



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 :

Kaupp, J., Frank, B. et A. Chen (2014), Évaluation de la pensée critique et de la solution de problèmes dans les grands groupes : activités suscitant des modèles pour le développement de la pensée critique, Toronto, Conseil ontarien de la qualité de l'enseignement supérieur.



Un organisme du gouvernement de l'Ontario

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, 2014.

Résumé

Le présent rapport décrit une étude ayant pour objet d'enquêter sur les retombées d'un programme d'études comportant des activités suscitant des modèles (ASM) sur le développement de la pensée critique (PC) dans le cadre d'un cours de génie donné en première année à l'Université Queen's. Le cours porte sur le développement des aptitudes en solution de problèmes, en modélisation et en pensée critique, au moyen, en partie, de problèmes contextualisés complexes, aussi appelés activités suscitant des modèles (ASM). De plus, les ASM permettent d'évaluer les aptitudes susmentionnées en contexte de cours, de manière rigoureuse, authentique et viable. L'étude, à laquelle 542 étudiants ont participé, s'est déroulée au cours de la session d'automne de l'année scolaire 2012-2013.

L'enseignement explicite de la pensée critique au moyen du modèle de Paul-Elder faisait partie du déroulement du cours, avec une série d'activités suscitant des modèles en vue de développer les aptitudes en pensée critique des étudiants. D'autres outils, dont plusieurs tests uniformisés, ont servi à évaluer l'aptitude en pensée critique au début et à la fin du déroulement du cours, notamment :

- 1) le test de la pensée critique Cornell : niveau Z (CLZ)
- 2) le test international de la pensée critique à questions ouvertes (ICTET)
- 3) l'évaluation de l'apprentissage au niveau collégial (CLA)
- 4) des sondages sur les cours
- 5) des protocoles de réflexion à voie haute : une séance d'entrevue où les participants devaient « réfléchir à voix haute » à leurs réponses à un problème ou à un scénario.

À la suite de la conclusion de l'étude et de l'analyse des données, nous n'avons pas constaté de gain significatif dans les aptitudes en pensée critique (APC) durant le cours en ce qui touche deux des outils uniformisés (le test de la pensée critique Cornell : niveau Z [CLZ] et le test international de la pensée critique à questions ouvertes [ICTET]). De plus, un autre outil n'était disponible que sous forme de test préalable à faire passer (l'évaluation de l'apprentissage au niveau collégial ([CLA]). Les groupes témoins participants ont fait face à des difficultés de recrutement et d'attrition, et leur participation n'a pas suffi à l'établissement de comparaisons. Toutefois, nous avons constaté une amélioration du rendement des étudiants quant aux résultats de la pensée critique intégrés aux activités des cours (les ASM) de même qu'aux exercices de réflexion à voix haute. Les réponses des étudiants aux questions du sondage portant sur leur perception du développement de leur pensée critique durant le cours ont permis de discerner l'utilité des éléments des cours et des ASM dans le développement de la pensée critique. De tous les outils uniformisés, c'est l'évaluation de l'apprentissage au niveau collégial (CLA) qui a affiché les plus fortes corrélations aux notes données par les évaluateurs en ce qui touche les activités suscitant des modèles, puisque l'application des aptitudes en pensée critique entre cet outil et les activités suscitant des modèles sont très semblables.

Nous avons discerné plusieurs éléments auxquels il convient, à notre sens, de prêter particulièrement attention dans les travaux ultérieurs :

- 1) Il y a lieu de prêter particulièrement attention à la concordance entre le cadre de pensée critique servant à l'enseignement, les activités intégrées de même que les outils uniformisés. Toute discordance entre ces éléments risque d'influer sur les résultats mesurés quant à la PC, de sorte qu'il y a lieu de faire preuve de la plus grande diligence dans le maintien de l'harmonisation à un degré élevé.
- 2) Il y a lieu de préserver la concordance des tâches entre les tests uniformisés et les activités intégrées. De façon précise, la mise en application de la pensée critique dans les tests uniformisés et sa mise en application dans les activités intégrées doivent être pratiquement identiques.
- 3) La motivation et la participation des étudiants ont posé considérablement problème. Les outils uniformisés ont suscité de la fatigue relativement aux tests, ce qui s'est traduit par des méthodes

- superficielles et des difficultés au chapitre du rendement. Pour minimiser de tels effets, il convient que les évaluations ayant trait aux tests préalables et postérieurs soient identiques par rapport au cheminement dans le cours.
- 4) Il est possible que les outils uniformisés n'aient pas la sensibilité requise pour évaluer les gains liés aux APC durant une seule session. Peut-être que de tels outils conviendront davantage à l'évaluation longitudinale du développement à long terme des APC tout au long du programme. Il y a lieu de procéder à l'évaluation du développement à court terme des APC à l'aide des artefacts produits par les étudiants et des pratiques authentiques.

Pour régler de tels problèmes, nous recommandons de choisir avec circonspection les cadres d'enseignement, les activités intégrées et les outils uniformisés de façon à assurer la concordance et la mise en application des cadres. À cette fin, les travaux ultérieurs quant à l'évaluation et au développement de la pensée critique en génie, y compris les travaux avec le Consortium sur l'évaluation des résultats d'apprentissage du COQES, se poursuivront à l'aide de certains outils uniformisés, mais également de rubriques à l'échelle des programmes (p. ex., la rubrique VALUE de l'association américaine des collèges et universités [AAC&U]) en vue de noter les artefacts produits par les étudiants et en lien avec les travaux scolaires, dans l'optique de faire une évaluation longitudinale de la pensée critique, allant de la première à la quatrième année.

Table des matières

Résumé	2
Introduction	7
Objectifs de recherche	7
Examen de la documentation	
Activités suscitant des modèles	3
Évaluations de la pensée critique	13
Contexte d'enseignement	16
Module 1 de l'APSC 100	16
Résultats d'apprentissage et structure des cours	18
Caractéristiques et résultats des ASM	18
Méthode et procédure	21
Aperçu de la conception et des variables de l'étude	21
Outils de l'étude	22
Méthodologie	25
Analyses des données et des statistiques	28
Résultats	31
Résultats des tests préalables et postérieurs	31
Résultats des ASM	34
Séances de réflexion à voix haute	38
Analyse du sondage	39
Fiabilité des tests de la pensée critique	44
Discussion	45
Recommandations et recherche à l'avenir	47
Bibliographie	52

Liste des graphiques

Graphique 1 : Le modèle Cornell-Illinois	11
Graphique 2 : Le modèle de Paul-Elder	
Graphique 3 : Le modèle CLA	13
Graphique 4 : Évaluation de la pensée critique à l'aide des ASM	
Graphique 5 : Cadre conceptuel employé pour le cours APSC100	21
Graphique 6 : Structure du groupe CLA	23
Graphique 7 : Structure du groupe CLZ	24
Graphique 8 : Structure du groupe ICTET	24
Graphique 9 : Division du groupe de réflexion à voix haute	25
Graphique 10 : Résultats des questions du sondage sur les efforts des étudiants	42
Graphique 11 : Perception des étudiants quant aux activités des cours sur le développement	
des APC	44
Graphique 12 : Perceptions des étudiants quant à l'intégration des connaissances, à la discussion du	contenu
et au tutorat	43
Graphique 13 : Classement par les étudiants des cheminements de première année qui contribuent au	ux APC
	44

Liste des tableaux

Tableau 1 : Caractéristiques de chacune des ASM	19
Tableau 2 : Résultats communs des ASM	20
Tableau 3 : Méthode de mesure, groupement des cohortes et outils d'étude	22
Tableau 4 : Éléments des sous-notes des tests de pensée critique	28
Tableau 5 : Descriptions des échelles Likert des sondages (annexes 7 et 8)	29
Tableau 6 : Sous-notes générales et particulières des ASM	30
Tableau 7 : Codage selon le modèle de Paul-Elder relatif à la pensée critique	30
Tableau 8 : Notes et sous-notes moyennes du groupe CLA	31
Tableau 9 : Groupe CLZ : Notes et sous-notes moyennes de la cohorte B	
Tableau 10 : Groupe CLZ : Notes moyennes de la cohorte C	32
Tableau 11 : Groupe ICTET : Moyenne des notes et des sous-notes de la cohorte D	33
Tableau 12 : Groupe ICTET : Moyenne des notes de la cohorte E	33
Tableau 13 : Validation de l'hypothèse : Comparaison des notes des cohortes au test postérieur	33
Tableau 14: ASM et comparaison des sous-notes aux ASM durant le cours APSC 100	34
Tableau 15 : Corrélations entre les éléments des ASM et de la CLA	35
Tableau 16 : Corrélations entre les éléments des ASM et de l'ICTET	
Tableau 17 : Corrélations entre les éléments des ASM et du CLZ	37
Tableau 18 : Mesures de fiabilité des ASM	38
Tableau 19 : Classement des participants à l'étude selon la motivation intrinsèque	40
Tableau 20 : Classement des participants à l'étude selon la motivation introjectée et externe	40
Tableau 21 : Classement des participants à l'étude selon la motivation interne et identifiée	40
Tableau 22 : Différences dans la note au test de la pensée critique selon le statut d'ALS	41

Introduction

Nombreuses sont les personnes qui jugent que la capacité de résoudre des problèmes et d'exercer une pensée critique constitue des résultats recherchés dans le système d'enseignement, tant de la maternelle à la 12^e année qu'au niveau de l'enseignement supérieur. Il s'agit d'aptitudes multiprésentes mesurées à l'aide d'un bon nombre d'accréditations dans les secteurs professionnels et de l'enseignement supérieur; elles figurent également de façon constante parmi les principales aptitudes et capacités recherchées chez les diplômés, selon des enquêtes auprès d'employeurs (Hart Research Associates, 2013). Or, malgré une telle importance, la pensée critique et la solution de problèmes sont souvent répertoriées par les employeurs dans les aptitudes auxquelles l'enseignement supérieur doit prêter avantage attention (Hart Research Associates, 2008; Arum et Roksa, 2011). De récentes données permettent de remettre en question la mesure dans laquelle l'enseignement au premier cycle appuie actuellement le développement des aptitudes en pensée critique et en solution de problèmes complexes (Arum et Roksa, 2011; Astin, 1993a; 1993b; Blaich & Wise, 2008; Klein et al., 2009; Pascarella, Blaich, Martin et Hanson, 2011). Le développement des aptitudes en pensée critique (APC) forme en soi un enjeu complexe, lequel se complique par une discordance dans la définition du concept de « pensée critique » et le cadre connexe à son développement (Ku, 2009). Parmi les cadres de pensée critique les plus répandus, il y a le modèle Cornell-Illinois (Ennis, Millman et Tomko, 1985), le modèle de Paul-Elder (Paul et Elder, 2005; Paul et Elder, 1996), le modèle de la CLA (Shavelson, 2008), le modèle Delphi de l'APA (Facione, 1990), de même que le modèle de pensée critique de Halpern (Halpern, 1999; Halpern et Riggio, 2002). Chacun de ces cadres ou modèles présente une définition différente de la pensée critique ainsi qu'un ensemble différent d'aptitudes, de traits et de capacités qui s'y trouvent. L'enseignement et l'évaluation des APC est également un domaine où les difficultés sont particulières, l'efficacité des stratégies pédagogiques en vue du développement de la pensée critique de même que l'authenticité de l'évaluation de la pensée critique faisant l'objet d'un examen très approfondi (Bensley et Murtagh, 2011; Solon, 2003).

En dépit de ces questions sous-jacentes, l'idée selon laquelle les APC sont d'une importance cruciale pour traiter des problèmes complexes et concrets fait généralement consensus. L'une des méthodes visant à développer la capacité de résoudre des problèmes complexes et concrets au sein de disciplines à forte teneur mathématique fait intervenir des activités suscitant des modèles (ASM), à savoir des problèmes réalistes servant à l'enseignement en classe et devant lesquels les apprenants doivent non seulement documenter leur solution aux problèmes, mais leur processus en vue de leur solution (Shuman, 2012; Shuman et Besterfield-Sacre, 2008). Dans le cadre de la solution, les ASM exigent la création d'une description, d'une procédure ou d'un système mathématique dont les étudiants se serviront pour mettre au point et peaufiner leur processus et solution (Chamberlin, 2004; Shuman et Besterfield-Sacre, 2008). Les ASM sont conçues et employées dans diverses matières, notamment les mathématiques, l'économie et le génie environnemental. Des études ont révélé le rôle utile joué par les ASM afin d'aider les étudiants à développer la compréhension conceptuelle, le transfert des connaissances ainsi que les aptitudes généralisables en solution de problèmes (Self, Shuman et Besterfield-Sacre, 2012; Yildirim, Shuman, Besterfield-Sacre et Yildirim, 2010).

Objectifs de recherche

Le présent rapport décrit une enquête sur le développement des APC dans un cours de génie (Module 1, Engineering Practise, de l'APSC 100) donné en première année à l'Université Queen's. L'objectif fondamental de notre étude consiste à enquêter sur les retombées d'un programme d'études comortant des ASM sur le développement des aptitudes en pensée critique chez les étudiants. En ce qui a trait à cet objectif, nous avons l'intention de définir des points de repère concernant les APC des étudiants de première année en génie au début de la session d'automne de l'année scolaire 2012-2013, d'exposer ces étudiants au programme d'études comportant des ASM, puis d'évaluer leur APC à la fin du déroulement du cours, une fois

terminée la session d'automne. De plus, nous comptons présenter une analyse approfondie en ce qui touche les outils de pensée critique employés dans l'étude au moyen des questions de recherche suivantes :

- 1) Y a-t-il une corrélation entre les notes des outils de pensée critique et celles des ASM?
- 2) Y a-t-il une corrélation entre les sous-notes des outils de pensée critique et celles des ASM?
- 3) Y a-t-il une corrélation entre la capacité de pensée critique et les facteurs motivants?
- 4) Y a-t-il une corrélation entre la capacité de pensée critique et certains cheminements dans les cours?
- 5) Y a-t-il une corrélation entre la capacité de pensée critique et certains facteurs intrinsèques?
- 6) Les outils de pensée critique utilisés sont-ils fiables et valides?

Le deuxième objectif de l'étude consiste à donner un aperçu de la façon d'évaluer la pensée critique en contexte de génie. Relativement à cet objectif, nous espérons fournir un point de départ pour les travaux ultérieurs des autres parties intéressées sur le développement et l'évaluation de la pensée critique. Nous proposons un court examen des outils de pensée critique, de concert avec les conclusions et recommandations en lien avec ces outils ainsi que d'autres observations découlant de l'étude. À cette fin, nous posons les questions suivantes :

- 1) Existe-t-il des données selon lesquelles les ASM comportent des retombées positives significatives sur les aptitudes en pensée critique des étudiants?
- 2) Quel cadre de pensée critique et quel outil de pensée critique reflètent la mise en application des aptitudes en pensée critique dans le règlement de problèmes complexes en génie?
- 3) Dans quelle mesure faut-il préserver la concordance des tâches entre l'outil de pensée critique et les problèmes complexes de génie?
- 4) Quelles sont les méthodes fructueuses pour évaluer les aptitudes en pensée critique dans le cadre d'un cours?

Trois outils ont servi, dans le contexte de la présente étude, à évaluer les APC des étudiants de première année en génie. Ces outils ont été employés dans le test préalable et postérieur, d'une part, en vue de définir des points de repère concernant les APC des étudiants qui font leur entrée en première année et, d'autre part, de déterminer l'efficacité de l'enseignement des ASM sur le développement de la capacité des étudiants en matière de pensée critique.

Examen de la documentation

Activités suscitant des modèles

Les ASM servent à l'enseignement au niveau universitaire en génie depuis 2004 (Diefes-Dux et al., 2004; Moore et Diefes-Dux, 2004; Shuman et Besterfield-Sacre, 2008). Conçues à l'origine comme outil d'évaluation en mathématique (Lesh, 1999; Lesh et Doerr, 2003), les ASM demeurent un thème à l'étude au niveau des écoles intermédiaires. Elles font également l'objet d'un projet de recherche de quatre ans dans sept universités américaines (projet MEDIA, s.d.), où leur utilisation est envisagée selon divers contextes dans des grands et petits groupes. Jusqu'à maintenant, les résultats des ASM se révèlent prometteurs en ce qui touche le développement des aptitudes des étudiants liées à la compréhension conceptuelle spécialisée, à la facilité de traiter l'information, au règlement des problèmes et à la communication. Les ASM exigent des étudiants qu'ils s'appuient sur leurs connaissances antérieures, de sorte qu'elles permettent souvent aux étudiants de déceler les idées fausses et de traiter celles-ci en contexte d'apprentissage. Elles favorisent également les liens entre l'information.

La conception des ASM repose sur un ensemble de six principes, adaptés à partir de leur contexte d'origine en mathématiques pour être employés dans un programme d'études en génie. Ces principes sont les suivants (Lesh & Doerr, 2000; Moore et Diefes-Dux, 2004) :

- 1) Construction du modèle : Cette activité nécessite la construction d'une description, d'une explication ou d'une procédure explicite en vue d'une situation mathématiquement significative.
- 2) Réalité : Elle exige de situer l'activité dans un contexte réaliste en matière de génie puis de la concevoir de telle sorte que les étudiants puissent interpréter utilement l'activité selon leurs divers niveaux de capacité mathématique et de connaissances générales.
- Autoévaluation : L'activité comporte des critères que les étudiants peuvent discerner puis utiliser pour évaluer leur mode de pensée actuel et réviser celui-ci.
- 4) Documentation du modèle : Les étudiants doivent créer de la documentation qui, sous une certaine forme, révélera explicitement leur mode de pensée quant à la situation qui pose problème.
- 5) Communicabilité et réutilisation : Ce volet exige des étudiants qu'ils en arrivent à des solutions communicables à autrui et adaptables à d'autres situations en génie.
- 6) Prototype efficace : Ce point permet de faire en sorte que le modèle produit soit le plus simple possible mais demeure mathématiquement significatif à des fins de génie.

L'enseignement des ASM accorde une place considérable au processus employé pour régler le problème de même qu'au raisonnement et à la pensée dont les étudiants se servent pour mettre au point leurs solutions plutôt qu'aux produits résultant de cette méthode. La solution relative à une ASM exige des participants qu'ils mettent en application et combinent plusieurs concepts en génie, en physique ou en mathématiques tirés des études qu'ils ont faites et de leurs antécédents pour formuler un modèle mathématique général pouvant servir à régler le problème. Habituellement, les étudiants ont recours à un processus itératif face à l'ASM : ils produisent d'abord un modèle, puis ils mettent celui-ci à l'essai et y apportent des révisions pour concevoir une solution convenable (Lesh et Doerr, 2003). Les étudiants traitent les ASM en groupe, ce qui permet de simuler le déroulement du travail en équipe propre à la pratique professionnelle. Les solutions des étudiants ayant trait à l'ASM prennent habituellement la forme d'un rapport exhaustif dans lequel est décrit le processus employé pour produire leur solution au problème.

Plusieurs études ont été réalisées pour enquêter sur les retombées de l'enseignement des ASM quant aux résultats d'apprentissage et au développement des aptitudes générales chez les étudiants. Voici ce que ces études ont révélé à propos des ASM :

- Elles favorisent un point de vue différent en ce qui concerne l'utilisation des concepts de génie : les étudiants mettent les concepts en application pour en arriver à une solution globale et de haut niveau plutôt qu'à une démarche de bas niveau, stéréotypée et machinale (Shuman et Besterfield-Sacre, 2008).
- 2) Elles invitent les étudiants à travailler en collaboration et en coopération au sein d'un groupe, contexte dans lequel ils affinent leurs aptitudes liées au travail d'équipe et à l'entregent et assurent la prestation d'une solution de qualité supérieure aux présentations individuelles (Gokhale, 1995).
- 3) Elles favorisent l'intégration et la synthèse de l'information et des concepts qui s'étendent au génie et à d'autres disciplines (Yildirim et al., 2010).
- 4) Elles sont propices au raisonnement et aux aptitudes de pensée de niveau supérieur par leur enseignement, mal structuré et de nature complexe (Chamberlin, 2002).

Les avantages susmentionnés des ASM mènent les étudiants à un cheminement d'apprentissage davantage significatif qui entraîne leur participation à un exercice fidèle à la pratique professionnelle en génie. Un tel cheminement d'apprentissage significatif contribue à favoriser à la fois les aptitudes de haut niveau et les résultats d'apprentissage générique à atteindre quant à la solution de problèmes complexes, à la communication, à leur bagage en matière d'information et à la pensée critique.

