforts concertés ont été déployés au niveau du projet pour reconnaître ces différences à l'appui de la validité apparente, tout en utilisant toujours un cadre constant pour l'évaluation et la conception des grilles. Lors des multiples séances d'évaluation hebdomadaires des facilitateurs, le gestionnaire de projet et les facilitateurs appliquaient les grilles VALUE aux échantillons d'évaluations de chacune des diverses disciplines en cause. Ces réunions de travail visaient à garantir la compréhension et l'application de concepts d'habiletés cognitives associés à leur domaine d'intérêt. Les facilitateurs de l'évaluation collaboraient avec les instructeurs pour reconcevoir les évaluations de cours et pour cocréer des grilles de cours afin de garantir que les descripteurs de performance reflétaient le contexte disciplinaire et s'alignaient avec les grilles VALUE. Dans la plupart des cas, ce travail constituait un processus itératif qui exigeait de constantes révisions.

La corrélation entre les mesures a été menée en tant que validation croisée des évaluations de cours. Les résultats fondés sur les cours étaient fortement coreliés aux résultats VALUE, mais le degré variait grandement. Les évaluations de cours comprenaient un certain nombre d'aspects qui n'étaient pas nécessairement saisis par la grille VALUE et ces aspects peuvent avoir confondu les résultats. À titre d'exemple, la plupart des grilles de cours incluait des dimensions pour la communication écrite et pour certaines habiletés techniques. Les domaines avec de faibles corrélations peuvent être spécifiques à un ou

deux cours; d'autres recherches seront menées pour examiner cette partie de l'analyse. Les corrélations au niveau des dimensions pourraient aussi être examinées, mais ceci aurait nécessiterait un exercice de mappage pour déterminer quelles dimensions des grilles de cours correspondent à quelles dimensions VALUE. Le fait que plusieurs des grilles de cours combinaient les langages de multiples dimensions compliquait davantage les choses. Des mesures ont été prises afin d'assortir les grilles de cours avec les codes de dimensions VALUE pertinents. À titre d'exemple, une dimension de grille de cours affiliée à plusieurs dimensions VALUE serait indiquée ainsi : « Identification de risque [PC1, PC2] ».

Compte tenu que les cours participants ont déterminé le point de mire de leur développement cognitif dans le cadre du projet, le pourcentage global des cours est étroitement lié à l'évaluation VALUE. À titre d'exemple, la corrélation entre le pourcentage du cours et la moyenne de la RP était  $r(284) = ,603$   $p < ,001$ . La MPC de session est dérivée à partir de toutes les évaluations que les étudiants ont terminées dans tous les cours pour cette session, dont certaines ne mettaient pas beaucoup l'accent sur l'évaluation spécifique d'habiletés cognitives. La diminution de la corrélation entre les résultats VALUE et les autres mesures de cours illustre la validité divergente de l'évaluation. La MPC cumulative constitue un regroupement de toutes les évaluations jusqu'ici au niveau du programme de l'étudiant et on pourrait dire qu'elle représente l'ensemble des habiletés et des efforts. Par conséquent, on peut dire que les habiletés et efforts transférables pris en compte dans le test HEIghten étaient modérément co-reliés à la MPC cumulative.

### **Valeur ajoutée entre l'acquisition des habiletés cognitives entre la première année et la quatrième année**

L'analyse des données quantitatives porte à croire en une importante croissance des habiletés cognitives entre la première année et la dernière année. La grosseur de l'effet Cohen de  $d = ,497$  au niveau de la performance en pensée critique dans le test HEIghten était reflétée dans des gains à deux niveaux le long du niveau médian de la performance en pensée critique de la grille VALUE. Par contre, il importe de reconnaître qu'il y a encore des étudiants en dernière année d'un programme de premier cycle qui se situent au niveau Benchmark. Le niveau de performance des étudiants par programme variait, mais tous les programmes démontraient une amélioration entre la première année et la dernière année (voir l'Annexe 2). L'effort constituait un important facteur et, aux fins de fiabilité des résultats, il y aurait lieu d'accorder une attention continue aux méthodes d'incitation des étudiants utilisées. Des méthodes comme l'octroi de notes de pourcentage pour l'achèvement du test se sont révélées efficaces en ce sens. Cette disposition dépend toutefois de l'engagement de l'instructeur.

