



Conseil ontarien
de la qualité de
l'enseignement
supérieur

Un organisme du gouvernement de l'Ontario

Élaboration de rubriques d'analyse pour mesurer les compétences

Gayle Lesmond, Susan McCahan
et David Beach, Université de Toronto



Publié par le

Conseil ontarien de la qualité de l'enseignement supérieur

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

Téléphone : 416 212-3893
Télécopieur : 416 212-3899
Site Web : www.heqco.ca
Courriel : info@heqco.ca

Citer ce document comme suit :

Lesmond, G., McCahan, S. et Beach, D. (2017), *Élaboration de rubriques d'analyse pour mesurer les compétences*, 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, 2017.

Synthèse

Le présent document décrit l'élaboration de rubriques d'analyse en vue d'un projet de mesure des compétences. Il a pour objet de décrire le processus d'élaboration d'un ensemble de rubriques générales d'analyse pour mesurer les compétences liées à la conception, à la communication et au travail en équipe, ainsi que d'un ensemble de résultats et d'indicateurs servant à mesurer l'analyse de problèmes et l'investigation.

Le travail d'élaboration des rubriques est structuré en trois grandes phases. Dans la première phase, dite de planification, nous avons procédé à une recension documentaire pour en arriver à une liste exhaustive de résultats de l'apprentissage dans les cinq domaines de compétence à l'étude. Nous avons également compilé une liste d'indicateurs, lesquels consistent en des résultats de l'apprentissage particuliers et mesurables. Puis nous avons épuré la liste exhaustive de résultats de l'apprentissage et d'indicateurs obtenue; pour ce faire, nous avons éliminé la redondance entre les systèmes, comblé les lacunes de contenu, et rassemblé les indicateurs en des catégories de résultats de l'apprentissage communs.

À la deuxième phase, nous avons ébauché les descripteurs de rubrique de la conception, de la communication et du travail en équipe puis modifié ceux-ci après avoir consulté les enseignants et les administrateurs des départements. Nous avons validé les résultats et indicateurs de l'analyse de problèmes et de l'investigation qui figurent dans le présent document au moyen d'une méthode de Delphes systématique. Le travail à un ensemble de descripteurs des indicateurs de l'analyse de problèmes et de l'investigation est actuellement en cours.

Dans la troisième et dernière phase, nous avons mis à l'essai les rubriques de conception, de communication et de travail en équipe. En ce qui touche la conception et la communication, nous avons mené des séances d'essais miroirs auprès d'étudiants de cycles supérieurs ayant l'expérience de la notation (les évaluateurs). À ces derniers en particulier, nous avons demandé d'évaluer à l'aide des rubriques certains échantillons de travaux présentés par des étudiants puis de donner une rétroaction au moyen des groupes de réflexion. La mise à l'essai de la rubrique du travail en équipe a consisté en une quasi-instauration auprès des adjoints à l'enseignement (AE) et en des séances de réflexion à voix haute auprès des enseignants (les experts). Cette mise à l'essai avait pour objectif de faire progresser la validation des résultats, des indicateurs et des descripteurs de rubrique, puis d'obtenir une rétroaction sur les moyens potentiels de les améliorer.

Une fois la mise à l'essai terminée, nous avons procédé à des analyses de données détaillées. L'analyse des discussions des groupes de réflexion avec les évaluateurs nous a révélé les constatations suivantes :

- a) le contenu des rubriques manquait de clarté;
- b) les niveaux des rubriques n'étaient pas suffisamment différenciés;
- c) des critères importants étaient omis dans les rubriques.

Après l'analyse des données issues des groupes de réflexion, nous avons effectué une révision poussée des rubriques pour tenir compte de la rétroaction présentée par les participants à la recherche. Il en a résulté un

ensemble de rubriques révisées qui peuvent s'appliquer à un vaste ensemble de cours et de types d'évaluation.

Le travail d'élaboration des rubriques nous a révélé trois leçons fondamentales :

- a) La méthode de Delphes est une solution de rechange judicieuse à la consultation individuelle dans ce type de travail.
- b) La formation des évaluateurs est cruciale.
- c) Les facultés du travail en équipe sont tout particulièrement difficiles à mesurer, notamment parce qu'elles nécessitent l'observation et l'interprétation du comportement.

La version définitive des rubriques de communication, de conception et de travail en équipe figure à l'annexe A. Les ensembles d'indicateurs en lien avec l'analyse de problèmes et l'investigation font partie de l'annexe B.

Table des matières

1. Introduction	9
2. Première phase : Planification	10
2.1 Terminologie.....	10
2.2 Ensemble de compétences.....	12
2.3 Recension documentaire	13
3. Deuxième phase : Construction	17
3.1 Élaboration des descripteurs de rubriques.....	17
3.2 Consultations auprès d'experts en contenu.....	21
4. Troisième phase : Essais et analyse.....	23
4.1 Essais miroirs	23
4.2 Analyse des données	25
5. Résultats et discussion	29
6. Conclusion	32
Bibliographie.....	34

Liste des graphiques

Graphique 1 : Illustration de la terminologie employée dans la présente étude.....	11
Graphique 2 : Caractérisation de la conception et de l'analyse de problèmes/résolution de problèmes par certitude de contraintes et de buts	16
Graphique 3 : Exemple de fiche des données de l'énoncé « Présenter une introduction claire qui énonce le thème et donne un aperçu de la matière »	27
Graphique 4 : Exemple de fiche de données de l'énoncé « Exposer avec exactitude le problème de conception technique puis résumer les détails fondamentaux (interpréter l'énoncé du problème si celui-ci est fourni) »	28
Graphique 5 : Exemple de fiche des données de l'énoncé « Relater avec exactitude l'apport des autres membres de l'équipe à l'activité de l'équipe »	28

Liste des tableaux

Tableau 1 : Structure de base d'une rubrique d'analyse	17
Tableau 2 : Niveaux de rendement.....	18
Tableau 3 : Extrait de la première version de la rubrique Communication.....	19
Tableau 4 : Modifications proposées à l'extrait de la rubrique Communication	19
Tableau 5 : Exemple d'une ligne de rubrique multidimensionnelle	20
Tableau 6 : Modifications proposées à une ligne de rubrique multidimensionnelle	20
Tableau 7 : Experts de l'investigation par programme.....	22
Tableau 8 : Experts de l'analyse de problèmes par programme	22
Tableau 9 : Mise à l'essai des rubriques par programme	25

1. Introduction

En 2012, la version révisée du document *Quality Assurance Framework* produit par l'Ontario Universities Council on Quality Assurance (OUCQA) a servi de fondement à l'assurance de la qualité des programmes d'études universitaires pour appuyer la mobilité universitaire des diplômés à l'échelle nationale et internationale [Ontario Universities Council on Quality Assurance (2012)]. La même année, des initiatives connexes ont porté sur le besoin en exigences internationales relativement aux modes et normes universitaires de qualification [Centre d'information canadien sur les diplômes internationaux (2012)]. Dans la foulée des Lignes directrices de l'UNESCO et de la Convention de Lisbonne sur la reconnaissance [Division de l'enseignement supérieur (2005); Conseil de l'Europe (1999)], on a compris que le besoin en assurance de la qualité visant les diplômés universitaires est d'une importance primordiale pour garantir la responsabilisation, la mobilité et la reconnaissance des programmes menant à un grade universitaire, tant au Canada qu'à l'étranger. Des questions et préoccupations du même ordre ont également convaincu en 2008 des ministres de l'éducation issus de pays membres de l'OCDE à envisager la mesure et l'évaluation de la qualité de l'enseignement supérieur. Les discussions à ce sujet ont débouché sur un intérêt accru envers l'élaboration et l'utilisation d'outils d'évaluation des résultats de l'apprentissage universels et généraux au niveau postsecondaire.

À la suite de telles initiatives, le Conseil ontarien de la qualité de l'enseignement supérieur (COQES) a lancé le projet de Consortium sur l'évaluation des résultats d'apprentissage (CERA) dans l'optique d'élaborer des outils de mesure des résultats de l'apprentissage liés aux facultés cognitives jugées pertinentes pour les diplômés des établissements d'enseignement supérieur. Dans le cadre de cette initiative, la FASE (faculté des sciences appliquées et du génie) de l'Université de Toronto a prévu l'élaboration de rubriques valides misant sur les travaux de mesure des résultats de l'apprentissage réalisés par l'instauration du système des qualités requises des diplômés du Bureau canadien d'agrément des programmes de génie (BCAPG). La question de recherche essentielle ayant motivé ce projet était la suivante : *Pouvons-nous créer des éléments valides et fiables d'une rubrique d'analyse qui donnent des renseignements sur les résultats de l'apprentissage dans les domaines choisis?* Le but consistait donc à élaborer un ensemble de rubriques universelles d'analyse qui permettraient de mesurer les résultats de l'apprentissage relatifs à la conception, à la communication, au travail en équipe, à l'analyse de problèmes (utilisée de façon interchangeable avec la résolution de problèmes) et l'investigation. Les rubriques devaient être universelles pour servir dans divers contextes (cours, programmes, disciplines, etc.) et, préférablement, fournir à tous les intervenants pertinents, y compris les étudiants, les enseignants et les administrateurs des universités, des renseignements sur l'apprentissage.

Nous avons structuré globalement le travail d'élaboration des rubriques en trois phases. Dans la première phase (planification), les étapes suivantes sont intervenues :

- L'élaboration d'un cadre commun de définition de la terminologie.
- La sélection et la définition des domaines fondamentaux de mesure.

- Une recension des travaux actuels pour en arriver à une liste exhaustive de facultés particulières et mesurables pour chaque domaine à l'étude.