Protocoles de réflexion à voix haute

Outre les outils d'évaluation de la pensée critique, la présente étude a recouru à des exercices de réflexion à voix haute. Ces protocoles émanent de la science cognitive (Ericsson et Simon, 1993; Fonteyn et al., 1993; Van Someren, Barnard et Sandberg, 1994) en guise de moyen d'observation et d'étude du raisonnement concomitant aux fins de l'analyse du raisonnement et du traitement de l'information par les participants ((Boren et Ramey, 2000; Ericsson et Simon, 1993; Ungson et Braunstein, 1982). Les protocoles de réflexion à voix haute ont servi à évaluer les activités cognitives telles que la solution de problèmes et la capacité de pensée critique en génie et dans d'autres domaines connexes (Daly, 2001; Ku et Ho, 2010a; Norris, 1990; Steif, Lobue, Kara et Fay, 2013). Dans les séances à cette fin, une tâche est confiée aux participants, un objectif leur est fixé et ils doivent « réfléchir à voix haute » dans leur processus de pensée en vue de trouver une solution. Un animateur, présent en tant qu'observateur passif, doit inciter les participants à, ou à leur rappeler de, verbaliser leurs processus de pensée. De telles séances sont enregistrées et les propos tenus sont transcrits. Ceux-ci sont la par suite codés au moyen de l'analyse de protocole en fonction d'un modèle cognitif particulier ou, dans le cas de la présente étude, d'un cadre de pensée critique.

Cadres de pensée critique

Chaque cadre de pensée critique décrit un point de vue différent quant à la construction complexe de la pensée critique. Chaque modèle est fondé sur une définition fonctionnelle de la pensée critique et présente un cadre quant aux aptitudes, aux attributs, aux normes et aux dispositions des éléments conformément à la définition fonctionnelle. Bon nombre de ces cadres ne comportent ni stratégie pédagogique explicite, ni séquence développementale pour les étudiants; ils fournissent tout simplement une définition succincte de la construction et de ses éléments. Cependant, la définition et le cadre constituent le fondement de l'intégration de la pensée critique au programme de cours, et jouent un rôle essentiel à ce chapitre. Les descriptions des modèles dans la section suivante font partie d'une introduction aux cadres de pensée critique et donnent un aperçu des évaluations des APC employées dans la présente étude. Pour obtenir d'autres cadres de pensée critique, veuillez consulter l'annexe 9.

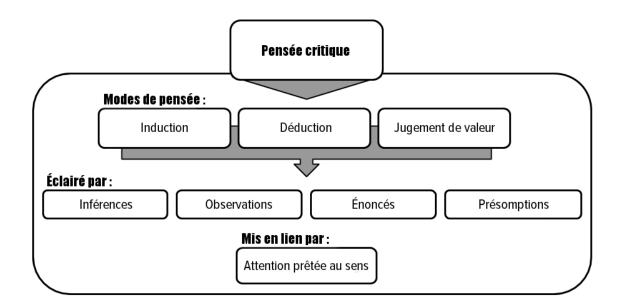
Modèle Cornell-Illinois

Le modèle Cornell-Illinois de la pensée critique a été mis au point puis peaufiné par Robert Ennis, d'après la définition fonctionnelle suivante de la pensée critique :

La pensée critique consiste en une pensée rationnelle et réfléchie portant sur les décisions quant à ce qu'il faut croire ou faire (Ennis et al., 1985).

Le modèle, exemplifié dans le graphique 1, fait l'objet d'une division et d'un sous-classement d'après trois modes de pensée critique (l'induction, la déduction et le jugement de valeur) et quatre méthodes sur lesquelles ils sont fondés : les résultats des inférences, des observations, des énoncés et des présomptions. Enfin, le modèle est mis en lien par un fil conducteur – l'attention prêtée au sens – qui s'entrelace dans les quatre méthodes et trois modes (Ennis et al., 1985).

Graphique 1 : Le modèle Cornell-Illinois



Modèle de Paul-Elder

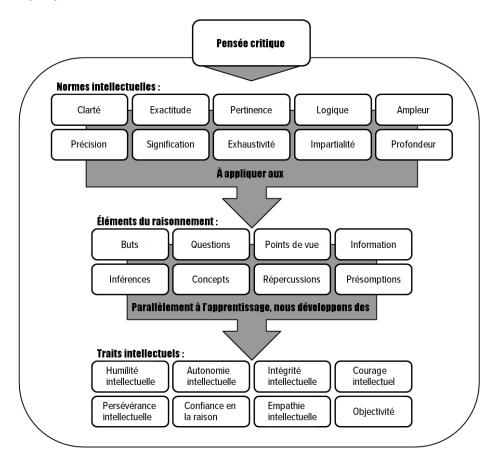
Le modèle de Paul-Elder, mis au point à l'origine par Paul (Paul, 1993; Paul et al., 1993) puis ensuite peaufiné par Paul et Elder (Paul et Elder, 2001), va de pair avec la Foundation for Critical Thinking (www.criticalthinking.org – site essentiellement en anglais), une organisation éducative sans but lucratif qui fait la promotion d'un changement essentiel dans l'enseignement et la société par le développement d'une pensée critique objective (Foundation for Critical Thinking, s.d.). Le modèle de Paul-Elder est fondé sur la définition fonctionnelle suivante de la pensée critique :

Le mode de pensée – quel que soit le thème, le contenu ou le problème en question – par lequel le penseur rehausse la qualité de sa pensée par son analyse, son évaluation et sa reconstruction savantes. La pensée critique est une pensée autoorientée, autodisciplinée, autosurveillée et autocorrigée. Elle présuppose l'assentiment à de rigoureuses normes d'excellence de même que l'orientation consciente de leur utilisation. Elle exige d'efficaces capacités en communication et en règlement de problèmes ainsi que l'engagement à surmonter notre égocentrisme et notre sociocentrisme innés. (Paul et Elder, 2005)

Le modèle de Paul-Elder divise la pensée critique en trois éléments fondamentaux : les éléments du raisonnement; les normes intellectuelles; les traits intellectuels. D'abord, les éléments du raisonnement correspondent à des éléments universels qui éclairent et décrivent l'ensemble du raisonnement ou de la pensée. Ensuite, les normes intellectuelles sont les normes appliquées aux éléments du raisonnement ou de la pensée dans l'optique d'interpréter ou d'évaluer la qualité. Enfin, les traits intellectuels constituent les caractéristiques ou traits recherchés chez les praticiens chevronnés de la pensée critique. Ces trois éléments sont interdépendants, et chacun joue un rôle quant au développement d'un penseur critique. Dans le modèle de Paul-Elder, les penseurs critiques mettent les normes intellectuelles en application aux éléments du raisonnement afin d'acquérir des traits intellectuels (graphique 2). Il existe deux dimensions essentielles de la pensée que les étudiants doivent maîtriser afin d'apprendre comment rehausser leur pensée. Ils doivent, d'une part, savoir discerner les parties composantes de leur pensée et, d'autre part, être en mesure d'évaluer

l'utilisation qu'ils font de celles-ci (Paul et al., 1996). Il est possible de structurer ces deux volets essentiels, de concert avec les normes intellectuelles, les éléments du raisonnement et les traits intellectuels, en une rubrique servant à l'évaluation de la pensée critique.

Graphique 2 : Le modèle de Paul-Elder



Modèle CLA

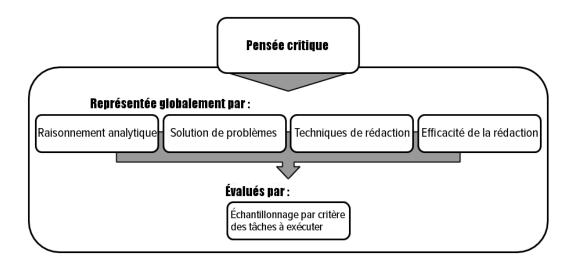
Le modèle CLA a été mis au point en vue d'une évaluation globale de la pensée critique par la solution de problèmes. Ce modèle, qui sert strictement à des fins d'évaluation, a été élaboré en vue de la Collegiate Learning Assessment (évaluation de l'apprentissage au niveau collégial, ou CLA), un test dont il sera question dans la section suivante. Le modèle CLA ne consiste pas en un cadre explicite, contrairement aux modèles de Paul-Elder ou Cornell-Illinois, qui ramènent la pensée critique à ses parties constituantes; il perçoit plutôt la pensée critique au sens large, comme le résume ci-après l'auteur Bok :

La capacité d'exercer une pensée critique – poser des questions pertinentes, déceler les problèmes et les définir, trouver des arguments de part et d'autre dans un débat, chercher des données pertinentes et s'en servir, puis arriver à la conclusion après un jugement soigneusement raisonné – constitue le moyen indispensable d'utiliser judicieusement l'information et le savoir (Bok, 2006).

D'après le modèle CLA, une méthode globale correspond à la meilleure évaluation de la pensée critique. Les tenants de ce modèle font valoir que la pensée critique ne peut être ventilée en des parties composantes puis

mesurée. Ils sont plutôt d'avis que c'est la construction élargie de la pensée critique qui est étroitement liée à, et représentée par, plusieurs critères ou aptitudes dont les étudiants se servent dans leurs réponses au test, comme le montre le graphique 3 :

Graphique 3 : Le modèle CLA



Le modèle CLA s'appuie sur une méthode d'échantillonnage par critère relativement simple et dont l'objet consiste à déterminer les capacités de l'étudiant par l'échantillonnage des tâches du domaine dans lequel il faut évaluer l'étudiant, l'observation de ses réponses, puis l'inférence de son rendement et de son apprentissage dans la construction élargie. L'auteur Shavelson (2008) explique l'échantillonnage par critère au moyen d'un exemple lié à la conduite automobile :

À titre d'exemple, dans le cas où vous cherchez à savoir si une personne est non seulement au fait du code de la route mais en mesure de bien conduire une voiture, il ne faut pas simplement lui faire passer un test à choix multiples. Il faudrait plutôt également lui faire passer un test de conduite qui comporte un échantillon de tâches à tirer du domaine de la conduite en général, comme démarrer la voiture, s'engager dans la circulation, tourner à droite et à gauche dans la circulation, reculer, et garer la voiture. En fonction de cet échantillon de rendement, il sera possible de tirer des inférences valides sur le rendement général de la personne en matière de conduite automobile (Shavelson, 2008).

L'évaluation de l'apprentissage au niveau collégial (CLA) s'inscrit dans la méthode d'échantillonnage par critère en exposant les étudiants à des problèmes globaux et concrets. Au moyen de ces problèmes, elle permet d'échantillonner les tâches et de recueillir les réponses des étudiants, lesquelles sont ensuite évaluées d'après un ensemble d'aptitudes génériques puis constituées en des rubriques. Pour en arriver à la bonne réponse à la tâche en question, les étudiants devront parvenir à appliquer la solution au problème, à raisonner de façon analytique et à rédiger de façon persuasive et efficace. Puisque ces facettes correspondent toutes à des éléments sous-jacents de la pensée critique, telle qu'elle est définie par le modèle CLA, la capacité en pensée critique peut donc être inférée à partir des réponses des étudiants aux questions du test.

Évaluations de la pensée critique

De nombreuses évaluations de la pensée critique sont disponibles, chacune étant construite à partir d'un cadre de pensée critique différent. Il en résulte une vaste gamme quant au type et à la mise en application de

chaque test comportant chacun des forces et des faiblesses qui leur sont propres. Dans ce contexte, le chargé de cours dispose d'une grande latitude pour choisir un test qui correspond le mieux à sa propre définition de la pensée critique, ou adopter un cadre d'enseignement qui comporte un test correspondant pour maintenir la concordance entre l'enseignement et l'évaluation. Les examens de ces évaluations dans la section suivante sont présentés pour fournir des renseignements quant aux forces et aux faiblesses de chaque évaluation, de même que commenter les évaluations choisies aux fins de la présente étude. Pour en savoir plus sur les autres évaluations de la pensée critique, veuillez consulter l'annexe 10.

Test de pensée critique Cornell : niveau Z

Le test de pensée critique Cornell : niveau Z (CLZ) consiste en un test à choix multiples comportant 52 éléments. Il est destiné aux élèves de niveau secondaire ainsi qu'aux étudiants de niveau universitaire qui sont doués. Le CLZ a été mis au point par Robert Ennis, Jason Millman et Thomas Tomko; l'édition la plus récente de ce test est parue en 2005 (Ennis et al., 1985). Le CLZ permet d'évaluer sept aspects de la pensée critique, conformément au modèle Cornell-Illinois (Ennis et al., 1985):

- 1) Induction
- 2) Déduction
- 3) Observation et crédibilité
- 4) Présomptions
- 5) Sens et erreurs

En matière de notation, le CLZ présente deux options : d'une part, la notation selon les « bonnes réponses seulement » où sont dénombrées les réponses justes et, d'autre part, la notation selon les « bonnes réponses moins la moitié des mauvaises réponses », où écopent les étudiants dont les réponses sont erronées. Si les deux options sont valables quant à la notation du test, les auteurs privilégient la deuxième option parce que le fait de deviner ne s'inscrit pas dans des habitudes significatives de pensée (Ennis et al., 1985). Le test peut être fait sur papier ou en ligne. D'autres sous-notes peuvent être calculées relativement aux sept aspects particuliers du modèle Cornell-Illinois, répertoriés ci-dessus. Plusieurs études ont permis de valider le CLZ, les mesures constatées à ce chapitre allant de $\alpha = 0,5$ à 0,87 (Ennis et al., 1985; Frisby, 1992).

Le recours à une évaluation à choix multiples des APC comporte des risques, lesquels émanent du fait que le test ne permet pas d'évaluer les aspects de disposition de la pensée critique. En règle générale, les évaluations à choix multiples de la PC ont fait l'objet de critiques, car les tests évaluent les connaissances verbales et quantitatives plutôt que la pensée critique, puisque le format des tests empêche les répondants de mettre en application leurs APC pour concevoir leur propre solution au problème (Abrami et al., 2008; Halpern, 2003; Ku, 2009). En outre, les tests à choix multiples ne permettent qu'une évaluation étroite d'un seul concept de pensée dans une question, alors que l'application concrète de la pensée critique fait habituellement appel à une vaste gamme de concepts et d'aptitudes (Bensley et Murtagh, 2011; Ku, 2009). De façon précise, les critiques relatives au CLZ sont dirigées vers les notes de faible validité en lien avec le test de même que le risque de biais sexiste ayant trait aux éléments du test (Stein et al., 2003).

Évaluation de l'apprentissage au niveau collégial

L'évaluation de l'apprentissage au niveau collégial (CLA) est composée d'un ensemble d'affectations de tâche Web à l'intention des étudiants de niveau universitaire allant de la première à la quatrième année. Cette évaluation, conçue et gérée par le Council for Aid to Education (CAE), est axée sur le modèle de la CLA relativement à la pensée critique et à la solution de problèmes. La notation de la CLA est faite au moyen d'un système automatisé faisant appel à une série de rubriques d'évaluation (Council for Aid to Education, n.d.; Shavelson, 2008). Les notes et sous-notes globales sont compilées à partir des dimensions de la pensée critique, notamment :

- 1) Le raisonnement analytique
- 2) La solution de problèmes
- 3) Les techniques de rédaction
- 4) L'efficacité de la rédaction

En règle générale, les étudiants remplissent l'une des deux tâches distinctes qui composent la CLA : une « tâche de rendement » et une « tâche de rédaction analytique » comportant deux sous-tâches, à savoir « formuler un argument » et « critiquer un argument ». La CLA fait état d'une validité élevée (α = 0,80) mais seulement à l'échelle des établissements, puisque sa validité est basse à l'échelle des étudiants (α = 0,45) (Klein et al., 2009; Klein, Benjamin, Shavelson et Bolus, 2007). De plus, les méthodes d'évaluation globale du test ont suscité certaines préoccupations, car elles ne permettent pas d'évaluer avec exactitude les aptitudes cognitives des éléments de la pensée critique. En outre, certains ont critiqué le mode d'évaluation de la CLA. Ces facteurs, auxquels s'ajoutent le coût de la CLA et les possibilités étroites d'application du test, constituent des obstacles éventuels à son utilisation (Possin, 2013). Malgré ces difficultés éventuelles, la CLA permet une évaluation exhaustive, assortie de tâches exigeant la recension, l'intégration et l'utilisation de plusieurs aptitudes et concepts de pensée critique dans l'une et l'autre des tâches. En outre, le CAE s'est récemment penché sur le problème de fiabilité à l'échelle des étudiants par la mise au point de la nouvelle CLA+, laquelle permet de rehausser la validité à l'échelle des étudiants pour la faire passer à α = 0,85-0,87 (Zahner, 2013).

Test international de la pensée critique à questions ouvertes

Le test international de la pensée critique à questions ouvertes (ICTET) a été mis au point par Richard Paul et Linda Elder de la Foundation for Critical Thinking. L'ICTET consiste en un test à questions ouvertes et à réponses longues ayant pour objet de présenter une évaluation des éléments fondamentaux de la pensée critique. L'ICTET comporte deux domaines d'intérêt particulier. Le premier consiste à présenter un moyen rationnel d'évaluer les APC, tandis que le deuxième consiste à fournir un outil de test qui incite les membres du corps professoral à enseigner leur discipline d'une façon qui favorise la pensée critique chez les étudiants (Paul et Elder, 2010). L'ICTET est réparti en deux formulaires distincts : d'une part, l'analyse d'un message-guide écrit et, d'autre part, l'évaluation du message-guide écrit. Dans le volet d'analyse (formulaire A) du test, l'étudiant doit discerner avec exactitude les éléments de raisonnement au sein d'un message-guide. Dans le volet d'évaluation du test (formulaire B), l'étudiant doit faire une analyse et une évaluation critiques du raisonnement employé dans le message-guide à l'origine. Les réponses de l'étudiant sont évaluées en fonction d'une rubrique axée sur les éléments du raisonnement qui composent le modèle de Paul relativement à la pensée critique (Paul et Elder, 2005):

- 1) But
- 2) Questions
- 3) Information
- 4) Conclusions
- 5) Concepts
- 6) Présomptions
- 7) Répercussion
- 8) Point de vue

Il est ensuite possible de calculer la note totale et les sous-notes connexes. L'ICTET a été créé pour comporter une validité corrélative élevée, de façon à ce que l'utilisation du test ait une corrélation significative et très visible pour les chargés de cours (Paul et Elder, 2007). Cette façon de faire favorise l'adoption de la pensée critique selon la discipline de même que le remaniement du programme de cours qui consiste à « enseigner en fonction des examens ». Du point de vue statistique, il reste à déterminer la validité des critères et concomitante de même que les mesures de fiabilité de l'ICTET, en raison de l'absence d'un critère universel pour évaluer les APC et du fait que l'outil en est, en quelque sorte, encore à ses débuts. Les questions de validité du contenu et de la construction de l'ICTET sont traitées par le recours à des messages-

guides propres à chaque discipline pour le test de même qu'à un modèle de pensée critique bien fondé. Autrement dit, le test permet d'évaluer ce qu'il est censé évaluer par le recours au modèle de Paul-Elder en guise de cadre et il permet d'évaluer avec exactitude les APC au moyen de messages-guides comportant de la matière pertinente à la discipline.

Dans l'utilisation d'un test de ce type, quelques difficultés éventuelles peuvent se manifester. D'abord, face aux messages-guides, les étudiants doivent discerner et évaluer les éléments de la pensée en fonction de ce qu'ils ont mémorisé. Bien que ces aptitudes soient d'une importance vitale dans le contexte de la pensée critique, les messages-guides en particulier ne peuvent permettre d'évaluer la mesure dans laquelle les étudiants mettent en application les APC en des situations concrètes (Bensley et Murtagh, 2011; Butler et Butler, 2012; Butler et al., 2012; Halpern, 2006). Ensuite, le caractère spécifique des questions risque de restreindre l'ampleur de la réponse des répondants au test, de sorte que ceux-ci auront moins tendance à s'investir dans la pensée critique (Taube, 1997). Enfin, la fiabilité interévaluateurs dans un test de ce type risque de poser problème, et il y a lieu d'en tenir compte lorsque vient le temps de faire passer ce test à grande échelle (Shavelson, Baxter et Gao, 1993).

Contexte d'enseignement

Module 1 de l'APSC 100

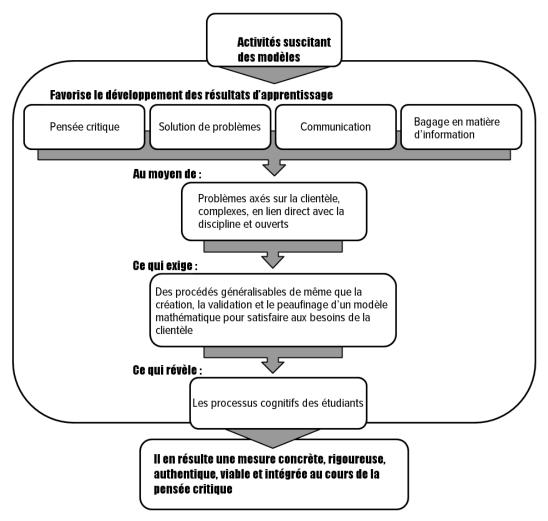
L'étude s'est déroulée dans le contexte d'un cours axé sur le projet et donné aux étudiants de première année dans le programme de premier cycle en génie à l'Université Queen's. Le programme comporte un tronc commun en première année : les 650 étudiants de première année en génie doivent prendre les mêmes cours avant de pouvoir choisir librement parmi dix programmes de génie en deuxième année. Au cours de la première session du programme, les étudiants suivent des cours portant sur la statique, la chimie, les systèmes terrestres, les graphiques et calculs en génie, de même que le cours APSC 100, réparti sur l'année complète et axé sur la conception et la pratique du génie. C'est dans ce dernier cours que la présente étude se situe (annexe 1).