### **Renforcement des capacités des instructeurs**

L'apprentissage des étudiants a été évalué dans le cadre de tous les travaux de cours reconçus, sauf deux, en utilisant un type de cadre de qualité comme des descripteurs de grille ou des descripteurs de note holistiques. Cette adaptation par les instructeurs était particulièrement impressionnante puisque 10 des

instructeurs n'avaient aucune expérience avec les grilles avant leur participation au projet de RMEC. De façon générale, les instructeurs ont trouvé les grilles conçues (reçues) très efficaces en vue de communiquer les attentes aux étudiants : « Je suis entièrement converti aux grilles et j'ai déjà recommandé la refonte des évaluations cognitives à un ami d'un autre département » (instructeur H).

Il y a dans chaque établissement des instructeurs qui sont considérés comme des adopteurs précoces et qui déblaient la piste pour les autres en vue de promouvoir le changement. Même si certains membres du groupe pouvaient être perçus comme des adopteurs précoces, avec 25 cours en cause, les participants n'étaient pas les « suspects habituels ». Les résultats du sondage sur l'enseignement révélèrent d'importantes différences au niveau des attitudes face à l'enseignement (voir la Figure 9). Les données de sondage portaient à croire en un important changement positif dans le cas de huit des instructeurs. La plupart de ceux qui ne voyaient aucun changement significatif étaient déjà fortement engagés dans l'enseignement et avaient déjà une attitude très positive à l'égard de l'enseignement. Les deux qui déclaraient une moins bonne attitude à l'égard de l'enseignement citaient des difficultés à mettre en œuvre les nouveaux régimes d'évaluation. Ils avaient apporté d'importants changements à la structure de leurs cours, ce qui avait exigé beaucoup de travail de leur part, et certains de leurs étudiants s'étaient plaints que les travaux étaient trop difficiles.

D'importantes réalisations ou découvertes quant aux approches pédagogiques ou à l'évaluation témoignaient également d'une croissance. Ces transformations étaient généralement précédées de défis. Tous les instructeurs ont fait face à divers défis, mais certains défis semblaient plus grands au niveau des cours de quatrième année, où les travaux devaient être ajustés sur mesure en fonction des sujets sur lesquels se concentraient les étudiants. Les instructeurs ont eu de la misère à appliquer une grille commune aux travaux des étudiants. À titre d'exemple, « la priorité consiste à réexaminer les travaux et à les actualiser en fonction du projet. Un défi qui exigeait une constante attention consistait à offrir aux étudiants de multiples choix de travaux tout en utilisant une grille universelle (instructeur D) ». De la même façon, les instructeurs qui donnaient de gros cours mentionnaient la nécessité de garantir la cohésion au niveau de la notation par plusieurs adjoints à l'enseignement (AE). Certains des cours ont procédé à une calibration de la notation avec leurs AE. L'aide du personnel affecté au projet a facilité l'exercice de calibration de la notation, ce qui a donné lieu à des commentaires positifs comme : « Je crois que cela a été très utile et m'a donné un bon aperçu de pourquoi et comment je peux faire ceci pour tous les travaux et dans tous les cours que j'enseigne » (instructeur N). Compte tenu de l'expérience positive avec les exercices de calibration, une plus grande adoption de ce processus de notation constant pourrait facilement être mise en œuvre.