La deuxième phase (construction des rubriques) a consisté en l'élaboration de rubriques en version provisoire, et en des consultations auprès d'experts pour réviser les composantes des rubriques. Nous avons mené la troisième et dernière phase (mise à l'essai) auprès de groupes d'évaluateurs pour voir à ce que leur interprétation des rubriques concorde avec le sens voulu et obtenir une rétroaction supplémentaire quant à la modification des rubriques. Par conséquent, la structure du rapport décrite dans les pages suivantes repose sur les trois grandes phases du projet. Les sections finales présentent en résumé les leçons tirées et fournissent des recommandations en vue d'un déploiement judicieux des rubriques.

2. Première phase : Planification

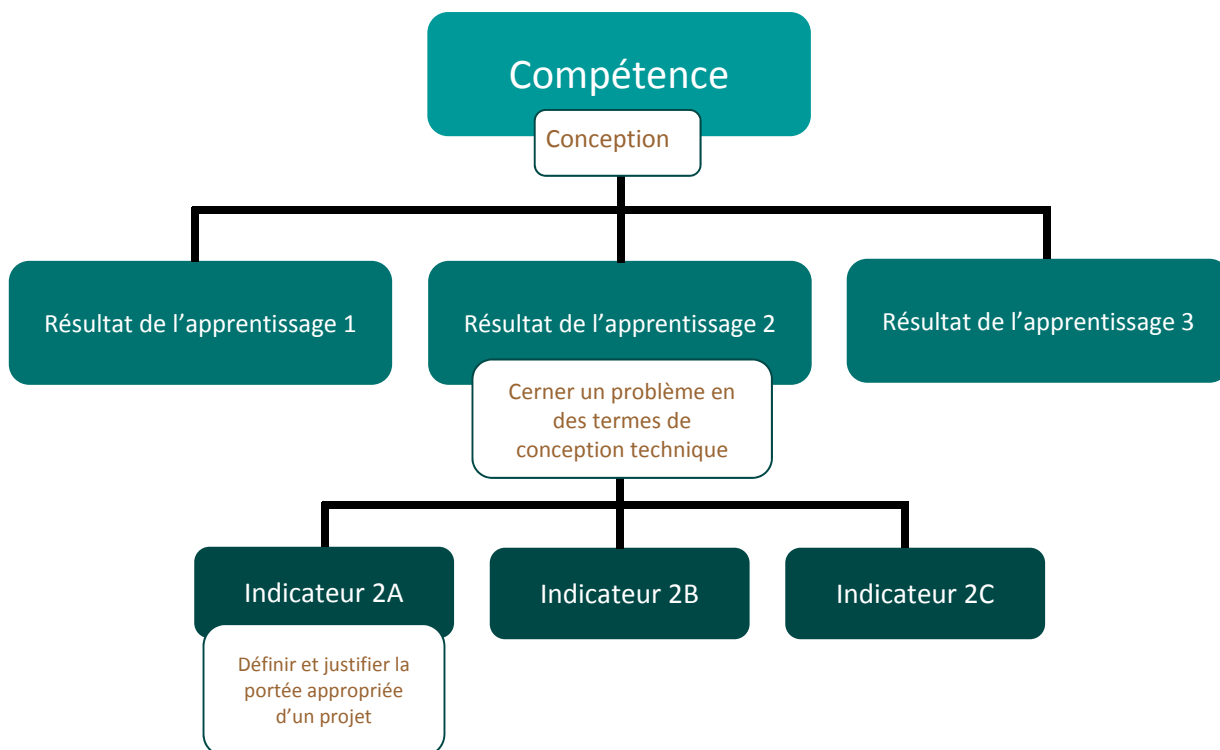
2.1 Terminologie

L'un des premiers obstacles auxquels nous avons fait face dans la recension documentaire découlait des différences terminologiques liées à la mesure des résultats de l'apprentissage. La première étape de l'élaboration des rubriques a donc consisté en la création d'un cadre commun de référence afin de décrire les buts du processus pédagogique. Par souci de clarté, nous avons différencié trois niveaux de capacités attestées. Le graphique 1 donne une illustration de ce cadre de référence.

- Le concept de « compétences » servira à faire allusion au niveau le plus global de désignation ou de catégorisation, et les compétences sont habituellement mesurées à l'échelle des programmes. Dans la présente recherche, nous avons étudié cinq domaines de compétences : la conception; la communication; le travail en équipe; l'analyse de problèmes; l'investigation.
- Le concept de « résultats de l'apprentissage » servira à faire allusion au niveau intermédiaire de catégorisation, et à préciser les capacités cognitives qui englobent les compétences. Les résultats de l'apprentissage décrivent habituellement l'apprentissage qui se déroule au niveau d'un cours. Par exemple, l'énoncé « Cerner un problème en des termes de conception technique » pourrait constituer un résultat de l'apprentissage en fonction de la compétence de conception. Les résultats de l'apprentissage sont exprimés à la fois de façon active et précise : autrement dit, leur expression débute souvent comme suit : « Les étudiants manifesteront la capacité de... ». Bien que cette exigence ne soit pas uniforme dans l'essentiel de la documentation recensée, elle s'inscrit dans un consensus largement répandu, selon lequel le travail relatif aux résultats de l'apprentissage pourrait être mis en comparaison et en phase avec la plupart des projets actuels ou futurs ayant trait aux résultats de l'apprentissage.
- Le concept d'« indicateurs » servira à faire allusion au niveau le plus détaillé de l'apprentissage, lequel permet de préciser des actions ou résultats particuliers qui sont mesurables et quantifiables et révèlent certains résultats de l'apprentissage habituellement liés à un devoir ou à un module d'apprentissage

d'un cours; par exemple, les énoncés « Définir et justifier la portée appropriée d'un projet » et « Définir les besoins appropriés et mesurables d'évaluation de solutions potentielles de conception technique » peuvent correspondre à deux indicateurs aux termes du résultat de l'apprentissage décrit ci-dessus. Un indicateur doit être suffisamment précis pour qu'il puisse être évalué au moyen d'une question unique ou d'un aspect d'un devoir. La précision des indicateurs répond aux mêmes exigences de formulation que la précision des résultats de l'apprentissage. De cette façon, il est possible d'évaluer la fiabilité des indicateurs et de déterminer la précision de mesures des indicateurs.

Graphique 1 : Illustration de la terminologie employée dans la présente étude



Habituellement, la plupart des auteurs qualifient les compétences de « résultats de l'apprentissage ». Tout au long du présent document, nous emploierons de façon constante la terminologie décrite ci-dessus. Dans les cas où nous devons faire allusion à la terminologie utilisée par autrui, nous énoncerons de façon explicite notre terminologie équivalente, laquelle remplacera la terminologie employée à l'origine par les auteurs, s'il y a lieu.

2.2 Ensemble de compétences

Nous avons défini au sens large un ensemble de compétences avant de procéder à une recension documentaire. Au départ, notre enquête s'est appuyée sur les cinq compétences recensées dans la proposition de recherche :

- la conception;
- la communication;
- le travail en équipe;
- l'analyse de problèmes (ou la résolution de problèmes);
- l'investigation.

Nous avons défini de façon exhaustive ces compétences après la recension documentaire, laquelle a permis d'approfondir les définitions proposées par d'autres chercheurs pour en arriver à un consensus. Histoire d'orienter la recension documentaire, nous avons employé les définitions provisoires suivantes :

- La conception désigne le processus d'arriver à « [...] la spécification d'un objet, manifesté par un agent, dans l'optique d'atteindre des buts, au sein d'un environnement particulier, à l'aide d'un ensemble de composantes primitives, conformément à un ensemble d'exigences, sous réserve de contraintes » [d'après Ralph et Wand (2009), p. 108]. Le résultat d'un travail de conception consiste en une représentation, un plan ou un procédé en vue de construire un objet ou un système. L'activité de conception peut déboucher, par exemple, sur une œuvre d'art, le bleu d'un immeuble, une maquette, un livre (plus précisément, un récit) ou un outil de sondage.
- La communication désigne les activités où il y a échange d'information d'une partie à l'autre.
- Le travail en équipe désigne les activités menées ou exécutées par des parties intéressées qui consistent en plus d'un agent, et où chaque partie intéressée tend, à un certain niveau, vers un objectif commun.
- L'analyse de problèmes (utilisée de façon interchangeable avec la résolution de problèmes) désigne un processus de définition et d'exécution d'une voie de solution vers l'atteinte d'un but, laquelle est habituellement jonchée de contraintes. Fréquemment, les contraintes sont telles que le nombre de solutions possibles est restreint. Le résultat de l'analyse de problèmes consiste en un produit mûri (p. ex., un nombre, un algorithme, un plan d'action ou un jugement décisif, un processus, une structure ou un lien) qui résulte de la synthèse du savoir actuel. Il est possible que l'analyse de problèmes débouche sur un algorithme de classement des données, la solution bien définie à un enjeu expérimental ou une équation (notamment d'ordre mathématique).
- L'investigation désigne le processus perçu en gros comme la « recherche » sur le plan scientifique ou technique, à savoir les activités ayant pour objet d'élargir l'ensemble du savoir et dans lesquelles interviennent l'hypothèse, l'expérimentation et la conclusion.

Ces cinq compétences s'inspirent des qualités requises des diplômés du BCAPG [Bureau canadien d'agrément des programmes de génie (2014)] et peuvent être mises en correspondance avec les UDLE

(attentes en matière d'études de premier cycle) de l'OCAV (conseil des vice-recteurs à l'enseignement de l'Ontario) [Ontario Universities Council on Quality Assurance (2012); Council of Ontario Universities (2011)].

2.3 Recension documentaire

Pour définir les résultats et indicateurs de compétences en particulier, nous avons dû non seulement enquêter sur les démarches antérieures de mesure des résultats de l'apprentissage, mais approfondir la définition générale du concept de compétences. Pour restreindre la portée de la recension documentaire requise, nous avons posé les questions suivantes relativement aux thèmes des résultats, de la mesure et des rubriques de l'apprentissage :

- Quels autres établissements d'enseignement ont procédé à la mesure des compétences ou des résultats de l'apprentissage?
 - Comment les autres établissements d'enseignement définissent-ils ceux-ci? Qu'est-ce qui est considéré comme un « ensemble étendu »?
 - En quoi les compétences ou résultats proposés se comparent-ils à l'ensemble de compétences ou de résultats compilés par l'équipe?
 - Existe-t-il des études de cas?