Le cours APSC 100, axé sur les équipes et les projets, sert à aiguiser la curiosité relativement au génie et à donner aux étudiants l'occasion de développer des aptitudes liées au jugement et à la solution de problèmes par le traitement de tâches où sont simulées des activités en génie. Le cours est réparti en trois modules : le module 1, qui traite de l'analyse des problèmes et de la modélisation; le module 2, lequel traite de l'expérimentation et de la mesure; ainsi que le module 3, sur la conception technique. Chacun de ces modules dure une session et correspond, sur le plan de la pondération, à un cours ordinaire d'une session en génie (Frank, Strong, Sellens et Clapham, 2012; Frank, Strong et Sellens, 2011). La présente étude s'insérait dans la prestation du module d'analyse de problème et de modélisation (APSC 100 Module 1). Celui-ci consiste en un cheminement d'intégration qui, tout au long de la session, fait appel à des concepts issus des cours en sciences d'ingénieur, en sciences naturelles et en mathématiques afin de résoudre des problèmes ouverts complexes. La structure du cours porte sur trois problèmes complexes appelés activités suscitant des modèles (ASM); celles-ci étaient traitées de façon séquentielle durant des blocs de trois semaines au cours de la session.

Compte tenu des situations décrites dans les ASM, les étudiants devaient créer puis utiliser un modèle mathématique dans un système physique au moyen d'un outil de calcul numérique (MATLAB), puis composer avec des problèmes d'ordre professionnel, dont les dilemmes éthiques, l'information contradictoire et l'information erronée ou manquante. Chaque ASM exige des étudiants qu'ils fassent appel à différentes connaissances du domaine, mais les étudiants apprennent à aborder chacune des trois ASM au moyen d'aptitudes en pensée critique. À titre d'exemple, les étudiants reçoivent une orientation pour dessiner des cartes conceptuelles, remettre en question la crédibilité des sources d'information, intégrer une gamme de facteurs à leur prise de décisions et tenir compte des répercussions de leurs conclusions. Ces aptitudes sont

appelées « éléments » de la pensée critique par l'auteur Paul, à savoir les processus de pensée d'une importance inestimable et qui interviennent dans toute activité complexe de solution de problèmes (Paul & Elder, 2005). Fait important, les ASM procurent un moyen concret et intégré au cours de mesurer de façon authentique, vigoureuse et viable le développement de la pensée critique et l'évaluation des aptitudes en pensée critique, comme le montre le graphique 4 ci-après.

Graphique 4 : Évaluation de la pensée critique à l'aide des ASM



Les ASM ont servi dans le cadre du cours pendant les trois dernières années (Frank et Kaupp, 2012). Durant l'année scolaire 2010-2011 à l'Université Queen's, des étudiants en génie ont fait l'objet d'observations en vue de rehausser leur capacité à résoudre des problèmes complexes et à satisfaire aux attentes liées au cours durant l'année. Dans l'ensemble, les notes des étudiants ayant trait aux éléments des rubriques des ASM que sont la définition des problèmes et la création d'un processus de solution ont augmenté au cours de l'année, et une analyse très détaillée de ces résultats est en voie de réalisation à d'autres fins. De plus, selon les étudiants, de telles activités favorisaient leur développement; la plupart d'entre eux ont relaté, dans un sondage sur le cours, que les ASM avaient eu pour effet d'améliorer leurs aptitudes en solution de problèmes ouverts.

17

Résultats d'apprentissage et structure des cours

Voici quels étaient les résultats d'apprentissage du module :

- Appliquer un processus prescrit servant à résoudre des problèmes complexes, contextualisés et axés sur la clientèle (mal définis, contraintes nombreuses, problématiques, information inconnue).
- 2) Créer un modèle et une analyse convenables sur le plan qualitatif puis les appliquer pour résoudre les problèmes.
- 3) Communiquer efficacement de l'information technique suivant un format prescrit et conformément à des techniques et des règles grammaticales normalisées.
- 4) Appliquer des concepts, notamment les principes de la santé et de la sécurité au travail, l'économie, le droit et l'équité, à des problèmes techniques.
- 5) Discerner un dilemme éthique simple puis le résoudre en appliquant des codes d'éthique professionnelle et des normes techniques.
- Appliquer des principes de pensée critique et créative à la solution de problèmes contextualisés.
- 7) Appliquer un outil de modélisation numérique (MATLAB) pour créer un modèle servant à régler des problèmes complexes.

Le module a été structuré pour aider les étudiants à acquérir de la confiance et des aptitudes dans la solution de problèmes techniques complexes, où toute l'information n'est pas connue, il y a ambiguïté et les buts ne sont pas nécessairement définis de façon claire. Au cours de la plupart des semaines, le cours d'une heure/semaine se déroulait selon une structure semblable à celle décrite ci-après :

- 1) Le chargé de cours présentait un problème récent ou un article d'actualité en lien avec l'objectif du cours; dans certains cas, les étudiants devaient en équipe donner suite au problème posé à l'aide d'un système de réaction de l'auditoire (SRA) en ligne.
- 2) Le chargé de cours présentait ou passait en revue le problème à résoudre durant la séance de trois semaines.
- 3) Le chargé de cours dirigeait une courte discussion sur un thème lié au problème à l'étude.
- 4) Les étudiants, en petits groupes, se penchaient sur un élément particulier du problème; dans de nombreux cas, il s'ensuivait une question ouverte à laquelle les étudiants répondaient à l'aide du SRA en ligne.

De plus, les étudiants ont assisté chaque semaine à un atelier de deux heures. Au début de cet atelier, il y avait une courte épreuve portant sur les lectures précédentes où les vidéos téléversées dans MATLAB, suivie d'une courte discussion à propos d'un concept de MATLAB, en particulier. L'atelier portait essentiellement sur un problème ayant contribué aux ASM en cours. Chaque semaine, les étudiants qui menaient à bien la tâche étaient notés, histoire de les encourager à demeurer au fait de la matière du cours.

Caractéristiques et résultats des ASM

Voici quelles étaient les trois ASM :

- 1) ASM 1 : Défectuosité d'un bac à câble (semaines 2-4) : ce problème portait sur la défectuosité d'un bac à câble (consultez l'annexe 2, objectifs de l'ASM).
- 2) ASM 2 : Conception d'une éolienne (semaines 6-8) : ce problème portait sur l'analyse et la conception d'une éolienne (consultez l'annexe 3, objectifs de l'ASM 2)

3) ASM 3 : Perte de chaleur d'un immeuble (semaines 9, 11, 12) : ce problème portait sur la conception de l'isolation d'une maison en vue d'une consommation énergétique nette zéro (consultez l'annexe 4, objectifs de l'ASM 3).

Dans chaque ASM, les étudiants devaient mettre au point le modèle d'un système physique pouvant résoudre le problème présenté par un client fictif, rédiger le code de MATLAB pour installer le modèle, puis évaluer leur rapport en fonction de trois à cinq des neuf éléments de la pensée critique répertoriés dans le modèle de la pensée critique des auteurs (Paul et Elder, 2005) (clarté, exactitude, pertinence, logique, ampleur, précision, signification, exhaustivité, impartialité et profondeur) (Paul et Elder, 2005). Le tableau 1 révèle les éléments intégrés à chaque ASM. Des éléments de la pensée critique étaient explicitement ciblés dans chacun des trois exercices par les discussions des principes en classe, le recours à des activités en classe et l'intégration de celles-ci aux exigences des ASM. Durant l'un des cours axés sur l'ASM 1, les étudiants ont dressé des listes des types de questions à poser au moment d'enquêter sur un accident, ce qui a débouché sur une discussion à propos des questions à poser. En guise de réalisation définitive attendue, les étudiants devaient répertorier les types de questions qu'ils doivent poser lorsqu'ils arrivent sur les lieux d'enquête d'un accident. Dans l'ASM 2, les étudiants devaient résumer l'information pertinente, notamment évaluer sa source de même que la crédibilité, les incertitudes et les biais s'y rapportant.

Tableau 1 : Caractéristiques de chacune des ASM

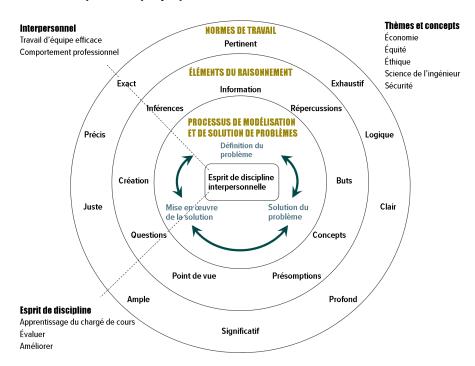
Catégorie	ASM 1	ASM 2	ASM 3
Technique	Contrainte-déformation, résistance	Écoulement des fluides, portance	Transfert de la chaleur
Conception	Définition du problème, schématisation du concept	Prise de décisions (p. ex., matrices d'évaluation pondérées)	Prise de décisions (p. ex., matrices d'évaluation pondérées)
Professionnelle	Sécurité, évaluation des risques, schémas des concepts	Associations, codes et normes	Économie, codes d'éthique, équité
Pensée critique	Poser des questions, incertitude dans l'information, repérage de l'information erronée ou contradictoire	Évaluer la crédibilité de l'information, l'argumentation, les présomptions, les inférences	Biais, inférences
Communications	Format de déclaration, utilisation de l'anglais, argumentation	Format de déclaration, utilisation de l'anglais, argumentation, concision	Format de déclaration, utilisation de l'anglais, argumentation, concision

Chaque ASM comportait cinq résultats communs ainsi que deux résultats particuliers aux tâches. Les résultats communs, exemplifiés ci-après dans le tableau 2, sont fondés sur le processus au moyen duquel les étudiants produisent leurs solutions, tandis que les résultats particuliers à chaque ASM portent sur les aspects de la pratique professionnelle. Qu'ils soient particuliers ou communs, les résultats sont structurés en une rubrique, comportant les critères de rendement établis en vue d'une évaluation cohérente selon divers scénarios.

Tableau 2 : Résultats communs des ASM

Résultat des ASM	Description
Résumé de l'information	Résume avec exactitude l'information pertinente se rapportant au problème (information sur les antécédents, le contexte, le contenu et la méthode), et englobe une évaluation de la crédibilité, de l'incertitude et des biais de l'information ainsi que de sa source.
Production de modèles	Crée des modèles quantitatifs, compare ceux-ci et les met en contraste dans MATLAB au moyen des approximations et présomptions produites à partir d'un processus justifié de solution de problèmes, étayé par de l'information.
Résultats des modèles	Évalue la validité du modèle et de ses résultats quant à l'erreur et à l'incertitude, en tirant des conclusions bien appuyées en vue de soutenir et de raffermir la solution.
Évaluation critique	Fait une évaluation critique des conclusions, compte tenu des normes intellectuelles de clarté, de précision, d'exactitude, de pertinence, de logique, d'ampleur, de profondeur, de signification, d'exhaustivité et d'impartialité.
Argumentation	Appuie de façon rationnelle les allégations et conclusions à l'aide des données et d'une description exhaustive du contexte dans lequel elles s'appliquent.
Communication	L'information est présentée de façon claire et concise; elle révèle le recours constant aux conventions de déclaration importantes en génie et d'ordre technique, dont les choix de la structure, du contenu, de la présentation et du style.

Le cadre conceptuel du module se trouve dans le graphique 5. La conception des activités du cours avait pour objet de présenter les aptitudes liées au travail d'équipe, d'encourager les étudiants dans leur apprentissage, et de recourir à des processus pour résoudre un problème et continuer d'améliorer ceux-ci (esprit de discipline). Ces notions se situent au cœur des cercles ci-après. Les étudiants étaient invités à appliquer les éléments du raisonnement aux problèmes à résoudre (dans l'anneau du milieu, ci-après), dont certains ont fait l'objet de discussions en classe. Les étudiants ont évalué leur propre présentation par rapport aux nombres de travail, lesquelles figurent dans l'anneau le plus grand.



Graphique 5 : Cadre conceptuel employé pour le cours APSC100

Méthode et procédure

Aperçu de la conception et des variables de l'étude

Au cours de la session d'automne de l'année scolaire 2012-2013, trois outils ont servi à évaluer les aptitudes en pensée critique (APC) des étudiants de première année en génie. Les évaluations de la pensée critique par l'ICTET et le CLZ ont servi de test préalable et postérieur pour définir des points de repère concernant les APC des nouveaux étudiants de première année et déterminer l'efficacité de l'enseignement des ASM en vue de développer les APC des étudiants dans le cadre du cours APSC 100. Dans notre étude, le programme d'études comportant des ASM constitue la variable indépendante et les APC des étudiants forment la variable dépendante.

Tous les étudiants de première année étaient invités à prendre part à l'étude élargie, laquelle avait obtenu l'approbation du conseil d'éthique en recherche générale (GREB) de l'Université Queen's. L'échantillonnage stratifié a servi à l'attribution de divers outils des tests préalables et postérieurs d'après une conception au sein de la matière. Ces évaluations des APC font partie des exigences du cours, si bien que le taux de participation frôlait 100 %.

Étant donné les similitudes dans les caractéristiques ou le milieu d'apprentissage des nouveaux étudiants, nous avions prévu collaborer avec des programmes de génie de trois autres universités pour évaluer la pensée critique à l'aide de l'un des outils, au début et à la fin de la session d'automne, en 2012-2013. Malheureusement, en raison de difficultés reliées au recrutement, nos collaborateurs n'ont pu obtenir des volontaires en nombre suffisant pour fournir un échantillon comparable sur le plan statistique, en guise de groupe témoin. Les volontaires étudiants de la première année au département de physique de l'Université Queen's, qui devaient suivre un programme d'études semblable, étaient invités à prendre part à l'étude en

tant que groupe témoin. Hélas, les difficultés de recrutement ont fait de nouveau surface : seul un nombre restreint de participants se sont portés volontaires. Ces derniers ont servi de groupe témoin dans le cadre des entretiens de réflexion à voix haute.

Outils de l'étude

Aux fins de notre étude, nous avons tenu compte avec circonspection de divers facteurs avant de sélectionner les outils d'évaluation des APC. Après avoir mûrement réfléchi aux enjeux touchant l'achat, l'application et l'évaluation de chaque test, nous avons opté pour les outils que sont le test de pensée critique Cornell : niveau Z (CLZ), le test international de la pensée critique à questions ouvertes (ICTET) et les protocoles de réflexion à voix haute afin d'évaluer les aptitudes en pensée critique. L'évaluation de l'apprentissage au niveau collégial (CLA) a également servi dans le cadre de cette étude, de concert avec une autre étude financée par le Conseil ontarien de la qualité de l'enseignement supérieur (COQES) à l'Université Queen's. Chaque test a été appliqué sous forme de test préalable et postérieur, exception faite de la CLA qui n'a servi que de test préalable. Pour chaque test, des groupes de nouveaux étudiants ont été formés de façon aléatoire puis répartis en cohortes selon différents jumelages en lien avec les tests préalables et postérieurs, dans l'optique d'enquêter sur les effets potentiels du coefficient test-retest, comme le montre le tableau 3 ci-après. En outre, certains étudiants devaient participer à des problèmes de réflexion à voix haute.

Tableau 3: Méthode de mesure, groupement des cohortes et outils d'étude

Méthode	Cohortes	N ^{bre}	Préalable	Postérieur
Évaluation de l'apprentissage au niveau collégial (CLA)	А	151	Sondage par test préalable	
Test de pensée critique Cornell : niveau Z (CLZ)	В	96	Sondage par test Cornell préalable, niveau Z	Sondage par test Cornell postérieur, niveau Z
	С	84	Sondage par test ICTET préalable	Sondage par test Cornell postérieur, niveau Z
Test international de la pensée critique à questions ouvertes (ICTET)	D	109	Sondage par test ICTET préalable	Sondage par test ICTET postérieur
	E	101	Sondage par test ICTET préalable, niveau Z	Sondage par test ICTET postérieur
Protocole de réflexion à voix haute	Témoin	3	Mini-ASM A	Mini-ASM B, entretiens de sortie
naute	Expérimental	2		Sortie

22

Notes des ASM	Tous	542	Évaluation en octobre, novembre et décembre par les évaluateurs	
---------------	------	-----	---	--

Au départ, la définition des points de repère concernant le test préalable des APC a eu lieu au début de la session d'automne en 2012-2013, avant l'enseignement de la pensée critique ou des ASM. L'évaluation définitive du test postérieur des APC s'est déroulée à la fin de la session d'automne en 2012-2013, une fois terminé l'enseignement de la pensée critique ou des ASM. De plus, les étudiants étaient invités à prendre part à deux activités de réflexion à voix haute, également tenues en guise de tests préalables et postérieurs afin d'évaluer les APC de ces étudiants.

Dans les graphiques suivants qui représentent les différents groupes à l'étude, **X** correspond à l'intervention (c.-à-d. le programme d'études comportant des ASM), alors que la mesure ou la constatation des APC (c.-à-d. un test préalable ou postérieur, une séance d'entretien) est explicitement formulée.

Groupe CLA

Le groupe CLA se compose de 151 étudiants qui ont passé la CLA en guise de test préalable sans passer de test postérieur connexe, comme le montre le graphique 6. Comme nous l'avons déjà mentionné, cette situation était attribuable à la CLA, laquelle intervenait dans une autre étude financée par le COQES, au temps considérable à consacrer pour recourir à la CLA, et au fait que la possibilité de test en lien avec la CLA ne s'est pas manifestée durant le délai d'exécution de l'étude. Cela dit, nous avançons l'hypothèse selon laquelle les notes de la CLA peuvent possiblement servir d'indicateur prévisionnel du rendement des ASM, et nous avons présenté des sous-échelles d'évaluation.

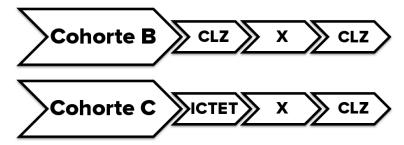
Graphique: 6: Structure du groupe CLA



Groupe CLZ

Le groupe CLZ est ensuite divisé en deux sections, la cohorte B et la cohorte C (lesquelles comprennent 97 et 84 étudiants, respectivement). La cohorte B passera le CLZ en guise de test préalable, puis de test postérieur. La cohorte C passera l'ICTET comme test préalable, puis le CLZ comme test postérieur (graphique 7). Les résultats du test postérieur de la cohorte B ne devraient pas différer de façon significative de ceux du test postérieur de la cohorte C, étant donné que les étudiants des deux sections présentent des caractéristiques comparables. Toutefois, s'il n'en était pas ainsi dans les résultats, il est possible que les effets du coefficient test-retest se fassent sentir dans le CLZ.

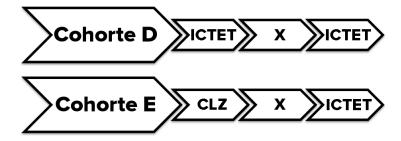
Graphique 7 : Structure du groupe CLZ



Groupe ICTET

Le groupe ICTET est ensuite divisé en deux sections, la cohorte D et la cohorte E (lesquelles comprennent 109 et 101 étudiants, respectivement). La cohorte D passera l'ICTET en guise de test préalable, puis de test postérieur. La cohorte E passera le CLZ comme test préalable, puis l'ICTET comme test postérieur (graphique 8). Les résultats du test postérieur de la cohorte D ne devraient pas différer de façon significative de ceux du test postérieur de la cohorte E, étant donné que les étudiants des deux sections présentent des caractéristiques comparables. Toutefois, s'il n'en était pas ainsi dans les résultats, il est possible que les effets du coefficient test-retest se fassent sentir dans l'ICTET.

Graphique 8: Structure du groupe ICTET

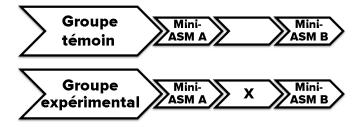


Groupes de réflexion à voix haute

Un groupe de trois étudiants, provenant du groupe d'étudiants de première année en génie, sera affecté à la condition expérimentale, c.-à-d. la participation au programme d'études comportant des ASM. Un autre groupe de trois étudiants, issu du groupe d'étudiants de première année en physique, sera affecté à la condition témoin, c.-à-d. aucune exposition au programme d'études comportant des ASM (graphique 9).

Les deux groupes d'étudiants devront résoudre une tâche ouverte (mini-ASM A, annexe 5) en guise de test préalable, puis une autre tâche ouverte à titre de test postérieur (mini-ASM B, annexe 6). De façon idéale, les membres soumis aux deux conditions demeureront les mêmes, à partir du test préalable jusqu'au test postérieur. Dans le cadre du test postérieur, l'observation du groupe expérimental devrait différer significativement de celle du groupe témoin en ce qui touche la qualité des APC des étudiants, car les étudiants soumis à la condition expérimentale auront reçu une formation leur permettant de recourir à leurs APC pour résoudre les tâches complexes.

