




Conseil ontarien
de la qualité de
l'enseignement
supérieur

Un organisme du gouvernement de l'Ontario



**L'extension de l'acquisition et de la mesure
des compétences transférables :
l'évaluation du potentiel de la notation par
rubrique dans le contexte de l'évaluation
par les pairs**

Steve Joordens, Dwayne Paré, Robin Walker
Université de Toronto, campus de Scarborough;
Jim Hewitt et Clare Brett
Institut d'études pédagogiques de l'Ontario

Publié par le

Conseil ontarien de la qualité de l'enseignement supérieur

1, rue Yonge, bureau 2402
Toronto (Ontario) 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 :

Joordens, S., Paré, D., Walker, R., Hewitt, J. et C. Brett (2019), *L'extension de l'acquisition et de la mesure des compétences transférables : l'évaluation du potentiel de la notation par rubrique dans le contexte de l'évaluation par les pairs*, Toronto, Conseil ontarien de la qualité de l'enseignement supérieur.



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

Table des matières

Remerciements.....	4
Synthèse.....	4
Terminologie.....	8
Introduction.....	10
Phase 1.....	15
Méthode.....	16
Résultats.....	20
Analyse.....	31
Phase 2.....	34
Méthodes : Les 53 premiers.....	37
Méthodes : Les 14 participants.....	41
Résultats.....	43
Analyse générale.....	47
Bibliographie.....	54

Liste des tableaux

Tableau 1 : Moyennes et écarts-types des notes fournies par les évaluateurs.....	22
Tableau 2 : Coefficients de corrélation et Kappa chez les experts et les AE	23
Tableau 3 : Coefficients de corrélation et Kappa entre la moyenne par les pairs et les moyennes des experts et des AE	25
Tableau 4 : Coefficients de corrélation et Kappa chez les paires d'experts et d'AE en fonction des cinq facteurs de rubrique	26
Tableau 5 : Coefficients de corrélation et Kappa entre la moyenne par les pairs et les moyennes des experts et des AE	27

Liste des graphiques

Graphique 1A : Fréquence des réponses de tous les participants dans chaque valeur de l'échelle à sept points de Likert (n=587).....	28
Graphique 1B : Valeurs « d'accord » et « en désaccord » pour chaque question (n=587)	29
Graphique 1C : Fréquence des réponses de tous les participants dans chaque valeur de l'échelle à sept points de Likert (n=200).....	Error! Bookmark not defined.
Graphique 1D : Valeurs « d'accord » et « en désaccord » pour chaque question (n=200)	Error! Bookmark not defined.
Graphique 2 : Réponses à la question : Avez-vous actuellement recours à l'évaluation par les pairs par voie technologique sous une forme ou une autre, ou aimeriez-vous que nous fournissions la technologie et le soutien? (n=53)	39
Graphique 3 : Réponses à la question : Laquelle des compétences transférables suivantes est tout particulièrement importante selon vous? (n=53).....	40
Graphique 4A : Fréquence des réponses de tous les participants dans chaque valeur de l'échelle à sept points de Likert (n=621).....	43
Graphique 4B : Valeurs « d'accord » et « en désaccord » pour chaque question (n=621)	44
Graphique 5A : Fréquence des réponses de tous les participants dans chaque valeur de l'échelle à sept points de Likert (n=14).....	45
Graphique 5B : Valeurs « d'accord » et « en désaccord » pour chaque question (n=14)	45

Remerciements

Les auteurs souhaitent remercier le Conseil ontarien de la qualité de l'enseignement supérieur (COQES) pour son appui à la présente recherche.

Synthèse

Les établissements d'enseignement postsecondaire sont responsables non seulement d'inculquer aux élèves les connaissances thématiques liées à leur champ d'études, mais de leur apprendre des compétences transférables. Celles-ci se transfèrent bien du milieu d'apprentissage au milieu professionnel, et d'un emploi à l'autre; la pensée critique ou créative, de même que la communication claire et efficace en sont des exemples. Si les connaissances thématiques font l'objet d'un enseignement en bonne et due forme, tel n'est pas souvent le cas des compétences transférables.

Dans le but de fournir des solutions qui permettraient aux pédagogues de concevoir des compétences transférables de base et de les mesurer, l'auteur Joordens (2018) préconise une démarche d'évaluation par les pairs; de façon précise, cette démarche fait intervenir le recours aux rubriques pour évaluer le travail entre les pairs. La démarche proposée s'appuie sur l'hypothèse suivante : si les élèves mettaient en application une rubrique axée sur les compétences au travail de leurs pairs et que le travail de chaque élève était évalué au moyen de cette rubrique par de nombreux pairs, la moyenne du nombre de notations attribuées par des pairs fournirait une bonne mesure de la compétence pour laquelle la rubrique a été créée à des fins d'évaluation.

La démarche proposée par l'auteur Joordens permet de répondre à deux questions relatives à l'acquisition des compétences transférables de base : Comment acquérons-nous ces compétences? De plus, comment savons-nous que ces compétences sont en cours d'acquisition? Le fait d'évaluer la présence d'une compétence transférable, comme la pensée critique dans le travail de nombreux pairs en soi, peut faire partie de l'amélioration de cette compétence chez les élèves. Dès que l'élève se sert d'une rubrique pour évaluer le travail des pairs, il peut transposer la connaissance s'y rapportant à son propre travail : ce concept peut être mis en application dans une vaste gamme de contextes d'apprentissage.

Phase 1

Pendant la Phase 1 de la présente étude, nous mettons à l'essai l'hypothèse directement au moyen de l'élément fondamental, à savoir l'intégration d'une rubrique validée à la phase d'évaluation, dans ce cas-ci la rubrique VALUE de pensée critique conçue par l'AAC&U (association des collèges et universités des États-Unis). Les résultats révèlent que les mesures fondées sur les notations attribuées par des pairs permettent bel et bien de saisir l'ordre d'importance de la qualité, mais qu'elles ont également tendance à surestimer l'ampleur réelle de la compétence présente, du moins lorsque les élèves sont en première année. Cet élément a été établi par la comparaison avec les notations moyennes attribuées par trois groupes : les AE, les évaluateurs experts de même que le professeur du cours.

De telles constatations ne posent pas fortement problème quant à la démarche proposée, mais elles débouchent sur des propositions de formation des élèves qui pourraient permettre d'optimiser la démarche. De plus, la première phase est source de données qualitatives, lesquelles montrent que les élèves comprennent et apprécient l'importance d'acquérir et de mesurer ces compétences, et qu'ils seraient prêts à exécuter le travail supplémentaire que comporte l'évaluation par les pairs si cela était possible.

Quatre ensembles différents de participants ont fourni des données :

- 694 élèves (recrutés dans un ensemble de 1 800 élèves inscrits à un cours d'initiation à la psychologie à l'UTSC);
- deux AE (sélectionnés à partir d'un ensemble d'AE actuels et antérieurs relativement au cours d'initiation à la psychologie);
- deux évaluateurs experts (recommandés par l'AAC&U, rompu à l'utilisation des rubriques VALUE);
- un professeur de cours (sans expérience ou formation préalable relativement à l'utilisation des rubriques VALUE).

peerScholar

La procédure employée dans la présente étude était gérée par un programme infonuagique appelé peerScholar, lequel a été conçu au campus de Scarborough de l'Université de Toronto (UTSC). Ce programme allie le processus d'évaluation par les pairs à l'autoévaluation et la rétroaction. Le programme peerScholar comporte trois phases : création, évaluation et réflexion.

Phase de création : Dans cette phase, les élèves soumettent un argumentaire qui découle de recherches et s'appuie sur des données probantes.

Phase d'évaluation : Le travail de chaque élève est évalué par six de leurs pairs, et chaque élève doit évaluer le travail de six pairs, puis donner de la rétroaction (dans la présente étude, après avoir regardé une vidéo éducative sur la façon de mettre en application la rubrique).

Phase de réflexion : Les élèves prennent connaissance de la rétroaction au sujet de leur propre travail puis doivent noter la qualité de la rétroaction reçue. Par la suite, les élèves ont la possibilité de réviser leur propre travail et rédiger une courte réflexion qui justifie leurs révisions.

Phase 2

La Phase 2 du projet s'appuie sur le recours à l'évaluation par les pairs dans la Phase 1 : on examine l'extensibilité de la méthode, puis on demande si la mise en application de rubriques par les élèves, pendant qu'ils procèdent à une évaluation par les pairs, pourrait fonctionner dans une vaste gamme de contextes, de niveaux et d'établissements d'enseignement, voire de pays.

L'équipe de recherche a pris contact avec des professeurs par plusieurs méthodes différentes et, au départ, elle a été en mesure d'attirer 53 professeurs. Toutefois, au fur et à mesure que le projet a évolué, il est devenu difficile d'aller de l'avant avec ces engagements au-delà de la première déclaration d'intérêt, de sorte qu'il a fallu adapter le processus de recrutement afin que le projet puisse se dérouler. Ultérieurement, un groupe de 14 professeurs ont convenu de participer à la Phase 2. Ce groupe était composé de professeurs de l'Université de Toronto (notamment l'Institut d'études pédagogiques de l'Ontario), de l'Université Lakehead, du Collège Humber, du Collège communautaire Assiniboine, de l'Université Urbana (aux États-Unis) et de l'Université d'Utrecht (Pays-Bas). Les 14 professeurs recrutés ont utilisé la méthode dans 20 cours différents puis sondé les opinions et les réactions des élèves.

Les données qualitatives des élèves ont permis de reproduire et d'étendre les résultats de la Phase 1, ce qui montre que les élèves aimaient la démarche et la jugeaient utile. Les données ont ensuite été complétées par celles des membres du corps professoral attestant la convivialité de la démarche, laquelle était non perturbatrice et facile à mettre en œuvre, et de leurs perceptions comme quoi la démarche favorisait l'apprentissage et qu'ils convenaient de l'utiliser à plus grande échelle. Une fois combinées, les constatations appuient le potentiel de cette démarche comme

moyen d'acquisition et de mesure de presque n'importe quelle compétence transférable dans presque n'importe quel contexte pédagogique.

Analyse et constatations

La validité des notes par rubrique décernées par les pairs suscite encore quelques points d'interrogation. Bien qu'ils aient reçu de courtes directives sur la façon de mettre en application la rubrique dans le processus d'évaluation, les élèves attribuaient néanmoins des notes supérieures à celles décernées par les AE, le professeur et les évaluateurs experts dans la présente étude. Une solution possible à ce problème consisterait à donner aux élèves une formation accrue à l'aide de rubriques. Une autre solution possible serait de jauger la mesure dans laquelle un biais est présent pour ensuite en neutraliser les effets sur le plan statistique.

Parmi d'autres difficultés qu'il faudra peut-être résoudre afin que cette démarche soit largement adoptée, il y a le fait d'amener les membres du corps professoral à faire l'essai d'une méthode qui leur est peut-être inconnue; la simplification du processus dans son utilisation, tant pour les membres du corps professoral que les élèves; le fait de convaincre les élèves que le « travail supplémentaire » perçu que fait intervenir ce processus en vaut la peine.

Conclusions

En général, les membres du corps professoral qui ont participé à la Phase 2 du projet évaluaient de façon très positive le potentiel de la démarche. Fait peut-être le plus pertinent, c'est que tous les membres du corps professoral ont convenu qu'ils continueraient de recourir à cette démarche et que celle-ci devrait être utilisée plus largement. Il est alors évident que la démarche proposée comporte absolument un potentiel élevé, malgré l'ajustement nécessaire pour garantir un niveau approprié de validité.

Terminologie

« **Les 53 premiers** » : Ce concept désigne un groupe de 53 membres du corps professoral (essentiellement du Canada, quelques-uns étant de l'Australie, du Royaume-Uni et de l'Afrique) qui ont rempli un formulaire Google attestant dans une certaine mesure leur intérêt envers la participation à un projet d'évaluation par les pairs. Au fur et à mesure que le projet a progressé, la plupart des membres de ce groupe n'ont pas continué, mais les données qu'ils ont fournies au moyen du formulaire Google ont révélé un certain aperçu du recours à l'évaluation par les pairs.

« **Les 14 participants** » : Ce concept désigne le groupe qui a pris part à l'étude d'évaluation de l'extensibilité de la démarche d'évaluation par les pairs. Il était essentiellement composé des autres professeurs de l'Université de Toronto, dans l'espoir de maintenir un niveau de participation élevé au projet.

Phase d'évaluation : La phase d'évaluation a fait intervenir des élèves qui ouvraient une session dans la plateforme peerScholar pour procéder à un cycle d'évaluation par les pairs et d'autoévaluation. De façon précise, ils ont eu accès à des compositions présentées par six de leurs collègues de classe, après quoi ils devaient évaluer chaque composition et donner de la rétroaction constructive. Les compositions étaient dépersonnalisées et choisies aléatoirement à partir de l'ensemble des soumissions.

Phase de création : La phase de création a fait intervenir des élèves qui ont présenté un argumentaire relativement à une requête. De façon précise, ils étaient tenus de trouver au moins trois sources fiables en lien avec le thème, de les lire puis de composer un argumentaire persuasif relativement à un enjeu sous-jacent des questions générales.

Pensée créative : La pensée créative allie à la fois la capacité de combiner des idées, des images ou un savoir-faire existants ou d'en faire la synthèse de façon originale, ainsi que l'expérience de la réflexion, de la réaction et du travail de façon imaginative, marquée par un niveau élevé d'innovation, de pensée divergente et de prise de risques.