- Quelle documentation universitaire porte sur la définition ou l'évaluation de compétences semblables aux compétences provisoires adoptées en gros du BCAPG ou des UDLE de l'OCAV?
 - Quel est le but du travail?
 - Quels sont les concepts utilisés (p. ex., les modèles, les théories, les définitions, etc.)?
 - Quels postulats particuliers sont avancés (p. ex., l'âge, le niveau de scolarité, la discipline)?
 - Quelles données ou expérimentations servent à appuyer le travail?
 - Quelles sont les conclusions du travail?
 - Quelles sont les retombées du travail?
 - Quelles caractéristiques générales s'appliquent, peu importe la discipline?
 - En quoi le travail se compare-t-il au cadre du GAC (comité des qualités requises des diplômés) à l'Université de Toronto?
 - Existe-t-il des rubriques? Si oui, ont-elles fait l'objet d'un essai de validité?

L'objectif de la recension documentaire consistait à créer une liste d'indicateurs et de résultats de l'apprentissage dans les cinq domaines de compétence à l'étude. La liste d'indicateurs et de résultats de l'apprentissage n'avait pas pour but d'en arriver à une liste exhaustive à utiliser exclusivement pour chaque évaluation ou cours, mais plutôt de servir en tant qu'ensemble d'indicateurs à partir duquel il serait possible de choisir et de compiler ceux qui sont pertinents à la matière des cours, aux résultats de l'apprentissage et à la stratégie pédagogique de l'enseignant pour créer une rubrique adaptée à un devoir donné.

Les sources principales employées dans l'élaboration de cette première liste d'indicateurs et de résultats figurent ci-dessous. Ces sources correspondent à des systèmes de résultats de l'apprentissage fréquemment employés dans la documentation, comme les rubriques VALUE, la NILOA, le projet Tuning, ainsi que les

systèmes « maison » de résultats de l'apprentissage. Une certaine mise en contexte pour chaque source se trouve également ci-dessous. Ces systèmes n'ont pas en soi de points forts ou de points faibles : ils correspondent simplement à diverses démarches pour composer avec la difficile description des buts de l'apprentissage.

- Les rubriques VALUE (évaluation valide de l'apprentissage dans l'enseignement au premier cycle) ont été mises au point par l'AACU (association des collèges et universités des États-Unis), dans le cadre de l'initiative LEAP (Liberal Education and America's Promise). Les rubriques VALUE, qui ne sont pas propres à une discipline en particulier, permettent d'évaluer l'apprentissage de l'étudiant à quatre niveaux distincts : le premier niveau (point repère) permet de mesurer le rendement de l'étudiant à son entrée à l'université, tandis que le dernier niveau (point culminant) permet de mesurer le rendement à l'achèvement d'un grade de premier cycle. Il existe 16 rubriques VALUE groupées en trois catégories principales : les facultés intellectuelles et concrètes; la responsabilité personnelle et sociale; l'apprentissage axé sur l'intégration et l'application. En ce qui nous concerne, les rubriques VALUE les plus pertinentes étaient la communication écrite, la communication verbale, le travail en équipe, la résolution de problèmes, de même que l'enquête et l'analyse [Association of American Colleges & Universities (2013)].
- Le projet Tuning a été mis au point par l'Union européenne. Le nom « Tuning » a été attribué au processus dirigé par le corps professoral afin de garantir la transparence et la qualité de l'enseignement supérieur par la stipulation des résultats (compétences) de l'apprentissage pour une discipline choisie à un niveau d'enseignement particulier. Ce projet tient compte de la variation dans les disciplines et tente de permettre de la souplesse en fonction d'un cadre et d'un processus communs. Il permet de différencier trois catégories de compétences génériques : les compétences instrumentales; les compétences interpersonnelles; les compétences systémiques [Bulgarelli, Lettmayr et Menéndez-Valdés (2009)].
- Le NILOA (institut national de l'évaluation des résultats d'apprentissage) a œuvré selon le cadre établi dans le DQP (profil des grades obtenus) de la Fondation Lumina pour donner aux établissements d'enseignement un moyen par lequel préciser ce que les étudiants devraient connaître et pouvoir faire, quelle que soit la discipline [Adelman, Ewell, Gaston et Schneider (2011)]. Dans le DQP, cinq grands axes des processus de l'apprentissage sont précisés : l'apprentissage appliqué; les facultés intellectuelles; les connaissances spécialisées; les connaissances générales; l'apprentissage communautaire. Il fournit un cadre global qui s'adapte à de nombreux types différents d'établissements d'enseignement en permettant de mettre en relief, dans une certaine mesure, un axe par rapport aux autres axes quant à un grade en particulier. De cette façon, les différents établissements d'enseignement peuvent préciser différentes exigences au premier cycle et, parallèlement, maintenir un niveau global de qualité.
- Le cadre des qualités requises des diplômés en génie a été conçu à l'Université de Toronto en guise de suite donnée au système d'agrément fondé sur les résultats que le BCAPG a mis en place. Voici les descriptions des qualités requises des diplômés du BCAPG pertinentes (c.-à-d. les compétences) :

- Conception : Capacité de concevoir des solutions à des problèmes d'ingénierie complexes et évolutifs et de concevoir des systèmes, des composants ou des processus qui répondent aux besoins spécifiés, tout en tenant compte des risques pour la santé et la sécurité publiques, des aspects législatifs et réglementaires, ainsi que des incidences économiques, environnementales, culturelles et sociales.
- Communication : Habileté à communiquer efficacement des concepts d'ingénierie complexes, au sein de la profession et au public en général, notamment lire, rédiger, parler et écouter, comprendre et rédiger de façon efficace des rapports et de la documentation pour la conception, ainsi qu'énoncer des directives claires et y donner suite.
- Travail individuel et en équipe : Capacité de fonctionner efficacement en tant que membre ou chef d'équipe, de préférence dans un contexte de travail multidisciplinaire.
- Analyse de problèmes : Capacité d'utiliser les connaissances et les principes appropriés pour identifier, formuler, analyser et résoudre des problèmes d'ingénierie complexes et en arriver à des conclusions étayées.
- Investigation : Capacité d'étudier des problèmes complexes au moyen de méthodes mettant en jeu la réalisation d'expériences, l'analyse et l'interprétation des données et la synthèse de l'information afin de formuler des conclusions valides.

[Bureau canadien d'agrément des programmes de génie (2014)]

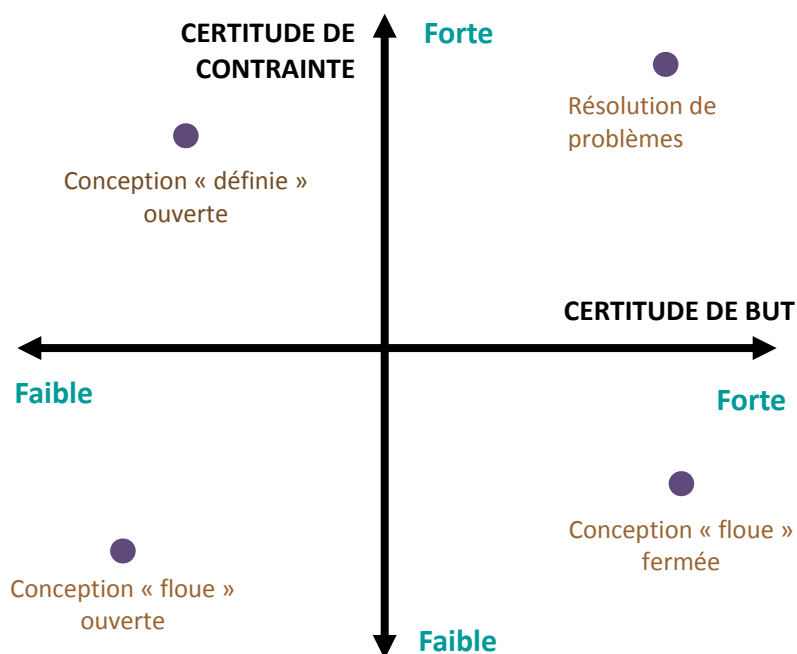
Notre recension documentaire nous a permis d'établir que les définitions provisoires des cinq compétences suffisaient à l'amorce de l'élaboration et de la validation des outils de mesure. De plus, grâce aux cinq compétences, nous avons pu miser sur l'expérience acquise au moyen du processus des qualités requises des diplômés du BCAPG, tout particulièrement par l'utilisation des indicateurs élaborés afin d'orienter la sélection des résultats et des indicateurs en vue des futurs travaux.

Fait à souligner, la vision entretenue par un établissement d'enseignement en particulier a influé en forte partie sur la sélection de l'ensemble complet de compétences. C'est ainsi que les établissements d'enseignement qui privilégiaient l'enseignement libéral général percevaient naturellement les compétences de façon différente de ceux où l'enseignement faisait l'objet d'une perspective ciblée (p. ex., les écoles professionnelles). Nous pouvions donc prévoir que les définitions des compétences propres à un établissement d'enseignement ou à un département en particulier seraient uniques en fonction des différences attendues entre les aspects pédagogiques privilégiés par les établissements d'enseignement.