Graphique 9 : Division du groupe de réflexion à voix haute



Création des sondages

Pour évaluer les facteurs intrinsèques et extrinsèques qui influent sur la pensée critique et enquêter sur les perceptions des étudiants quant au développement des APC, nous avons eu recours à deux sondages distincts, lesquels furent mis en application durant les tests préalables et postérieurs. Les réponses aux sondages ont été recueillies puis codées au moyen de cartes de réponses du test Scantron et d'un numériseur EZData doté du logiciel de numérisation Remark OMR.

Le sondage du test préalable, lequel a permis d'évaluer la motivation des étudiants et leur maîtrise de l'anglais, consistait en dix questions liées à la motivation et trois questions ayant trait à la maîtrise de l'anglais (annexe 7). Les dix questions liées à la motivation provenaient de l'échelle de motivation en éducation (Vallerand et al., 1992), laquelle comporte trois questions en lien avec la motivation intrinsèque, trois questions se rapportant à la motivation intégrée et identifiée, de même que trois questions ayant trait à la motivation extrinsèque introjectée. Ces questions ont été groupées d'après la théorie de l'autodétermination des auteurs Deci et Ryan (Deci et Ryan, 1985; 2010; Ryan et Deci, 2000). Dans les trois questions portant sur la maîtrise de l'anglais, les étudiants devaient préciser s'ils avaient dû passer un examen relatif à la maîtrise de l'anglais, le test qu'ils avaient eu à passer de même que les résultats obtenus, car la maîtrise de l'anglais s'est révélée un facteur influant sur les aptitudes en pensée critique (Dunham, 1997; Rashid et Hashim, 2008).

Le sondage du test postérieur a permis d'évaluer les facteurs externes influant sur la pensée critique comme la charge de travail, de même que la perception des étudiants relative au développement des APC en ce qui touche le cheminement général à l'université et le cheminement propre au module 1 du cours APSC 100 (annexe 8). Trois questions portaient sur la charge de travail de l'étudiant dans les travaux individuels, les travaux d'équipe en groupe et les travaux individuels en groupe. Cinq questions étaient liées au cheminement à l'université, au cheminement dans le module 1 du cours APSC 100 quant aux travaux individuels et en groupe, aux séances de cours et à l'interaction avec le personnel des cours en lien avec les ASM. Quatre questions se rapportaient à la participation des étudiants relativement à la transformation et à la synthèse des idées issues des différents cours, au tutorat de même qu'à la discussion des idées provenant de différents cours avec le corps professoral, la famille ou les pairs. Le dernier élément de ce sondage avait pour objet les perceptions des étudiants au sujet des cours de première année et des activités parascolaires en lien avec le développement des APC; les étudiants devaient classer, par ordre décroissant d'importance, les cheminements en question.

Méthodologie

Application des tests

De concert avec l'évaluation des APC intégrée au cheminement dans le cours, l'application des tests s'est déroulée durant les séances prévues en laboratoire à la semaine 1 et à la semaine 12 de la session d'automne. Hormis la CLA, les tests n'ont pas été chronométrés, de sorte que les étudiants disposaient de la

plage horaire au complet pour finir de répondre aux tests et aux sondages. Tous les participants ont mené à bien le CLZ en moins d'une heure et l'ICTET en moins de 90 minutes. Des surveillants supervisaient le déroulement des tests et avaient reçu pour consigne d'aider les étudiants seulement en ce qui concerne les questions de base, sans donner suite aux questions qui risquaient d'influer sur les réponses de ces derniers.

Notation des tests

La notation de la CLA s'est faite à l'aide de la méthode de notation automatisée mise au point par le CAE, assortie des données résultantes fournies sur les notes et les sous-échelles d'évaluation. La note maximale qu'un étudiant pouvait obtenir s'établissait à 1 400 pour l'ensemble de la CLA, et à 6 pour chacune des sous-échelles d'évaluation.

La notation des tests de CLZ s'est faite au moyen des systèmes EZdata du test Scantron. La note maximale qu'un étudiant pouvait obtenir s'établissait à 52, compte tenu des suggestions des auteurs et de la méthode des « bonnes réponses seulement » privilégiée pour établir la note globale (Ennis et Weir, 1985). Cette méthode de notation a été retenue parce qu'elle est celle que les étudiants connaissent le mieux. La notation des données des sous-échelles d'évaluation s'est faite à l'aide de la même méthode des « bonnes réponses seulement » : chaque sous-échelle d'évaluation comportait une note maximale distincte qu'un étudiant pouvait obtenir, en fonction du nombre de questions lié à chaque échelle respective (déduction : 10; observations et crédibilité : 4; sens et erreurs : 11; induction : 17; présomptions : 10).

La notation de l'ICTET a nécessité une formation exhaustive, assurée par la Foundation for Critical Thinking et conformément au modèle de « formation du formateur », afin d'établir la fiabilité interévaluateurs (FI). La FI a été évaluée à l'aide du coefficient alpha de Krippendorff (Hayes et Krippendorff, 2007), puis jaugée à la fin des séances de la Foundation for Critical Thinking et après les séances de formation des évaluateurs, lesquelles précédaient l'évaluation de l'ICTET. La note maximale qu'un étudiant pouvait obtenir s'établissait à 80 pour l'ensemble de l'ICTET, et à 10 pour chacune des sous-échelles d'évaluation.

Notation des ASM

Étant donné le grand nombre d'étudiants inscrits au module 1 de l'APSC 100, il y avait neuf évaluateurs chargés de la notation des ASM; ces évaluateurs devaient donner une note aux mêmes étudiants tout au long des ASM. Afin d'assurer durablement la fiabilité des évaluateurs, une séance d'étalonnage s'est déroulée avant l'évaluation de chaque ASM. Au cours de cette séance, le chargé de cours et les évaluateurs ont évalué un sous-ensemble de présentations par les étudiants jusqu'à ce qu'ils atteignent un consensus et que chaque évaluateur se sente à l'aise d'accorder une note en fonction des critères formulés dans la rubrique. Chaque ASM a été notée au moyen d'une rubrique aux ASM; celles-ci comportaient six résultats communs ainsi que deux résultats propres à chaque ASM (annexes 2, 3 et 4).

Après la séance d'étalonnage, les évaluateurs ont noté chaque présentation des étudiants dans les ASM en toute autonomie. La moyenne et l'erreur-type des notes et sous-notes de chaque évaluateur relativement aux ASM ont fait l'objet d'un suivi : les évaluateurs devaient valider auprès du chargé de cours les présentations des étudiants qui constituaient des valeurs aberrantes pour, d'une part, garantir la cohérence et l'exactitude et, d'autre part, éviter l'inflation des notes.

Séances de réflexion à voix haute

Dans le cadre de notre étude, nous avons mis au point une tâche simulant les ASM dans le cadre du cours APSC 100, mais à une échelle nettement moindre. Ces mini-ASM présentent un scénario ouvert et mal défini dont le système est axé sur un problème de physique fondamentale. Dans ce scénario, les sujets devaient évaluer le système puis présenter des solutions en vue d'une demande particulière, ainsi que

d'autres recommandations qu'ils jugeaient opportunes en matière de sécurité. Les sujets avaient obtenu de l'information supplémentaire sur divers niveaux d'authenticité, de fiabilité et de crédibilité pour les aider à formuler leurs recommandations. Ils disposaient d'une heure pour résoudre le problème tout en « réfléchissant à voix haute » à leurs solutions. Une fois l'heure écoulée, ils devaient présenter leurs recommandations.

Au début des séances, les étudiants ont rencontré l'animateur et pris connaissance des attentes relatives aux exercices de réflexion à voix haute. Ils sont ensuite passés par un exercice d'échauffement, où ils devaient apporter des améliorations à un appareil électroménager ordinaire ou à une machine simple (p. ex., une bicyclette, une machine à laver) puis présenter le tout à l'animateur. Une fois terminé l'exercice d'échauffement, la séance de réflexion à voix haute s'est amorcée.

Test préalable de réflexion à voix haute

Dans le test préalable de réflexion à voix haute, les sujets devaient fournir des recommandations en matière de sécurité à un conseil municipal quant à la pente pour toboggan proposée par celui-ci en vue d'un festival hivernal. L'équipe avait reçu un courriel du conseil municipal dans lequel celui-ci décrivait sa demande de même que les détails du problème. Les sujets avaient obtenus de la documentation supplémentaire pour les aider à traiter le problème :

- 1) Une opinion indépendante sur la sécurité du toboggan
- 2) Un article de journal portant sur la sécurité dans la pratique du toboggan
- 3) Une liste des coefficients de frottement, dressée par des étudiants
- 4) L'extrait d'un ouvrage à propos des coefficients de frottement
- 5) De l'information sur la masse moyenne des enfants et adultes américains
- 6) Un article scientifique sur la tolérance des êtres humains et leur capacité de survie à un accident
- 7) Une fiche d'équations en physique

Dans ce scénario, les principaux point de contestation se rapportaient à la provenance de l'information supplémentaire et au fait que les paramètres initiaux fournis se traduisaient par des conditions très risquées dans la pente. Le premier courriel en provenance du conseil municipal comportait la référence non sollicitée d'un conseiller municipal ayant des antécédents militaires en ce qui touche la tolérance de l'être humain relativement aux collisions. Les sujets auraient dû déterminer, d'une façon ou d'une autre, que cette référence, de même que l'opinion indépendante sur la sécurité en toboggan, des extraits de l'article de journal et la liste des coefficients de frottement dressée par des étudiants, était potentiellement incertains. Au bout du compte, les sujets auraient dû se rendre compte que la pente, en ses conditions initiales, était très dangereuse, et ils auraient dû proposer des recommandations résultant d'une analyse éclairée.

Test postérieur de réflexion à voix haute

Dans le test postérieur de réflexion à voix haute, les sujets devaient présenter des recommandations quant à la sécurité d'un parc d'attractions – FunZone Amusements – relativement à la montagne russe prototype proposée. L'équipe avait reçu un courriel de l'entreprise, décrivant sa demande et les détails du problème. Les sujets avaient obtenu de la documentation supplémentaire pour les aider à régler le problème :

- 1) Un résumé des normes de la société américaine de mise à l'essai et des matériaux (ASTM) relativement à la conception des appareils dans les parcs d'attractions
- 2) Des articles scientifiques sur les montagnes russes et la force G
- 3) Des rapports et articles sur la sécurité des montagnes russes
- 4) Une fiche d'équations en physique

27

Dans ce scénario, les principaux points de contestation se rapportaient à l'information conflictuelle de la documentation supplémentaire, à l'interprétation d'une norme professionnelle de même qu'aux paramètres des montagnes russes. Le courriel initial reçu de l'entreprise comportait certaines lignes directrices non sollicitées quant aux dimensions des rails et à la vélocité moyenne. De telles dimensions de référence risquaient d'entraîner des conditions dangereuses pour les passagers. Après avoir analysé en profondeur la documentation supplémentaire, les étudiants auraient dû intégrer les nombreuses références à la création d'un argument convaincant. De plus, la question liée à la réutilisation des chariots existants et des contraintes a entravé la capacité des étudiants à interpréter l'information dans un diagramme technique puis à classer leur solution conformément à une norme professionnelle.

Analyses des données et des statistiques

Une fois rendues anonymes, toutes les données de l'étude ont été analysées au moyen de l'outil IBM SPSS Statistics 21 et de diverses techniques paramétriques et non paramétriques. Les renseignements détaillés sur les techniques d'analyse des éléments particuliers de l'étude figurent dans leurs sections respectives.

Tests de la pensée critique

Pour évaluer les gains des APC, les notes et sous-notes préalables et postérieures permettant de mesurer certains éléments de la pensée critique (tableau 4) ont été mises en comparaison à l'aide de différents tests t jumelés pour tous les groupes à l'essai. Des tests t indépendants, permettant de comparer les notes des tests postérieurs et la différence des notes (postérieures et préalables) entre les cohortes ayant passé des tests postérieurs semblables, étaient appliqués en guise de mesure de validation de la conception expérimentale et à titre d'élément de la fiabilité des tests. Parmi les autres mesures de fiabilité des tests, il y avait le coefficient alpha de Cronbach et la corrélation des notes préalables et postérieures chez les groupes ayant passé des tests préalables et postérieurs uniformes.

Les sous-notes relatives au CLZ ont mis en évidence le dédoublement des mesures quant aux éléments d'induction et de présomptions du modèle de pensée critique utilisé dans le cadre de ce test. Le coefficient alpha de Cronbach a été calculé afin d'évaluer la fiabilité interne de ce groupement.

Test	Éléments des sous-notes							
CLA	Raisonnement analytique	Efficacité de la rédaction	Techniques de rédaction	Solution de problèmes				
CLZ	Induction	Déduction	Observation et crédibilité	Sens et erreurs	Présomptions			
ICTET	But	Questions	Information	Conclusions	Concepts	Présomptions	Répercussions	Point de vue

Sondages

Les résultats des sondages ont servi à évaluer la motivation des étudiants qui passent le test (au moyen des questions relatives à l'échelle de motivation en éducation [Vallerand et al., 1992]) de même qu'à évaluer les perceptions des étudiants quant à certains cheminements dans les cours et facteurs externes permettant de développer les APC. Les questions des sondages en lien avec la motivation étaient réparties en des catégories permettant de jauger la motivation intrinsèque, externe et intégrée. Par la suite, les réponses

obtenues à l'aide de l'échelle de Likert à cinq points ont été recodées selon une échelle à trois points pour approfondir la comparaison et l'analyse. Les questions ayant trait aux facteurs externes qui influent sur la pensée critique ont fait l'objet d'un traitement semblable : elles ont été ramenées à une échelle de deux points là où c'était nécessaire. Toutes les échelles et réponses figurent ci-après dans le tableau 5. Les étudiants devaient également classer certains cheminements dans le cours, au sein du programme d'études de première année en génie, et des activités parascolaires, selon leur importance dans le développement des APC (« À votre sens, qu'est-ce qui a favorisé le développement du type de pensée utilisé dans le test postérieur de la pensée critique au cours des trois derniers mois? »).

Un test *t* indépendant a servi à enquêter sur les différences entre le statut d'anglais langue seconde (ALS) et d'anglais langue maternelle (ALM) relativement aux APC, en fonction du rendement dans le test postérieur. En ce qui touche les échelles recodées, le coefficient alpha de Cronbach a servi à évaluer la cohérence par rapport à l'échelle principale à cinq points. Le test H de Kruskal-Wallis a servi à établir les différences entre les APC et les sous-notes respectives (évaluées d'après la note obtenue dans le test postérieur) de même que la motivation ou les facteurs influant sur les APC.

Tableau 5 : Descriptions des échelles Likert des sondages (annexes 7 et 8)

Échelle de Likert	1	2	3	4	5
à 5 points	tout à fait en désaccord	en désaccord	neutre	d'accord	tout à fait d'accord
à 5 points, charge de travail	de 0 à 5 heures	de 6 à 10 heures	de 11 à 20 heures	de 20 à 30 heures	30 heures et +
à 4 points, participation	très souvent	souvent	parfois	jamais	
à 3 points	en désaccord	neutre	d'accord		
à 2 points	en désaccord	d'accord			

Activités suscitant des modèles (ASM)

Une analyse de la variance des mesures répétées a été faite pour déterminer les différences dans les trois ASM et les sous-notes respectives, comme le montre le tableau 6, au cours de la session d'automne. Les facteurs motivants et supplémentaires ayant une incidence sur le rendement des ASM ont également été approfondis, au moyen des catégories de motivation recodées à trois points, à l'aide du test H de Kruskal-Wallis. Le lien entre les notes des ASM et celles des tests de la pensée critique (notes du test postérieur pour le CLZ et l'ICTET; note du test préalable pour la CLA) de même qu'entre les sous-notes des tests de la pensée critique et des ASM a été évaluée au moyen du coefficient de corrélation de produit-moment de Pearson. Les rapports entre les sous-notes et les notes des ASM ont également été approfondis afin d'évaluer la cohérence et la corrélation à l'interne des sous-notes et notes des ASM.

Tableau 6 : Sous-notes générales et particulières des ASM

					Sous-notes			
ASM			Éléme	ents commur	ıs		Éléments propre	es aux ASM
1	Résumé de l'information	Production de modèle	Résultats du modèle	Évaluation critique	Argumentation	Communication	Processus proposé	Analyse de sécurité
2	Résumé de l'information	Production de modèle	Résultats du modèle	Évaluation critique	Argumentation	Communication	Lettre d'accompagnement	Solution de rechange à la production d'électricité
3	Résumé de l'information	Production de modèle	Résultats du modèle	Évaluation critique	Argumentation	Communication	Lettre d'accompagnement	Raisonnement éthique

Séances de réflexion à voix haute

Les vidéos des séances préalables et postérieures d'entretien ayant trait à la réflexion à voix haute ont fait l'objet d'une retranscription et d'annotations. L'équipe de recherche a réparti les retranscriptions de la réflexion à voix haute en cinq segments, chacun ayant trait à un enjeu particulier sur lequel le groupe s'est penché. L'équipe de recherche a choisi le segment des recommandations de sécurité pour en faire une analyse approfondie, car elle estimait que celui-ci révélait la plus importante quantité d'éléments relatifs à la pensée critique. Elle a ensuite codé ce segment en fonction des éléments de la pensée critique, conformément au modèle de Paul-Elder, comme le montre le tableau 7. Elle a évalué la qualité de chaque élément, en répartissant les codes selon leur qualité acceptable ou inacceptable. L'évaluation de la qualité s'est faite par la comparaison de chaque code avec l'indicateur de l'élément correspondant; si la réponse était négative par rapport à l'indicateur, l'élément était jugé de basse qualité, et il était jugé de grande qualité dans les cas où la réponse était positive.

Une fois le codage terminé, l'équipe de recherche a analysé le segment des recommandations de sécurité afin d'y déceler les thèmes communs et les différences notables entre les tests préalables et postérieurs des groupes expérimental et témoin. Ces deux groupes ont également faits l'objet d'une comparaison et d'une mise en contraste.

Tableau 7 : Codage selon le modèle de Paul-Elder relatif à la pensée critique

Éléments	Indicateurs
But	Les participants ont-ils clarifié le but de la tâche confiée? Les participants se sont-ils renseignés sur le but de la documentation supplémentaire fournie?
Questions	Les participants ont-ils clarifié les questions auxquelles ils étaient censés répondre? Les participants ont-ils un plan d'action pour répondre aux questions répertoriées?
Points de vue	Les participants ont-ils demandé s'il existait d'autres points de vue pertinents à prendre en compte? Les participants se sont-ils renseignés sur les points de vue exprimés dans la documentation supplémentaire fournie?

Présomptions	Les participants ont-ils recensé ou remis en question leurs propres présomptions? Les participants ont-ils cherché à connaître la mesure dans laquelle leurs propres présomptions étaient valides? Les participants ont-ils décelé les présomptions des auteurs de la documentation supplémentaire fournie?
Information	Les participants ont-ils décelé l'information manquante? Les participants ont-ils cherché à savoir comment ils pourraient obtenir l'information dont ils avaient besoin? Les participants ont-ils remis en question la source de l'information à l'appui?
Concepts	Les participants ont-ils cherché à savoir si l'hypothèse ou le concept pris en compte s'appliquait à la situation donnée? Les participants ont-ils cherché à savoir s'il y avait d'autres théories ou principes permettant une explication améliorée de la situation donnée?
Conclusions	Les participants ont-ils cherché à savoir si leurs conclusions étaient appuyées par leur analyse ou la documentation supplémentaire? Les participants ont-ils cherché à savoir s'il y avait d'autres conclusions qui pourraient également concorder avec les données?

Résultats

Résultats des tests préalables et postérieurs

Groupe CLA: Cohorte A

Le résumé des notes obtenues par la cohorte A au test préalable de la pensée critique figure ci-après dans le tableau 8. Les étudiants de la cohorte (M = 1~204; ET = 161) ont obtenu une note supérieure à la moyenne des écoles américaines (M = 1~050; ET = 97) participant à l'application de la CLA (156 établissements). De plus, les résultats du groupe CLA relativement aux quatre sous-notes en lien avec le raisonnement analytique, la solution de problèmes, les techniques de rédaction et l'efficacité de la rédaction surpassaient ceux des écoles américaines (CAE, 2012).