Pensée critique : La pensée critique consiste en une habitude mentale caractérisée par l'exploration exhaustive des enjeux, des idées, des artefacts et des événements avant l'acceptation ou la formulation d'une opinion ou d'une conclusion.

Évaluateurs experts : Dans le présent document, ce concept désigne les évaluateurs recommandés et favorisés par l'AAC&U. Ces évaluateurs sont passés par une formation rigoureuse quant à la mise en application de la rubrique VALUE de pensée critique aux travaux des élèves.

peerScholar : peerScholar consiste en une plateforme de gestion de l'apprentissage qui allie l'évaluation par les pairs à deux autres pratiques pédagogiques fondées sur les données probantes : les autorités d'évaluation et le recours à la rétroaction à des fins de formation. Il s'agit d'une plateforme infonuagique dont le processus d'évaluation par les pairs est composé de trois phases : la phase de création, la phase d'évaluation et la phase de réflexion.

Phase 1 : Au moyen de la Phase 1, nous avons essentiellement cherché à savoir si les moyennes dérivées des notes par rubrique fournies par les pairs pouvaient servir à prédire les notes données par les experts (c.-à-d. celles présentées par les AE et les professeurs).

Phase 2 : La Phase 2 correspond à une tentative d'évaluation directe des affirmations d'extensibilité préalablement formulées, essentiellement par la demande aux membres du corps professoral de faire l'essai de l'évaluation par les pairs dans leur classe, quel que soit le niveau enseigné. La Phase 2 comporte également les questionnaires destinés aux élèves et aux membres du corps professoral et qui ont été remplis à la suite de l'exercice d'évaluation par les pairs. La remise des soumissions s'est faite au moyen de la plateforme peerScholar.

Phase de réflexion : La phase de réflexion a fait intervenir des élèves qui ont accédé à la rétroaction reçue de six de leurs pairs à propos de leur propre composition. Par la suite, ces élèves devaient noter la rétroaction en fonction de son utilité, et du ton donné à la présentation. Une fois la rétroaction notée, les élèves avaient l'occasion de réviser leur composition en fonction de la rétroaction reçue. Une fois les révisions apportées, les élèves ont rédigé une courte réflexion dans laquelle ils ont justifié les révisions apportées de même que les révisions proposées qu'ils ont choisi de ne pas mettre en œuvre.

Rubriques VALUE : L'AAC&U (association des collèges et universités des États-Unis) a mis au point 16 rubriques en lien avec les différentes compétences transférables. Il appert que ces rubriques présentent toute des niveaux acceptables de fiabilité, une validité apparente et qu'elles sont publiquement accessibles.

Introduction

Les pédagogues et les établissements d'enseignement ont deux responsabilités distinctes à l'égard de leurs élèves. D'une part, ils doivent fournir l'information qui est cruciale au champ d'études de l'élève. À titre d'exemple, en psychologie, on attend des pédagogues qu'ils enseignent les théories primaires et les expérimentations sur le terrain, les graphiques importants qui s'y rapportent de même que l'état actuel des « connaissances ». D'autre part, ils doivent permettre aux élèves de s'exercer à répétition au moyen des compétences transférables de base, comme la pensée critique, la pensée créative et la communication efficace, ainsi que la sensibilité métacognitive dont ils ont besoin pour comprendre leurs forces et leurs faiblesses à un moment donné.

Les pédagogues universitaires réussissent très bien à enseigner l'information, en partie parce qu'ils savent comment et en partie parce qu'ils peuvent mesurer la réussite en toute impartialité, mais surtout parce qu'il est relativement facile de le faire en fonction des contraintes des établissements d'enseignement. Toutefois, lorsqu'il est question d'enseignement des compétences en bonne et due forme, les choses se compliquent. Les pédagogues universitaires ne disposent d'aucun moyen de mesurer l'acquisition des compétences, ni ne sont en mesure de percevoir actuellement comment bien réussir à ce chapitre, étant donné les contraintes de leurs établissements d'enseignement en matière de temps et de ressources. Par conséquent, ils enseignent et mesurent officiellement l'information, mais ils adoptent une démarche nettement moins officielle pour ce qui est de l'acquisition des compétences. Autrement dit, ils s'appuient sur de superbes cours magistraux ou lectures pour stimuler la pensée critique ou créative, puis ils donnent aux élèves des devoirs qui, espèrent-ils, sauront les aider à acquérir des compétences en communication. S'ils évaluent le rendement par rapport à ces devoirs, ils ne reconnaissent pas officiellement que les notes données témoignent de l'acquisition des compétences. Habituellement, les notes s'ajoutent plutôt à une note finale sur laquelle la capacité de l'élève à mémoriser l'information exerce une influence considérable.

Récemment, l'auteur Joordens (2018) a proposé un processus qui pourrait apporter une solution aux problèmes en lien avec les notions « Comment? » et « Comment savons-nous? », et ce, sans pour autant métamorphoser la démarche type du pédagogue dans l'enseignement à son groupe. La notion « Comment? » se rapporte à la question posée quant à savoir comment permettre optimalement l'acquisition de compétences transférables de base, comme la pensée critique. La notion « Comment savons-nous? » se rapporte au besoin d'évaluer de telles compétences, de façon à confirmer leur acquisition et à permettre le peaufinement poussé du processus d'acquisition de manière factuelle. La démarche proposée par l'auteur Joordens (2018) allie l'évaluation par les pairs par voie technologique à l'utilisation des rubriques créées précisément pour évaluer les compétences transférables de base.

La section suivante décrit en détail le processus proposé : elle met en relief comment ce processus procure une expérience d'apprentissage marquante et, parallèlement, elle permet la quantification directe des compétences dans le contexte d'activités en classe. Par la suite, il est question du principal point de mire de la recherche relatée ici et qui est le suivant : la démarche proposée part du principe que l'évaluation par les pairs procure une mesure valide des compétences transférables. De façon précise, la moyenne des notes d'évaluation par les pairs qui découle de la mise en application par les élèves de rubriques relatives aux compétences transférables au travail de leurs collègues de classe permet de mesurer avec exactitude les compétences en question; or, est-ce le cas concrètement? Plusieurs études ont révélé que la moyenne des notes d'évaluation par les pairs débouche sur une statistique qui se rapproche des notes décernées par les professeurs [Cho, Schunn et Wilson (2006); Paré et Joordens (2008)], mais cela vaut-il également pour les notes fondées sur les rubriques?

Allier l'évaluation par les pairs aux rubriques validées

Un corpus de publications en perpétuelle croissance révèle que le fait de demander aux élèves d'évaluer le travail de leurs pairs (évaluation par les pairs) constitue une démarche pédagogique très puissante à de nombreux égards [Dochy, Segers et D. Sluijsmans (1999); Liu et Carless (2006); Topping (1998)]. Fait peut-être le plus pertinent à ce travail, il s'agit de l'un des rares processus pédagogiques qui comporte un fondement factuel expérimental à l'appui de sa capacité à favoriser l'acquisition de la pensée critique et à améliorer la sensibilité métacognitive [Joordens, Paré et Collimore (2014)]. De façon précise, les auteurs Joordens et al. (2014) présentent des données qui révèlent que, dans chaque expérience d'examen du travail de 6 pairs, les notes données par les élèves relativement à la qualité s'apparentent davantage à celles décernées par les experts et celles fondées sur les moyennes des notes données par les pairs. Les auteurs Joordens et al. attribuent cet état de choses aux améliorations à l'« évaluation de la qualité », laquelle se situe, à leur sens, au cœur de la pensée critique. De plus, ils démontrent que l'évaluation par les élèves de la qualité de leur propre travail a gagné considérablement en exactitude à la suite d'une seule analyse du travail par six de leurs pairs. Étant donné que l'évaluation par les pairs exige des élèves qu'ils consacrent beaucoup plus de temps et d'efforts à examiner le travail des pairs, il est également pertinent de souligner qu'un tel processus procure une possibilité d'apprentissage approfondie que les élèves aiment et à laquelle ils attachent de l'importance [Collimore, Paré et Joordens (2015); Hanrahan et Isaacs (2001)].

La liste de technologies d'évaluation par les pairs qui permettent aux pédagogues de bonifier les devoirs donnés aux élèves au moyen d'étapes d'évaluation par les pairs ne cesse de s'allonger. Nous avons mis au point dans notre laboratoire un tel système, appelé peerScholar (consulter le vision.peerScholar.com pour obtenir des précisions à ce chapitre), lequel allie l'évaluation par les pairs à deux autres pratiques pédagogiques factuelles, à savoir les autoévaluations [Boud (2013)] et le recours à la rétroaction à des fins

de formation [Bennett (2011); Nicol et Macfarlane-Dick (2006)]. Notre exposé ici se rapportera au processus qu'il est destiné à appuyer, étant donné que c'est celui que nous connaissons le mieux, et que c'est le système qui servira dans les recherches courantes. Cela dit, l'idée de base suivante pourrait vraisemblablement être implantée dans de nombreuses solutions d'évaluation par les pairs par voie technologique.

Dans le système peerScholar, les élèves passent par trois phases pour mener à bien le travail. Dans la première phase, dite de « création », les élèves soumettent tout simplement par voie numérique une composition qui donne suite à un message-guide et à la matière qui leur sont fournis par leur pédagogue. Cette partie peut concorder avec un devoir auquel le pédagogue fait déjà appel dans son cours. L'évaluation par les pairs se déroule à la deuxième phase, dite d'« évaluation ». Dans cette phase, les élèves lisent un ensemble de compositions soumises par leurs pairs mais qui sont dépersonnalisées et choisies aléatoirement. Dans notre utilisation, les élèves lisent habituellement six compositions par leurs pairs, après quoi ils doivent mener à bien un exercice d'évaluation conformément aux questions ou aux rubriques définies par leur pédagogue dans l'outil. Cette phase fera l'objet de précisions un peu plus loin; pour l'instant, pendant qu'un élève évalue ses pairs, le même nombre de pairs évaluera son travail, après quoi à la troisième phase, dite de « réflexion », il prendra connaissance de la rétroaction mise en application à son travail, il notera cette rétroaction puis il sera autorisé à réviser son travail en vue d'une présentation finale par l'intégration de la rétroaction qu'il juge utile.

Les auteurs Joordens, Paré et Collimore (2014) présentent une analyse pédagogique détaillée de ce processus ainsi que de la façon dont celui-ci appuie l'acquisition des compétences transférables de base. En bref, le processus exige des élèves qu'ils exercent à répétition la pensée critique, la pensée créative, la communication orale et la communication réceptive, et ce, dans un contexte riche en rétroaction qui contribue à rehausser la sensibilité métacognitive. Outre une répartition pédagogique, le rapport montre également en détail des expérimentations dans lesquelles il est attesté que le processus décrit comporte des effets positifs considérables sur les mesures invoquées pour refléter la pensée critique et la sensibilité métacognitive telles qu'elles sont décrites précédemment.

Aux fins actuelles, l'attention sera prêtée à une augmentation potentielle du processus ou, de façon précise, à l'intégration raisonnée des rubriques validées à la phase d'évaluation. Par conséquent, l'idée centrale, c'est que pendant que les élèves évaluent le travail de leurs pairs, ils procèdent entre autres à la notation du travail au moyen d'une rubrique validée en lien avec la compétence transférable qu'on cherche à mesurer. Étant donné que chaque composition est évaluée par six pairs, la moyenne des notes qui y sont décernées (décrites ci-après comme la « moyenne par les pairs ») fournira une mesure de cette compétence. En ce qui touche précisément le recours à six pairs, il ressort d'études préalables que les moyennes par les pairs des

notes globales de la qualité sont valides et fiables lorsqu'elles s'appuient sur aussi peu que quatre pairs dans un rapport [Cho et al. (2006)], et aussi peu que cinq pairs comme en témoigne un autre rapport [Paré et Joordens (2008)]. Par conséquent, le nombre six est privilégié lorsqu'on s'appuie sur des moyennes par des pairs en tant que prestation légèrement excédentaire à ce qui est nécessaire.

Lorsque nous faisons allusion aux rubriques validées, les rubriques VALUE mises au point par l'AAC&U en sont un exemple évident [Rhodes (2016)]. L'AAC&U a conçu des rubriques ayant trait à 16 compétences transférables différentes, et toutes ces rubriques ont révélé des niveaux acceptables de fiabilité ainsi qu'une validité apparente nette [Finley (2011)]. Par exemple, si une enseignante donnée souhaitait que ses élèves acquièrent la pensée créative, elle verrait alors à ce que les élèves comprennent que la pensée créative manifestée dans leur composition est le but de base et que, dans l'évaluation du travail des autres, ces derniers puissent recourir à la rubrique de pensée créative de VALUE pour quantifier la pensée créative manifeste dans une composition donnée.