Le renforcement des capacités des instructeurs était influencé, en partie, par les conversations entre diverses disciplines démontrant une reconnaissance et un respect des similarités et différences. Lors d'une des dernières réunions du réseau d'instructeurs de la RMEC, deux instructeurs ont dit trouver que les grilles étaient trop longues ou fastidieuses pour être pratiques. Ces instructeurs étaient prêts à discuter ouvertement de ces défis avec d'autres instructeurs de la RMEC et ont discuté, en groupe, d'options

positives pour rectifier le problème. Par conséquent, ces deux instructeurs ont décidé de participer de nouveau au projet l'année suivante. Même si le projet offrait de l'aide par l'entremise des facilitateurs de l'évaluation, il s'avérait tout aussi important de mettre sur pied un réseau de soutien par les pairs. Trois des instructeurs de cours ont fait la transition vers un rôle de mentor officiel. Ce rôle a été désigné « encadreur en matière d'évaluation » et les prochaines étapes de recherche incluent un examen pour déterminer en quoi ce rôle peut être maintenu. Il s'agit d'un rôle « d'ami essentiel » qui agit comme encadreur, mentor ou intervenant et qui aide à élaborer du matériel de cours et à examiner le matériel d'évaluation. Le temps investi à titre d'encadreur en matière d'évaluation dans le cadre de ce projet a été officiellement reconnu par le corps professoral comme faisant partie de la charge administrative. En tirant profit de réseaux personnels comme ceux des facilitateurs de l'évaluation et des encadreurs, les instructeurs ont été en mesure de demander des conseils et des rétroactions, d'où un enrichissement de leurs connaissances pédagogiques (Kreber et Cranton, 2000).

### Propagation de l'initiative

Trois réunions du réseau ont été organisées, l'une à l'automne, l'une en hiver et l'autre sous forme de mini-symposium d'un jour au printemps présidé par le conférencier invité Randy Bass, Ph. D.<sup>1</sup>. Ces réunions avaient pour but de partager des expériences et de créer une communauté de la pratique. Dans le cadre du projet de RMEC, on s'attendait à ce que les instructeurs participants fassent part de leurs initiatives et de leurs expériences au niveau du projet. Ce partage d'idées et d'expériences constituait le but premier de la réunion, à l'instar des buts de communautés de la pratique (Wenger, 1998). Les présentations suscitaient des conversations au sein de tous les groupes disciplinaires. Avant la réunion, les instructeurs croyaient généralement que les défis auxquels ils faisaient face étaient propres à leur discipline. Suivant les réunions et les présentations, ils ont constaté que des défis comme les hypothèses concernant l'aptitude des étudiants à intégrer des perspectives au-delà de leur discipline étaient relativement universelles. Les conversations dépassaient souvent les confins des réunions et avaient même des incidences sur les départements. À titre d'exemple, « la participation au projet a engendré une discussion élargie sur la nature des évaluations dans les laboratoires de cours. Nous n'avons pas encore résolu la question, mais la discussion est riche et importante » (instructeur B).

Une stratégie utilisée pour encourager la participation des instructeurs consistait à les inclure en tant que chercheurs. Sur demande, le gestionnaire de la recherche apportait des modifications à l'éthique pour

---

1 Randy Bass est vice-recteur adjoint, Teaching and Learning Initiatives, et professeur d'anglais à l'Université Georgetown. Il a été « Principal's Distinguished Visitor for the Enhancement of Learning » et le thème de son discours liminaire était « Assessment Matters: Integrative Learning in a Dis-integrative Era ».

décrire des concepts de recherche individuels, ce qui favorisait encore plus des enquêtes savantes et des articles par coauteurs, au besoin. L'initiative a fourni une riche occasion de perfectionnement professionnel à plusieurs instructeurs dont certains ont mené leurs propres enquêtes savantes (voir les références en ce sens à l'Annexe 4). Il y a eu sept présentations de conférence de la part des instructeurs de cours directement liées à leur reconception de l'évaluation. Beaucoup des instructeurs de cours continuent d'apporter des améliorations à leurs cours. Certains collaboraient déjà au niveau de leurs efforts de reconception; les instructeurs en sciences infirmières qui ont instauré des travaux de lectures essentielles au niveau de la première année et de la quatrième année ont mis au point des grilles de cours similaires pour examiner l'évolution de la pensée critique dans le cadre du programme de sciences infirmières. Il y a eu une réponse positive face aux défis d'avenir à relever. À titre d'exemple, un instructeur a dit « nous avons l'intention de fournir aux étudiants plus d'exemples du type de travail auquel nous nous attendons de leur part et aussi de les inciter à passer plus de temps à évaluer à fond d'autres études (par exemple, publiées et par des pairs) pour mieux comprendre le processus à l'avance » (instructeur M).