L'analyse des systèmes de résultats d'apprentissage a révélé une confusion quant à la démarcation entre la conception et l'analyse des problèmes. Plusieurs cadres d'analyse des problèmes ont été étudiés, dont le modèle de Woods [Woods et al. (2001)] et l'outil d'évaluation de l'analyse des problèmes de Deek [Deek et al. (1999)]. Histoire d'en conserver le caractère distinct, ces deux compétences ont fait l'objet d'une distinction artificielle. Nous avons prévu que l'intégration de l'analyse des problèmes et de la conception à l'ensemble des compétences susciterait certaines difficultés, tout particulièrement en ce qui touche la distinction entre ces deux compétences, mais aussi en ce qui concerne le rajustement selon les différences

entre programmes comme le génie, les sciences physiques, les arts, les sciences sociales ou les lettres et sciences humaines. Cependant, la caractérisation des activités de conception et d'analyse de problèmes en fonction d'axes orthogonaux fondés sur la certitude des buts et des contraintes ou le recours à un système semblable pour différencier les deux compétences permet de constater que l'ampleur des activités couvertes par ces deux compétences s'est révélée élargie et davantage générale (voir le graphique 2). Une fois l'analyse des problèmes et la conception intégrées à la compilation des indicateurs et le chevauchement autorisé, la liste a permis d'englober le spectre plutôt que de tenter de découvrir une démarcation claire entre les deux compétences en question.

Graphique 2 : Caractérisation de la conception et de l'analyse de problèmes/résolution de problèmes par certitude de contraintes et de buts



Une fois la recension documentaire menée à bien, nous avons assemblé des listes exhaustives d'indicateurs. Après avoir discuté avec l'équipe de recherche, nous avons épuré la liste exhaustive qui en a résulté; pour ce faire, nous avons éliminé la redondance entre les systèmes et rassemblé les indicateurs en des catégories de résultats de l'apprentissage communs. L'équipe de recherche a analysé cette longue liste de résultats et d'indicateurs pour y discerner plusieurs lacunes. Fait particulier, nous avons décelé un manque d'indicateurs dans le domaine de la prise de conscience des compétences. L'auteur Woods et ses collègues (2001; 2002) insistent sur l'importance chez les étudiants de la prise de conscience du processus d'analyse des problèmes, en particulier. Habituellement, une telle prise de conscience permet de décrire la quantité de

savoir que possède l'étudiant non seulement à propos de la terminologie, mais au sujet des composantes de base des compétences particulières. Les étudiants dont la prise de conscience est accrue sont bien préparés à vivre des moments d'introspection au sujet de leurs capacités, à procéder à une réflexion fructueuse sur celles-ci et à en faire une évaluation critique. De plus, il est prévu qu'un rehaussement de la prise de conscience permet aux étudiants de tirer davantage parti de la rétroaction et de l'évaluateur de leur rendement, puisqu'ils peuvent alors établir davantage des liens entre la rétroaction et certains domaines à améliorer. Nous avons donc déterminé que la prise de conscience des compétences doit être évaluée en tant que dimension de chaque compétence.

3. Deuxième phase : Construction

3.1 Élaboration des descripteurs de rubriques

Après la compilation des résultats et des indicateurs, nous avons procédé à l'élaboration de certains outils (rubriques) de mesure en version provisoire relativement aux compétences de conception, de communication et de travail en équipe. Bien que la documentation fournie ne permette pas de dégager un consensus clair quant au nombre idéal de niveaux à intégrer dans une rubrique, le BCAPG a proposé des rubriques à quatre niveaux dans les exemples qu'il a publiés à l'intention des établissements d'enseignement. Pour cette raison, nous avons élaboré des descripteurs de rubrique (voir le tableau 1 pour obtenir les composantes de base d'une rubrique d'analyse) au moyen d'une échelle à quatre niveaux (échec; en deçà des attentes; attentes comblées; au-delà des attentes).

Tableau 1 : Structure de base d'une rubrique d'analyse

Indicateurs/Critères de rubrique	Échelle (niveau de maîtrise)			
	Échec	En deçà des attentes	Attentes comblées	Au-delà des attentes
Indicateur 1	Descripteur 1a	Descripteur 1b	Descripteur 1c	Descripteur 1d
Indicateur 2	Descripteur 2a	Descripteur 2b	Descripteur 2c	Descripteur 2d

Au bout du compte, nous avons divisé le niveau « Échec » en deux catégories distinctes mais connexes, conformément à la définition dans le tableau 2. Nous avons exécuté cette mesure pour montrer que la maîtrise de l'indicateur n'est pas attestée si l'étudiant ne parvient pas à achever le travail nécessaire ou s'il n'y a pas suffisamment de travail à mesurer. Toutefois, il se peut que l'étudiant ait effectué le travail, mais de telle sorte que la qualité de ce travail soit insuffisante ou que sa conception se révèle erronée, ce qui le situe dans la catégorie « Échec » relativement à l'indicateur en question. Ces deux types différents d'échecs dans l'attestation font l'objet d'une distinction au sein des rubriques pour permettre au responsable de la

notation de tirer au clair la rétroaction attribuée à l'étudiant quant à la nature de l'échec. Les définitions provisoires suivantes ont servi à orienter l'élaboration des rubriques.

Tableau 2 : Niveaux de rendement

Échec		En deçà des attentes	Attentes comblées	Au-delà des attentes
<i>Aucune attestation / Conception erronée</i>				
L'indicateur n'est pas attesté à cause du travail insuffisant à mesurer.	Il y a un manque total de qualité, une incompréhension fondamentale du concept, ou les deux.	La qualité fait défaut; le travail doit être revu en profondeur pour être acceptable.	La définition de la qualité. Le travail est acceptable et atteste une maîtrise dans une certaine mesure.	L'étudiant surpasse les attentes habituelles pour produire un travail de qualité supérieure.

Dans l'élaboration des descripteurs de rubrique, l'équipe de recherche a prêté tout particulièrement attention à deux principes fondamentaux. Le premier principe se rapporte aux différences dans les niveaux de rendement, en particulier la mise en relief des différences qualitatives plutôt que quantitatives. Parfois, les différences dans les niveaux des rubriques font état essentiellement d'une différence dans la quantité; par exemple, la note de l'étudiant(e) montre qu'il (ou elle) a surpassé les attentes parce qu'il (ou elle) en a fait plus. À titre d'exemple, dans une première version de notre rubrique communication, l'atteinte du niveau le plus élevé de rendement dans l'indicateur « créer un "flux" de communication (verbale ou écrite) par l'organisation » signifiait que toutes les idées (par opposition à la plupart des idées) étaient présentées selon une progression claire et logique. Or, le fait de s'appuyer sur des termes comparatifs ou quantitatifs comme « aucun », « la plupart » ou « tous » pose problème, car ceux-ci nuisent à la qualité et à la précision des descripteurs, de sorte que le discernement clair des démarcations entre les niveaux de rubrique par les évaluateurs s'en trouve compliqué. C'est ainsi que la notation qui correspond au niveau de rendement « au-delà des attentes » est supérieure à celle des niveaux « attentes comblées » ou « en deçà des attentes », non seulement parce qu'un devoir correspond à un nombre supérieur (ou inférieur) de critères en particulier (p. ex., le nombre de fautes grammaticales et d'orthographe) mais parce que le niveau de rendement « au-delà des attentes » diffère fondamentalement et qualitativement des deux autres. À titre d'exemple précis, la façon par laquelle les fautes grammaticales compromettent la compréhension du document se situe à un niveau approfondi d'évaluation de la qualité de la rédaction, plutôt qu'à un simple nombre absolu de fautes. Dans le tableau 3, nous présentons en guise d'exemple une ligne tirée d'une version antérieure de la rubrique communication, où des termes quantitatifs servent exclusivement à distinguer les niveaux de rendement. Le tableau 4 montre les modifications proposées pour tirer au clair les différences entre les niveaux.

Tableau 3 : Extrait de la première version de la rubrique Communication

Indicateur	Échec	En deçà des attentes	Attentes comblées	Au-delà des attentes
[...] formuler, par écrit, visuellement ou verbalement, un appui crédible et persuasif à une affirmation	Aucune affirmation n'est formulée, la plupart des affirmations ne sont pas crédibles en soi ou ne sont pas appuyées logiquement par des données de sources fiables	Certaines affirmations et conclusions sont crédibles en soi ou sont appuyées logiquement par des données de sources fiables	La plupart des affirmations et conclusions sont crédibles en soi ou sont appuyées logiquement par des données de sources assez fiables.	Toutes les affirmations et conclusions sont crédibles en soi ou sont appuyées logiquement par des données de sources très fiables.

Tableau 4 : Modifications proposées à l'extrait de la rubrique Communication

Indicateur	Échec		En deçà des attentes	Attentes comblées	Au-delà des attentes
	<i>Aucune att.</i>	<i>Conc. erronée</i>			
[...] formuler, par écrit, visuellement ou verbalement, un appui crédible et persuasif à une affirmation	<input type="checkbox"/> Les affirmations ne sont pas appuyées	<input type="checkbox"/> Les affirmations sont appuyées par des données et un raisonnement hors de propos <input type="checkbox"/> Aucun lien entre les données et les affirmations	<input type="checkbox"/> Les affirmations sont appuyées par des données et un raisonnement faibles <input type="checkbox"/> Les affirmations sont appuyées de façon inconstante <input type="checkbox"/> Les liens sont faibles entre les données et les affirmations	<input type="checkbox"/> Les affirmations sont appuyées par des données crédibles (ou l'utilisation de données crédibles) et un raisonnement valable <input type="checkbox"/> Les affirmations sont bien appuyées de façon constante <input type="checkbox"/> Les liens sont suffisants entre les données et les affirmations	<input type="checkbox"/> Attentes plus que comblées <input type="checkbox"/> Les affirmations reposent sur un traitement des données et un raisonnement ingénieux <input type="checkbox"/> Les données et affirmation contraires sont prises en compte et un contrepoint raisonnable est présenté

Le second principe de l'élaboration des descriptions des rubriques se rapporte à la dimensionnalité des rubriques. Pour qu'une rubrique soit valide, chaque ligne doit être unidimensionnelle, ce qui est parfois appelé « parallélisme ». L'unidimensionnalité désigne l'existence d'un attribut sous-jacent unique à chaque niveau du continuum (c.-à-d. allant du niveau « échec » au niveau « au-delà des attentes »). La ligne unidimensionnelle d'une rubrique n'instaure ni des nouveaux concepts, ni des concepts qui ne font pas partie des indicateurs ou des critères de cette rubrique. Elle n'instaure pas non plus de nouveaux concepts

au fur et à mesure que le niveau change. Le tableau 5 montre un exemple de ligne multidimensionnelle d'une rubrique dans laquelle plusieurs difficultés sont introduites, notamment être critique à la première cellule et manifester de l'indifférence à la deuxième cellule, deux attributs qui ne sont pas parallèles à l'idée d'afficher un « état d'esprit positif » relevée dans l'indicateur.