Cette démarche correspond à beaucoup plus qu'un moyen d'arriver à une fin. Outre la prestation de mesures de compétences transférables, elle peut également fournir un moyen amélioré de les enseigner. Dans le domaine de la musique, les élèves ne tardent pas à concevoir qu'il est crucial de « développer son oreille » pour s'améliorer à titre d'interprète [d'après Woody (2012)]. Autrement dit, ils doivent apprendre comment écouter avec exactitude pour ensuite apprécier les choses subtiles mais importantes que les autres musiciens exécutent lorsqu'ils jouent. Ils doivent aller au-delà du fait qu'il s'agit d'« un air agréable » ou « qui les rend heureux » pour en arriver à s'intéresser à « l'utilisation des silences » ou « à la création de séquences ou à la façon de jouer les notes au-delà de ce qui est attendu ». Dès qu'ils perçoivent ces éléments dans l'interprétation d'autrui de même que l'effet que comportent ces techniques, ils sont à même de percevoir comment ils peuvent intégrer ces éléments à leur propre interprétation.

La mise en application des rubriques en lien avec les compétences transférables constitue, en quelque sorte, un moyen de « développer son oreille » en lien avec cette compétence. À titre d'exemple, plutôt que de tenter de produire une pensée critique, les élèves doivent la détecter dans le travail des autres, trouver les cas dans lesquels l'élève fait quelque chose de bien, le reconnaître puis évaluer les effets qui sont produits. Par la suite, tout comme le musicien, ils sont à même de recourir à de telles techniques dans leur propre travail.

Une autre façon de concevoir ce processus, c'est d'ouvrir les élèves à la capacité d'intervertir le point de vue entre celui de l'interprète et celui du public. Chaque bon interprète est en mesure de se situer au-delà de soi pour envisager son interprétation sous l'angle du public. Habituellement, dans les universités, les élèves doivent essentiellement exécuter des œuvres, ou du moins en composer. Rarement doivent-ils évaluer des

œuvres, encore moins de façon structurée comme dans la mise en application de rubriques. Dès que l'élève est en mesure de percevoir les compositions sous l'angle de l'évaluation, il peut s'approprier cette compétence puis l'activer dans sa composition, ce qui devrait en faire un meilleur compositeur.

Ce processus de mise en application des rubriques, d'une importance cruciale à l'extension, est souvent qualifié d'indifférent au contenu. Toute tentative d'officialiser et de mesurer l'apprentissage de ces compétences d'une façon permettant de concurrencer l'apprentissage de l'information doit décrire un processus qui pourra être mis en application globalement dans une vaste gamme de contextes d'apprentissage. Étant donné que l'enseignement détermine le type de composition que les élèves rédigeront au départ, ces compositions peuvent se trouver dans n'importe quel domaine d'études et les élèves peuvent se situer à n'importe quel niveau d'études. À titre d'exemple, à l'Université de Toronto, une évaluation par les pairs par voie technologique est désormais employée dans plus de 50 cours par session, allant du cours d'initiation à la psychologie donné en première année aux cours donnés en médecine à de petits groupes aux niveaux supérieurs. En outre, puisqu'une rubrique peut être adaptée à n'importe quelle compétence transférable – par exemple, les rubriques de l'AAC&U se rapportent à 16 compétences transférables possibles – le point de mire du devoir peut également varier très librement.

À première vue, le processus proposé semble alors comporter plusieurs avantages. Il procure aux élèves un exercice structuré permettant d'évaluer les compétences transférables, lequel peut se révéler tout particulièrement utile compte tenu de l'ajout de la mise en application de rubriques aux compositions par les pairs. Cet exercice est conçu de telle sorte qu'il favorise le traitement approfondi, et par conséquent, l'apprentissage efficace. De plus, il présente pour chaque élève une mesure de la vigueur d'une compétence donnée telle qu'elle se manifeste dans le travail de l'élève. Enfin, il peut être mis en application dans presque chaque contexte de classe, quel que soit le niveau, la matière ou le mode de prestation des cours.

Si le potentiel de l'alliance de l'évaluation par les pairs par voie technologique à l'utilisation raisonnée des rubriques validées semble impressionnant, il pourrait manifestement tirer parti d'une évaluation directe de ce potentiel. L'étude est conçue pour fournir une telle évaluation, et, dans les deux phases, elle le fait par l'évaluation de deux hypothèses sur lesquelles repose ce potentiel; il y a, d'une part, l'hypothèse selon laquelle les moyennes fondées sur les notes des rubriques fournies par les pairs sont valides et, d'autre part, l'hypothèse comme quoi le processus est suffisamment souple pour appuyer une mise en application à vaste échelle dans l'ensemble de l'écosystème de l'enseignement supérieur.

Phase 1

Le processus proposé part du principe que la mesure fournie par la moyenne des notations attribuées par les pairs procure une mesure valide de la compétence en question. Il ressort de recherches menées au préalable que lorsque des notations d'au moins quatre ou cinq pairs sont combinées, elles procurent un bon paramètre de prévisions des notations qui seraient attribuées par des assistants à l'enseignement et qui se comparent à celles attribuées par un autre assistant à l'enseignement [Cho, Schunn et Wilson (2006); Paré et Joordens (2008)]. Dans une autre étude auprès d'élèves du secondaire, il ressort que les notations attribuées par les pairs ont également permis de prédire les notes d'experts à un niveau élevé [Tseng et Tsai (2007)]. Toutefois, dans ces études faites au préalable, les élèves devaient seulement noter globalement la qualité; ils n'évaluaient pas précisément la validité des notes par rubrique, encore moins les rubriques axées sur les compétences transférables quelque peu complexes auxquelles nous nous intéressons essentiellement ici.

Par conséquent, le point de mire fondamental de la Phase 1 consiste à évaluer directement la validité des estimations des compétences transférables d'après les moyennes par les pairs, dans un contexte où les élèves appliquent la rubrique VALUE de pensée critique au travail de leurs pairs. De telles estimations fondées sur les moyennes par les pairs seront comparées aux estimations fournies par les assistants à l'enseignement de même que les experts formés par l'AAC&U pour mettre en application les rubriques VALUE. Dans l'évaluation de la validité, nous examinons à la fois la concordance quant à l'ordre de l'importance de la qualité (c.-à-d. les coefficients de corrélation de Pearson) et en ce qui touche la concordance réelle des notes (c.-à-d. mesures à l'aide du coefficient Kappa), après quoi nous comparons la concordance des moyennes par les pairs avec celles des assistants à l'enseignement et des experts de même que les niveaux de concordance entre les deux groupes d'experts. Nos constatations semblent indiquer que les moyennes par les pairs procurent de bons paramètres de prévisions des notes attribuées par les assistants à l'enseignement et les experts formés, mais que les résultats ne sont pas aussi probants en ce qui touche la concordance réelle, essentiellement parce que les moyennes par les pairs ont tendance à être quelque peu gonflées par rapport aux notes données par les deux autres groupes.

De plus, nous avons recueilli quelques données qualitatives en lien avec les évaluations par les pairs au moyen des données du questionnaire. De façon précise, sitôt après avoir achevé l'activité, les élèves devaient répondre à plusieurs questions ayant trait au déséquilibre proposé entre la formation relative aux renseignements et la formation relative aux compétences ainsi qu'à des questions ayant trait au processus auquel ils venaient tout juste de participer en tant que méthode pour corriger ce déséquilibre (voir l'annexe B). La nécessité de fournir un certain contexte a rendu ces questions un peu plus compliquées que celles qui figurent habituellement dans un questionnaire. Par exemple, l'une des questions allait comme

suit : « dans la plupart des cours, vous recevez beaucoup de renseignements à propos de certaines matières, mais il est également important que vous puissiez acquérir des compétences transférables, comme la pensée critique, la pensée créative ou la communication claire et efficace. Dans quelle mesure souscrivez-vous à l'énoncé suivant : "J'estime que les cours donnés habituellement à l'université me poussent à penser de façon critique et créative, et qu'ils me donnent de l'expérience ayant trait à la communication efficace de mes idées"? ». À notre connaissance, personne n'avait tenté auparavant d'évaluer le point de vue des élèves quant à ces enjeux et, étant donné que nous propulsons nos élèves en plein dans les nouvelles démarches pour résoudre ces enjeux, nous étions d'avis que leurs points de vue seraient intéressants et potentiellement utiles.

Méthode

Participants

Quatre ensembles distincts de participants ont apporté des données à la présente étude.

Élèves – Les élèves participant à la présente étude comptaient parmi 1 800 élèves inscrits au cours d'initiation à la psychologie offerts au campus de Scarborough de l'Université de Toronto. Tous les élèves ont mené à bien l'activité d'évaluation par les pairs par voie technologique dans le cadre de leur cours. Sitôt achevée, la phase d'évaluation par les pairs de l'activité en ligne, une demande est apparue à l'écran pour demander aux élèves leur consentement à utiliser leurs données à des fins de recherche. Cette demande insistait sur le fait que la participation des élèves était volontaire et que non seulement leur participation à l'étude était indépendante du cours, mais que l'enseignant du cours n'apprendrait, ni ne saurait, si l'élève en question consentait ou pas à y prendre part. Les élèves devaient choisir l'une de deux options au moyen d'un menu déroulant : ou bien « Je consens à ce que mes données servent à des fins de recherche », ou bien « Je préférerais que mes données NE SERVENT PAS à des fins de recherche ». Des 1 683 élèves ayant mené à bien l'activité, 694 participants ont consenti à nous laisser utiliser leurs données, ce qui a donné un taux d'acceptation de 41 %.

Assistants à l'enseignement (AE) – Deux assistants à l'enseignement chevronnés, choisis parmi l'ensemble d'assistants actuels et préalables du cours d'initiation à la psychologie, devaient noter le travail de l'élève au moyen de la rubrique VALUE de la pensée critique. Ces AE ont été sélectionnés en fonction de leur assiduité évaluée et de leur travail de grande qualité dans le cadre d'évaluations préalables du rendement des élèves. Ils ont touché une rémunération horaire pour leur participation à l'étude.

Évaluateurs experts – Deux évaluateurs experts recommandés et favorisés par l'AAC&U ont été embauchés pour prendre part à l'étude. Ces experts sont passés par une formation rigoureuse quant à la mise en application de la rubrique VALUE de la pensée critique au travail des élèves.

Professeur de cours – Outre les groupes ou la paire d'évaluateurs ci-dessus, il a été convenu de l'intérêt de demander au professeur du cours de mettre également en application la rubrique VALUE de la pensée critique dans la présente étude. Malgré son activité de recherche dans l'étude de la pensée critique, le professeur n'avait préalablement acquis ni expérience, ni formation dans la mise en application des rubriques VALUE.

Procédure

Les données des élèves ont été recueillies dans le contexte d'une activité en ligne d'évaluation par les pairs appuyée par peerScholar, une technologie éducative infonuagique. Cette activité comptait pour 12 % de la note finale du cours et, à l'étape de la composition, les élèves devaient prendre position à propos d'un débat d'actualité et présenter de bons arguments : ce faisant, il leur fallait révéler le plus grand nombre possible de données probantes ayant trait à la pensée critique.

Chaque activité de peerScholar est composée de trois phases : une phase de création, une phase d'évaluation et une phase de réflexion. Chacune des phases s'est déroulée comme suit.

Phase de création – Chaque élève devait soumettre un argumentaire d'une ou à deux pages (à double interligne) en réponse à la question suivante : « La double norme entre l'éthique de recherche sur les animaux par rapport à celle sur les êtres humains est-elle justifiable? ». Les élèves étaient précisément tenus de trouver et lire au moins trois sources fiables en lien avec le sujet, puis de rédiger un argumentaire persuasif ayant trait à un sous-enjeu de cette vaste question. Par exemple, un élève pouvait au bout du compte faire valoir que la double norme était justifiable parce que les animaux sont dépourvus d'auto-conscience. Cet élève devait ensuite appuyer son argumentaire à l'aide d'une recherche explicite et d'arguments convaincants. Une fois l'argumentaire rédigé, l'élève ouvrait une session dans peerScholar puis soumettait son travail avant la date butoir de cette phase.

Phase d'évaluation – Une fois toutes les compositions remises par les élèves, ces derniers devaient de nouveau ouvrir une session pour procéder à un cycle d'évaluation par les pairs et d'auto-évaluation. De façon précise, ils avaient accès aux compositions soumises par six de leurs collègues de classe, puis ils devaient évaluer chaque composition et donner une rétroaction constructive s'y rapportant. Les compositions étaient dépersonnalisées et choisies aléatoirement à partir de l'ensemble des soumissions. Les

évaluations particulières qu'ils devaient exécuter étaient déterminées par l'enseignant. Dans ce cas-ci, ces évaluations consistaient en ce qui suit : a) les élèves pouvaient joindre des commentaires en ligne directement aux travaux par les pairs pour signaler des difficultés de rédaction minimales comme, entre autres, des corrections grammaticales ou des corrections de citations; b) les élèves devaient remplir une matrice mise en correspondance avec la rubrique VALUE de la pensée critique (voir l'annexe A); c) les élèves devaient donner à chacun des pairs des conseils constructifs en insistant sur l'élément qui – s'il était amélioré – contribuerait à rehausser essentiellement l'ensemble du travail, après quoi ils devaient indiquer à leurs pairs une façon par laquelle ils pouvaient améliorer le tout. Aux fins qui nous concernaient, les notes de la matrice VALUE étaient les plus cruciales.

Avant d'exécuter l'évaluation de leurs pairs au moyen de la rubrique, les élèves devaient regarder une vidéo de 24 minutes (<http://heqco.s3-website.ca-central-1.amazonaws.com>) qui les renseignait sur la façon de mettre optimalement en application la rubrique. Après la vidéo, ils étaient tenus de répondre avec exactitude à cinq questions relatives au contenu avant de passer à la phase d'évaluation. Cette étape consistait en une tentative de déterminer en quelque sorte la capacité de nos élèves à mettre en application la rubrique.