L'un des principaux extrants de la recherche au niveau « macro » était l'élaboration d'un guide institutionnel pour l'évaluation des habiletés cognitives (Simper, Fostaty Young, Frank et Scott, 2018). Le guide a été rédigé sous forme de manuel pour aider le projet à faire la transition d'une recherche subventionnée à une pratique fondée sur les données probantes permanente et viable.

## Conclusions

La refonte des méthodes d'évaluation cognitive s'est avérée un projet institutionnel fondé sur un réseau qui mettait l'accent sur le développement et l'évaluation d'habiletés cognitives (pensée critique, pensée créative et résolution de problème). Le projet visait deux grands objectifs : aider les instructeurs à mettre au point des évaluations axées sur les cours et évaluer le développement des habiletés cognitives des étudiants à l'aide de mesures normalisées. Grâce à l'instauration d'un réseau, 24 instructeurs de 15 départements ont été appuyés pour les aider à aligner de façon constructivement l'évaluation des habiletés cognitives, en utilisant efficacement des mesures de qualité. Fait intéressant, 40 % des instructeurs participants ont utilisé pour la première fois des grilles pour évaluer les artéfacts des étudiants dans leurs cours. Une analyse qualitative des rapports narratifs sur les cours a été menée pour mettre au point des indicateurs de croissance et de renforcement des capacités. Avec des résultats prometteurs démontrés au niveau de plusieurs de ces grilles. En outre, les instructeurs ont décrit la signification du réseau, y compris l'aide des facilitateurs de l'évaluation et les réseaux de soutien par les pairs. D'autres améliorations s'avèreraient nécessaires au niveau de l'adoption à plus grande échelle de procédés formels pour la notation cohésive des cours. Les données des sondages auprès des instructeurs ont démontré les défis inhérents liés à l'instauration d'un changement de culture au niveau des pratiques d'évaluation, mais aussi le degré de

renforcement des capacités atteint. Il est recommandé que ces mesures fassent l'objet d'un suivi à l'avenir pour déterminer les effets à long terme sur le projet.

Pour évaluer l'apprentissage des étudiants, trois modes d'évaluation ont été utilisés (grilles de cours, grilles VALUE et test HEIghten). L'analyse des données d'évaluation des étudiants a démontré des corrélations significatives entre les notes axées sur le cours et notes des grilles VALUE et a révélé d'importants gains au niveau de la pensée critique et de la résolution de problème entre la première année et la dernière année. Ces données ont des incidences au niveau institutionnel, faisant valoir la performance accrue des étudiants sur le plan des habiletés cognitives, influençant les décisions pédagogiques et les améliorations de cours. La notation avec les grilles VALUE a démontré qu'entre 20 % et 40 % des étudiants de quatrième année affichaient des niveaux de performance Benchmark 2 au mois au niveau des dimensions associées à la résolution de problèmes et à la pensée critique, et la notation des tests HEIghten indiquait que 13 % des étudiants de quatrième année performaient au niveau développement tandis que 59 % performaient au niveau compétent et 28 % au niveau avancé. Des mesures d'évaluation comme celles-ci donnent l'occasion de chercher des cohortes d'étudiants qui se situent en deçà des niveaux des études de premier cycle. La prochaine itération du projet de RMEC continuera de mobiliser et d'appuyer un réseau d'évaluations institutionnelles visant à développer et évaluer ces habiletés, mettant l'accent sur le rôle de facilitateurs de l'évaluation et d'encadreurs associés à une discipline spécifique.