Tableau 8 : Notes et sous-notes moyennes du groupe CLA

	N ^{bre}	Note du test		Raisonnement analytique		Solution de problèmes		Techniques de rédaction		Efficacité de la rédaction	
		Moyenne	ET	Moyenne	ET	Moyenne	ET	Moyenne	ET	Moyenne	ET
Cohorte A	151	1 204	161	3,6	0,9	3,5	0,9	3,8	0,8	3,7	0,9
Établissements américains		1 050	97	2,9	0,8	2,7	0,8	3,2	0,9	2,9	0,8

Groupe CLZ: Cohorte B

Tableau 9 : Groupe CLZ : Notes et sous-notes moyennes de la cohorte B

	N ^{bre}	Note du test		Déduction		Sémantique et sens		Crédibi	lité	Induction		Présomptions	
		Moyenne	ET	Moyenne	ET	Moyenne	ET	Moyenne	ET	Moyenne	ET	Moyenne	ET
Test préalable	96	30,90	4,51	7,63*	1,12	4,94*	1,77	2,62	1,04	10,42	1,71	5,29	1,94
Test postérieur	96	30,47	5,77	7,09*	1,39	5,48*	1,95	2,70	1,02	9,91	2,43	5,29	1,86

Groupe CLZ: Cohorte C

Les étudiants de la cohorte C ont passé l'ICTET en guise de test préalable et le CLZ en guise de test postérieur. Chez ce groupe, le but principal consistait à évaluer les effets du coefficient test-retest relativement au CLZ. Le résumé des notes au test préalable de la pensée critique (M = 4,78; ET = 1,41) et du test postérieur de la pensée critique (M = 30,13; ET = 4,99) figure ci-après dans le tableau 10; au total, 84 étudiants ont mené à bien les tests préalables et postérieurs. Bien que les sous-notes de chaque test aient été compilées, elles ne figurent pas ici en raison des différences dans la mesure des sous-notes en lien avec chaque test.

Tableau 10 : Groupe CLZ : Notes moyennes de la cohorte C

	N ^{bre}	Notes du test					
		Moyenne	ET				
Test préalable (ICTET)	84	4,78	1,41				
Test postérieur (CLZ)	84	30,13	4,99				

Au moyen des normes de l'utilisateur du CLZ, quant à l'évaluation des bonnes réponses seulement, la note moyenne du test postérieur déclarée dans la présente étude (M = 30,47; ET = 5,77) était comparable à deux des notes moyennes dans des études évaluant la capacité des nouveaux étudiants de premier cycle sur le plan des APC, et supérieure aux notes moyennes dans deux autres études (M = 29,8; ET = 4,4; M = 27,8; ET = 4,7; M = 25,9; ET = 4,2; M = 31,2; ET = 3,9) (Ennis et al., 1985).

Groupe ICTET: Cohorte D

Les étudiants de la cohorte D ont passé l'ICTET à la fois en guise de test préalable et de test postérieur. Le résumé des notes obtenu au test préalable de la pensée critique (M = 4,78; ET = 1,06) et au test postérieur de la pensée critique (M = 4,60; ET = 1,32) figure ci-après dans le tableau 11; au total, 109 étudiants ont mené à bien les tests préalables et postérieurs. L'évaluation des sous-notes relative à l'ICTET ne révèle pas de différence significative entre le test préalable et le test postérieur.

Tableau 11 : Groupe ICTET : Moyenne des notes et des sous-notes de la cohorte D

	N ^{bre}	Note au t	est	But		Questi	ons	Informa	ation	Conclus	sions	Conce	epts	Présomp	tions	Répercus	ssions	Points vue	
		Moyenne	ET	Moyenne	ET	Moyenn	e ET	Moyenn	e ET	Moyenn	e ET	Moyenn	e ET	Moyenne	ET	Moyenne	ET	Moyenn	e ET
Test préalable.	109	4,78 1	,06	6,77	1,73	5,48	2,16	5,52	2,13	5,70	1,93	4,00	1,92	2,44	1,19	3,60	2,04	4,78	1,94
Test postérieur	109	4,60 1	,32	6,28	1,89	5,59	1,95	5,47	1,70	5,32	2,09	3,90	1,88	2,68	1,69	3,24	2,14	4,32	2,28

Groupe ICTET: Cohorte E

Les étudiants de la cohorte E ont passé le CLZ en guise de test préalable et l'ICTET en guise de test postérieur. Chez ce groupe, le but principal consistait à évaluer l'incidence du coefficient test-retest en lien avec l'ICTET. Le résumé des notes obtenues au test préalable de la pensée critique (M = 30,13; ET = 4,25) et au test postérieur de la pensée critique (M = 4,67; ET = 1,21) figure ci-après dans le tableau 12. La moyenne de la note obtenue au test préalable s'établissait à 30,17 ($\pm 4,25$), contre 4,67 ($\pm 1,21$) au test préalable; au total, 101 étudiants ont mené à bien les tests préalables et postérieurs. Bien que les sous-notes de chaque test aient été compilées, elles ne figurent pas ici en raison des différences dans la mesure des sous-notes en lien avec chaque test.

Tableau 12 : Groupe ICTET : Moyenne des notes de la cohorte E

	N ^{bre}	Note du test					
		Moyenne	ET				
Test préalable (CLZ)	101	30,17	4,25				
Test postérieur (ICTET)	101	4,67	1,21				

Ni les mesures ni la comparaison du rendement n'ayant trait à l'ICTET n'ont été établies, faute d'un nombre suffisant d'études parues faisant appel à l'ICTET.

Validation de l'hypothèse

Après comparaison au moyen des tests *t* jumelés, il n'y avait pas de différence significative entre les APC des cohortes ayant le même test postérieur, comme le montre le tableau 13 ci-après. Cet état de choses tend à indiquer la solidité de la mesure des APC au moyen du CLZ ou de l'ICTET; il est possible que l'incidence éventuelle du coefficient test-retest ne constitue pas un facteur aussi important que nous l'avions imaginé.

Tableau 13 : Validation de l'hypothèse : Comparaison des notes des cohortes au test postérieur

Cohorte	Note au test de	Note au test de la pensée critique			
	Moyenne	ET			
Cohorte B (CLZ-CLZ)	Postérieur	30,47	5,77		
Cohorte C (ICTET-CLZ)	Postérieur	30,13	4,99		
Cohorte D (ICTET-ICTET)	Postérieur	4,60	1,32		
Cohorte E (CLZ-ICTET)	Postérieur	4,67	1,21		

Résultats des ASM

Les notes des ASM pour tous les étudiants, examinées par des analyses de la variance à mesures répétées au moyen d'un test ultérieur rajusté Bonferroni, ont produit des hausses statistiquement significatives au cours de la session à p<0,005, comme le montre le tableau 14. Les sous-notes relatives aux ASM communes que sont le résumé de l'information, la production de modèles, l'évaluation critique et l'argumentation ont révélé des hausses statistiquement significatives au cours de la session. Les résultats des modèles et la communication ont révélé une hausse statistiquement significative dans les deux premières ASM, examinées par des analyses de la variance à mesures répétées au moyen d'un test ultérieur rajusté de Bonferroni. Aucune différence n'a été constatée entre la deuxième et la troisième ASM.

Tableau 14 : ASM et comparaison des sous-notes aux ASM durant le cours APSC 100

ASM	SM N ^{bre}		Note aux ASM Résumé de l'information					Résultats des modèles		Évaluation critique		Argumentation		Communication	
		Moy.	E.T.	Moy.	E.T.	Moy.	E.T.	Moy.	E.T.	Moy.	E.T.	Moy.	E.T.	Moy.	E.T.
1	536	61,65	12,22	5,30	1,37	4,72	1,47	4,88	1,41	4,30	1,68	4,66	1,30	5,49	1,14
2	541	70,74	9,27	5,69	1,17	5,82	1,12	5,67	1,07	5,50	1,12	5,64	0,97	5,89	1,02
3	542	73,85	7,75	5,98	1,00	6,16	1,09	5,71	0,92	5,74	0,88	5,81	0,76	5,89	0,92

Corrélations entre les ASM et les tests de la pensée critique

Le lien entre les notes des ASM et les notes des sous-échelles d'évaluations distinctes ainsi que les tests de la pensée critique et les sous-notes s'y rapportant ont fait l'objet d'une enquête. Dans trois cas, les corrélations significatives étaient faibles entre certaines notes et sous-notes des tests de la pensée critique ainsi que les notes et sous-notes des ASM. C'est la CLA qui fait état du plus grand nombrer d'éléments en corrélation avec les notes des ASM et points des sous-échelles d'évaluation, comme le montre le tableau 15. Fait intéressant, les ASM 1 et 2 ont fait état du plus grand notre de corrélations, tandis que les ASM 3 ont fait état du plus petit nombre d'éléments en corrélation. L'ICTET a fait état d'un nombre moindre de corrélations avec les notes et sous-notes des ASM, inférieures au nombre de corrélations entre les ASM et la CLA, comme le révèle le tableau 16. Le CLZ a fait état d'un nombre restreint de corrélations significatives et négatives avec les notes et sous-notes des ASM, ainsi que du nombre le plus restreint d'éléments en corrélation avec les échelles des ASM comparativement aux autres tests de la pensée critique, comme le révèle le tableau 17.

Tableau 15 : Corrélations entre les éléments des ASM et de la CLA

		Note à la CLA	Raisonnement analytique	Solution de problèmes	Techniques de rédaction	Efficacité de la rédaction	
	Résumé de l'information	0,048	0,044	0,024	-0,045	0,085	
	Production de modèles	0,130	0,108	0,093	0,068	0,149	
	Résultats des modèles	0,178*	0,197*	0,168*	0,053	0,201*	
ASM 1 (N ^{bre} = 148)	Évaluation critique	0,277**	0,250**	0,257**	0,170*	0,280**	
	Argumentation	0,154	0,139	0,112	0,135	0,204*	
	Communication	0,202*	0,185 [*]	0,166*	0,131	0,177*	
	Note globale	0,248**	0,233**	0,206*	0,143	0,281**	
	Résumé de l'information	0,165*	0,137	0,079	0,077	0,155	
	Production de modèles	0,119	0,089	0,050	-0,095	0,067	
	Résultats des modèles	0,231**	0,214 0,183		0,099	0,206*	
ASM 2 (N ^{bre} =150)	Évaluation critique	0,217**	0,278**	0,163*	0,104	0,201*	
	Argumentation	0,275**	0,289**	0,219**	0,039	0,211**	
	Communication	0,276**	0,205*	0,194*	0,244**	0,223**	
	Note globale	0,258**	0,240**	0,167*	0,073	0,212**	
	Résumé de l'information	0,039	0,037	0,011	-0,022	0,044	
	Production de modèles	-0,174 [*]	-0,128	-0,160 [*]	-0,109	-0,130	
40	Résultats des modèles	-0,090	-0,098	-0,064	-0,143	-0,092	
ASM 3 (N ^{bre} = 151)	Évaluation critique	0,115	0,060	0,089	0,099	0,079	
	Argumentation	0,123	0,161 [*]	0,079	0,045	0,100	
	Communication	-0,044	-0,053	-0,056	-0,080	-0,032	
	Note globale	0,002	0,007	-0,016	-0,026	0,010	

^{**} La corrélation est significative à l'échelle de 0,01 (à deux queues)

^{*} La corrélation est significative à l'échelle de 0,05 (à deux queues)

Tableau 16 : Corrélations entre les éléments des ASM et de l'ICTET

		Note à l'ICTET	But	Questions	Information	Conclusions	Concepts	Présomptions	Répercussions	Points de vue
	Résumé de l'information	0,033	-0,024	0,054	-0,059	0,004	0,023	0,144	0,025	0,017
	Production de modèles	0,011	-0,024	-0,056	0,005	-0,023	0,034	-0,052	0,041	0,109
ASM 1	Résultats des modèles	-0,174	-0,149	-0,226*	-0,098	-0,110	-0,037	-0,121	-0,117	-0,087
(N ^{bre} = 82)	Évaluation critique	-0,082	-0,097	-0,158	-0,105	0,036	-0,009	-0,028	-0,152	0,051
	Argumentation	-0,061	-0,202*	-0,145	-0,013	0,036	0,065	-0,086	-0,049	0,042
	Communication	0,215	0,029	0,117	0,181	0,207*	0,227*	0,207*	0,023	0,186
	Note globale	-0,070	-0,148	-0,105	-0,035	-0,026	0,042	-0,013	-0,105	0,012
	Résumé de l'information	0,177	0,127	0,191	0,086	0,159	0,153	0,194	0,051	0,030
	Production de modèles	0,044	0,122	0,236*	0,122	0,014	0,005	-0,034	-0,070	-0,114
ASM 2	Résultats des modèles	0,147	0,066	0,175	0,057	0,041	0,102	0,061	0,204*	0,080
$(N^{bre} = 84)$	Évaluation critique	0,037	0,088	0,175	0,078	0,063	-0,011	0,004	-0,028	-0,134
	Argumentation	0,206*	0,121	0,181	0,119	0,151	0,136	0,121	0,125	0,156
	Communication	0,315**	0,244	0,292**	0,113	0,277**	0,163	0,378**	0,126	0,139
	Note globale	0,212*	0,172	0,293**	0,128	0,151	0,090	0,175	0,105	0,057
	Résumé de l'information	-0,013	-0,086	-0,046	-0,015	-0,092	0,111	-0,044	0,060	0,029
	Production de modèles	0,191	0,201*	0,198*	0,192	0,138	0,087	0,063	0,147	0,024
ASM 3	Résultats des modèles	-0,091	-0,083	-0,014	-0,145	0,012	0,050	-0,032	-0,179	-0,096
$(N^{bre} = 84)$	Évaluation critique	-0,093	-0,106	-0,045	-0,034	-0,081	0,045	-0,081	-0,040	-0,143
	Argumentation	0,058	0,026	0,067	0,096	0,111	0,113	0,025	-0,099	-0,001
	Communication	0,044	0,055	0,177	0,153	0,041	-0,013	-0,031	-0,115	-0,002
	Note globale	0,031	0,007	0,086	0,077	0,047	0,097	-0,019	-0,092	-0,015

^{**} La corrélation est significative à l'échelle de 0,01 (à deux queues)

^{*} La corrélation est significative à l'échelle de 0,05 (à deux queues)

Tableau 17 : Corrélations entre les éléments des ASM et du CLZ

		Note au CLZ	Déduction	Induction	Sémantique et sens	Observations et crédibilité	Présomptions
	Résumé de l'information	-0,138	-0,053	-0,122	-0,152	-0,067	-0,034
	Production de modèles	0,039	-0,034	0,104	0,028	0,047	-0,043
	Résultats des modèles	0,146	0,109	0,155	0,041	0,133	0,052
ASM 1 (<i>N</i> ^{bre} = 97)	Évaluation critique	-0,025	-0,022	0,057	-0,169	0,080	-0,004
	Argumentation	-0,028	-0,217 [*]	0,086	-0,018	0,012	-0,027
	Communication	0,291**	0,180	0,172	0,150	0,221*	0,265**
	Note globale	0,058	-0,025	0,091	-0,062	0,071	0,104
	Résumé de l'information	0,165	0,070	0,123	0,173	0,009	0,114
	Production de modèles	0,303**	0,098	0,163	0,311**	0,161	0,242*
	Résultats des modèles	-0,069	-0,098	-0,072	0,022	0,016	-0,078
ASM 2 (N ^{bre} = 97)	Évaluation critique	0,048	-0,059	0,097	0,045	-0,067	0,056
	Argumentation	0,053	-0,101	0,042	0,072	0,061	0,077
	Communication	0,062	-0,099	0,113	-0,004	-0,030	0,138
	Note globale	0,092	-0,081	0,094	0,101	0,003	0,116
	Résumé de l'information	0,091	-0,044	0,149	0,222*	-0,158	-0,026
	Production de modèles	0,162	0,065	0,129	0,183	0,110	0,035
	Résultats des modèles	0,140	0,069	0,092	0,119	0,091	0,088
ASM 3 (<i>Nbre</i> = 97)	Évaluation critique	0,002	-0,155	0,003	0,072	-0,107	0,101
	Argumentation	-0,017	-0,078	0,006	0,023	-0,116	0,038
	Communication	0,145	0,079	0,113	0,077	0,066	0,127
	Note globale	0,185	0,020	0,195	0,170	0,018	0,117

^{**} La corrélation est significative à l'échelle de 0,01 (à deux queues)

Fiabilité des ASM

La fiabilité de chaque ASM a été évaluée par les corrélations totales des éléments entre les sous-notes et la note globale des ASM ainsi que le coefficient alpha de Cronbach. Les résultats des corrélations totales des éléments (sous-notes par rapport à la note totale) et de la cohérence interne, mesurée par le coefficient alpha

^{*} La corrélation est significative à l'échelle de 0,05 (à deux queues)

de Cronbach, figurent ci-après dans le tableau 18. Chaque ASM a révélé une cohérence interne élevée $(0,7<\alpha<0,9)$ ainsi que de fortes corrélations significatives avec la note totale $(r>0,5;\ p<0,005)$. Comme nous l'avons mentionné précédemment, nous nous servons des ASM en guise de mesure relative au cours quant au développement et à l'évaluation des aptitudes en pensée critique et en solution de problèmes. De telles statistiques révèlent dans quelle mesure chaque sous-note contribue bien à la note globale des ASM (corrélations totales des éléments) et en quoi chaque sous-note est fiable et étroitement liée par rapport à la note globale de l'ASM (coefficient alpha de Cronbach).

Tableau 18 : Mesures de fiabilité des ASM

	Coefficient alpha de Cronbach	N ^{bre}	Résumé de l'information	Production de modèles	Résultats des modèles	Évaluation critique	Argumentation	Communication
ASM 1	0.813	536	0,623**	0,647**	0,756**	0,754**	0,797**	0,621**
ASM 2	0.802	540	0,689**	0,620**	0,762**	0,677**	0,754**	0,622**
ASM 3	0.771	542	0,602**	0,690**	0,601**	0,600**	0,732**	0,701**

^{**}La corrélation est significative à l'échelle de 0,005 (à deux queues).

Séances de réflexion à voix haute

Les résultats qui figurent dans la présente section de notre étude donnent un aperçu du codage et de l'analyse du segment des recommandations de sécurité, à partir de la retranscription des séances de réflexion à voix haute préalables et postérieures chez le groupe témoin (n = 3) et expérimental (n = 2). Ce segment a été codé en fonction des éléments de la pensée critique qui correspondent au modèle de Paul-Elder. La qualité de chaque élément a été évaluée, à l'aide de codes distingués en fonction de la qualité acceptable ou inacceptable. Une fois codé, le segment des recommandations de sécurité a fait l'objet d'une analyse quant aux thèmes communs et aux différences notables entre les tests préalables et postérieurs chez le groupe expérimental et le groupe témoin. Les deux groupes ont également fait l'objet de comparaisons et de mises en contraste qui sont présentées ci-après.

Comparaison des réflexions à voix haute préalables et postérieures : groupe témoin

Les trois sujets issus de la physique et qui formaient le groupe témoin ont fait état d'une amélioration considérable entre le test préalable et le test postérieur. Dans le test préalable, les participants abordaient le problème à l'envers : ils examinaient l'information supplémentaire puis ils formulaient des recommandations au début de la séance, avant toute analyse. Ces conclusions, présentées à titre de recommandations de sécurité, résultaient essentiellement des présomptions ancrées dans le cheminement personnel des étudiants plutôt que d'une analyse et d'une solution exhaustives quant aux principaux éléments du problème. Les sujets ont tenté par la suite de faire une analyse du scénario, mais ils ont éprouvé de la confusion et il leur manquait une méthode ou un plan clair pour régler le volet de la physique en lien avec le problème. Les sujets sont demeurés coincés à l'étape de l'analyse, au-delà du temps alloué, et ils n'ont pu présenter de conclusions ni de recommandations de sécurité découlant de cette analyse.

Dans le test postérieur, les sujets ont abordé le problème d'une façon semblable à celle manifestée par le groupe en génie : ils ont fondé essentiellement leurs conclusions sur des présomptions, en se servant peu des concepts ou de l'information supplémentaire. Les sujets ont fait appel à leur cheminement personnel et

empirique pour formuler des conclusions plutôt que d'appuyer ces conclusions sur l'information supplémentaire. La remise en question de leurs conclusions présumées faisait l'objet d'une faible validation, et ils ont formulé leurs recommandations en ne faisant référence ni aux limites de leurs connaissances et de leur savoir-faire, ni aux retombées sociétales. Cependant, ces éléments de la pensée critique figuraient dans leurs conclusions, ce qui peut être considéré comme une amélioration par rapport au rendement dans le test préalable, malgré la qualité relative de tels éléments.

Comparaison des réflexions à voix haute préalables et postérieures : groupe expérimental

Le groupe expérimental, composé de deux étudiants en génie ayant pris part au programme d'études comportant des ASM, a montré des améliorations dans les éléments de la pensée critique, notamment les domaines des concepts, de l'information et des répercussions. Dans le test préalable, les sujets ont pondéré leurs conclusions par des présomptions, tirées de leur cheminement personnel ou empirique ou de leurs connaissances préalables des concepts de physique, complétées à l'occasion par des données et de la documentation. Ils ont adopté et utilisé la documentation d'appui parallèlement à leur raisonnement présumé pour constituer leurs conclusions, sans approfondir davantage ce volet. Les sujets n'ont remis en question ni la validité, ni l'exactitude de leurs propres présomptions, et ils n'ont tenu compte que succinctement des répercussions éventuelles de leurs solutions.