Phase de réflexion – Pendant que chaque élève donnait de la rétroaction à six pairs au cours de la phase d'évaluation, six pairs donnaient également de la rétroaction en lien avec la composition de chaque élève donné. Dans la phase de réflexion, plusieurs choses se produisent. D'abord, les élèves obtiennent l'accès à l'ensemble complet de la rétroaction donnée par leurs pairs. Par la suite, pour chaque rétroaction, ils devaient attribuer une note à l'utilité de même qu'au ton de la présentation. Après avoir évalué chacune des rétroactions, ils devaient alors avoir l'occasion de réviser leur composition en vue d'une présentation finale d'après la rétroaction utile reçue. Une fois la révision terminée en vue de la soumission, ils ont rédigé une courte réflexion dans laquelle ils ont justifié les révisions apportées de même que les révisions proposées par les pairs mais qu'ils ont choisi de ne pas mettre en œuvre.

Au bout du compte, les élèves étaient évalués en fonction de quatre composantes. Leur composition en version provisoire valait 2 %, leur composition en version définitive valait 4 %, la qualité de leur rétroaction aux pairs valait 3 %, pendant que leur capacité à justifier les révisions qu'ils ont apportées ou pas compte tenu de la rétroaction reçue par les pairs valait 3 %. Les composantes étaient notées par les AE, une fois l'activité achevée.

À titre de dernière étape à l'issue de la phase de réflexion, les élèves ont reçu un ensemble de questions (voir l'annexe B) dans le même contexte de devoirs. Les questions auxquelles ils devaient répondre servaient

à saisir leurs opinions sur a) l'importance d'exercer et de mesurer les compétences transférables; b) la mesure dans laquelle ils souscrivaient à la démarche particulière à l'étude dans le cadre de ce travail.

Structure, documents et contexte

L'activité fondamentale à la présente étude comptait parmi plusieurs activités en ligne qui font partie normalement du cours d'initiation à la psychologie. Toutes ces activités sont décrites dans le syllabus de cours, lequel englobe également des détails quant aux dates butoirs et à l'utilisation du système. L'application est très intuitive, et aucun élève n'a signalé de problème quant à la navigation comme souhaité.

La collecte des données des élèves (c.-à-d. la procédure décrite ci-dessus) s'est déroulée en ligne, et le système peerScholar en a assuré la gestion. Chaque phase comportait une date d'ouverture et de fermeture en particulier. Les élèves étaient libres d'achever chaque phase de l'activité au cours de nombreuses séances à l'aide du dispositif et du navigateur de leur choix. Puisque l'activité était en ligne, ils pouvaient la mener à bien à partir du domicile, de l'école ou du lieu de leur choix.

L'un des documents cruciaux employés dans la présente étude est une rubrique conçue pour évaluer les compétences en pensée critique : la rubrique VALUE de la pensée critique (voir l'annexe A). Elle résulte de recherches considérables menées par l'AAC&U (association des collèges et universités des États-Unis) en vue d'optimiser la fiabilité, la validité et les possibilités de mise en application à des contextes généraux [Finley (2011)]. L'AAC&U offre également des programmes de formation que les membres du corps professoral peuvent suivre pour rehausser leur savoir-faire et leur constance dans la mise en application des rubriques. Nos élèves n'avaient pas reçu une telle formation, mais comme nous l'avons mentionné précédemment, nous avons produit un module par voie électronique pour déterminer à peu près leur capacité, lequel est accessible en ligne (<http://heqco.s3-website.ca-central-1.amazonaws.com>).

Nous avons également créé un questionnaire conçu précisément pour la présente étude. Les discussions axées sur la recension des compétences transférables de base et des méthodes en vue de leur acquisition comportent rarement le point de vue manifeste des élèves. Étant donné le caractère véritablement exceptionnel du processus évalué ici, il nous a semblé opportun de sonder les élèves tant en ce qui concerne le processus particulier que les buts généraux liés à l'acquisition des compétences transférables. Puisqu'un tel questionnaire n'existait pas, nous l'avons créé nous-mêmes : il figure à l'annexe B. À cette étape-ci, ce questionnaire avait simplement pour but d'obtenir un premier aperçu du point de vue des élèves.

Les données des AE et des experts ont été recueillies à l'aide d'un outil en ligne que nous avons constitué, appelé Rubricaun, en vue d'optimiser le processus de mise en application de la rubrique aux travaux des élèves. Cet outil montre aux utilisateurs le travail à évaluer de même que la rubrique sous une forme interactive (voir l'annexe C). Un tel processus a permis de diminuer considérablement le temps et les efforts requis à l'évaluation de vastes collections de documents (comparativement à l'évaluation de documents en format papier ou à l'aide de tableurs). De plus, les AE ont regardé la même vidéo de formation que les élèves avant de mener à bien les évaluations.

Les AE et les experts ont évalué le même ensemble aléatoirement choisi de 200 compositions tirées de l'ensemble de 665 compositions : les 694 élèves qui ont consenti à participer, à l'exclusion des élèves qui n'ont pas fait six évaluations de pairs ou n'ont pas été évalués par six pairs. Autrement dit, si un élève ne parvenait pas à achever la phase d'évaluation, jusqu'à six élèves recevaient alors une révision de moins que ce qui était prévu. Ce phénomène s'est rarement manifesté, tout particulièrement parce que les élèves recevaient la note zéro à cette activité lorsque pareille situation se produisait sans motif. Dans l'activité en cours, 29 des 694 élèves qui avaient donné leur consentement (4,2 %) n'ont pas obtenu un ensemble complet d'examen par les pairs.

Résultats

Nous avons analysé les résultats de ce travail en deux sections. La première porte sur l'évaluation de la fiabilité des notes des rubriques fournies par les élèves, et elle englobe les données strictement issues des 200 compositions choisies aléatoirement qui sont décrites dans la section précédente. La deuxième a trait aux constatations découlant du questionnaire et englobe des données provenant de l'offre complète à l'automne 2017 du cours Initiation à la psychologie, au campus de Scarborough de l'Université de Toronto.

Données sur la fiabilité

Au bout du compte, notre but fondamental consiste à évaluer la fiabilité des notes des rubriques obtenues par la mise en moyenne des notes fournies par des élèves de premier cycle qui évaluent leurs pairs sans formation en la matière. Rappelons que 200 compositions ont été choisies aléatoirement à partir des compositions offertes suivant ce qui a été décrit au préalable, et que chaque composition choisie a fait l'objet d'une évaluation par six pairs, une paire d'AE aux cycles supérieurs de même qu'une paire d'experts à l'échelle du corps professoral qui ont suivi une formation fournie par l'AAC&U. De plus, le professeur du cours a noté toutes les compositions simplement pour donner une idée de la fourchette dans laquelle les notes d'un membre du corps professoral pourraient se situer, et de la mesure dans laquelle ils peuvent se comparer à d'autres dans le continuum du savoir-faire. Après avoir examiné les niveaux de concordance

entre des ensembles particuliers d'évaluateurs, nous pouvons percevoir clairement la fiabilité dans l'évaluation des rubriques par les pairs.

À titre de première étape, nous avons examiné les statistiques descriptives en lien avec les notes liées à chaque groupe, suivant ce qui est présenté dans le tableau 1. Les moyennes dans le tableau 1 sont fonction de la note des rubriques au total en lien avec chaque composition. Chaque note de rubrique reflétait concrètement cinq facteurs distincts de la pensée critique, plus précisément : l'explication des enjeux; les données probantes; l'influence exercée par le contexte et les hypothèses; la position de l'élève; de même que les conclusions et les résultats connexes. Chacun de ces facteurs pouvait valoir jusqu'à quatre points au maximum. La moyenne reflète donc l'addition des cinq facteurs, divisée ensuite par cinq, pour en arriver à une note moyenne globale.

Il convient de souligner que la formation donnée par l'AAC&U autorise la note 0, tandis que les évaluateurs dans la présente étude étaient tenus de donner une note minimale de 1 (c.-à-d. la catégorie la plus basse à chaque composante des rubriques). Il s'agissait tout simplement du résultat de notre propre confusion, car les rubriques mêmes (voir l'annexe A) ne montrent pas de note 0, de sorte que nous sommes partis du principe que les notes les plus basses correspondaient à 1. Avant de joindre les experts et d'apprendre que notre présomption était erronée, nos élèves étaient déjà passés par le processus sans l'option 0. Nous avons donc exigé des autres évaluateurs qu'ils procèdent de la même manière que nos élèves, par souci de cohérence.

Comme les moyennes liées aux quatre groupes (c.-à-d. les quatre colonnes du tableau à partir de la droite) semblent l'indiquer, les faibles niveaux de savoir-faire sont en corrélation avec les notes moyennes supérieures. Une analyse ANOVA unidimensionnelle entre les facteurs a confirmé que ces trois moyennes affichent une différence fiable ($F(3, 796) = 199,18; p < 0,001$). Des tests-t de suivi à échantillons indépendants ont révélé que la note moyenne donnée par les experts était considérablement inférieure à celle donnée par les assistants à l'enseignement ($t(398) = 13,14; p < 0,001$), ce qui est en soi passablement inférieur à la note moyenne donnée par les pairs ($t(398) = 16,18; p < 0,001$). Selon ce qui est sous-entendu, la note moyenne donnée par les experts était aussi considérablement inférieure à la note moyenne des pairs ($t(398) = 31,35; p < 0,001$). Les notes données par le professeur ne différaient pas des notes moyennes des AE ($t(398) < 1$), mais étaient considérablement supérieures à la note moyenne des experts ($t(398) = -8,80; p < 0,001$), et considérablement inférieures à la note moyenne par les pairs ($t(398) = 12,17; p < 0,001$).

Tableau 1 : Moyennes et écarts-types des notes fournies par les évaluateurs

	Expert 1	Expert 2	AE 1	AE 2	Moyenne des experts	Professeur	Moyenne des AE	Moyenne des pairs
Moyenne	1,68	1,59	2,35	1,82	1,64	2,08	2,09	2,73
Écart-type	0,35	0,30	0,44	0,48	0,24	0,29	0,39	0,40
Nota : Les notes qui composent ces moyennes se situent dans une fourchette allant de 1 à 4								

Ce que de telles constatations semblent indiquer, c'est que plus une personne acquiert de l'expérience dans la mise en application de la rubrique, plus les notes qu'elle décerne diminuent. Rappelons que les experts ont suivi un cours complet parrainé par l'AAC&U, conçu pour leur enseigner comment mettre en application les rubriques de façon à ce que leur travail soit constant et valide. Les AE et les élèves ont regardé une courte vidéo ayant pour objet de rehausser la constance et la compréhension en ce qui touche la mise en application numérique. Le professeur du cours a produit la courte vidéo. Par conséquent, tout repose sur un continuum de savoir-faire qui va de l'élève et de l'AE jusqu'à l'expert, en passant par le professeur. La tendance des résultats perçue ici est vraisemblablement attribuable aux capacités accrues de discernement qui accompagnent le savoir-faire, de sorte que l'expert est en mesure de percevoir les différences en matière de qualité qui échappent encore à l'œil du novice. Plus la sensibilité d'une personne à de telles différences augmente, moins cette personne est encline à décerner une note élevée à une composition qui est tout juste passable [MacLean et al. (2010)].

Une fois notées ces différences dans les notes globales des groupes, nous nous tournons maintenant vers l'examen de la fiabilité des notes par rubrique qui sont fondées sur la moyenne des notes attribuées par les pairs. Ce faisant, nous calculons deux statistiques pour chaque analyse. La première consiste en un simple coefficient de corrélation de Pearson qui permet essentiellement d'évaluer la capacité d'un groupe de notation (p. ex., les élèves, les AE, etc.) à prédire la notation qu'une composition obtiendra d'un autre groupe de notation. Les coefficients de corrélation de Pearson permettent de saisir les associations dans l'ordre d'importance des compositions et ne subissent pas l'influence des valeurs particulières des notations en soi. Ensuite, nous calculons également les notes du coefficient Kappa, lesquels permettent d'évaluer la véritable concordance entre les évaluateurs et donc de saisir la mesure dans laquelle certaines notes décernées par des groupes d'évaluateurs vont de pair.

Pour bien comprendre la façon par laquelle ces statistiques saisissent différemment la « prédiction », un réexamen rapide du tableau 1 se révèle utile. Il convient de souligner que si les experts ont attribué en moyenne des notes inférieures à celles des AE, chez les experts (tout particulièrement) de même que les AE

(dans une moindre mesure), les notes étaient semblables. Par conséquent, lorsque nous calculons les coefficients Kappa, nous devons nous attendre à des notes intergroupes supérieures à celles entre les groupes, parce que le niveau différent des notes dans l'ensemble entraînera une diminution du coefficient Kappa. Cependant, dans la mesure où une composition ayant obtenu une note relativement élevée dans le contexte de l'évaluateur particulier obtient également une note élevée dans le contexte de différents évaluateurs, la corrélation pourrait être élevée. La corrélation n'est pas sensible au niveau des notes globales, mais porte plutôt sur l'ordre d'importance de la note dans le contexte d'un évaluateur particulier.