Tableau 5 : Exemple d'une ligne de rubrique multidimensionnelle

Indicateur	Échec	En deçà des attentes	Attentes comblées	Au-delà des attentes
Transmet un état d'esprit positif quant à l'équipe et à son travail (p. ex., par son ton et ses expressions)	<input type="checkbox"/> Transmet un état d'esprit négatif quant à l'équipe et à son travail par un ton (verbalement ou à l'écrit), des expressions faciales ou un langage corporel négatifs <input type="checkbox"/> Critique souvent en public l'équipe et son travail	<input type="checkbox"/> Transmet de l'indifférence quant à l'équipe et à son travail	<input type="checkbox"/> Utilise un ton (verbalement ou par écrit), des expressions faciales ou un langage corporel positifs de façon à transmettre un état d'esprit positif quant à l'équipe et à son travail	<input type="checkbox"/> Attentes plus que comblées <input type="checkbox"/> Invite les autres membres de l'équipe à avoir un état d'esprit positif quant à l'équipe et à son travail

Dans la version révisée de la rubrique (voir le tableau 6), il y a davantage de parallélisme dans les énoncés; nous avons modifié l'indicateur afin que celui-ci témoigne de la notion de « constructif ». Ce concept est ensuite plus ou moins répété à chaque niveau de la rubrique.

Tableau 6 : Modifications proposées à une ligne de rubrique multidimensionnelle

Indicateur	Échec	En deçà des attentes	Attentes comblées	Au-delà des attentes
Transmet un état d'esprit constructif quant à l'équipe et à son travail	<i>AA*</i> <input type="checkbox"/> Transmet un état d'esprit négatif quant à l'équipe et à son travail, ce qui nuit à la cohésion et à l'efficacité de l'équipe	<input type="checkbox"/> Tente de transmettre un état d'esprit constructif quant à l'équipe et à son travail, mais le fait de façon inconstante	<input type="checkbox"/> Transmet constamment un état d'esprit constructif; manifeste une compréhension du moment où il faut être positif quant à l'équipe et à son travail ou critiquer adéquatement ceux-ci	<input type="checkbox"/> Attentes plus que comblées <input type="checkbox"/> Invite les autres membres de l'équipe à adopter un état d'esprit constructif quant à l'équipe et à son travail

*AA : Aucune attestation

3.2 Consultations auprès d'experts en contenu

La consultation auprès d'experts s'est révélée une facette importante de la validation des rubriques. Si la plupart des experts¹ souscrivaient à l'idée que les résultats, les indicateurs et les descripteurs de rubrique étaient formulés avec clarté et exactitude et qu'ils étaient représentatifs de la compétence pertinente, il était alors possible d'avancer que les rubriques étaient valides sur les plans de la forme (c.-à-d. elles semblaient mesurer de façon subjective ce qu'elles sont censées mesurer) et du contenu (c.-à-d. elles témoignaient des facultés et du comportement qu'elles devaient mesurer).

Une fois les descripteurs de rubrique rédigés, l'équipe de recherche a mené des discussions individuelles auprès d'experts en contenu de la conception, de la communication et du travail en équipe en provenance de la FASE (faculté des sciences appliquées et du génie) de l'Université de Toronto. Les experts en conception consistaient en des membres du personnel enseignant qui donnent des cours axés sur la conception. Les experts en communication étaient essentiellement des membres du corps professoral du ECP (programme de communication en génie) pendant que les experts du travail en équipe, quoique très peu nombreux, étaient des professionnels en milieu universitaire et ayant l'expérience de la recherche dans le travail en équipe. Ces séances avaient pour objectif principal d'obtenir de la rétroaction sur les rubriques en version provisoire et d'y apporter des révisions si nécessaire. Les questions suivantes ont servi à orienter les consultations auprès des experts :

- Les résultats reflètent-ils les grands aspects du processus de conception, de communication ou de travail en équipe? Dans la négative, quels sont les résultats ou indicateurs manquants?
- Les résultats et indicateurs sont-ils classés dans la bonne catégorie?
- Les indicateurs sont-ils tous nécessaires à chaque résultat? Y a-t-il des redondances?
- En quoi la formulation employée pour décrire les résultats et les indicateurs peut-elle être améliorée?
- En quoi la formulation employée pour décrire les niveaux de rendement peut-elle être améliorée?

En ce qui touche les compétences Analyse de problèmes et Investigation (pour lesquelles les descripteurs demeurent en cours d'élaboration), les consultations auprès d'experts, lesquelles ont porté sur les résultats et les indicateurs, se sont déroulées selon une démarche différente : la méthode de Delphes. Celle-ci consiste en un outil d'enquête systématique et interactif qui sert à susciter la participation des experts et des intervenants par une série de tours de questionnaire [Linstone et Turoff (1975)]. La méthode de Delphes s'appuie sur le principe selon lequel le savoir collectif d'un groupe d'experts est davantage valide que celui d'un particulier. Au moyen de cette méthode, l'objectif de l'étude consistait à discerner les facultés, connaissances et comportements dont les experts convenaient de l'importance afin de mesurer les facultés

1 Les experts consultés dans la présente étude sont des membres du corps professoral qui donnent des cours relatifs aux compétences en question. Tant en ce qui concerne le travail en lien avec la méthode de Delphes relaté ici, de même que les consultations moins formalisées, les experts consistent en des personnes ayant une expérience de plusieurs années dans la création d'évaluations (p. ex., tests, devoirs) et la prestation de rétroaction et de notes aux étudiants de premier cycle à l'Université de Toronto.

en analyse de problèmes et en investigation. Les tableaux 7 et 8 indiquent le nombre d'experts par programme selon la méthode de Delphes. Nous avons mené l'étude d'après la méthode de Delphes en deux tours de questionnaire. Dans le premier tour, les experts devaient fournir une liste complète d'indicateurs servant à mesurer l'analyse de problèmes ou l'investigation. Dans le second tour, nous avons présenté aux experts les indicateurs compilés à partir du premier tour. Ces experts devaient ensuite noter dans quelle mesure ils allaient probablement se servir de chaque indicateur dans leur enseignement (1 : très peu probable; 2 : peu probable; 3 : probable; 4 : très probable) ainsi que l'importance de chaque indicateur par rapport au cursus dans l'ensemble (1 : pas du tout important; 2 : quelque peu important; 3 : important; 4 : très important). Par la suite, nous avons combiné la liste des indicateurs générés à partir de la méthode de Delphes à la liste d'origine découlant de la recension documentaire menée au cours de la première phase du projet.

Tableau 7 : Experts de l'investigation par programme

Programme	Nombre d'experts
Biomatériaux et génie biomédical	4
Génie chimique	1
Génie électrique et informatique	2
Science et génie des matériaux	1
Génie mécanique et industriel	2
Physique	1

Tableau 8 : Experts de l'analyse de problèmes par programme

Programme	Nombre d'experts
Génie chimique	5
Génie civil	3
Génie électrique et informatique	2
Génie mécanique et industriel	5

La rétroaction obtenue par la consultation auprès d'experts a entraîné des changements appréciables à la formulation et à la structure des résultats, des indicateurs et des descripteurs. La liste définitive des résultats et indicateurs ayant trait à l'analyse de problèmes et à l'investigation se trouve dans l'annexe B.

Une fois les rubriques de conception, de communication et de travail en équipe modifiées selon la rétroaction des experts, elles ont été jugées prêtes en vue des essais.

4. Troisième phase : Essais et analyse

4.1 Essais miroirs

Nous avons évalué les rubriques par des essais miroirs, une méthode dans laquelle des travaux déjà présentés par des étudiants et notés ont été réévalués par des évaluateurs au-delà du contexte propre à un cours donné actuellement. Le groupe des évaluateurs (appelé groupe de réflexion) a donné de la rétroaction sur les rubriques. Dans le cadre de ce processus, les évaluateurs consistaient en des étudiants de cycles supérieurs ayant acquis au préalable une expérience d'adjoint à l'enseignement dans des cours en lien avec les compétences mesurées. Les essais miroirs avaient pour objectif de mesurer l'utilité des rubriques, leur clarté et leur constance entre évaluateurs.

Deux raisons justifient le recours aux essais miroirs plutôt que l'instauration dans un cours donné en temps réel. D'une part, il s'est révélé utile de compter sur plusieurs personnes pour évaluer le même travail, ce qui permettait d'examiner la fiabilité entre évaluateurs. Dans une situation normale de notation, chaque travail présenté par les étudiants fait l'objet d'une seule évaluation, ce qui ne permettrait pas la tenue d'essais entre évaluateurs. D'autre part, grâce à l'utilisation de travaux déjà présentés par des étudiants et notés, nous avons pu sélectionner des exemples aux niveaux supérieur, intermédiaire et inférieur de l'échelle de qualité. Cette façon de faire s'est révélée utile pour examiner l'efficacité de la gamme de descripteurs des rubriques, tous niveaux confondus. Au bout du compte, le but consistait à raffiner davantage les rubriques à déployer dans un cours en temps réel.

La première étape des essais miroirs a consisté à recenser les cours dans lesquels les compétences sont enseignées et mesurées. À partir de ces cours, nous avons ensuite constitué des échantillons d'étudiants correspondant à une gamme de connaissances approfondies. Après la recension des cours pertinents et la compilation des exemples de devoirs, nous avons recruté des étudiants de cycles supérieurs ayant l'expérience de la notation en tant qu'évaluateurs.