Dans le test postérieur, les sujets ont fondé leurs conclusions sur leur analyse initiale, complétée par des concepts et de l'information supplémentaire. Il y avait quelques conclusions fondées sur des présomptions formées à la première personne du singulier, ou d'après le cheminement ou les données empiriques. Les concepts de la physique et des forces ont été employés de façon progressive, les sujets élaborant des rapports mathématiques entre les variables plutôt que de simplement recourir aux équations. La validité de l'information supplémentaire a été remise en question, lorsque les sujets ont tenu compte de l'autorité relative à l'information fournie. Les sujets ont également pris en compte les répercussions de leurs conclusions : dans celles-ci, ils ont mis en évidence les limites et les domaines pouvant susciter des préoccupations à traiter à l'avenir, et ils ont même décelé l'éventualité d'un conflit relativement aux codes du bâtiment.

Les sujets ont manifesté une amélioration par rapport au test préalable, notamment dans les éléments que sont les concepts, les présomptions, l'information et les répercussions. Ils ont remis en question la validité et les sources de l'information supplémentaire, de même que l'exactitude et la validé de leurs propres présomptions et conclusions, par contraste avec le test préalable où ces aptitudes se révélaient mauvaises. Les concepts ont servi au-delà de leur simple application, puisque les sujets y ont recouru de concert avec les formules mathématiques pour en arriver à un modèle mathématique mettant en lien les paramètres de conception où des recommandations étaient requises. Enfin, les sujets ont révélé une amélioration marquée dans la prise en compte des répercussions : ils ont su déceler les codes et exigences supplémentaires audelà de la portée des recommandations, et ils ont décelé aisément les lacunes éventuelles à combler dans leurs connaissances.

Analyse du sondage

Sous-échelles de motivation

Les dix questions sur la motivation préalables au sondage ont été transformées pour les faire passer de questions composantes à trois sous-échelles, lesquelles permettaient à chacune de sonder un type différent de motivation : la motivation intrinsèque (MI), la motivation introjectée et externe (IJ/ER), de même que la motivation interne et identifiée (IN/ID). Les mesures de fiabilité (cohérente interne) de chaque sous-échelle sont les suivantes : MI : α = 0,769; IJ/ER : α = 0,706; IN/ID : α = 0,712. Les résultats des groupements par sous-échelle figurent ci-après aux tableaux 19, 20 et 21. Il ressort des réponses des étudiants que ces

derniers sont majoritairement d'accord ou tout à fait d'accord avec toutes les questions de motivation, ce qui les place dans un groupement à motivation élevée. Une analyse de corrélation a été effectuée pour enquêter sur les liens entre la motivation et le rendement dans les ASM de même que le rendement au test postérieur de la pensée critique et les sous-notes ayant trait à chaque outil. Aucune corrélation entre la motivation, les ASM et le rendement au test de la pensée critique n'a été constatée.

Tableau 19 : Classement des participants à l'étude selon la motivation intrinsèque

Motivation intrinsèque							
		Fréquence	Pourcentage	Pourcentage valide			
	Faible motivation	47	8,7	8,8			
	Neutre	44	8,1	8,3			
	Forte motivation	441	81,4	82,9			
	Total	532	98,2	100,0			
	Données manquantes	10	1,8				
Total		542	100,0				

Tableau 20 : Classement des participants à l'étude selon la motivation introjectée et externe

Motivation introjectée et externe							
		Fréquence	Pourcentage	Pourcentage valide			
	Faible motivation	53	9,8	10,0			
	Neutre	31	5,7	5,8			
	Forte motivation	448	82,7	84,2			
	Total	532	98,2	100,0			
	Données manquantes	10	1,8				
Total		542	100,0				

Tableau 21 : Classement des participants à l'étude selon la motivation interne et identifiée

Motivation interne et identifiée							
		Fréquence	Pourcentage	Pourcentage valide			
	Faible motivation	9	1,7	1,7			
	Neutre	7	1,3	1,3			
	Forte motivation	516	95,2	97,0			
	Total	532	98,2	100,0			
	Données manquantes	10	1,8				
Total		542	100,0				

Statut d'ALS

La majorité des étudiants participant à l'étude avaient l'anglais comme langue maternelle (ALM) (n = 498), tandis qu'un faible pourcentage de ces étudiants avaient l'anglais comme langue seconde (ALS) (n = 34) et que dix étudiants ont refusé de se prononcer à ce sujet. L'échantillonnage stratifié a eu pour effet de réduire ces nombres dans les groupements selon le même test de pensée critique et la même maîtrise de l'anglais. La CLA est le seul test où il y avait une différence significative dans le rendement du test de la pensée critique d'après le statut d'ALS, (t (142) = -5,364; p<0,005), les notes des étudiants de la catégorie ALS (t (t (t (t (t (t (t))).

979,5; ET = 40,36) étant inférieures à celles des étudiants de la catégorie ALM (M = 1218,92; ET = 12,94), comme le montre le tableau 22 ci-après.

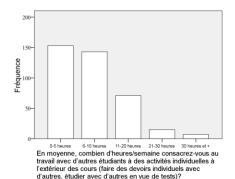
Tableau 22 : Différences dans la note au test de la pensée critique selon le statut d'ALS

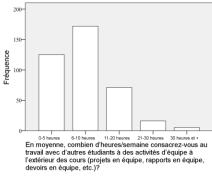
	Statut d'ALS	N ^{bre}	Moyenne	E.T.
Résultats, CLA	Anglais langue seconde	12	979,50	139,81
	Anglais langue maternelle	132	1 218,92	148,71
Résultats, test postérieur CLZ	Anglais langue seconde	10	29,60	5,56
	Anglais langue maternelle	171	30,27	5,50
Résultats, test postérieur ICTET	Anglais langue seconde	12	45,36	10,11
	Anglais langue maternelle	199	46,35	12,76

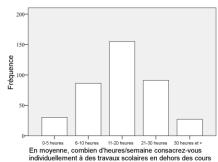
Facteurs externes influant sur les APC

Les résultats des questions postérieures à l'enquête en ce qui touche l'effort des étudiants (trois questions, graphique 10), les perceptions des étudiants quant au cheminement à l'université et dans le cours sur le développement des APC (six questions, graphique 11), de même que les perceptions des étudiants quant à l'intégration des connaissances, à la discussion et à l'apprentissage par les pairs (quatre questions, graphique 12) figurent ci-après. Les classements des étudiants dans les activités de première année axées sur les cours et parascolaires s'y trouvent également (graphique 13). Bien qu'il n'y ait pas de lien significatif entre ces questions et les APC, il est possible d'en apprendre sur les cheminements dans les cours jugés essentiels par les étudiants pour développer le type de pensée utilisé au cours du test postérieur de la pensée critique. D'après les étudiants, le travail aux ASM en contexte individuel et collectif de même que les séances de cours données dans le cadre de l'APSC 100 ont favorisé le type de pensée utilisé dans leur test postérieur de la pensée critique. De plus, lorsqu'ils devaient classer les cheminements dans les cours en première année selon le développement du type de pensée utilisé dans leur test postérieur de la pensée critique, ils ont considéré le module 1 du cours APSC 100 comme le plus important pour le développement des APC (à 45,4 %). D'autres cheminements de cours, à savoir le cours APSC 111 (Physique) (à 22,9 %) et le cours APSC 131 (Chimie) (à 22,4 %), suivaient dans le classement des cheminements dans les cours les plus importants.

Graphique 10 : Résultats des questions du sondage sur les efforts des étudiants

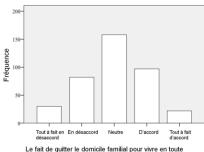




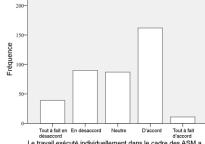


En moyenne, combien d'heures/semaine consacrez-vous individuellement à des travaux scolaires en dehors des cours (faire des devoirs individuels, étudier en vue de tests, réviser les notes de cours, etc.)?

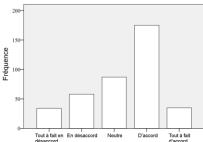
Graphique 11 : Perception des étudiants quant aux activités des cours sur le développement des APC



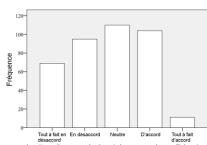
Le fait de quitter le domicile familial pour vivre en toute autonomie a contribué au développement du type de pensée auquel j'ai recouru pour passer le test postérieur de la pensée critique



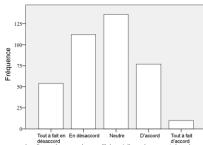
Le travail exécuté individuellement dans le cadre des ASM a contribué à développer le type de pensée auquel j'ai recouru pour passer le test postérieur de la pensée critique



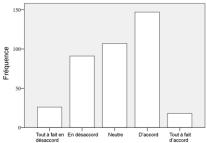
Le travail exécuté avec d'autres étudiants dans le cadre des ASM a contribué à développer le type de pensée auquel j'ai recouru pour passer le test postérieur de la pensée critique



desaccord Neutre D'accord Tout fait d'accord La rétroaction reçue du chargé de cours ou des auxiliaires à l'enseignement dans le cadre de l'APSC 100 a contribué à développer le type de pensée auquel j'ai recouru pour passer le test postérieur de la pensée critique

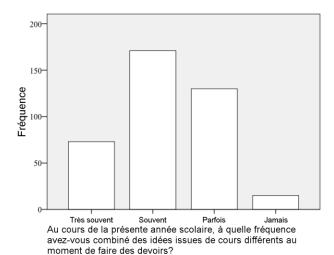


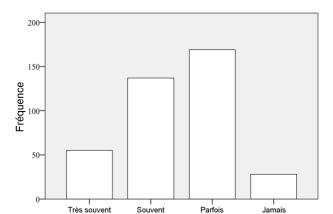
rivit a tait en En désaccord Neutre D'accord Tout à fait desaccord La discussion avec les auxiliaires à l'enseignement dans le cadre du module 1 de l'APSC 100 a contribué à développer le type de pensée auquel j'ai recour



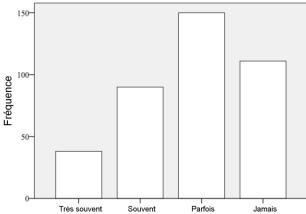
rout a fait en En désaccord Neutre D'accord Tout à fait désaccord
Les activités et la discussion en lien avec les séances de cours du module 1 de l'APSC 100 ont contribué à développer le type de pensée auquel j'ai recouru pour passer le test postérieur de la pensée critique

Graphique 11 : Perceptions des étudiants quant à l'intégration des connaissances, à la discussion du contenu et au tutorat

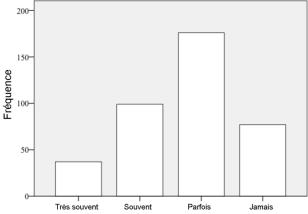




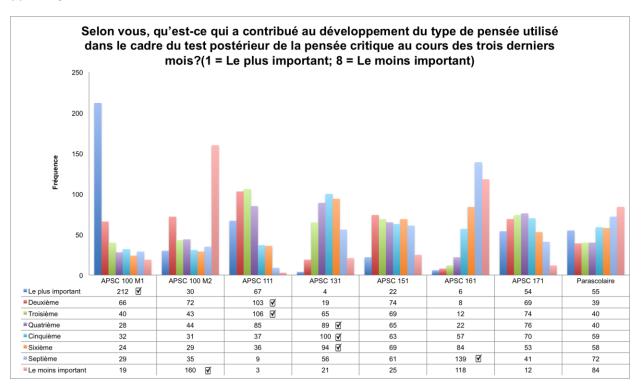
Au cours de la présente année scolaire, à quelle fréquence avezvous discuté d'idées provenant de vos lectures ou séances de cours avec d'autres à l'extérieur des cours (étudiants, membres de la famille, collègues de travail, etc.)?



Au cours de la présente année scolaire, à quelle fréquence avez-vous discuté d'idées provenant de vos lectures ou séances de cours avec d'autres membres du corps professoral?



Au cours de la présente année scolaire, à quelle fréquence avez-vous joué le rôle de tuteur ou d'enseignant auprès des autres étudiants (moyennant une rémunération ou à titre bénévole)?



Graphique 12 : Classement par les étudiants des cheminements de première année qui contribuent aux APC

Fiabilité des tests de la pensée critique

La fiabilité de chaque test a été calculée à partir des résultats de l'étude au moyen de nombreuses mesures. La fiabilité du coefficient test-retest a été évaluée au moyen du coefficient de corrélation de produit-moment de Pearson et des notes aux tests préalables et postérieurs. La cohérence a été mesurée par le coefficient alpha de Cronbach au moyen des sous-notes de test.

CLA

Puisque la CLA n'a servi qu'à titre de test préalable, il était impossible de déterminer les statistiques de fiabilité du coefficient test-retest. Le coefficient alpha de Cronbach de la CLA était $\alpha = 0,920$, calculé à l'aide des sous-notes de la CLA relativement à la solution de problèmes, au raisonnement analytique, aux techniques de rédaction et à l'efficacité de la rédaction. Nous avons pu déduire de cette mesure élevée de cohérence interne que les sous-échelles d'évaluation et les notes globales de la CLA sont très étroitement liées.

CLZ

La fiabilité du coefficient test-retest, mesurée par le coefficient de corrélation de produit-moment de Pearson, a été établie à r (95) = 0,471; p<0,005. Une corrélation de cette ampleur révèle que le coefficient test-retest n'a pas d'incidence chez la cohorte B, un fait appuyé par la comparaison des notes du test postérieur avec la cohorte E. Le coefficient alpha de Cronbach, à l'aide des sous-notes ayant trait à la déduction, à la sémantique et au sens, à l'observation et la crédibilité, à l'induction et aux présomptions, a été établi à $\alpha = 0,645$. Cette mesure modérée de cohérence interne porte à croire que les sous-échelles d'évaluation et

les notes globales du CLZ sont étroitement liées, mais elles suscitent certaines questions quant à la fiabilité de l'outil.

Test international de la pensée critiques à questions ouvertes (ICTET)

Une autre statistique de fiabilité a fait l'objet d'une enquête dans le cas de l'ICTET, en raison des nombreux évaluateurs chargés d'évaluer le test. La FI a été calculée à l'aide du coefficient alpha de Krippendorff, mesuré comme nous l'avons mentionné préalablement (Hayes et Krippendorff, 2007), à l'issue des séances avec la Foundation for Critical Thinking et après les séances de formation aux évaluateurs. Les mesures de la FI s'établissaient à $\alpha = 0.932$ et à $\alpha = 0.974$, respectivement, ce qui révèle un niveau élevé de concordance entre les évaluateurs et garantit une évaluation cohérente. Après avoir recueilli les feuilles notées pour les attribuer aux évaluateurs, un seul évaluateur a noté toutes les feuilles de la cohorte D.

La fiabilité du coefficient test-retest a été établie comme suit : r = 0,150 (n = 101). Cette faible corrélation révèle une incidence possible du coefficient test-retest relativement à l'ICTET. Toutefois, une comparaison entre les cohortes montre que le coefficient alpha de Cronbach, au moyen des sous-notes de l'ICTET quant au but, aux questions, à l'information, aux conclusions, aux concepts, aux présomptions, aux répercussions et aux points de vue a été établie à $\alpha = 0,828$, ce qui tend à indiquer une forte relation entre les sous-échelles d'évaluation et la note globale de l'ICTET.

Discussion

Les tests de fiabilité des outils des APC présentés ci-dessus s'inscrivent dans les études précédentes dont fait état la documentation, mais il n'y a pas de changement significatif dans la moyenne entre le test préalable et le test postérieur. De tels résultats ne révèlent pas la réalisation de gains sur le plan de la pensée critique au cours de la session, ce qui contraste avec les données tirées des ASM mêmes, des résultats du sondage et des analyses de la réflexion à voix haute, lesquels révèlent bel et bien des gains en matière d'apprentissage. Les auteurs estiment que les éléments d'information mettent en évidence des problèmes liés à l'enseignement, à la concordance, à l'évaluation de même qu'à l'application des tests uniformisés.

Les commentaires empiriques des surveillants portent à croire que les étudiants étaient très enthousiastes durant le test préalable et qu'ils étaient plutôt fatigués à l'issue d'une première session stressante dans un programme intense de génie en première année. Dans le contexte des sondages auprès des étudiants, lesquels ont eu lieu deux semaines avant le test postérieur au cours d'une prestation normale du programme, nous avons constaté que la moitié des étudiants avaient peine à suivre le rythme des exigences du programme. Au cours du test postérieur, les surveillants ont constaté d'importants problèmes de motivation, dont de nombreux commentaires du genre « Pourquoi faut-il passer le même test encore une autre fois? », de même qu'un grand nombre d'étudiants qui ont semblé consentir peu d'efforts au test postérieur.

Les ASM ont révélé des corrélations totales élevées par élément entre les sous-échelles d'évaluation et les notes, de même qu'une bonne cohérence interne des sous-échelles d'évaluation. Par contraste avec les résultats tirés des outils des APC, des gains significatifs au chapitre du rendement ont été constatés dans les moyennes des notes des ASM, notamment des ASM 1 aux ASM 2, tandis qu'il y avait relativement peu de changement entre les ASM 2 et les ASM 3. Des améliorations significatives ont été constatées sur les plans de l'argumentation, de l'évaluation de l'information, et de la capacité des étudiants à faire une évaluation critique de leur propre travail.

Il subsiste une question de concordance entre les tests de la pensée critique et les méthodes d'évaluation de la pensée critique en contexte de génie par le truchement des ASM. La corrélation entre les notes des ASM de même que le CLZ et l'ICTET n'était pas significative, bien que le modèle Paul-Elder de la pensée critique ait servi de cadre d'enseignement de la pensée critique dans le cours et que l'ICTET ait été établi selon un

cadre apparenté. Cependant, le modèle s'appuyait sur des tâches où les étudiants devaient discerner des points de vue, des inférences, etc., à partir d'un message-guide, plutôt que d'exiger un argument logique construit comme dans le cas de la CLA et des ASM. La CLA, laquelle fait appelle à des tâches en lien avec des réponses construites, comportait la corrélation la plus marquée avec les notes des ASM. Dans la présente étude, la similitude des tâches permet une meilleure concordance que la similitude du cadre de la pensée critique.

Le contexte, le cadre et la tâche jouent un rôle important dans l'évaluation des aptitudes générales comme la pensée critique, la solution de problèmes, etc. Or, le CLZ n'était pas conçu dans un contexte de génie. Il faisait appel à un cadre différent de celui de Paul et Elder et, sous sa forme de test à choix multiples, proposait une tâche très différente des réponses construites attendues en ce qui touche les problèmes contextuels et complexes présentés dans les activités de cours. Fait intéressant, plusieurs des échelles du CLZ étaient en corrélation négative avec le rendement dans les ASM. La pensée critique est une construction multidimensionnelle où interviennent les aptitudes, la disposition et la métacognition; le CLZ et l'ICTET proposent des messages-guides très structurés ou des choix multiples, lesquels ne prennent en compte ni la disposition, ni la métacogition.

Il vaut la peine de souligner que d'autres chercheurs n'ont pas décelé de gains d'apprentissage significatifs au cours de la session, même lorsqu'un cours porte exclusivement sur la pensée critique (ce qui n'était pas le cas dans la présente étude). Les auteurs Nieto et Saiz (2008) révèlent des améliorations du rendement dans les activités de cours (les notes), mais aucune différence en ce qui touche les notes des tests préalables et postérieurs relativement au CLZ. Les APC et les activités étaient évaluées selon le modèle de pensée critique de Halpern, ce qui montre que la différence entre le modèle des tests et le modèle d'enseignement (la concordance) comporte une incidence.

Il n'y avait pas de corrélation significative entre les échelles des sondages sur la motivation et les outils. En règle générale, les étudiants ont répondu de façon positive (« d'accord » ou « tout à fait d'accord ») à l'ensemble des éléments des sondages sur la motivation et ils n'étaient pas tenus de choisir entre les facteurs motivants. Voilà une faiblesse des éléments des sondages sur la motivation scolaire employés dans la présente étude.

En ce qui touche les problèmes de la réflexion à voix haute, il y avait des différences très perceptibles entre le groupe témoin issu du génie et celui provenant de la physique. Les sujets en génie ont eu davantage recours aux volets de la pensée critique dans leurs conclusions relatives aux recommandations de sécurité. Au cours des tests préalables, les deux groupes se sont appuyés de façon significative sur leurs cheminements personnels présumés pour tirer des conclusions. Toutefois, les sujets en génie ont utilisé d'autres éléments, par l'intégration de concepts et d'information supplémentaire à leur analyse et, au bout du compte, à leurs conclusions. Pour sa part, le groupe témoin en physique s'est fié strictement au raisonnement présumé. Malgré de telles améliorations dans la manifestation d'éléments de la pensée critique et leur mise en application, les deux groupes n'ont ni remis en question la crédibilité ou la validité de l'information supplémentaire, ni pris en compte les répercussions éventuelles de leurs recommandations.