Notre point de mire sera au bout du compte la qualité des moyennes par les pairs en tant que mesures de la pensée critique, mais les mesures au sein des évaluateurs en lien avec le groupe d'évaluateurs experts et les groupes des AE procurent des points de référence utiles à des fins de comparaison. Les évaluateurs experts nous donnent une idée de ce qui est vraisemblablement la norme la plus élevée possible, pendant que les AE fournissent un point de comparaison qui témoignent de la pratique normative actuelle dans l'enseignement supérieur, laquelle consiste à compter habituellement sur l'assistance à l'enseignement pour effectuer les évaluations. Les coefficients de corrélation et de Kappa ayant trait à chacun de ces groupes figurent dans le tableau 2.

Tableau 2 : Coefficients de corrélation et Kappa chez les experts et les AE

	Experts	Assistants à l'enseignement
Coefficient de corrélation de Pearson	0,57**	0,45**
Coefficient Kappa de Cohen	0,40**	0,15**
<p>Nota : Les coefficients Kappa relatés sont pondérés sous forme de coefficients Kappa de Cohen. * Indique un résultat significatif à <0,05 ** Indique un résultat significatif à p <0,01</p>		

Les experts formés par l'AAC&U ont révélé le niveau le plus élevé de concordance dans leur évaluation du travail, tant en ce qui concerne les niveaux de concordance relatifs (c.-à-d. le coefficient de corrélation) et absolu (c.-à-d. le coefficient Kappa). Les AE ont révélé un niveau élevé de concordance relative, mais un coefficient Kappa bien inférieur, ce qui témoigne du fait qu'un AE a attribué des notes nettement supérieures à l'autre. L'une et l'autre des corrélations sont en phase avec ou, dans le cas des experts, quelque peu supérieures aux corrélations relatées au préalable dans le travail subjectif de notation des AE [Cho, Schunn et Wilson (2006); Paré et Joordens (2008)]. En ce qui touche la concordance absolue, le coefficient Kappa de Cohen chez les experts est en phase avec, sinon supérieur à, celui préalablement observé chez les experts formés par l'AAC&U. À titre d'exemple, dans les recherches préalables sur la

fiabilité de la rubrique VALUE de la pensée critique, on relate un coefficient Kappa de Cohen de 0,29 entre une paire d'évaluateurs experts [Finley (2011)].

Une fois obtenues ces valeurs à titre de points de référence, nous nous penchons maintenant sur les moyennes par les pairs. Une complexité statistique supplémentaire justifie une analyse à cette étape-ci. Lorsque les AE et les experts ont noté 200 compositions, les mêmes deux AE et les deux mêmes experts ont noté toutes les 200 compositions. Toutefois, lorsque les pairs ont évalué ces compositions, un ensemble différent de six pairs a contribué à chaque moyenne par les pairs. Voilà un élément sous-optimal sur le plan statistique, comme c'est parfois le cas lorsque les recherches sont faites sur le travail qui a véritablement eu lieu dans une véritable classe. Toutefois, il importe d'attirer l'attention sur l'effet que cela pourrait comporter sur les résultats. Les analyses que nous employons se basent sur un ensemble constant d'évaluateurs, comme c'est le cas en ce qui touche nos données sur les AE et les experts. Lorsque les évaluateurs changent, comme c'est le cas en ce qui touche les données des pairs, cela introduit un autre facteur de variation qui, éventuellement, diminuera la mesure dans laquelle les moyennes par les pairs permettront de prédire, ou bien relativement (c.-à-d. le coefficient de corrélation), ou bien exactement (c.-à-d. le coefficient Kappa), les autres ensembles de données. Par conséquent, dans la mesure où ce phénomène influe sur nos données, il en résultera une sous-estimation des rapports réels, ce qui rendra les données probantes des relations que nous percevons encore plus fiables.

Avant de nous pencher sur les coefficients de corrélation et Kappa qui témoignent des niveaux de concordance relatifs et absolus des moyennes par les pairs avec les notes des AE et des experts, nous avons d'abord quantifié la concordance au sein de chacune des notes attribuées par les pairs englobées dans chaque moyenne. Il est utile de documenter cette valeur, car les études à l'avenir pourront s'intéresser, disons, au suivi de la façon dont la concordance des valeurs, qui composent une moyenne, fluctue au fur et à mesure que le savoir-faire est acquis. Afin de calculer cette valeur, nous avons calculé les coefficients Kappa de Cohen pondérés pour chacune des 15 paires distinctes d'évaluateurs, pour ensuite établir la moyenne de ces valeurs. La mesure qui en résulte, quant à la concordance chez les élèves, s'établit à 0,95. Il n'est pas étonnant de constater que, parmi les trois groupes que nous avons examinés, ce sont les élèves de premier cycle en première année qui révèlent le niveau de concordance absolu le plus faible.

Bien entendu, le processus proposé en est un qui s'appuie sur les notes attribuées par les pairs en moyenne, plutôt que sur les notes attribuées par les pairs en particulier. Les coefficients de corrélation et de Kappa se rapportant aux notes attribuées par les pairs en moyenne, comparativement aux notes fournies par nos autres groupes (les moyennes s'y rapportant), figurent au tableau 3. Une fois de plus, comme le montre le tableau 1, les pairs ont été plus généreux dans les notes attribuées, et cet élément à lui seul comportera de fortes retombées positives sur la concordance absolue saisie par les coefficients Kappa. Par conséquent, les

valeurs presque nulles que nous constatons ici n'ont rien d'étonnant, quoiqu'elles reflètent clairement un biais dans les notes qui, idéalement, ne se produirait pas.

Tableau 3 : Coefficients de corrélation et Kappa entre la moyenne par les pairs et les moyennes des experts et des AE

	Expert 1	Expert 2	AE 1	AE 2	Moyenne des experts	Professeur	Moyenne des AE
Corrélation	0,22**	0,42**	0,35**	0,49**	0,36**	0,35**	0,49**
Kappa	-0,05**	-0,06**	-0,05**	-0,04**	-0,03*	-0,02	-0,03*

Nota : Les calculs particuliers des coefficients Kappa ont été pondérés en tant que coefficients Kappa de Cohen et coefficients Kappa de Fleiss en ce qui touche les paires d'évaluateurs et les groupes d'évaluateurs, respectivement.
 * Indique un résultat significatif à $p < 0,05$ ** Indique un résultat significatif à $p < 0,01$

Lorsqu'il est question des notes relatives, (c.-à-d. les coefficients de corrélation), les moyennes par les pairs se révèlent plutôt bonnes. Lorsque l'attention est prêtée aux coefficients de corrélation avec les moyennes des experts, les AE et la note présentée par le professeur, les valeurs constatées ne se situent pas dans la même fourchette que celles discernées au sein des groupes d'experts et d'AE. En réalité, la corrélation entre la moyenne par les pairs et la moyenne des AE est légèrement supérieure à celle entre les deux ensembles de notes des AE.

Rappelons que les analyses ci-dessus ont toutes été menées en fonction de la note totale par rubrique pour chaque composition. Par souci d'exhaustivité, nous avons également analysé les résultats répartis selon les cinq facteurs distincts de la pensée critique : l'explication des enjeux; les données probantes; l'influence exercée par le contexte et les hypothèses; la position de l'élève; de même que les conclusions et résultats connexes. Plutôt que de décrire toutes ces constatations dans le texte, nous les avons résumées dans les tableaux 4 et 5, lesquelles correspondent aux tableaux 2 et 3 mais répartissent les données selon les facteurs. Dans l'ensemble, les tendances générales manifestes au sein des données globales sont représentées et fiables à l'échelle des facteurs également. Nul facteur ne se démarque par une réaction différentielle d'une façon qui justifierait un approfondissement.

Tableau 4 : Coefficients de corrélation et Kappa chez les paires d'experts et d'AE en fonction des cinq facteurs de rubrique

	Experts	Assistants à l'enseignement
Explication des enjeux	R : 0,60** K : 0,59**	R : 0,26** K : 0,13**
Données probantes	R : 0,54** K : 0,60**	R : 0,37** K : 0,11**
Contexte et hypothèses	R : 0,52** K : 0,54**	R : 0,28** K : 0,21**
Position de l'élève	R : 0,50** K : 0,58**	R : 0,35** K : 0,16**
Conclusions et résultats connexes	R : 0,59** K : 0,59**	R : 0,27** K : 0,08**
<p>Nota : Les coefficients Kappa relatés consistent en des coefficients Kappa de Cohen pondérés. * Indique un résultat significatif à $p < 0,05$ ** Indique un résultat significatif à $p < 0,01$</p>		

Tableau 5 : Coefficients de corrélation et Kappa entre la moyenne par les pairs et les moyennes des experts et des AE

	Expert 1	Expert 2	AE 1	AE 2	Professeur	Moyenne des experts	Moyenne des AE
Explication des enjeux	R : 0,16* K : -0,10	R : 0,27** K : -0,10	R : 0,29** K : -0,11	R : 0,38** K : -0,16	R : 0,28** K : -0,12	R : 0,23** K : -0,07	R : 0,42** K : -0,11
Données probantes	R : 0,09 K : -0,14	R : 0,20** K : -0,15	R : 0,24** K : -0,17	R : 0,39** K : -0,15	R : 0,35** K : -0,04	R : 0,15* K : -0,11	R : 0,39** K : -0,08
Contexte et hypothèses	R : 0,05 K : -0,16	R : 0,12 K : -0,22	R : 0,27** K : -0,18	R : 0,36** K : -0,14	R : 0,27** K : -0,1	R : 0,09 K : -0,16	R : 0,40** K : -0,10
Position de l'élève	R : 0,02 K : -0,12	R : 0,13 K : -0,16	R : 0,18* K : -0,08	R : 0,25** K : -0,12	R : 0,27** K : -0,13	R : 0,07 K : -0,11	R : 0,26** K : -0,06
Conclusions et résultats connexes	R : 0,09 K : -0,14	R : 0,28** K : -0,15	R : 0,28** K : -0,13	R : 0,36** K : -0,12	R : 0,32** K : -0,11	R : 0,19** K : -0,11	R : 0,41** K : -0,09
<p>Nota : Les calculs particuliers des coefficients Kappa ont été pondérés en tant que coefficient Kappa et Cohen et des coefficients Kappa et Fleiss en ce qui touche les paires d'évaluateurs et les groupes d'évaluateurs, respectivement. * Indique un résultat significatif à $p < 0,05$ ** Indique un résultat significatif à $p < 0,01$</p>							

Données du questionnaire

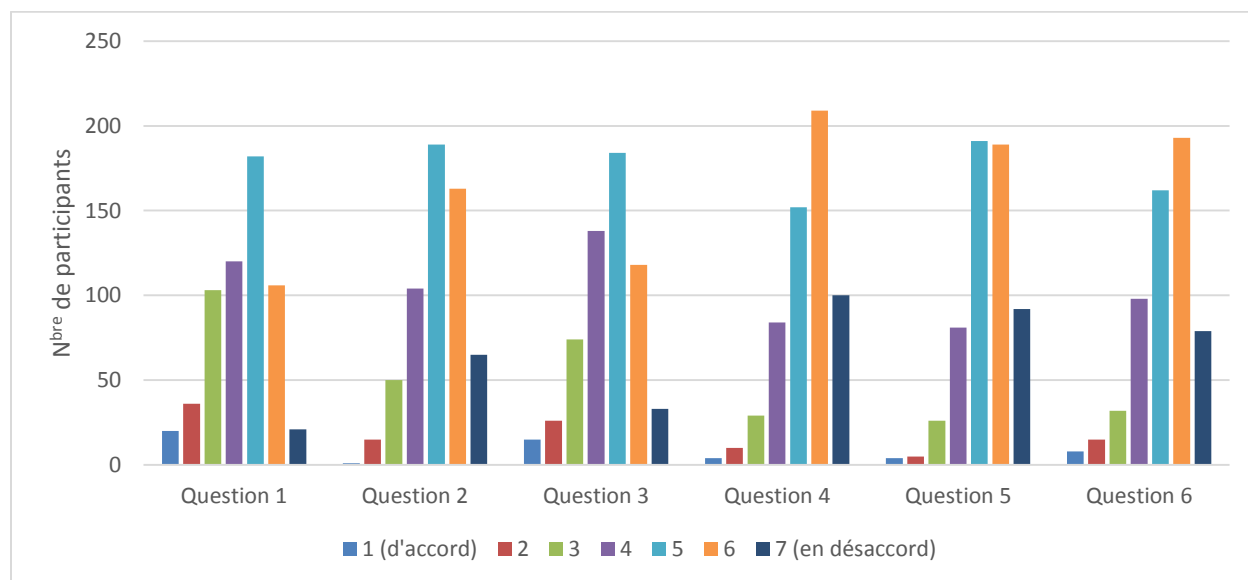
Dès qu'ils ont achevé l'activité, toujours dans la plateforme peerScholar, les élèves ont pris connaissance des six énoncés répertoriés à l'annexe B pour lesquels ils devaient, sur une échelle à sept points de Likert, noter dans quelle mesure ils y souscrivaient. Étant donné que les enjeux au sujet desquels nous leur posons des questions sont relativement nouveaux à leurs yeux, les énoncés présentent souvent un certain contexte avant d'exposer l'énoncé crucial sur lequel l'élève doit se pencher. Les questions sont donc relativement longues et difficiles à représenter dans la cellule d'un tableau. Par conséquent, nous exposons les numéros des questions ici ainsi que les questions intégrales à l'annexe B.