Les essais miroirs ont eu lieu à l'aide d'une série de groupes de réflexion. Ceux-ci ont donné aux chercheurs l'occasion d'obtenir des renseignements approfondis par la discussion ciblée et l'interaction des groupes. Dans les groupes de réflexion, les participants ont mené une activité d'évaluation dans laquelle ils ont mesuré des échantillons semblables de travaux présentés par des étudiants au moyen d'une rubrique élaborée à partir des lignes de la rubrique Communication, de la rubrique Conception, ou des deux. Nous avons également sondé les participants pour obtenir leur rétroaction quant aux indicateurs, aux descripteurs et à la rubrique dans son ensemble. La dernière activité a consisté en une discussion orientée par les questions ou messages-guides qui suivent :

- Répertoriez les indicateurs qu'il convient, selon vous, de mesurer dans ce devoir mais qui sont manquants.
- Répertoriez les indicateurs qui font partie de cette rubrique mais ne sont pas à-propos pour ce devoir.
 - Veuillez expliquer pourquoi, selon vous, ils ne sont pas à-propos pour le devoir.
- Répertoriez les indicateurs qui prêtent à confusion.
 - Qu'est-ce qui prête à confusion dans ces indicateurs? Y a-t-il des mots qui, selon vous, sont flous?
- Répertoriez les descripteurs qui, selon vous, prêtent à confusion.
 - Qu'est-ce qui prête à confusion dans ces indicateurs? Y a-t-il des mots qui, selon vous, sont flous?
 - Le descripteur ne semblait-il pas lié à l'indicateur?
 - Ne fonctionnait-il pas dans le cadre du devoir?
- En matière de rubriques, quelle formation avez-vous suivie (si c'est le cas) par le passé?
 - Quelles ressources étaient fournies?
 - Qu'est-ce qui était utile? Qu'est-ce qui était inutile?
- De quoi les AE ont-ils besoin en ce qui touche le matériel de formation (p. ex., des exemples de travail qui montrent les attentes liées au rendement, les définitions de la terminologie, des astuces générales pour employer les rubriques dans l'évaluation)?
- Si vous deviez vous servir de cette rubrique de nouveau, signalez-nous une modification que vous y apporteriez.

Au total, nous avons mené 17 séances de groupe de réflexion avec 46 évaluateurs pour mettre à l'essai les rubriques Conception et Communication.

Toutefois, en ce qui touche la rubrique Travail en équipe, la démarche des essais miroirs n'a pu être utilisée. Contrairement aux rubriques Communication et Conception, où il était facile de recueillir le produit de travaux présentés par des étudiants comme des dissertations et des rapports de conception, le travail en équipe est essentiellement axé sur le rendement et ne peut être mesuré que par les personnes en mesure d'observer le processus. Par conséquent, plutôt que de procéder à des essais miroirs, l'équipe de recherche a adopté deux modes de mesure du travail en équipe. Le premier mode – la quasi-instauration – a consisté en des étudiants de cycles supérieurs qui : 1) ont observé trois équipes qu'ils supervisaient à intervalles réguliers (en tant qu'AE) tout au long de la session; 2) ont utilisé la rubrique Travail en équipe afin d'évaluer ces équipes; 3) ont donné de la rétroaction détaillée en fonction de leur expérience d'utilisation de la rubrique. En tout, six étudiants de cycles supérieurs ont pris part à ce processus. Le second mode – les séances de réflexion à voix haute – a fait intervenir des discussions avec les enseignants responsables des cours et des activités parascolaires où le travail en équipe joue un rôle fondamental. Dans ces séances, les enseignants devaient présenter des revues détaillées des activités d'évaluation types à l'aide de la rubrique Travail en équipe. Ils devaient tout particulièrement discerner les indicateurs à-propos quant aux devoirs des étudiants, expliquer pourquoi ces indicateurs étaient à-propos (ou pourquoi ils étaient hors de propos) et décrire comment ils interprétaient les éléments particuliers de la rubrique. Trois enseignants ont participé aux séances de réflexion à voix haute.

Le tableau suivant expose le nombre total de fois où les rubriques ont été mises à l'essai dans chaque programme ou département de la FASE (faculté des sciences appliquées du génie). Suivant ce qui est indiqué, la mise à l'essai des rubriques s'est déroulée dans environ la moitié des programmes de génie offerts à l'Université de Toronto.

Tableau 9 : Mise à l'essai des rubriques par programme

Programme ou département	Total du nombre de mises à l'essai
Études aérospatiales	s.o.
Biomatériaux et génie géomédical	s.o.
Génie chimique et chimie appliquée	s.o.
Génie civil	s.o.
Génie électrique et informatique	1
Sciences du génie	6
Programme de première année	7
ILUE (institut du leadership en enseignement de premier cycle)	2
Sciences des matériaux	s.o.
Génie mécanique et industriel	4

4.2 Analyse des données

Deux questions fondamentales ont orienté le processus d'analyse : 1) En quoi la rétroaction des évaluateurs peut-elle servir à modifier les rubriques? 2) Que peut nous révéler la notation des évaluateurs au sujet de la fiabilité des rubriques entre évaluateurs? Par conséquent, nous avons créé des fiches de données pour chaque indicateur et devoir. Celles-ci ont mis en évidence la notation quantitative et la rétroaction qualitative des évaluateurs à propos de l'indicateur, des descripteurs et de la rubrique dans l'ensemble, selon ce qui a été obtenu par les enquêtes et les discussions auprès des groupes de réflexion. L'équipe de recherche a ensuite analysé en détail les fiches de données.

L'analyse des données d'essai s'est amorcée par la revue de la notation des évaluateurs pour chaque indicateur. Si les évaluateurs avaient choisi plusieurs niveaux de rendement (c.-à-d. si leurs sélections s'étaient étendues collectivement à trois ou quatre niveaux de la rubrique), nous pouvions alors entrevoir

une notation inconstante et le risque d'une piètre concordance entre évaluateurs. Nous avons ensuite révisé la rétroaction qualitative. Souvent, les données qualitatives permettaient d'expliquer pourquoi la notation d'un indicateur en particulier semblait inconstante. Les graphiques 3 et 4 montrent des exemples de fiches de données de certains indicateurs provenant des rubriques Communication et Conception, respectivement.

Les chiffres dans le tableau correspondent au nombre d'évaluateurs ayant noté les travaux au niveau attribué. Par exemple, dans le graphique 3, deux évaluateurs ont constaté que le travail comblait les attentes relatives à l'indicateur Éthique 3. Un demi-point était employé lorsque l'évaluateur jugeait que le devoir se situait entre deux niveaux adjacents. Par exemple, dans le graphique 3, un évaluateur a constaté que le travail se situait entre les niveaux « en deçà des attentes » et « attentes comblées » quant à l'indicateur Éthique 1.

Dans le graphique 3, la notation des évaluateurs quant à l'indicateur de la communication « Présenter une introduction claire qui énonce le thème et donne un aperçu de la matière » est présentée dans la catégorie « Éthique et technologie », laquelle correspond à un devoir où les étudiants devaient sélectionner et analyser un instrument technologique. Les fiches de données ont révélé que les notes attribuées étaient relativement constantes, à l'exception du devoir n° 5.

Dans le graphique 4, la notation des évaluateurs quant à l'indicateur de conception « Exposer avec exactitude le problème de conception technique puis résumer les détails fondamentaux (interpréter l'énoncé du problème si celui-ci est fourni) » est présentée. Les notes étaient attribuées relativement au « PRPMP » (plan des exigences du projet et de la gestion du projet), dans lequel les étudiants devaient présenter le plan exhaustif des activités de conception qu'ils proposaient. Par contraste avec le graphique 3, la répartition des notes était étendue, ce qui témoigne d'une notation inconstante et d'une piètre concordance entre évaluateurs. Les commentaires donnaient à penser qu'une telle inconstance était vraisemblablement attribuable, du moins en partie, à l'incapacité des évaluateurs à distinguer les niveaux de rendement de même qu'au caractère multidimensionnel de l'indicateur.

En ce qui concerne le travail en équipe, la constance dans la notation n'a pas été mesurée parce que les étudiants de cycles supérieurs ont choisi les indicateurs en fonction des besoins de leurs équipes particulières. L'analyse des données du travail en équipe a donc porté sur la rétroaction qualitative donnée par les évaluateurs dans les enquêtes et les séances de réflexion à voix haute. Le graphique 5 constitue un exemple de fiches de données relativement à l'indicateur du travail en équipe « Relater avec exactitude l'apport des autres membres de l'équipe à l'activité de l'équipe ».

Nous avons également examiné la fréquence de sélection de chaque indicateur, car celle-ci a donné un certain aperçu des critères de rubriques perçus comme les plus importants dans la mesure du travail en équipe. Chez les AE, les indicateurs du travail en équipe les plus sélectionnés (c.-à-d. à 15 reprises ou plus) étaient les suivants :

- T1A Communiquer de façon respectueuse avec les membres de l'équipe, au moyen du ton, du langage corporel et des expressions faciales qui conviennent.
- T1G Proposer de nouvelles suggestions qui s'appuient sur les idées d'autrui.

- T1H Énoncer les mérites des idées parallèles exprimées par autrui.
- T2D Assister aux réunions de l'équipe à intervalles réguliers et à temps.
- T2E Mener à bien toutes les tâches attribuées avant les délais (à l'externe ou à l'interne).
- T2F Produire du travail de qualité qui fait progresser l'équipe.
- T2G Apporter des contributions individuelles qui font progresser le projet, directement ou indirectement.

Une fois toutes les feuilles de données examinées, les rubriques ont fait l'objet de modifications pour tenir compte de la rétroaction des évaluateurs.