L'un et l'autre des groupes ont révélé une amélioration dans le recours aux éléments de la pensée critique durant les tests postérieurs. Celui provenant de la physique a continué de se fier essentiellement aux cheminements personnels présumés pour tirer des conclusions, mais il s'est mis à intégrer des éléments conceptuels aux recommandations, en se servant de méthodes et de prolongements de la théorie en combinaison pour enrichir ses conclusions. De plus, il s'est mis à utiliser des documents supplémentaires pour appuyer ses conclusions, ce qui s'apparentait en quelque sorte à l'utilisation des documents supplémentaires par le groupe issu du génie dans son test préalable.

Pour sa part, le groupe en génie a révélé une diminution du raisonnement présumé, axé sur le cheminement personnel. Les conclusions présentées par ce groupe étaient étayées par des concepts, de l'information, des

répercussions ainsi qu'une évaluation des présomptions. Les sujets en génie ont fait appel à plusieurs concepts pour créer un rapport mathématique entre les paramètres de conception en vue de résoudre le problème, ce qui pouvait déboucher au bout du compte sur un modèle descriptif. Un tel recours aux concepts s'apparentait à la méthode enseignée de même qu'aux activités suscitant des modèles auxquelles les sujets étaient exposés dans le cours APSC 100. Comparativement au test préalable, les sujets en génie ont tenu compte de la crédibilité de l'information supplémentaire et fait appel à plusieurs sources pour appuyer leurs solutions et constituer le fondement des recommandations de sécurité. Le groupe en génie, au cours du test postérieur, a intégré de nouveaux éléments de la pensée critique qui n'étaient pas constatés au cours du test préalable; ses membres ont tenu compte des répercussions de leurs solutions en ce qui touche les règlements de zonage éventuels, et ils ont remis en question leurs propres présomptions et limites quant aux connaissances.

Il est possible que les différences entre ces groupes résultent des ASM et de l'enseignement de la pensée critique dans le cadre du cours APSC 100. Il se peut également qu'elles soient attribuables aux divers antécédents scolaires propres aux différents groupes. En guise de moyen pour déterminer les raisons potentielles, un court entretien de cessation a eu lieu à la fin du test postérieur de réflexion à voix haute. Les sujets devaient préciser si quelque chose à l'université ou dans leur cheminement de vie les avait aidés en ce qui touche la réflexion dans ce type de problème. Le groupe témoin a répondu que le test préalable, parallèlement à une question semblable axée sur des montagnes russes dans leur cours de physique, les avait aidés à passer le test postérieur. Les étudiants du groupe expérimental ont répondu que, selon eux, le cheminement dans le cadre du cours APSC 100 les avait aidés, et ils ont fait allusion à l'établissement d'un modèle et d'un rapport entre les variables et une solution autodéfinie ayant trait à ce type de raisonnement. Il est ressorti de la réflexion à voix haute à l'issue du cours que les étudiants se fiaient moins à leurs présomptions personnelles, qu'ils prenaient en compte d'autres points de vue et qu'ils évaluaient la validité de l'information davantage qu'au début de la réflexion à voix haute sur les cours. Les thèmes issus des études sur la réflexion à voix haute concordent bien avec les buts de l'enseignement et les gains de rendement constatés dans les sous-notes des ASM.

Le sondage auprès des étudiants a révélé que ces derniers avaient, en très grande majorité, déterminé que les activités du cours, essentiellement les ASM, constituaient le cheminement le plus déterminant dans le développement de la pensée critique. Compte tenu des gains constatés dans le rendement aux ASM et des observations tirées des entretiens de réflexion à voix haute, les auteurs en viennent à la conclusion qu'il y a discordance entre les deux outils des APC servant à l'évaluation préalable et postérieure au cours, à savoir le test de pensée critique Cornell : niveau Z (CLA) et le test international de la pensée critique à questions ouvertes (ICTET), et le cours en tant que tel.

Recommandations et recherche à l'avenir

Il résulte de la présente étude plusieurs recommandations, à partir des objectifs de recherche décrits au début du rapport :

Y a-t-il une corrélation entre les notes des outils de pensée critique et celles des ASM?

Aucune corrélation n'était constatée entre les notes des étudiants ayant trait aux outils uniformisés de la pensée critique et les notes aux ASM. D'après les auteurs, ce défaut de corrélation peut s'expliquer par la concordance entre les outils uniformisés et les ASM de même que les difficultés quant à la sensibilité des outils. La concordance sera traitée dans une recommandation ultérieure. En ce qui touche la sensibilité, il se peut que les évaluations uniformisées soient dépourvues de la sensibilité ou de la résolution en quantité suffisante pour mesurer le développement des APC au cours d'une session. Les auteurs de la CLA ne recommandent pas l'utilisation de leur test sous forme de jumelage préalable et postérieur durant la session, puisqu'ils ont constaté qu'il faut alors davantage de temps pour discerner le développement. Les auteurs de

l'ICTET ont manifesté une préoccupation semblable, quant à l'éventualité que les gains dans les APC mesurés par un outil uniformisé au cours d'une session soient seulement à titre temporaire, puisque le développement des APC se déroule habituellement au moyen d'une mise en application, d'une pratique et d'une réflexion continues (Halpern, 2002; Paul et Elder, 2005). Quant au CLA, il a servi dans plusieurs études selon des modalités préalables et postérieures, et il a révélé les gains des étudiants sur le plan des APC, mesurées par l'évaluation. Toutefois, les critiques résumées dans le présent rapport suscitent des préoccupations quant à la validité et à l'exactitude des évaluations à choix multiples des APC.

Y a-t-il une corrélation entre les sous-notes des outils de pensée critique et celles des ASM?

Certaines corrélations étaient constatées entre les sous-notes des outils uniformisés de la pensée critique et les sous-notes aux ASM. Au chapitre du plus grand nombre d'éléments en corrélation entre les sous-notes, c'est la CLA qui figurait au premier rang, suivie de l'ICTET, tandis que le CLZ occupait le dernier rang à ce chapitre. Un peu à l'exemple de ce qui a trait aux notes globales, les auteurs sont d'avis qu'il est possible d'expliquer les différences dans les éléments en corrélation par la concordance entre les outils uniformisés en particulier et les ASM, point qui sera traité dans une recommandation ultérieure.

Y a-t-il une corrélation entre la capacité de pensée critique et les facteurs motivants?

Aucune corrélation n'a été constatée entre l'aptitude en pensée critique, telle qu'elle est mesurée par les notes des étudiants aux tests uniformisés de la pensée critique, et les facteurs motivants. Les réponses des étudiants à l'outil de sondage ont révélé que la motivation de ces derniers était élevée pour toutes les sous-échelles d'évaluation de l'outil. Il est possible que cette situation soit attribuable à la seule utilisation d'un certain groupe de questions pour mesurer une construction complexe comme la motivation, plutôt que de recourir à l'outil dans son ensemble. De plus, les étudiants n'étaient tenus ni de classer, ni de comparer les facteurs motivants, ce qui a peut-être contribué à la grande motivation déclarée par les étudiants quant aux différents types de motivation dont font état les sous-échelles d'évaluation.

Y a-t-il une corrélation entre la capacité de pensée critique et certains cheminements dans les cours? Y a-t-il une corrélation entre la capacité de pensée critique et certains facteurs intrinsèques?

Ces deux questions sont traitées conjointement. À partir des déclarations volontaires des étudiants, des entretiens de réflexion à voix haute et de l'information sur le sondage, une corrélation s'établissait entre l'aptitude en pensée critique et des cheminements dans les cours en particulier. D'après un grand nombre d'étudiants, le développement de leurs aptitudes en pensée critique est le fruit du module 1 de l'APSC 100 et des ASM mêmes. À leur sens, les travaux dans les ASM et les séances de cours du module 1 de l'APSC 100 étaient bénéfiques quant à la capacité de la pensée critique. En ce qui touche les facteurs extrinsèques, les étudiants ont précisé que leur capacité de pensée critique s'était développée dans les cours de physique (APSC 11) et de chimie (APSC 131) en première année. Ils ont souligné que les activités parascolaires et le cheminement à l'université n'avaient relativement pas d'importance dans le développement de leur capacité de pensée critique.

Les outils de pensée critique utilisés sont-ils fiables et valides?

Les outils uniformisés révèlent des forces et des faiblesses différentes en ce qui touche la fiabilité et la validité. La fiabilité a été évaluée au moyen du coefficient alpha de Cronbach, de la fiabilité du coefficient test-retest et, dans la mesure du possible, de la fiabilité interévaluateurs. La CLA a révélé une cohérence interne élevée, mais la fiabilité du coefficient test-retest ne pouvait être déterminée parce que la CLA a servi strictement de test préalable. Le CLZ révèle une fiabilité marquée du coefficient test-retest mais une cohérence interne modérée, ce qui suscite certaines préoccupations quant à son exactitude. En ce qui concerne l'ICTET, il s'est caractérisé par une piètre fiabilité du coefficient test-retest, mais une fiabilité interévaluateurs élevée de même qu'une grande cohérence interne. La faible fiabilité du coefficient test-retest

ne constitue pas un problème inhérent à l'ICTET en soi, mais elle révèle qu'il n'est peut-être pas souhaitable de recourir au même message-guide dans un test préalable et postérieur.

Sous leur forme classique, les mesures de validité des outils, quels qu'ils soient, sont difficiles à mettre en place. En ce qui touche la présente étude, un outil valide est celui dont les tâches reflètent la mise en application des APC décrites par les objectifs du cours et les ASM. Deux méthodes intervenaient à ce chapitre : d'une part, la mesure dans laquelle les tâches de l'outil reflétaient bien les objectifs du cours, où était envisagée l'application des APC en vue de régler des problèmes complexes en génie et, d'autre part, les corrélations entre les sous-échelles d'évaluation des ASM et les sous-échelles d'évaluation de l'outil. Les outils axés sur le rendement, comme la CLA, dans lesquels l'évaluation globale des APC est recherchée. semblent convenir à l'évaluation des APC en génie. Parmi les trois outils, c'est la CLA qui a révélé le plus grand nombre de points en corrélation entre les sous-échelles d'évaluation et, malgré son contexte générique, ces tâches vont de pair avec les objectifs du cours. Il y a lieu d'éviter les outils à choix multiples comme le CLZ, car la dimension de reconnaissance et de mémorisation des évaluations s'y rapportant ne reflète pas fidèlement les objectifs du cours. Quant à l'ICTET, bien qu'il permette le maintien de certaines connaissances propres au domaine au moyen d'un message-quide relatif au génie, il ne reflète pas la mise en application des APC décrite dans les objectifs du cours. Ces différences sont d'autant plus exemplifiées par le nombre de corrélations, allant de faible à modérée, entre les ASM et chacune des sous-échelles d'évaluation respectives des outils.

Les objectifs secondaires de l'étude avaient pour but de fournir une méthode d'évaluation de la pensée critique en contexte de génie, puis de fournir un point de départ aux autres parties intéressées au développement et à l'évaluation de la pensée critique. Il résulte de ces objectifs les conclusions suivantes :

Existe-t-il des données selon lesquelles les ASM comportent des retombées positives significatives sur les aptitudes en pensée critique des étudiants?

Malgré l'absence de gains aux APC des étudiants selon les mesures des outils uniformisés, les auteurs en viennent à la conclusion que les ASM comportent une incidence positive significative sur le développement des APC. Cette conclusion est appuyée par la hausse des notes aux ASM et la progression des résultats relatifs à la pensée critique durant le cheminement de cours. Compte tenu des résultats du sondage en lien avec le test postérieur et du classement des cours, les ASM ainsi que les séances de cours et ateliers du module 1 de l'APSC 100 sont, aux yeux des étudiants, les principaux facteurs ayant contribué au développement des APC. Enfin, il ressort après analyse des entretiens de réflexion à voix haute que le groupe en génie a davantage recouru aux APC que leurs homologues en physique, et que ce groupe est d'avis que l'ASM et le module 1 de l'APS 100 constituent les facteurs ayant contribué au développement de telles aptitudes.

Quel cadre de pensée critique et quel outil de pensée critique reflètent la mise en application des aptitudes en pensée critique dans le règlement de problèmes complexes en génie?

Il convient d'axer le choix d'un cadre d'enseignement sur la mesure dans laquelle celui-ci reflète bien les objectifs du cours quant à la mise en application des APC. De plus, les dimensions de ce cadre doivent s'inscrire dans le mode d'application des aptitudes en pensée critique au sein de la pratique propre à la discipline. Tout modèle ou cadre de pensée critique conviendra à des fins propres à la discipline, pourvu que soit rendue explicite la façon dont les dimensions du cadre et les APC sont mises en application dans l'exercice de la discipline.

Le choix d'un outil uniformisé a nécessité une attention et une enquête soutenues, et celui-ci ne peut être adapté aussi facilement qu'un cadre. Les messages-guides de chaque outil sont très spécifiques et conçus de façon à évaluer les dimensions du cadre sur lequel ils s'appuient. Une évaluation ainsi structurée ne peut servir qu'à évaluer dans quelle mesure l'étudiant comprend ou reconnaît bien les dimensions du cadre; elle

ne permet pas d'évaluer dans quelle mesure les APC sont mises en application pour résoudre des problèmes complexes en génie.

Dans quelle mesure faut-il préserver la concordance des tâches entre l'outil de pensée critique et les problèmes complexes de génie?

Il convient d'agir avec la plus grande circonspection lorsque vient le temps d'instaurer et de maintenir la concordance des tâches entre un outil de pensée critique et le mode d'application de la pensée critique en vue de résoudre des problèmes complexes en génie, ou des problèmes liés à une discipline en particulier. La mauvaise concordance des tâches occasionne des résultats inexacts et suscite des questions quant à la fiabilité, à la validité et à l'authenticité de l'outil. Mettre en place la concordance des tâches au moyen d'outils uniformisés se révèle éprouvant, comme en font foi les outils uniformisés ayant servi dans la présente étude, nommément le CLZ et l'ICTET, puisque les tâches dans le contexte de ces outils ne permettaient pas de refléter avec exactitude les problèmes complexes en génie. Bien que les tâches qu'elle présente soient de nature générique, la CLA comporte une mesure supérieure de la concordance des tâches en lien avec les APC en génie. Les outils conçus sur mesure, comme les ASM, peuvent être mis au point pour en arriver à une concordance poussée des tâches se traduisant par un moyen valide et fiable d'évaluation des APC.

Quelles sont les méthodes fructueuses pour évaluer les aptitudes en pensée critique dans le cadre d'un cours?

Après mûre réflexion quant aux résultats et aux difficultés en lien avec l'étude, les méthodes efficaces d'évaluation des aptitudes en pensée critique dans le contexte d'un cours sont celles qui sont intégrées aux résultats du cours et qui concordent bien avec ceux-ci. Il s'agit de méthodes bien intégrées, harmonieuses et pratiquement impossibles à distinguer des autres activités liées au cheminement dans le cours; elles procurent également aux étudiants un lien significatif entre les APC et leur discipline. Or, les outils uniformisés se situent habituellement à l'extérieur des activités de cours, ce qui risque d'entraîner chez les étudiants qui doivent mettre à exécution les outils de façon significative une participation et une motivation amoindries, dont il résultera des mesures non fiables. De plus, il est difficile de mettre ces outils en concordance aux fins de la PC propre à la discipline, comme nous l'avons décrit dans le paragraphe précédent. Le recours aux outils en guise de test préalable et postérieur nuit également à la motivation et est attribuable précisément à la fatique relative au test. Le prolongement de l'intervalle entre les mesures préalables et postérieures donnera aux étudiants suffisamment de temps pour réfléchir aux APC et les intégrer, se remettre de la fatique et de l'attrition causées par les tests, s'investir dans des tâches et obtenir de la rétroaction, et éviter un ordonnancement trop rapproché des examens finaux. De telles questions peuvent être réglées par le recours à une tâche authentique d'enseignement, de développement et d'évaluation des APC. Les auteurs estiment que les ASM constituent un exemple d'une telle méthode. Les ASM sont conçues à titre d'application concrète de problèmes propres à la discipline et elles fournissent un moyen d'enseignement et d'évaluation en simultané des APC facilement intégrés, bien mis en concordance, rigoureux et authentiques.

Une fois rassemblés, les résultats de l'étude et les conclusions traitant des objectifs de recherche permettent la mise en place d'un ensemble de thèmes communs relativement à l'utilisation d'outils uniformisés en vue d'une évaluation de la pensée critique dans le contexte d'un grand cours donné au sein d'un établissement d'enseignement supérieur :

1) Motivation: En règle générale, le recours aux outils uniformisés en guise de test préalable et postérieur sera perçu comme étant dissocié des activités du cours, ce qui risque d'amoindrir la motivation liée à l'exécution avec soin des outils, particulièrement à la fin du cours. Dans les cas où ces outils ne sont pas intégrés aux activités du cours, d'importants biais d'auto-sélection posent souvent problème; dans les cas où les outils sont intégrés, c'est la motivation qui risque de poser problème.

- 2) Concordance : Il y a lieu que les tâches requises par un outil correspondent à une évaluation authentique de la pensée critique, conformément aux objectifs du cours. Dans la majorité des outils permettant d'évaluer les APC en guise d'aptitude générique, la mise au point et le maintien de la concordance des tâches revêtent une importance considérable dans la collecte de données valides et fiables.
- 3) Sensibilité: L'outil devra comporter une résolution suffisante pour mesurer les changements aux APC au fil d'un cours type donné durant une session. L'outil dont la résolution est insuffisante sera peu utile pour déterminer l'efficacité d'une intervention, en vue d'une amélioration souhaitée dans le cours ou le programme d'études. Dans les cas où les outils uniformisés sont dépourvus de la résolution permettant d'évaluer les gains en APC durant un cours, leur utilisation conviendra seulement à la mesure du développement à long terme pour fins d'évaluation et d'amélioration du programme.

Une méthode qui peut permettre de régler un bon nombre de ces problèmes consiste à évaluer le développement des APC par la notation de certains artefacts des étudiants créés au début et à la fin d'une séance scolaire donnée (une session, une année, ou un programme complet) au moyen d'une rubrique de notation constante. À titre d'exemple, les rubriques d'évaluation valide de l'apprentissage dans l'enseignement au premier cycle (VALUE), créées par l'association américaine des collèges et universités (AAC&U) (« Value : Valid Assessment of Learning in Undergraduatee Education », s.d.), suscitent un intérêt croissant. Ces rubriques sont propices à l'évaluation longitudinale des artefacts produites par les étudiants en fonction des crédits universitaires, ce qui atténue les problèmes posés par la motivation et la concordance. Cette méthode a servi à évaluer la réussite d'initiatives, dont la conception de cours en collaboration, le développement des aptitudes propres à l'apprentissage continu et les résultats de l'apprentissage en enseignement à titre général (Finley, 2012; Pusecker, Torres, Crawford, Levia et Lehman, 2012; Rhodes, 2012; Siefert, 2012).

Cette méthode sera privilégiée à l'avenir à l'Université Queen's, dans le cadre du Consortium sur l'évaluation des résultats d'apprentissage du COQES, de concert avec une tâche de réponse construite, comme la CLA+, en vue de mesurer la pensée critique telle qu'elle se manifeste dans les artefacts produits par les étudiants, y compris les ASM, dans l'ensemble des corps professoraux participants, notamment celui en génie. Des évaluateurs formés peuvent, à un moment donné en particulier, noter ces éléments, ce qui atténue la dérive des notes en raison de la disposition des évaluateurs au cours d'une session ou d'une année. Les rubriques VALUE proposent une méthode largement utilisée et validée quant à l'évaluation de processus cognitifs génériques, dont les APC, et il est possible de mettre au point des variantes de ces rubriques qui soient axées sur la discipline (p. ex., afin d'évaluer les aptitudes dans les processus de conception en génie). Un échantillon représentatif des présentations des étudiants pourra être choisi aux fins de l'évaluation du développement, ce qui allégera le nombre important de notes qui, selon d'autres modalités, serait requis au sein de grands groupes. Peut-être sera-t-il possible de déterminer un outil uniformisé permettant l'utilisation d'une rubrique identique à celle employée pour la notation de la PC au moyen d'artefacts produits par les étudiants dans les cours universitaires, de façon à garantir la concordance entre les constructions.

En conclusion, le recours aux ASM constitue une méthode particulière quant au développement et à l'évaluation d'aptitudes de catégorie supérieure de façon rigoureuse, authentique et viable. La méthode d'évaluation des aptitudes de pensée critique en génie sera, selon toute vraisemblance, très différente des méthodes employées dans d'autres disciplines pour évaluer la pensée critique dans le contexte d'activités d'amélioration d'un cours ou d'un programme régulier. Quelle que soit la méthode privilégiée, il y a lieu de prêter particulièrement attention aux conclusions décrites dans le présent rapport. Les processus de l'amélioration continue font de plus en plus l'objet de discussions ayant trait à l'amélioration de la qualité et à l'accréditation, de sorte qu'il convient d'accorder une grande priorité à la détermination de méthodes viables et fiables pour intégrer l'évaluation des résultats de l'apprentissage générique et des aptitudes de niveau supérieur, comme la pensée critique.