Les données résultantes sont représentées aux graphiques 1A à 1D. Les graphiques 1A et 1C englobent la fréquence des réponses dans chaque valeur de l'échelle de Likert. Les graphiques 1B et 1D révèlent la répartition des réponses qui pourraient être classées dans les catégories « d'accord » ou « en désaccord ». Le pourcentage de réponses « en désaccord » est le reflet de la somme des réponses 1, 2 et 3, divisée par les réponses totales à cette question. Le pourcentage de réponses « d'accord » témoigne de la somme des réponses 5, 6 et 7, divisée par le nombre total de réponses à cette question. Les graphiques 1A et 1B

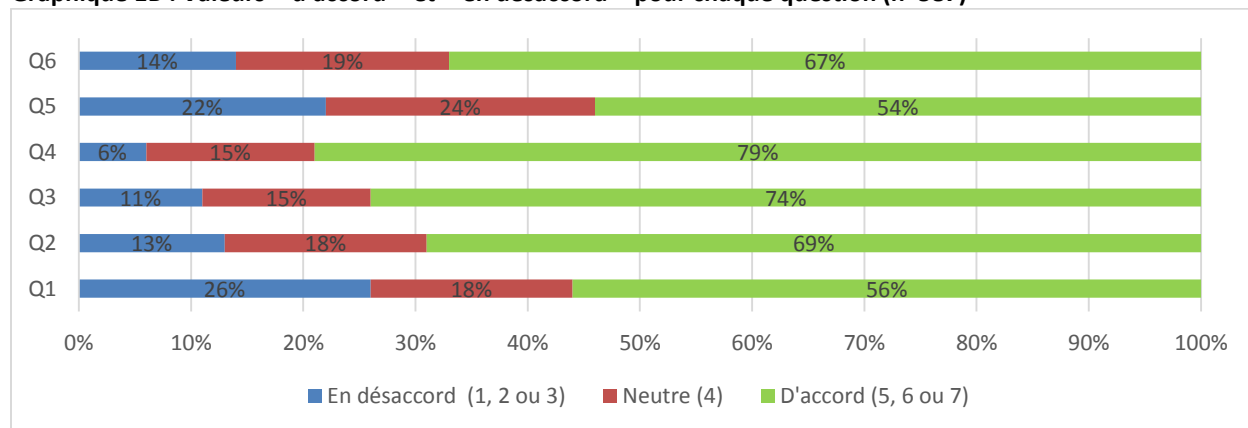
montrent les données des élèves ayant mené à bien l'activité, consenti à ce que leurs données soient divulguées et répondu à la question. Les élèves étaient autorisés à ignorer les questions s'ils le souhaitaient, ce qui explique pourquoi le nombre total de points se situe dans la fourchette de 587 à 588 la plupart du temps (en fonction d'un taux de questions ignorées d'environ 12 %). La constance de ce nombre porte à croire que ce sont les mêmes élèves qui n'ont pas répondu aux questions plutôt que la conséquence d'un biais pour éviter une question en particulier.

Les graphiques 1C et 1D ne révèlent que les réponses en provenance des élèves ayant rédigé les 200 compositions, lesquelles formaient le point de mire de nos analyses précédentes. Les deux ensembles de données présentent une constance élevée et, à ce titre, notre commentaire n'établira pas de distinction entre les deux.

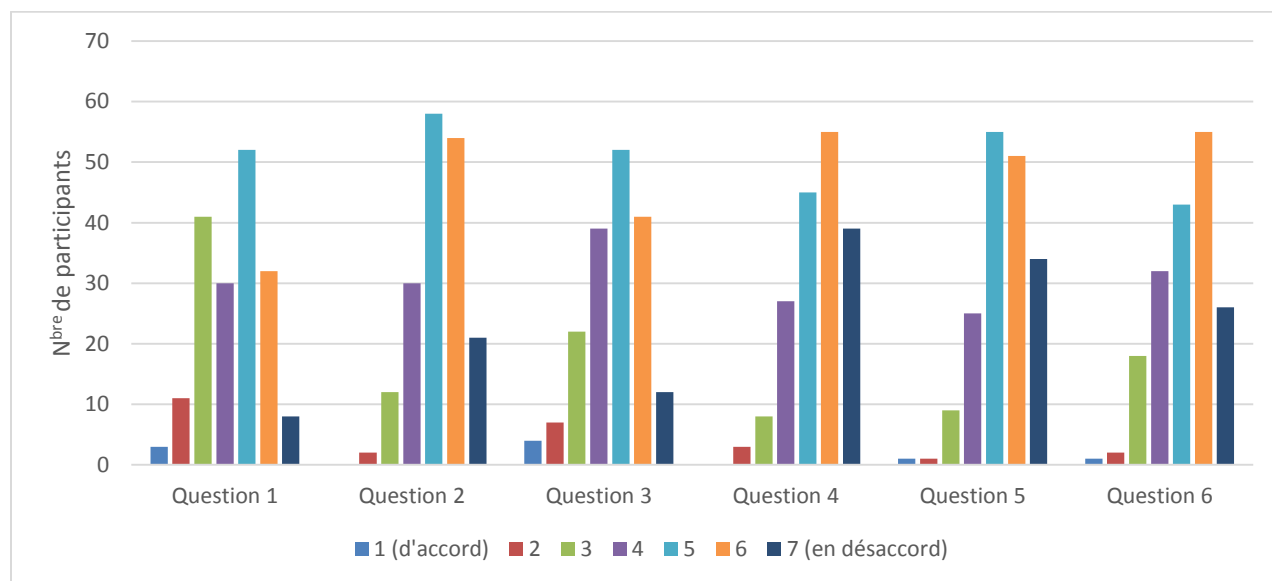
Graphique 1A : Fréquence des réponses de tous les participants dans chaque valeur de l'échelle à sept points de Likert (n=587)



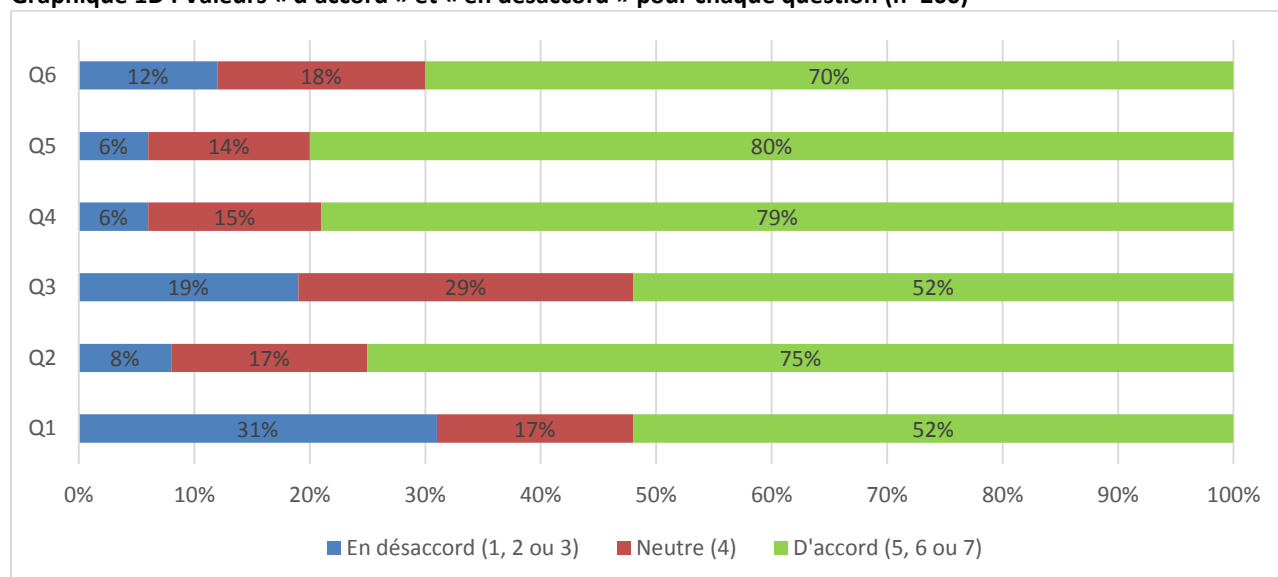
Graphique 1B : Valeurs « d'accord » et « en désaccord » pour chaque question (n=587)



Graphique 1C : Fréquence des réponses de tous les participants dans chaque valeur de l'échelle à sept points de Likert (n=200)



Graphique 1D : Valeurs « d'accord » et « en désaccord » pour chaque question (n=200)



Un court résumé de ces constatations : d'abord, les élèves étaient généralement d'accord (52 %) avec l'énoncé comme quoi les notes reçues dans les cours étaient le reflet de leur apprentissage, mais ce taux d'accord était parmi les plus faibles en réponse aux questions posées. Ils étaient davantage d'accord (71 %) avec l'énoncé selon lequel les cours universitaires types, tels qu'ils sont donnés actuellement, les poussent à penser de façon critique et créative, de même qu'à faire part à autrui de leurs idées. Si nous combinons ces deux questions, nous ne constatons pas de profond mécontentement par rapport au *modus operandi* actuel en contexte universitaire. Les deux questions suivantes se rapportaient à la nécessité d'une pratique structurée pour acquérir des compétences comme la pensée critique. D'après la question 3, les élèves étaient quelque peu d'accord (57 %) avec l'énoncé selon lequel ils ont actuellement l'occasion de mettre ces compétences en pratique, mais le résultat de la question 4 portent à croire qu'ils sont fortement d'accord (78 %) avec l'énoncé selon lequel ils consentiraient à faire des travaux supplémentaires, du type appuyé par peerScholar, afin d'acquérir ces compétences de façon davantage officielle et structurée. Les deux dernières questions avaient davantage trait à la nécessité de mesurer ces compétences. Les élèves étaient fortement d'accord (80 %) avec l'énoncé selon lequel le but qui consiste à mesurer et à documenter les niveaux des compétences transférables est nécessaire. La plupart était également d'accord (71 %) avec la démarche, pour ce faire, qui consiste à passer par le processus sur lequel le présent travail est axé (à savoir l'évaluation par les pairs).

Dans les réponses à ces questions, il ne faut pas percevoir la conviction définitive des élèves à propos de ces enjeux. Le questionnaire n'a pas fait l'objet d'une validation de fiabilité, la formulation de certaines des questions aurait pu être davantage neutre et, en règle générale, mieux vaut intégrer certaines questions dites de « réflexion ». Néanmoins, les réponses nous donnent un aperçu du point de vue des élèves, et il est gratifiant d'apprendre que, lorsqu'on prend le temps de leur expliquer et qu'ils en font l'expérience pour la toute première fois, les élèves perçoivent l'utilité de mesurer les compétences transférables en général et la démarche particulière évaluée dans le présent document.

Analyse

L'auteur Joordens (2018) avance que le système d'éducation actuel est déséquilibré, qu'il privilégie l'assimilation d'information au détriment de l'apprentissage des compétences de base. Il propose, en outre, une démarche particulière qui permettrait de corriger ce déséquilibre et pourrait servir dans n'importe quel contexte de cours et à n'importe quel niveau. Dans la démarche décrite, les élèves doivent mettre en application des rubriques validées en lien avec les compétences transférables de base aux compositions présentées par leurs pairs. Au dire de l'auteur Joordens, cette démarche pourrait être exécutée au moyen de n'importe quelle technologie qui appuie l'évaluation par les pairs.

À la Phase 1, nous prêtons essentiellement attention à une hypothèse centrale de la démarche proposée : les notes par rubrique tirées de la moyenne d'un ensemble de réponses par les pairs procureraient une mesure acceptable de la compétence en question. Pour évaluer cette hypothèse, nous avons comparé les notes selon les moyennes des pairs par rapport aux notes (et aux points de référence) présentés par deux groupes d'experts : les assistants à l'enseignement et les évaluateurs formés. Dans la mesure où les notes fondées sur les moyennes des pairs permettent de prédire les notes des experts, sur les plans relatif ou absolu, un appui supplémentaire à la démarche proposée serait alors fourni.

Nos résultats révèlent que dans les cas où les évaluateurs sont, à tout le moins, des élèves de première année dans le cours d'initiation à la psychologie, les notes fondées sur les moyennes des pairs procurent des notes valables sur le plan relatif, mais des notes médiocres sur le plan absolu. Autrement dit, comme les corrélations semblent l'indiquer, les compositions auxquelles les élèves de premier cycle donnent de fortes notes sont également celles auxquelles les autres groupes donnent de fortes notes, si bien que les élèves sont certainement sensibles aux différences dans les compositions et les notes qu'ils attribuent en témoignent. Toutefois, les faibles taux de concordance absolue révélés par nos analyses des coefficients Kappa met en relief le fait que les élèves de premier cycle sont tout simplement moins sévères. Ils attribuent des notes considérablement élevées, les plus élevées parmi tous les groupes. La conclusion à tirer alors,

c'est que les notes reflètent l'ampleur de la pensée critique présente, mais elles le font de telle sorte qu'il y a un biais de surestimation.