Graphique 3 : Exemple de fiche des données de l'énoncé « Présenter une introduction claire qui énonce le thème et donne un aperçu de la matière »

	Échec		En deçà des attentes	Attentes comblées	Au-delà des attentes	Pas de sélection
	A.A.	C.E.				
Éthique 1			0,5	4,5		
Éthique 2			1,5	1,5		
Éthique 3			1	2		
Éthique 4		1,5	3,5			
Éthique 5		2	0,5	0,5		
Éthique 6			5			

Commentaires à propos de l'indicateur

- Difficile de rédiger des introductions et des conclusions qui comblent les exigences décrites dans la rubrique.
- Deux facteurs sont confondus : le thème et les descriptions (aperçu)

Commentaires à propos des descripteurs

- Usage inconstant des cases à cocher
- Usage inconstant de la formulation d'une ligne à l'autre
- Formulation subjective dans l'énoncé « L'introduction permet de capter et de maintenir l'intérêt du lecteur » du niveau de rendement « Au-delà des attentes ».

Graphique 4 : Exemple de fiche de données de l'énoncé « Exposer avec exactitude le problème de conception technique puis résumer les détails fondamentaux (interpréter l'énoncé du problème si celui-ci est fourni) »

	Échec		En deçà des attentes	Attentes comblées	Au-delà des attentes	Pas de sélection
	A.A.	C.E.				
PRPMP1			5	3,5	0,5	
PRPMP2			4	1	1	1
PRPMP3	1	1	4	2,5	0,5	
PRPMP4		1,5	3,5	2		
PRPMP5			5	4	2	

Commentaires à propos de l'indicateur

- On devrait le répartir en plusieurs indicateurs.

Commentaires à propos des descripteurs

- Dans de nombreux cas, le descripteur relatif au niveau de rendement « En deçà des attentes » correspond à une note insuffisante.
- Difficile d'établir la distinction entre les niveaux de rendement « Attentes comblées » et « Au-delà des attentes »
- Le niveau de rendement « Attentes comblées » est trop général.

Graphique 5 : Exemple de fiche des données de l'énoncé « Relater avec exactitude l'apport des autres membres de l'équipe à l'activité de l'équipe »

Rétroaction à propos de l'indicateur

- Prête à confusion; qu'est-ce qu'on entend par « relater »?

Rétroaction à propos des descripteurs

- Le niveau de rendement « Attentes comblées » pose problème, car il est souvent difficile pour les étudiants de donner des exemples particuliers : ils ne prêtent habituellement pas attention à cet élément au complet.
- Peut-être faudrait-il ajouter un indicateur permettant de mesurer la capacité des étudiants à « saisir les données des pratiques en équipe »

5. Résultats et discussion

Au moyen du présent projet, nous avons cherché à élaborer des rubriques d'analyse valides afin de mesurer les résultats de l'apprentissage dans cinq domaines de compétence clés : la conception; la communication; le travail en équipe; l'analyse de problèmes; l'investigation. Les résultats de ce travail sont catalogués aux annexes A et B. L'annexe A comporte la version définitive des rubriques de la conception, de la communication et du travail en équipe. L'annexe B englobe la liste des indicateurs de l'analyse de problèmes et de l'investigation, laquelle a été compilée et modifiée au moyen de la méthode de Delphes. Toute cette matière figure également dans un site Web public :

<https://sites.google.com/site/uoftlearningoutcomesproject/> (site en anglais seulement)

Le but fondamental consistait à créer des rubriques universelles et adaptables : universelles en ce sens qu'elles puissent servir à mesurer l'apprentissage des étudiants en divers contextes, et adaptables de façon à ce que les enseignants puissent les rajuster en fonction des besoins et préférences qui leur sont propres. L'équipe a d'abord élaboré des descripteurs liés aux indicateurs de la conception, de la communication ou du travail en équipe afin d'obtenir de la rétroaction sur les moyens possibles d'améliorer les rubriques, les chercheurs ont mené une série de discussions structurées en groupe avec les étudiants de cycles supérieurs et les enseignants (les évaluateurs). Ces discussions nous ont considérablement éclairés quant aux perceptions des évaluateurs sur la clarté, la pertinence et l'exhaustivité des rubriques. De plus, elles nous ont révélé la nécessité d'apporter des changements considérables aux rubriques avant leur déploiement. Les principales constatations de notre analyse des séances des groupes de réflexion sont résumées ci-dessous :

- La terminologie des rubriques n'était pas suffisamment claire. Les évaluateurs devaient nommer les indicateurs ou descripteurs de rubrique qui étaient, selon eux, difficiles à comprendre. Souvent, ils nommaient des parties de la rubrique qui étaient trop générales comme l'indicateur de communication « Sélectionner le contenu et la démarche qui conviennent aux destinataires et au but » ou des termes particuliers qui n'étaient pas clairs. À titre d'exemple d'un terme manquant de clarté, il y avait « diverses » dans l'indicateur de communication « Intégrer les données de diverses sources », dans lequel ou bien les évaluateurs étaient incertains de sa signification (« Je ne pouvais dire avec certitude ce que ça signifie[ait] »), ou bien ils avançaient plusieurs interprétations conflictuelles. Un autre exemple se trouvait dans l'indicateur de conception « Extraire et intégrer l'information provenant d'intervenants et de sources appropriées autres (fiables, diverses, crédibles) pour rehausser la compréhension du problème », où les évaluateurs ont fait part de leur confusion quant au terme « intervenants » (« Ce terme désigne-t-il les différents groupes qui s'en servent... ou les différentes parties en cause dans l'utilisation d'une seule application? »). Pour s'employer à résoudre le caractère incompréhensible de la terminologie des rubriques, l'équipe de recherche a discerné tous les indicateurs et descripteurs qui manquaient de clarté pour ensuite les modifier afin d'en accroître la simplicité et la clarté. Dans l'exemple de l'indicateur de communication susmentionné, celui-ci est devenu à la suite des changements « Intégrer les données à partir d'une *gamme* de sources différentes », ce qui témoigne clairement du sens voulu (c.-à-d. la portée du type de sources utilisées). Des notes de bas de page et des exemples se sont également ajoutés pour tirer

au clair les termes dont l'interprétation risquait d'être erronée, ainsi qu'un guide exhaustif des rubriques à l'intention des enseignants et des AE (voir l'annexe C).

- Les niveaux de rendement n'étaient pas suffisamment différenciés. Pour les évaluateurs, il était souvent difficile d'établir aisément la distinction entre les niveaux de rendement de la rubrique qui étaient à proximité, par exemple entre « échec » et « en deçà des attentes » ou entre « attentes comblées » et « au-delà des attentes » (« En ce qui touche les niveaux " attentes comblées " et " au-delà des attentes ", je ne pouvais vraiment pas établir la différence »). De plus, ils nous ont dit que le niveau « au-delà des attentes » était souvent défini au moyen d'une formulation excessivement subjective, laquelle ne permettait pas de le distinguer convenablement du niveau « attentes comblées » (« J'ai également constaté que la différence entre les niveaux " attentes comblées " et " au-delà des attentes " n'était pas claire... qualifier quelque chose d'ingénieux est très subjectif, sans être très spécifique »). Pour améliorer la distinction entre les niveaux de rendement, l'équipe de recherche a procédé à un examen poussé des rubriques et, avec la participation de l'équipe au complet, elle a ajouté d'autres qualificatifs distincts pour établir une distinction claire entre les catégories de rendement.

De même, les évaluateurs ont également souligné l'importance de la souplesse dans la sélection des niveaux de rendement :

Souvent, j'ai de la difficulté à choisir une catégorie, par exemple entre « en deçà des attentes » et « attentes comblées ». La description est parfois claire et parfois floue. J'avais cette notion à l'esprit lorsque je m'occupais de la notation [...] Prenons l'exemple d'une ligne, où il y a le niveau « échec » et où il y a le niveau « au-delà des attentes »; il pourrait peut-être y avoir une flèche quant à l'endroit où aller pour pouvoir apporter des changements, pour qu'il y ait en réalité quelque chose entre les catégories.

Les rubriques étaient conçues de façon à ce que les enseignants puissent, par exemple, ajouter de l'espace destiné aux commentaires ou aux éléments pour donner aux évaluateurs la possibilité non seulement de choisir au sein des niveaux de rendement, mais entre ceux-ci. Les chercheurs insistent donc sur l'importance de l'adaptation aux enseignants.

- Des critères importants étaient omis dans les rubriques. Les évaluateurs ont présenté des exemples d'indicateurs qui, à leur sens, étaient nécessaires à l'évaluation, mais n'étaient pas compris dans la rubrique. Parmi ces exemples, il y avait entre autres les techniques de rédaction et le professionnalisme dans la communication, l'autonomie des solutions dans la conception, de même que la répartition de la charge de travail entre membres de l'équipe dans le travail en équipe. Dans de nombreux cas, les indicateurs jugés manquants étaient compris dans la rubrique (sans toutefois faire partie de l'évaluation d'une séance de groupe de réflexion en particulier à des fins pratiques). Dans les cas où un indicateur important manquait effectivement à une compétence, il a été ajouté au bout du compte à la rubrique avec ses descripteurs connexes. Par exemple, à la suite de la

rétroaction des évaluateurs quant au contenu manquant de la rubrique de communication, les chercheurs ont ajouté plusieurs indicateurs servant à évaluer précisément le résumé ou la synthèse d'un rapport de même que l'exactitude et la clarté grammaticales.

La rétroaction susmentionnée que les évaluateurs ont fournie a débouché sur d'importantes améliorations aux rubriques. Compte tenu de ces constatations fondamentales (et de l'expérience liée à la direction du projet), l'équipe de recherche a tiré des leçons importantes qui pourront se révéler utiles aux autres chercheurs qui s'intéressent à l'élaboration et à la mise en place d'outils universels pour l'évaluation des résultats d'apprentissage. La liste suivante décrit les leçons apprises dans l'élaboration des rubriques relatives à la conception, à la communication et au travail en équipe, ainsi que dans la mise au point définitive de la liste des résultats et indicateurs en lien avec l'analyse de problèmes et l'investigation.