## Références

- Ashford-Rowe, K., Herrington, J. et Brown, C. (2013). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education*, 39(2), 205–222.  
<https://doi.org/10.1080/02602938.2013.819566>
- Banta, T. W. et Palomba, C. A. (2014). *Assessment essentials: Planning, implementing, and improving assessment in higher education*. John Wiley & Sons.  
[https://books.google.ca/books?hl=en&lr=&id=rO\\_sBQAAQBAJ&oi=fnd&pg=PA13&dq=Assessment+Essentials:+Planning,+Implementing+and+Improving+Assessment+in+Higher+Education&ots=NiNDO4gSAq&sig=I\\_DgF-oXEE\\_dLOSzVfFX0w2pX-s](https://books.google.ca/books?hl=en&lr=&id=rO_sBQAAQBAJ&oi=fnd&pg=PA13&dq=Assessment+Essentials:+Planning,+Implementing+and+Improving+Assessment+in+Higher+Education&ots=NiNDO4gSAq&sig=I_DgF-oXEE_dLOSzVfFX0w2pX-s)
- Bernstein, D. et Greenhoot, A. F. (2014). Team-designed Improvement of writing and critical thinking in large undergraduate courses. *Teaching and Learning Inquiry: The ISSOTL Journal*, 2(1), 39–61.  
<https://doi.org/10.1353/iss.2014.0008>
- Biggs, J. (2014). Constructive alignment in university teaching. *HERDSA Review of Higher Education*, 1(1), 5–22.
- Boud, D. et Dochy, F. (2010). *Assessment 2020. Seven propositions for assessment reform in higher education* (pp. 1–4).
- Chasteen, S. V. et Code, W. (2018). *The science education initiative handbook. A practical guide to fostering change in university courses and faculty by embedding discipline-based education specialists within departments*. Boulder: The University of Colorado Boulder.
- Chun, M. (2010). Taking teaching to (performance) task: Linking pedagogical and assessment practices. *Change: The Magazine of Higher Learning*, 42(2), 22–29.
- Corbo, J. C., Reinholz, D. L., Dancy, M. H., Deetz, S. et Finkelstein, N. (2014). Sustainable Change: A Model for Transforming Departmental Culture to Support STEM Education Innovation. *ArXiv Preprint ArXiv:1412.3034*. <http://arxiv.org/abs/1412.3034>
- Council for Aid to Education. (2016). *CLA+ National Results, 2015-16* (pp. 1–21). New York, NY.  
[http://cae.org/images/uploads/pdf/CLA\\_National\\_Results\\_2015-16.pdf](http://cae.org/images/uploads/pdf/CLA_National_Results_2015-16.pdf)
- Dawson, P. (2017). Assessment rubrics: Towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*, 42(3), 347–360.
- Drezek McConnell, K. et Rhodes, T. L. (2017). *On Solid Ground, Value Report 2017* (pp. 1–52).
- Frank, B. M., Simper, N. et Kaupp, J. A. (2016). How we know they're learning: Comparing approaches to longitudinal assessment of transferable learning outcomes. In *ASEE Annual Conference and Exposition, Conference Proceedings* (Vol. 2016-June).
- González, J. et Wagenaar, R. (2003). *Tuning educational structures in Europe*. University of Deusto Final report. Phase one. Bilbao.

[http://eua.be/eua/jsp/en/upload/TUNING\\_Announcement\\_Closing\\_Conference.1084282515011.pdf](http://eua.be/eua/jsp/en/upload/TUNING_Announcement_Closing_Conference.1084282515011.pdf)