Compte tenu de ces constatations, notre hypothèse est la suivante : plus une personne acquiert de l'expérience ou de la formation quant à la mise en application d'une rubrique, plus elle est à même de percevoir les éléments où il y a matière à amélioration. Par conséquent, les notes relatives à la qualité globale diminueraient avec le savoir-faire. Le fait que les experts dans notre étude aient révélé à la fois des notes moindres, mais des coefficients Kappa élevés au sein du groupe, porte à croire que les notes qu'ils attribuent sont, selon toute vraisemblance, beaucoup plus exactes. Autrement dit, la concordance accrue donne à penser que les experts ne sont pas tout simplement plus sévères de façon aléatoire dans leurs notes, mais qu'il y a plutôt une systématisme quant aux raisons pour lesquelles ils donnent des notes moindres aux compositions, ce qui semble de nouveau indiquer une sensibilité accrue à la compétence en cours d'évaluation.

Dans des travaux précédents, les auteurs Paré et Joordens (2008) ont également constaté que les élèves de premier cycle ont tendance à décerner au travail de leurs pairs des notes légèrement supérieures à celles que décernent les AE, du moins à la Phase 1. Toutefois, une simple description aux élèves de la répartition souhaitée des notes et de la justification appuyant celle-ci permettrait de mettre les notes données par les élèves de premier cycle en phase avec celles décernées par les AE (c.-à-d. l'expérimentation 2). Par conséquent, il se peut qu'une démarche semblable consistant à préciser la répartition voulue d'une façon semblable à celle manifestée par nos experts puisse diminuer de la même façon les notes accordées par les élèves de premier cycle et, de ce fait, accroître leur niveau de concordance absolue avec les experts.

Cela dit, le savoir-faire en soi constitue vraisemblablement un enjeu important avec cette tendance des résultats. Une fois de plus, au fur et à mesure qu'une personne acquiert de l'expérience dans un contexte de discernement, les facultés de discernement de cette personne s'affinent et gagnent en nuance [MacLean et al. (2010)]. Ces personnes perçoivent des problèmes qu'elles ne voyaient pas lorsqu'elles étaient moins expérimentées, et c'est vraisemblablement pourquoi nos experts accordent des notes inférieures au même travail comparativement aux AE et aux élèves de premier cycle. Par conséquent, peut-être que les élèves de premier cycle pourraient seulement atteindre des niveaux élevés de concordance absolue avec les experts par l'acquisition d'une expérience accrue dans la mise en application de rubriques.

Si la mise en application de rubriques dans le contexte des autoévaluations devenait répandue, l'augmentation du savoir-faire des élèves augmenterait de concert avec l'expérience acquise au fur et à mesure de leur progression dans leurs cours. Nous devrions donc attendre des coefficients Kappa supérieurs si nous devons comparer les élèves de deuxième année aux experts, voire supérieurs en troisième année et

encore supérieurs en quatrième année. Toutefois, ce type de démarche passive qui consiste à attendre l'acquisition du savoir-faire suscite une inquiétude. Cela dit, si l'expérience se traduit par une diminution des notes attribuées, alors même si la qualité de, disons, la pensée critique s'améliore chez les élèves de la première à la quatrième année, si ces élèves sont constamment évalués par leurs pairs, et si les évaluations par les pairs révèlent des notes moindres au fur et à mesure que s'acquiert l'expérience, il se peut que les gains effectués en pensée critique soient masqués par la diminution des notes en lien avec l'expérience. Ces deux facteurs seront peut-être différenciés par une prise en compte raisonnée des coefficients Kappa, mais il s'agit d'un exemple qui met bel et bien en relief les complexités à ne pas oublier lorsqu'il vient le temps de mesurer les compétences transférables à l'aide des notes décernées par les élèves.

Une démarche différente consisterait à assurer, d'emblée et de façon active, la prestation d'une formation aux élèves. Nous avons tenté de donner une certaine formation dans l'expérimentation actuelle, mais peut-être qu'il faut une expérience de formation beaucoup plus officielle, qui comporte des exemples de référence préalablement notés et exige des élèves de façon plus formelle de déterminer la capacité de notation au moyen de tels exemples de référence avant d'évaluer le travail des pairs. D'une certaine façon, cet élément plaide en faveur du type de formation qui s'apparente à celle que nos évaluateurs experts de l'AAC&U ont reçue. Comme nous l'avons avancé précédemment, ce type de formation ainsi que l'exercice de mise en application des rubriques correspond vraisemblablement à une démarche pédagogique efficace en vue de l'enseignement de ces compétences, de sorte que la nécessité de donner une formation peut en réalité se révéler une bonne chose à ajouter à la démarche proposée.

Un dernier élément important à propos de ce biais des élèves de premier cycle qui donne des notes par rubrique relativement élevées, c'est que cette tendance même pourrait être mesurée, voire corrigée, moyennant de la prévoyance. De façon précise, pendant que les élèves évaluent les compositions de leurs pairs, il serait possible d'inclure des compositions préalablement notées par des experts. Les différences entre les notes données par les élèves aux compositions par rapport à celles décernées par les experts pourraient servir à corriger concrètement les notes des élèves. Autrement dit, l'ampleur du biais pourrait être quantifiée et corrigée, et cette démarche permettrait – c'est, du moins, ce qui est souhaitable – de percevoir explicitement une diminution du biais au fur et à mesure que l'élève acquiert de l'expérience.

Les notes moyennes, fondées sur les notes par rubrique données par les pairs, permettent-elles de mesurer de façon valide et fiable les compétences transférables en lien avec la rubrique en question? Les travaux actuels portent à croire qu'elles présentent certainement le potentiel pour ce faire, mais il y a dans nos données un biais chez les élèves de premier cycle qui donnent des notes légèrement supérieures, même lorsqu'elles correspondent bien aux notes des experts. Les travaux supplémentaires qui permettraient ou

bien de trouver des moyens de supprimer ces biais, ou bien d'en faire l'estimation et de le corriger, seraient très utiles pour optimiser la validité des moyennes par les pairs.

Outre l'évaluation de la fiabilité des moyennes par les pairs à titre de mesures des compétences transférables, nous avons décrit certaines données du questionnaire conçues pour éclaircir les opinions des élèves à propos du besoin d'acquérir et de mesurer les compétences transférables de façon plus dynamique que nous le faisons actuellement. À une échelle très générale, les données résultantes portent à croire que les élèves ne sont pas complètement insatisfaits de leurs expériences pédagogiques actuelles, mais qu'ils perçoivent l'utilité d'acquérir les compétences transférables et de les mesurer, qu'ils avalisent la démarche qu'ils ont suivie, et qu'ils consentent à travailler davantage de la sorte si, au bout du compte, leur acquisition des compétences s'en trouve améliorée.

Phase 2

Dans la Phase 1, nous cherchions essentiellement à savoir si les moyennes découlant des notes par rubrique données par les pairs pouvaient servir ou non à prédire les notes données par les experts. Les résultats étaient encourageants dans l'ensemble, car la capacité de prédiction des notes données par les experts à l'aide de la moyenne par les pairs se situait dans la même fourchette que la capacité des AE en la matière et de prédiction des notes d'autrui. Certes, quelques enjeux demeurent relativement à un biais de notation élevée qu'il reste à résoudre, mais la démarche semble présenter dans l'ensemble un potentiel favorable.

La Phase 2 que nous relatons ici porte sur une question différente mais tout aussi pertinente : cette démarche où des élèves mettent en application des rubriques dans leur travail d'évaluation de leurs pairs peut-elle fonctionner au sein d'une vaste gamme de contextes pédagogiques? Autrement dit, cette démarche est-elle véritablement extensible à presque n'importe quel contexte pédagogique de la manière requise en vue de l'utilisation répandue au sein de nos systèmes éducatifs actuels? Une telle extensibilité est nécessaire s'il faut percevoir cette démarche comme un moyen de rééquilibrer les systèmes éducatifs au complet comme le préconise l'auteur Joordens (2018), si bien qu'il s'agit d'une question à examiner.

Peut-être vaut-il mieux considérer ce qui suit comme une « démonstration de faisabilité ». La démarche est pure et simple. Les auteurs du présent rapport, essentiellement Joordens, ont fait appel à leurs collègues pour les implorer de faire l'essai du processus décrit dans leur classe, quelle que soit celle-ci. L'objectif consistait à obtenir le plus grand nombre possible de collègues au sein des établissements d'enseignement, voire des pays, pour faire l'essai du processus et leur demander, tant à leurs élèves qu'à eux, de relater les expériences vécues s'y rapportant. Les élèves devaient répondre aux mêmes questions posées à ceux ayant

pris part à la Phase 1. Les membres du corps professoral devaient répondre à un ensemble distinct de questions (voir l'annexe D).

L'un des principaux enjeux sur lesquels nous comptions nous pencher était l'extensibilité sur les plans de l'adoption de la démarche et de son utilisation continue. Autrement dit, pour qu'une nouvelle démarche soit véritablement extensible au sein d'un système d'éducation, certaines vérités doivent être attestées. D'abord, du point de vue des membres du corps professoral, la nouvelle démarche doit être facile à mettre en œuvre. Hormis tout l'engouement que suscitent les technologies perturbatrices, les membres du corps professoral sont déjà bousculés par le temps et, s'ils doivent envisager sérieusement un changement à leur mode d'enseignement, celui-ci doit être facile à mettre en œuvre. En outre, ils doivent être persuadés que le temps investi en vaut la peine, que les élèves apprennent quelque chose d'utile ou qu'un autre but pertinent (p. ex., la mesure d'une compétence) est atteint. Voilà les enjeux que nous espérions mesurer dans notre questionnaire destiné aux membres du corps professoral.

La réaction des élèves face au processus constitue également un élément pertinent de l'extensibilité. Les pédagogues adorent mettre en relief les mérites de l'apprentissage actif [Bean (2011)]; à ce compte-là, pourquoi l'apprentissage actif ne fait-il pas déjà partie intégrante de chaque classe? Pourquoi tant de classes demeurent caractérisées par un manuel de cours, des cours magistraux, des examens à choix multiples et rien d'autre? Cette situation s'explique en partie par le fait que, du point de vue des élèves, l'apprentissage actif exige temps et efforts, souvent dans un contexte relativement nouveau. Il s'agit d'un processus moins connu qui exige une gestion du temps et pour lequel les élèves doivent mener à bien quelque chose d'une façon qui diffère de l'étude faite en vue d'un examen final. Si les élèves n'aiment pas le processus, leurs évaluations des membres du corps professoral risquent alors d'en souffrir, de sorte que les membres du corps professoral seront moins enclins à adopter les démarches s'y rapportant.

Tout cela pour dire que la réaction des élèves à l'expérience d'apprentissage est importante. Si les élèves aiment l'apprentissage sous une certaine forme ou qu'ils en perçoivent l'utilité, ils peuvent alors compter parmi les rouages qui permettent de propager cette forme d'apprentissage. Par contre, si l'apprentissage sous une certaine forme leur semble fastidieux, difficile, désagréable ou, pour quelque raison que ce soit, indésirable, ils risquent alors de compter parmi les obstacles à la propagation d'un tel type d'apprentissage. Par conséquent, le fait d'en apprendre davantage sur la réaction des élèves au processus qui consiste à leur demander de mettre en application des rubriques en lien avec les compétences transférables aux travaux de leurs pairs est pertinent quant à la question de l'extensibilité possible de la démarche.

À la Phase 1, nous avons cherché à obtenir certaines données initiales du point de vue des élèves, et ces données étaient essentiellement positives, ce qui portait à croire que les élèves aimaient le processus et en

percevaient l'utilité dans leur éducation. De telles constatations sont encourageantes, mais peut-être est-il important de souligner notre investissement auprès de ces élèves particuliers et la possibilité que ces derniers aient donné des réponses qui, croyaient-ils, auraient l'heur de plaire à leur professeur. L'obtention de données du même type mais en provenance d'élèves moins directement liés à un défenseur de la démarche pourrait nous révéler la mesure dans laquelle les élèves jugent véritablement utile le processus.

Dans tout ce contexte, les pages suivantes du présent document retraceront la démarche progressive pour en arriver ultérieurement à concrétiser les buts qui viennent d'être décrits. Pour la première fois, les membres de l'équipe ont tenté de mener de la recherche avec des membres du corps professoral de plusieurs établissements d'enseignement, dont une grande majorité qu'ils ne connaissaient pas intimement. Procéder ainsi peut se révéler très éprouvant; une partie du présent document mettra en relief certaines des difficultés en la matière qui ont fait partie d'un parcours à l'issue duquel ont été obtenus différents ensembles de données pertinents au processus à l'étude.

Aperçu du processus de recrutement

Peut-être que nous sommes irrésistiblement portés à croire que les autres, qui jouent un rôle semblable au nôtre, nourrissent la même passion que nous à propos de certaines choses. C'est dans cet état d'esprit que nous avons établi le contact avec d'autres membres du corps professoral au moyen d'affichage dans les groupes LinkedIn, d'annonces dans *Academica* (un bulletin électronique destiné quotidiennement aux universitaires canadiens), de groupes dans Listserv comme celui qui sert la SAPES (Société pour l'avancement de la pédagogie dans l'enseignement supérieur), du rayonnement dans les médias sociaux, du bouche-à-oreille et de pratiquement tout autre moyen possible à partir de l'été 2018. Les participants intéressés devaient remplir un formulaire dans Google pour manifester leur intérêt et répondre à plusieurs questions.