Bibliographie

Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R. et D. Zhang (2008), « Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis », dans *Review of Educational Research*, vol. 78 n° 4, p. 1102-1134.

Adams, M. H., Whitlow, J. F., Stover, L. M. et K.W. Johnson (1996), « Critical thinking as an educational outcome: An evaluation of current tools of measurement », dans *Nurse Educator*, vol. 21 n° 3, p. 23.

Arum, R. et J. Roksa (2011), *Academically Adrift: Limited Learning on College Campuses*, Chicago (IL), University of Chicago Press.

Astin, A. W. (1993a). What matters in college?: Four critical years revisited, San Francisco (CA), Jossey-Bass.

Astin, A. W. (1993b), « What Matters in College », dans Liberal Education, vol. 79 n° 4, p. 4-15.

Bensley, D. A. et M.P. Murtagh (2011), « Guidelines for a Scientific Approach to Critical Thinking Assessment », dans *Teaching of Psychology*, vol. 39 n° 1, p. 5-16. doi:10.1177/0098628311430642.

Blaich, C. et K. Wise (2008), Overview of findings from the first year of the Wabash National Study of Liberal Arts Education, manuscrit inédit.

Bok, D. (2006), Our underachieving colleges: A candid look at how much students learn and why they should be learning more, Princeton (NJ), Princeton University Press.

Bondy, K. N., Koenigseder, L. A., Ishee, J. H. et B.G. Williams (2001), « Psychometric Properties of the California Critical Thinking Tests, » dans *Journal of Nursing Measurement*, vol. 9 n° 3, p. 09-328.

Boren, T. et J. Ramey (2000), « Thinking aloud: reconciling theory and practice », *IEEE Transactions on Professional Communication*, vol. 43 n° 3, p. 261-278. doi:10.1109/47.867942.

Butler, H. A. (2012), « Halpern Critical Thinking Assessment Predicts Real-World Outcomes of Critical Thinking », dans *Applied Cognitive Psychology*, vol. 26 n° 5, p. 721-729, doi:10.1002/acp.2851.

Butler, H. A., Dwyer, C. P., Hogan, M. J. et al. (2012), « The Halpern Critical Thinking Assessment and real-world outcomes: Cross-national applications », dans *Thinking Skills and Creativity*, vol. 7 n° 2, p. 112-121. doi:10.1016/j.tsc.2012.04.001.

Council for Aid to Education, 2011-2012 CLA Institutional Report: Queen's University. (2012). 2011-2012 CLA Institutional Report: Queen's University, New York, Council for Aid to Education.

Council for Aid to Education (s.d.). Architecture of the CLA Tasks, extrait de http://www.collegiatelearningassessment.org/files/Architecture_of_the_CLA_Tasks.pdf

CAT Instrument Technical Information (2010), *Tennessee Tech University*, extrait de http://www.tntech.edu/files/cat/reports/CAT_Technical_Information_V7.pdf

Chamberlin, M. T. (2004), « Design principles for teacher investigations of student work », dans *Mathematics Teacher Education and Development*, vol. 6 n° 1, p. 61-72.

Chamberlin, S. (2002), Analysis of interest during and after model eliciting activities: A comparison of gifted and general population students, dissertation inédite de doctorat, extrait de http://docs.lib.purdue.edu/dissertations/AAI3099758/

Chan, N.-M., Ho, I. T. et K.Y.L. Ku (2011), « Epistemic beliefs and critical thinking of Chinese students», dans *Learning and Individual Differences*, vol. 21 n° 1, p. 67-77. doi:10.1016/j.lindif.2010.11.001.

Daly, W. M. (2001), « The development of an alternative method in the assessment of critical thinking as an outcome of nursing education», dans *Journal of Advanced Nursing*, vol. 36 n° 1, p. 120-130. doi:10.1046/i.1365-2648.2001.01949.x.

Deci, E. L. et R.M. Ryan (1985), *Intrinsic Motivation and Self-Determination in Human Behavior* (réimpression), New York, Springer.

Deci, E. L. et R.M. Ryan (2010), Self-Determination, *onlinelibrary.wiley.com*, Hoboken (NJ), John Wiley & Sons, Inc. doi:10.1002/9780470479216.corpsv0834.

Diefes-Dux, H. A., Moore, T., Zawojewski, J. et al. (2004), A framework for posing open-ended engineering problems: model-eliciting activities (p. 455-460), présenté à la 34^e édition de l'événement Annual Frontiers in Education, 2004, FIE 2004, IEEE, doi:10.1109/FIE.2004.1408556.

Dunham, R. A. (1997), « Assessing EFL Student Progress in Critical Thinking With the Ennis-Weir Critical Thinking Essay Test! » dans *Launch Your Career*, vol. 19 n° 1, p. 43.

Ennis, R. H. (1987), A taxonomy of critical thinking dispositions and abilities, *psycnet.apa.org*, W H Freeman/Times Books/Henry Holt & Co.

Ennis, R. H. et E.E. Weir (1985), The Ennis-Weir Critical Thinking Essay Test: An Instrument for Teaching and Testing, extrait de http://faculty.education.illinois.edu/rhennis/tewctet/Ennis-Weir Merged.pdf

Ennis, R. H., Millman, J. et T.N. Tomko (1985), Cornell Critical Thinking Tests Level X & Level Z: Manual, Boise (ID), Midwest Publications.

Ericsson, K. A. et H.A. Simon (1993), *Protocol Analysis: Verbal Reports as Data,* édition révisée, Cambridge (MA), MIT Press.

Facione, P. A. (1990), « The delphi report », dans *Committee on pre-college philosophy,* American Philosophical Association.

Facione, P. A., Facione, N. C. et S.W. Blohm (2007), *The California Critical Thinking Skills Test: CCTST*, San Jose (CA), California Academic Press.

Fawkes, D., O'meara, B., Weber, D. et D. Flage (2005), « Examining the exam: a critical look at the California critical thinking skills test », dans *Science & Education*, vol. 14 n° 2, p. 117-135.

Finley, A. (2012), « Reliable Are the VALUE Rubrics? », dans *Peer Review*, vol. 13/14 nº 4/1.

Fonteyn, M. E., Kuipers, B. et S.J. Grobe (1993), A Description of Think Aloud Method and Protocol Analysis », dans *Qualitative Health Research*, vol. 3 n° 4, p. 430-441. doi:10.1177/104973239300300403.

Foundation for Critical Thinking, (s.d.), extrait de http://www.criticalthinking.org

- Frank, B. M., Strong, D. S., Sellens, R. et L. Clapham (2012), Progress with the Professional Spine: A Four-Year Engineering Design and Practice Sequence, présenté aux actes de la 8^e conférence internationale de l'initiative CDIO, Brisbane (Australie).
- Frank, B. et J. Kaupp (2012), Evaluating Integrative Model Eliciting Activities in First Year Engineering, présenté aux actes de la conférence de l'Association canadienne de l'éducation engénie en 2012 (ACEG12), Winnipeg (Manitoba).
- Frank, B., Strong, D. et R. Sellens (2011), The Professional Spine: Creation of a Four-year Engineering Design and Practice Sequence, actes de l'Association canadienne de l'éducation en génie.
- Frisby, C. L. (1992), Construct Validity and Psychometric Properties of the Cornell Critical Thinking Test (Level Z): a Contrasted Groups Analysis », dans *Psychological Reports*, vol. 71 n° 1, p. 291-303. doi:10.2466/pr0.1992.71.1.291.
- Gasper, B. J. et S.M. Gardner (2013), Engaging Students in Authentic Microbiology Research in an Introductory Biology Laboratory Course is Correlated with Gains in Student Understanding of the Nature of Authentic Research and Critical Thinking », dans *Journal of Microbiology & Biology Education*, vol. 14 n° 1, p. 25-34. doi:10.1128/jmbe.v14i1.460.
- Gokhale, A. A. (1995), « Collaborative Learning Enhances Critical Thinking », dans *Journal of Technology Education*, vol. 7 n° 1, extrait de http://scholar.lib.vt.edu/ejournals/JTE/v7n1/gokhale.jte-v7n1.html
- Gottesman, A. J. et S.G. Hoskins (2013), « CREATE Cornerstone: Introduction to Scientific Thinking, a New Course for STEM-Interested Freshmen, Demystifies Scientific Thinking through Analysis of Scientific Literature », dans *CBE-Life Sciences Education*, vol. 12 n° 1, p. 59-72.
- Halpern, D. F. (2002), « Teaching for Critical Thinking: Helping College Students Develop the Skills and Dispositions of a Critical Thinker », dans *New Directions for Teaching and Learning*, vol. 1999 n° 80, p. 69-74. doi:10.1002/tl.8005.
- Halpern, D. F. (2000), Thinking critically about critical thinking: Lessons from cognitive psychology (p. 22), présenté aux actes de l'atelier Training Critical Thinking Skills for Battle Command de l'ARI.
- Halpern, D. F. (2003a), « The "how" and "why" of critical thinking assessment », dans D. Fasko (éd.), *Critical thinking and reasoning: Current research, theory and practice*, Cresskill (NJ), Hampton Press.
- Halpern, D. F. (2006), Halpern critical thinking assessment using everyday situations: Background and scoring standards, rapport inédit.
- Halpern, D. F. et; H.R. Riggio (2002), Thinking critically about critical thinking: Workbook to accompany Thought and knowledge: An introduction to critical thinking, 4^e éd., Mahwah (NJ), Erlbaum.
- Hart Research Associates (2008), How should colleges assess and improve student learning? Employers' views on the accountability challenge. Association of American Colleges and Universities Report. Washington (DC), American Association of Colleges and Universities et Hart Research Associates.
- Hart Research Associates (2013), It Takes More Than a Major: Employer Priorities for College Learning and Student Success. Association of American Colleges and Universities Report. Washington (DC), American Association of Colleges and Universities et Hart Research Associates.
- Hayes, A. F. et K. Krippendorff (2007), « Answering the call for a standard reliability measure for coding data », dans *Communication Methods and Measures*, vol. 1 n° 1, p. 77-89.

- Jacobs, S. S. (1995), « Technical characteristics and some correlates of the California critical thinking skills test, forms A and B », dans *Research in Higher Education*, vol. 36 n° 1, p. 89-108. doi:10.1007/BF02207768.
- Klein, S., Benjamin, R., Shavelson, R. et R. Bolus (2007), « The Collegiate Learning Assessment Facts and Fantasies », dans *Evaluation Review*, vol. 31 n° 5, p. 415-439.
- Klein, S. C., Liu, O. L. E., Sconing, J. A., Bolus, R. C., Bridgeman, B. E., Kugelmass, H. C. et J.C. Steedle (2009), rapport Test Validity Study (TVS).
- Ku, K. Y. (2009), « Assessing students' critical thinking performance: Urging for measurements using multiresponse format », dans *Thinking Skills and Creativity*, vol. 4 n° 1, p. 70-76.
- Ku, K. Y. L. et I.T. Ho (2010a), « Metacognitive strategies that enhance critical thinking », dans *Metacognition* and *Learning*, vol. 5 n° 3, p. 251-267. doi:10.1007/s11409-010-9060-6.
- Ku, K. Y. L. et I.T. Ho (2010b), « Dispositional factors predicting Chinese students' critical thinking performance », dans *Personality and Individual Differences*, vol. 48 n° 1, p. 54-58. doi:10.1016/j.paid.2009.08.015.
- Leppa, C. J. (1997), « Standardized Measures of Critical Thinking: Experience with the California Critical Thinking Tests », dans *Nurse Educator*, vol. 22 n° 5, p. 29.
- Lesh, R. (1999), « The development of representational abilities in middle school mathematics », dans I. E. Sigel (éd.), *Development of Mental Representation: Theories and Application* (p. 323-350), Hillsdale (NJ), Lawrence Erlbaum Associates.
- Lesh, R. A. et H.M. Doerr (2003), Beyond Constructivism, Hillsdale (NJ), Lawrence Erlbaum Associates.
- Lesh, R. et H.M. Doerr (2000), « Symbolizing, communicating, and mathematizing: Key components of models and modeling », dans In P. Cobb, E. Yackel & K. McClain (éd.), *Symbolizing and communicating in mathematics classrooms: Perspectives on discourse, tools, and instructional design* (p. 361-384), Oxford, Routledge.
- Marin, L. M. et D.F. Halpern (2011), « Pedagogy for developing critical thinking in adolescents: Explicit instruction produces greatest gains », dans *Thinking Skills and Creativity*, vol. 6 n° 1, p. 1-13. doi:10.1016/j.tsc.2010.08.002.
- MEDIA Project (s.d.), Collaborative Research: Improving Engineering Students' Learning Strategies through Models and Modeling, extrait de http://www.modelsandmodeling.pitt.edu/
- Moore, T. et H. Diefes-Dux (2004), *Developing model-eliciting activities for undergraduate students based on advanced engineering content*, présenté à la 34^e conférence Frontiers in education de l'ASEE/IEEE, Savannah (GA), doi: 10.1109/FIE.2004.1408557.
- Nieto, A. M. et C. Saiz (2008), « Evaluation of Halpern's "structural component" for improving critical thinking », dans *The Spanish Journal of Psychology*, vol. 11 n° 1, p. 266-274.
- Norris, S. P. (1990), « Effect of Eliciting Verbal Reports of Thinking on Critical Thinking Test Performance », dans *Journal of Educational Measurement*, vol. 27 n° 1, p. 41-58. doi:10.1111/j.1745-3984.1990.tb00733.x.
- Pascarella, E. T., Blaich, C., Martin, G. L. et J.M. Hanson (2011), « How Robust Are the Findings of Academically Adrift? », dans *Change: The Magazine of Higher Learning*, vol. 43 n° 3, p. 20-24. doi:10.1080/00091383.2011.568898.

Paul, R. et L. Elder (2001), « *The Miniature Guide to Critical Thinking* », dans Foundation for Critical Thinking, extrait de http://www.criticalthinking.org/

Paul, R. et L. Elder (2005), « *A guide for educators to critical thinking competency standards: Standards, principles, performance indicators, and outcomes with a critical thinking master rubric »*, dans Foundation for Critical Thinking, extrait de http://www.criticalthinking.org/

Paul, R. et L. Elder (2007), « Consequential validity: using assessment to drive instruction », dans Foundation for Critical Thinking, extrait de http://www.criticalthinking.org/

Paul, R. et L. Elder (2010), « International Critical Thinking Test », dans Foundation for Critical Thinking.

Paul, R. (1993), « The Logic of Creative and Critical Thinking », dans *American Behavioral Scientist*, vol. 37 n° 1, p. 21-39. doi:10.1177/0002764293037001004.

Paul, R. et L. Elder (1996), « *Using Intellectual Standards to Assess Student Reasoning* », dans Foundation for Critical Thinking, extrait de http://www.criticalthinking.org/

Paul, R., Willsen, J. et A.J.A. Binker (1993), *Critical thinking*, Santa Rosa (CA), Foundation for Critical Thinking.

Possin, K. (2013), A Serious Flaw in the Collegiate Learning Assessment [CLA] Test (p. 1-15), Winona (MN), The Critical Thinking Lab.

Pusecker, K., Torres, R., Crawford, I., Levia, D. et D. Lehman (2012), « Increasing the Validity of Outcomes Assessment », dans *Peer Review*, vol. 13/14 n° 4/1, p. 27-30.

Rashid, R. A. et R.A. Hashim (2008), « The Relationship between critical thinking and language proficiency of Malaysian and undergraduates », dans *Proceeding of the EDU-COM 2008 International Conference*, *Symposia and Campus Events*, 19-21 novembre 2008, Université Edith Cowan, Perth (Australie-Occidentale).

Rhodes, T. L. (2012), « Emerging Evidence on Using Rubrics », dans Peer Review, vol. 13/14 n° 4/1, p. 4-5.

Ryan, R. M. et E.L. Deci (2000), « Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being », dans *American Psychologist*, vol. 55 n° 1, p. 68. doi:10.1037/0003-066X.55.1.68.

Schufried GmbH. (s.d.), Vienna Test System, schuhfried.com, Moedling (Autriche) Schufried GmbH.

Self, B., Shuman, L. J. et M. Besterfield-Sacre (2012), Model Eliciting Activities: Lessons Learned From a Five-Year, Seven-Institution Collaboration, présenté à la 6^e conférence de l'International Academy of Technology, Education and Development.

Shavelson, R. J. (2008), « The collegiate learning assessment », dans In *Ford Policy Forum 2008: Forum for the Future of Higher Education*, extrait de http://net.educause.edu/ir/library/pdf/fp085.pdf

Shavelson, R. J., Baxter, G. P.et X. Gao (1993), « Sampling variability of performance assessments », dans *Journal of Educational Measurement*, vol. 30 n° 3, p. 215-232.

Shuman, L. J. (2012), AC 2012-3847: CCLI: Model eliciting activities, présenté aux actes de la conférence annuelle de l'ASEE.

Shuman, L. J. et M.E. Besterfield-Sacre (2008), The model eliciting activity (MEA) construct: moving engineering education research into the classroom, présenté à la 9^e conférence biennale de l'ASME sur la conception et l'analyse des systèmes de génie, Haïfa (Israël).

Siefert, L. (2012), « Assessing General Education Learning Outcomes », dans *Peer Review*, vol. 13/14 n° 4/1, p. 9-11.

Solon, T. (2003), « Teaching critical thinking: The more, the better », dans *The Community College Enterprise*, vol. 9 n° 2, p. 25-38.

Steif, P. S., Lobue, J. M., Kara, L. B. et A.L. Fay (2013), « Improving Problem Solving Performance by Inducing Talk about Salient Problem Features », dans *Journal of Engineering Education*, vol. 99 n° 2, p. 135-142. doi:10.1002/j.2168-9830.2010.tb01050.x.

Stein, B. et A. Haynes (2011), « Engaging Faculty in the Assessment and Improvement of Students' Critical Thinking Using the Critical Thinking Assessment Test », dans *Change: The Magazine of Higher Learning*, vol. 43 n° 2, p. 44-49. doi:10.1080/00091383.2011.550254.

Stein, B., Haynes, A. et M. Redding (2008), *Project CAT: Assessing Critical Thinking Skills Final Report*, Université Tennessee Tech, extrait de http://www.tntech.edu/files/cat/reports/Project CAT Final Report.pdf

Stein, B., Haynes, A. et M. Redding (2006), *Project CAT: Assessing Critical Thinking Skills* (p. 290-299), présenté aux actes de la National STEM Assessment Conference, Washington (DC).

Stein, B., Haynes, A., Redding, M., Ennis, T. et M. Cecil (2007), « Assessing critical thinking in STEM and beyond », dans In M. Iskander (éd.), *Innovations in E-learning, Instruction Technology, Assessment and Engineering Education* (p. 7982), New York (Springer).

Stein, B., Haynes, A. et J. Unterstein (2003), Assessing critical thinking skills, présenté à l'assemblée annuelle de la SACS/COC, Nashville.

Taube, K. T. (1997), « Critical thinking ability and disposition as factors of performance on a written critical thinking test », dans *The Journal of General Education*, vol. 46 n° 2, p. 129-164.

Université Tennessee Tech, (s.d.), Successful Projects | Tennessee Tech University, extrait de http://www.tntech.edu/cat/links-to-successful-projects

Ungson, G. R. et D.N. Braunstein (éd) (1982), *Decision Making: An Interdisciplinary Inquiry*, Boston (MA), Kent Publishing Company.

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Briere, N. M., Senecal, C. et E.F. Vallieres (1992), « The Academic Motivation Scale: A measure of intrinsic, extrinsic, and amotivation in education », dans *Educational and Psychological Measurement*, vol. 52 n° 4, p. 1003-1017. doi:10.1177/0013164492052004025.

VALUE: Valid Assessment of Learning in Undergraduate Education (s.d.), VALUE: Valid Assessment of Learning in Undergraduate Education. Association of American Colleges and Universities, extrait de http://www.aacu.org/value/rubrics/index_p.cfm?CFID=8319302&CFTOKEN=61210529

Van Someren, M. W., Barnard, Y. F. et J.A. Sandberg (1994), *The think aloud method: A practical guide to modelling cognitive processes*, Londres, Academic Press.

Évaluation de la pensée critique et de la solution de problèmes dans les grands groupes : Activités suscitant des modèles pour le développement de la pensée critique

Yildirim, T. P., Shuman, L., Besterfield-Sacre, M. et T. Yildirim (2010), « Model eliciting activities: assessing engineering student problem solving and skill integration processes », dans *International Journal of Engineering Education*, vol. 26 n° 4, p. 831-845.

Zahner, D. (2013), « Reliability and Validity – CLA », dans *Council for Aid to Education*, extrait de http://cae.org/images/uploads/pdf/Reliability_and_Validity_of_CLA_Plus.pdf