- La méthode de Delphes peut se révéler une solution de rechange judicieuse à la consultation individuelle. La méthode de Delphes a servi à répertorier et à peaufiner une liste exhaustive de résultats et d'indicateurs permettant d'étoffer les critères des rubriques en lien avec l'analyse de problèmes et l'investigation. En ce qui touche la conception, la communication et le travail en équipe, la méthode de Delphes n'a pas servi et la consultation auprès d'experts s'est déroulée essentiellement par des séances en personne et individuelles. Bien que cette démarche ait permis d'obtenir des données utiles qui ont contribué à une importante modification des rubriques, seuls quelques experts pouvaient en réalité prendre part au processus. À titre d'exemple, la consultation auprès d'experts en ce qui touche la communication s'est déroulée sur une période de six mois, mais a résulté en des réunions avec seulement cinq experts. Par contre, l'étude au moyen de la méthode de Delphes a eu lieu pendant une période semblable auprès de 10 et de 12 experts, respectivement, en ce qui concerne l'investigation et l'analyse de problèmes. La méthode de Delphes nous a donc permis de consulter un nombre élargi d'experts durant une période raisonnablement courte.
- La formation des évaluateurs est cruciale. Suivant ce qui est décrit au préalable, les discussions des groupes de réflexion ont révélé un grand nombre de conceptions erronées et de points prêtant à confusion en ce qui touche la terminologie des rubriques. La tendance des participants à faire une interprétation erronée des termes des rubriques donne à penser que la rubrique, en soi, ne suffit pas. De façon précise, nous proposons la tenue de séances comparatives afin que la rubrique soit interprétée de la même façon par tous ceux qui évaluent les travaux présentés par les étudiants. La formation aidera les évaluateurs à bien comprendre les directives de l'affectation, l'objectif de la rubrique, et les termes particuliers au sein de la rubrique, de sorte que la constance dans la notation s'en trouvera accrue. Pour s'atteler à cette tâche, un manuel de formation en version provisoire a été élaboré.
- Le travail en équipe est difficile à évaluer en raison de la complexité du comportement humain. Le travail en équipe s'est révélé très éprouvant en ce qui touche l'élaboration de rubriques. Selon ce qui a été mentionné au préalable, le travail en équipe est une compétence axée sur les processus et ancrée dans l'interaction sociale, de sorte que l'observation faite par ceux à l'extérieur de l'équipe

s'en trouve compliquée. De fait, l'évaluation par les seuls évaluateurs externes de nombreux indicateurs compris dans notre propre rubrique est presque impossible. À titre d'exemple, il y a l'indicateur du travail en équipe « Prêter assistance aux membres de l'équipe si nécessaire ou selon les besoins ». Pour traiter cet enjeu, nous recommandons l'évaluation du travail en équipe au moyen de plusieurs sources, tout particulièrement les propres réflexions des étudiants, les réflexions de leurs collègues d'équipe, ainsi que les évaluations d'observateurs externes [Association of American Colleges and Universities (2013)]. Une évaluation triangulée fera en sorte que chaque source de données puisse apporter un éclairage différent et unique au processus du travail en équipe.

6. Conclusion

L'évaluation des résultats de l'apprentissage devient de plus en plus une partie intégrante de la politique éducative au niveau postsecondaire. Son attrait réside dans sa capacité à fournir une information concrète sur la qualité et l'efficacité de l'enseignement supérieur. À cet égard, nous avons cherché dans le présent projet à élaborer une réserve de rubriques universelles afin d'évaluer l'apprentissage des étudiants dans cinq domaines clés, nommément la conception, la communication, le travail en équipe, l'analyse de problèmes, et l'investigation. Le processus d'élaboration des rubriques s'est amorcé par une recension documentaire sur l'évaluation des résultats de l'apprentissage afin de dégager une liste de résultats et d'indicateurs pour chaque compétence. Cette étape a ensuite fait place à la construction des rubriques, où nous avons consulté des experts, rédigé en version provisoire les descripteurs de la conception, de la communication et du travail en équipe, puis peaufiné les résultats et indicateurs ayant trait à l'analyse de problèmes et à l'investigation. À la troisième et dernière phase, les évaluateurs et enseignants ont donné leur rétroaction sur les rubriques par les discussions des groupes de réflexion, et la « quasi-institution » dans le cas du travail en équipe. Au bout du compte, le travail du projet a débouché sur un ensemble de rubriques et d'analyses validées pour évaluer la conception, la communication et le travail en équipe (voir l'annexe A), une liste de résultats et d'indicateurs validée en ce qui touche l'analyse de problèmes et l'investigation (voir l'annexe B), de même qu'une série d'articles pour aider les utilisateurs à élaborer leurs propres rubriques adaptées au moyen de la banque de rubriques (voir l'annexe C).

Ce qu'il importe de retenir de notre travail, c'est que les rubriques ne sont qu'aussi valables que la formation s'y rapportant. Nous insistons donc sur le fait que les évaluateurs suivent une formation quant au mode d'utilisation de la rubrique afin d'évaluer efficacement les travaux présentés par les étudiants. En ce qui touche les équipes d'enseignants qui évaluent le travail dans le même cours, ou les résultats à l'échelle du programme, l'analyse comparative joue un rôle crucial dans la fiabilité entre évaluateurs. Dans l'analyse comparative, des groupes d'évaluateurs procèdent à l'évaluation d'échantillons de travaux présentés par des étudiants, dont certains correspondent clairement à chaque niveau de la rubrique, pendant que d'autres n'ont pas un lien aussi direct. Les évaluateurs se penchent sur les échantillons ensemble. Cet exercice leur permet d'en arriver à une compréhension commune non seulement de la terminologie des rubriques mais de la qualité du travail qui illustre chaque niveau de rendement, si bien que la constance dans la notation s'en trouve grandement facilitée. Outre l'analyse comparative avec les évaluateurs, nous recommandons l'intégration de la banque de rubriques à des moteurs de gestion de l'apprentissage (LME) actuellement

employés ou à d'autres outils de technologie éducative servant à gérer l'évaluation et la rétroaction. Une fois intégrée la technologie éducative, l'établissement d'enseignement – non seulement les étudiants – obtiendra de l'information détaillée sur l'apprentissage et le rendement des étudiants, ce qui contribuera à la transparence des éléments à valeur ajoutée de l'enseignement supérieur.

Les versions les plus actualisées des rubriques sont affichées dans un site Web public : [\(https://sites.google.com/site/uoftlearningoutcomesproject/\)](https://sites.google.com/site/uoftlearningoutcomesproject/) (site en anglais seulement). Au fur et à mesure que d'autres travaux seront exécutés pour peaufiner ces outils, les versions mises à jour seront affichées. Pour obtenir la liste des publications connexes à la présente étude, veuillez consulter l'annexe C.

Bibliographie

- Adelman, C., Ewell, P., Gaston, P. et C. G. Schneider (2011), *The Degree Qualifications Profile*. Extrait de : http://degreeprofile.org/advantage/publication/The_Degree_Qualifications_Profile.pdf
- Association of American Colleges & Universities (2013), *VALUE : Valid Assessment of Learning In Undergraduate Education*. Extrait de : <http://aacu.org/value/rubrics/index.cfm>
- Bulgarelli, A., Lettmayr, C. et J. Menéndez-Valdés (2009), *La transition vers les acquis de l'apprentissage : Politiques et pratiques en Europe*. Centre européen pour le développement de la formation professionnelle.
- Bureau canadien d'agrément des programmes de génie (2014), *Normes et procédures d'agrément*. Extrait de : http://www.engineerscanada.ca/sites/default/files/2014_accreditation_criteria_and_procedures_v06.pdf
- Centre d'information canadien sur les diplômes internationaux (2012), *Normes pancanadiennes de qualité pour l'évaluation des diplômes d'études internationales – Phase II : Rapport final*. Conseil des ministres de l'éducation (Canada).
- Conseil de l'Europe (1999), *Convention sur la reconnaissance des qualifications relatives à l'enseignement supérieur dans la région européenne*.
- Council of Ontario Universities (2011), *Ensuring the Value of University Degrees in Ontario*. Council of Ontario Universities.
- Deek, F. P., Hiltz, S. R., Kimmel, H. et N. Rotter (1999), « Cognitive Assessment of Students' Problem Solving and Program Development Skills », dans *Journal of Engineering Education*, vol. 88 n° 3, p. 317–326.
- Division de l'enseignement supérieur (2006), *Lignes directrices pour des prestations de qualité dans l'enseignement supérieur transfrontalier*, Organisation des Nations Unies pour l'éducation, la science et la culture.
- Linstone, H. A. et M. Turoff (éd.) (1975), *The Delphi Method : Techniques and Applications* (vol. 29), Reading (Massachusetts), Addison-Wesley.
- Ontario Universities Council on Quality Assurance (2012), *Quality Assurance Framework Technical Report*, Council of Ontario Universities.
- Ralph, P. et Y. Wand (2009), « A proposal for a formal definition of the design concept », dans K. Lyytinen, P. Loucopoulos, J. Mylopoulos et W. Robinson (éd.), *Design Requirements Workshop* (p. 103–136), Berlin, Springer-Verlag.
- Woods, D. R., Kourti, T., Wood, P. E., Sheardown, H., Crowe, C. M. et J. M. Dickson (2001), « Assessing Problem-Solving Skills—Part 1 : The Context for Assessment », dans *Chemical Engineering Education*, numéro d'automne, p. 300–307.

Woods, D. R., Kourti, T., Wood, P. E., Sheardown, H., Crowe, C. M. et J. M. Dickson (2002), « Assessing Problem-Solving Skills—Part 2 : Assessing the Process of Problem Solving », dans *Chemical Engineering Education*, numéro d'hiver, p. 60–66.