- Hannah, S. T. et Lester, P. B. (2009). A multilevel approach to building and leading learning organizations. *The Leadership Quarterly*, 20(1), 34–48. <https://doi.org/10.1016/j.leaqua.2008.11.003>
- Hanson, M. (2001). Institutional Theory and Educational Change. *Educational Administration Quarterly*, 37(5), 637–661. <https://doi.org/10.1177/00131610121969451>
- Hathcoat, J. D. et Penn, J. D. (2012). Generalizability of student writing across multiple tasks: A challenge for authentic assessment. *Research & Practice in Assessment*, 7(2), 16–28.
- Henderson, C., Beach, A. et Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature. *Journal of Research in Science Teaching*, 48(8), 952–984. <https://doi.org/10.1002/tea.20439>
- Hutchings, P., Huber, M. T. et Ciccone, A. (2011). *The scholarship of teaching and learning reconsidered: Institutional integration and impact* (Vol. 21). John Wiley & Sons. Extrait de <https://books.google.ca/books?hl=en&lr=&id=GPbdPMVOyAIC&oi=fnd&pg=PT6&dq=Hutchings,+Huber,+and+Ciccone,+2011&ots=xiottWlwMO&sig=FtBxtufGR5dCIC4oKMmRj4Ykfn0>
- Jankowski, N., Hutchings, P., Ewell, P., Kinzie, J. et Kuh, G. (2013). The degree qualifications profile: What it is and why we need it now. *Change: The Magazine of Higher Learning*, 45(6), 6–15.
- Jonassen, D. H. (1997). Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology Research and Development*, 45(1), 65–94. <https://doi.org/10.1007/BF02299613>
- Jonsson, A. et Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144.
- Kezar, A. (2013). "Theories of change: Change agent guides." In *How colleges change understanding, leading, and enacting change* (1st Edition, pp. 20–40). Taylor & Francis Group. <https://www.taylorfrancis.com/>
- Kreber, C. et Cranton, P. A. (2000). Exploring the scholarship of teaching. *The Journal of Higher Education*, 71(4), 476–495. doi:10.1080/00221546.2000.11778846
- Kinzie, J. (2010). Perspectives from campus leaders on the current state of student learning outcomes assessment. *Assessment Update*, 22(5), 14–15.
- Koretz, D. M. (2008). *Measuring up*. Harvard University Press. <https://books.google.ca/books?hl=en&lr=&id=Nw-qBPCTcYYC&oi=fnd&pg=PA1&dq=Koretz+2008+high+stakes+testing&ots=k8YvOqRftq&sig=wljSkJZC Kvh-iumQ53Gxc-WK73I>

- Liu, O. L., Bridgeman, B., et Adler, R. M. (2012). Measuring learning outcomes in higher education: Motivation matters. *Educational Researcher*, 41(9), 352–362.  
<https://doi.org/10.3102/0013189X12459679>
- Madaus, G. F. et Clarke, M. (2001). The adverse impact of high stakes testing on minority students: Evidence from 100 years of test data. <http://eric.ed.gov/?id=ED450183>
- Mathers, C. E., Finney, S. J. et Hathcoat, J. D. (2018). Student learning in higher education: A longitudinal analysis and faculty discussion. *Assessment & Evaluation in Higher Education*.  
<https://srhe.tandfonline.com/doi/abs/10.1080/02602938.2018.1443202>
- Miller, E. R., Fairweather, J. S., Slakey, L., Smith, T. et King, T. (2017). Catalyzing institutional transformation: Insights from the AAU STEM initiative. *Change: The Magazine of Higher Learning*, 49(5), 36–45.
- Popham, W. J. (1997). What's wrong — and what's right — with rubrics. *Educational Leadership*, 55, 72–75.
- Popham, W. J. (1999). *Classroom assessment: What teachers need to know*. ERIC.
- Purmton, E. F. et Alexander, E. R. (2013). Obtaining faculty motivation and "Buy In" to a major program change: A case study in assessment. *Business Education Innovation Journal*, 5(1), 53–57.
- Rhodes, T. L. (2011). Emerging evidence on using rubrics. *Peer Review*, 13(4/1), 4–5.
- Rhodes, T. L. et Finley, A. P. (2013). *Using the VALUE rubrics for improvement of learning and authentic assessment*. Washington, DC: Association of American Colleges and Universities.
- Roxa, T. et Martensson, K. (2015). Microcultures and informal learning: A heuristic guiding analysis of conditions for informal learning in local higher education workplaces. *International Journal for Academic Development*, 20(2), 193–205. <https://doi.org/10.1080/1360144X.2015.1029929>
- Roxa, T., Martenson, K., Alveteg, M., (2011). Understanding and influencing teaching and learning cultures at university: A network approach. *Higher Education*, 62(1), 99–111.  
[doi:10.1007/s10734-010-9368-9](https://doi.org/10.1007/s10734-010-9368-9)
- Simper, N., Fostaty Young, S., Frank, B. et Scott, J. (2018). Guide for Institutional Assessment of Cognitive Skills. Université Queen's.  
<http://www.queensu.ca/qloa/sites/webpublish.queensu.ca/qloa/www/files/files/laRMEC%20handbook-Web.pdf>
- Simper, N., Frank, B., Kaupp, J., Mulligan, N. et Scott, J. (2018). Comparison of standardized assessment methods: Logistics, costs, incentives and use of data. *Assessment and Evaluation in Higher Education*, 44(6), 1–16.
- Simper, N., Frank, B., Scott, J. et Kaupp, J. (2018). *Learning Outcomes Assessment and Program Improvement at Queen's University* (pp. 1–55). Ontario: Toronto: Higher Education Quality Council of Ontario.  
[http://www.heqco.ca/SiteCollectionDocuments/Formatted%20Queens\\_LOAC\\_report.pdf](http://www.heqco.ca/SiteCollectionDocuments/Formatted%20Queens_LOAC_report.pdf)

- Stein, B. et Haynes, A. (2011). Engaging faculty in the assessment and improvement of students' critical thinking using the critical thinking assessment test. *Change: The Magazine of Higher Learning*, 43(2), 44–49.
- Tremblay, K. (2013). OECD Assessment of Higher Education Learning Outcomes (AHELO). In *Modeling and measuring competencies in higher education* (pp. 113–126). Springer.  
[http://link.springer.com/chapter/10.1007/978-94-6091-867-4\\_8](http://link.springer.com/chapter/10.1007/978-94-6091-867-4_8)
- Tremblay, K., Lalancette, D., & Roseveare, D. (2012). *Assessment of Higher Education Learning Outcomes feasibility study report: Volume 1 - design and implementation*. OECD.
- Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge, U.K;New York, N.Y.;: Cambridge University Press.
- Wenger, E. (2000). Communities of practice and social learning systems. *Organization*, 7(2), 225–246.
- Wenger, E. (2011, October). *Communities of practice: A brief introduction*. Presented at the STEP Leadership Workshop, University of Oregon. <http://scholarsbank.uoregon.edu/xmlui/handle/1794/11736>
- Wenger, E., McDermott, R. A. et Snyder, W. (2002). *Cultivating communities of practice: A guide to managing knowledge*. Harvard Business Press.  
[https://books.google.ca/books?hl=en&lr=&id=m1xZuNq9RygC&oi=fnd&pg=PR9&dq=communities+of+practice+wenger&ots=ZUadhlae9\\_&sig=KCT1b83mzJB-dm4O0zAO26DunRg](https://books.google.ca/books?hl=en&lr=&id=m1xZuNq9RygC&oi=fnd&pg=PR9&dq=communities+of+practice+wenger&ots=ZUadhlae9_&sig=KCT1b83mzJB-dm4O0zAO26DunRg)
- Wieman, C. (2007). Why not try a scientific approach to science education? *Change: The Magazine of Higher Learning*, 39(5), 9–15.
- Wieman, C., Deslauriers, L. et Gilley, B. (2013). Use of research-based instructional strategies: How to avoid faculty quitting. *Physical Review Special Topics-Physics Education Research*, 9(2), 023102.
- Wiggins, G. P. et McTighe, J. (2005). *Understanding by design*. Alexandria: Association for Supervision and Curriculum Development.