L'intérêt a semblé élevé chez les 53 différents membres du corps professoral qui ont pris le temps de remplir le premier formulaire dans lequel ils devaient, entre autres, fournir leur adresse courriel. Tous ceux qui ont rempli le formulaire ont reçu un protocole d'expérimentation dans lequel étaient expliquées les étapes nécessaires à la participation, et chacun a été informé de prendre les contacts nécessaires pour favoriser la prestation du logiciel nécessaire, de même qu'obtenir l'aide requise en ce qui touche les formalités exigées par leur établissement d'enseignement, notamment le formulaire d'éthique.

Une fois bien amorcée la session d'automne 2018, la situation est demeurée tranquille... même un peu trop! Lorsque nous nous sommes mis à les joindre individuellement, la plupart n'ont tout simplement pas répondu. Chez ceux qui l'ont fait, certains ont affirmé ne pas avoir complètement compris les exigences ou,

pour diverses autres raisons, ne plus être disposés à participer ou en mesure de le faire. Les quelques-uns qui demeuraient intéressés ne semblaient pas trop savoir quoi faire, malgré tous les renseignements qu'ils avaient reçus. À cette étape-là, il était évident que la plupart des participants n'allaient pas concrètement mettre en œuvre le processus ou même en faire l'essai, de sorte qu'une nouvelle démarche était nécessaire. Cela dit, les données recueillies au départ auprès des 53 personnes demeuraient recueillies et portaient le descriptif suivant : « Les 53 premiers ».

Nous étions découragés, mais nous avons révisé notre démarche et nous nous sommes mis une fois de plus en mode recrutement; cette fois-ci, nous avons d'emblée énoncé clairement ce que nous attendions, puis nous avons consacré davantage de temps dès le début du processus à établir des liens avec les participants potentiels et à éclaircir les attentes liées à la participation à tous les niveaux. Nous avons également intensifié les communications avec les collègues de notre propre établissement d'enseignement, dans l'espoir qu'ils seraient davantage enclins à mener à terme leur engagement, une fois celui-ci pris. Ces participants, 14 professeurs en tout, constituent le principal groupe d'employés pour évaluer l'extensibilité de la démarche. Ils ont également rempli un formulaire Google qui permettait de fournir certains renseignements, mais ils l'ont fait à l'issue de leur participation. Leurs données porteront la désignation suivante : « Les 14 participants », étant donné qu'ils ont concrètement pris part à l'étude comme prévu.

Méthodes : Les 53 premiers

Participants

Parmi « les 53 premiers », 45 provenaient du Canada, 3 de l'Australie, 2 du Royaume-Uni, 2 de l'Afrique, et 1 de l'Inde. Ces participants ont donné des cours à l'université, tous niveaux confondus (y compris au premier cycle) et dans presque toutes les disciplines. Outre la prestation de caractéristiques démographiques de base, ils ont répondu à deux questions et leurs réponses ont semblé suffisamment intéressantes pour faire partie du présent document.

Documents

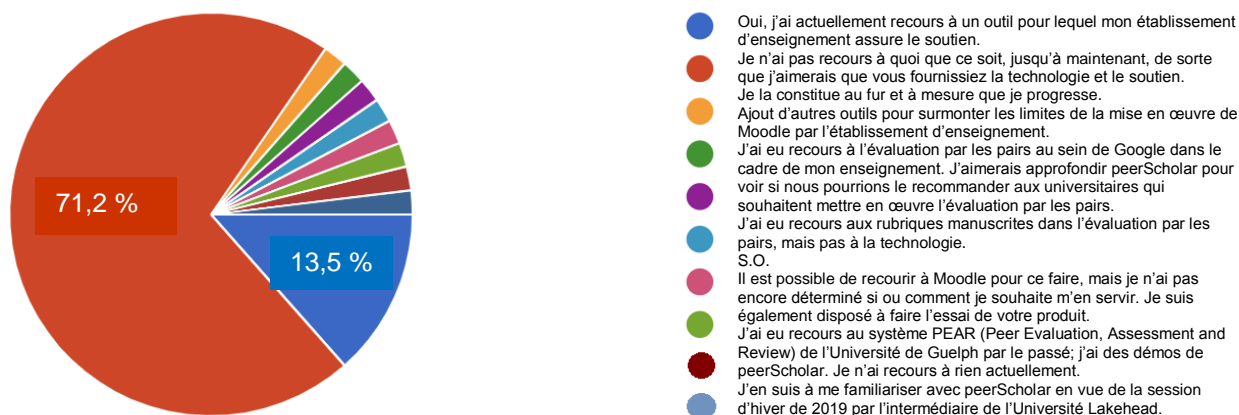
Tous les répondants ont présenté leurs données au moyen du formulaire Google. Il s'agissait d'un formulaire simple, essentiellement conçu pour aider les enquêteurs à établir un lien avec les participants. Toutefois, il s'y trouvait un formulaire de consentement à utiliser les données et, au-delà des caractéristiques démographiques de base, les deux questions suivantes s'y trouvaient :

1. Avez-vous actuellement recours à une évaluation par les pairs par voie technologique sous une forme ou une autre, ou aimeriez-vous que nous fournissions la technologie et le soutien?
2. Laquelle des compétences transférables suivantes est tout particulièrement importante selon vous?
 - la pensée critique;
 - la pensée créative;
 - la sensibilité métacognitive;
 - la communication écrite;
 - la communication réceptive;
 - la communication orale;
 - la littératie mathématique;
 - la littératie scientifique;
 - autre.

Résultats et analyse

De ce que nous savions, aucun n'avait publié de données en lien avec les deux questions non démographiques posées, de sorte que nous avons décidé d'intégrer le tout aux constatations du présent document. En ce qui touche la première question, les résultats se trouvent dans le graphique 2. Environ 13,5% des membres du corps professoral ayant répondu à notre invitation à participer avaient actuellement recours à l'évaluation par les pairs par voie technologique sous une forme ou une autre, au sein de leur établissement d'enseignement. Ce graphique est pertinent par rapport à l'enjeu de l'extensibilité du processus à l'étude, car il est nécessaire de recourir à un système d'évaluation par les pairs par voie technologique sous une forme ou une autre, afin de gérer le processus.

Graphique 2 : Réponses à la question : Avez-vous actuellement recours à l'évaluation par les pairs par voie technologique sous une forme ou une autre, ou aimeriez-vous que nous fournissions la technologie et le soutien? (n=53)

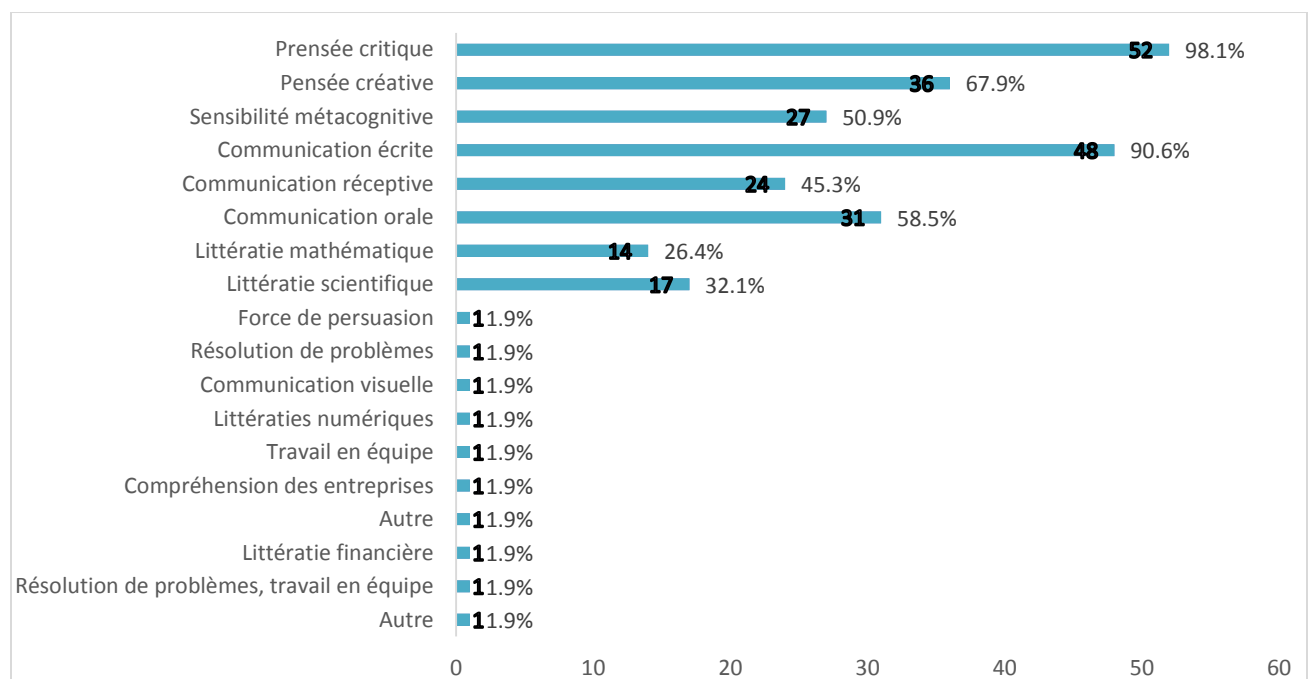


Dans la deuxième question, nous demandions aux membres du corps professoral en question laquelle des compétences transférables était la plus pertinente à leurs yeux. Nous avons autorisé les membres du corps professoral à cocher autant de réponses qu'ils le souhaitaient à partir d'une liste de huit compétences fournies, et nous leur avons également permis de rédiger leurs propres compétences si celles-ci n'étaient pas répertoriées. Les résultats se trouvent dans le graphique 3. De toute évidence, la pensée critique et la communication écrite étaient en tête de liste: plus de 90 % des répondants ont considéré chacune de ces compétences comme pertinentes. Elles étaient suivies de la pensée créative, de la communication orale, de la sensibilité métacognitive et de la communication réceptive. Dans la liste, les deux autres compétences, à savoir la littératie scientifique et la littératie mathématique, comptaient les taux de « pertinence » les plus faibles, mais cela résulte vraisemblablement du fait que seuls les cours dans certains domaines – tout particulièrement les sciences – prêtent attention à ces compétences. Tous les éléments ayant une fréquence de 1 étaient inclus par les répondants.

Une fois de plus, l'intention était que tous ces participants feraient l'essai du processus proposé : demander à leurs élèves de mettre en application une rubrique pertinente à la compétence transférable au travail de leurs pairs dans le contexte de l'évaluation par les pairs, mais la plupart du temps, cette étape n'a pas été suivie. En réalité, certains des membres du corps professoral font partie des 14 participants dont nous

ferons l'analyse, mais tel n'est pas le cas pour la plupart. Néanmoins, leurs réponses aux deux questions posées mettent en relief deux points. S'il faut que ce processus fasse l'objet d'une utilisation à grande échelle, il sera crucial entre autres d'obtenir les technologies à l'appui de cette forme d'évaluation par les pairs et d'y recourir de façon plus fréquente qu'à l'heure actuelle. De plus, s'il faut prêter attention à la mise en œuvre de cette démarche quant à seulement deux ou trois compétences transférables au départ, la pensée critique et la communication écrite sont les deux qui semblent les plus pertinentes d'après ceux qui s'intéressent potentiellement à la démarche.

Graphique 3 : Réponses à la question : Laquelle des compétences transférables suivantes est tout particulièrement importante selon vous? (n=53)



Méthodes : Les 14 participants

Participants

Les 14 professeurs qui ont mis en œuvre le processus décrit dans leur classe l'ont tous fait à la session d'hiver 2019. Ce sont huit professeurs de l'Université de Toronto (dont quatre issus de l'Institut d'études pédagogiques de l'Ontario), deux de l'Université Lakehead, un du Collège Humber, un du Collège communautaire Assiniboine, un de l'Université Urbana (États-Unis), de même qu'un de l'Université Utrecht (Pays-Bas).

De ces 14 enseignants, 9 s'en sont servi dans une classe, 4 l'ont utilisé dans deux classes et 1 y a eu recours dans trois classes, de sorte que le processus a servi en tout dans 20 classes. Ces classes allaient de la première année jusqu'au niveau de la maîtrise dans des matières telles que la psychologie, la neuroscience, l'audit, l'audit de niveau avancé, la gestion des ressources humaines, le droit, la criminalistique, le français oral, l'éducation, la recherche et la conception architecturale.

Documents

Un protocole de recherche a été créé pour orienter la participation des enseignants tout au long de la mise en place de l'activité. Ce protocole décrivait la nécessité d'intégrer trois éléments à toute activité : a) une « rubrique validée » concordante; b) un message-guide d'adhésion ou d'abstention quant au consentement; c) les six questions posées aux élèves (annexe A). Le message-guide de consentement et les questions aux élèves figuraient dans le protocole.

Bien entendu, les professeurs étaient libres de définir à leur guise la composition que leurs élèves devaient rédiger, mais ils étaient fortement invités à choisir d'abord la rubrique à laquelle ils allaient recourir, pour ensuite faire en sorte que le devoir à faire par les élèves concordait avec cette rubrique. À titre d'exemple, s'ils souhaitaient se servir de la rubrique de l'AAC&U de la pensée critique, ils devaient ensuite voir à ce que la tâche confiée aux élèves exige que ces derniers fassent montre de pensée critique. Le message-guide de consentement et le questionnaire feront l'objet d'une analyse en détail à la section « Procédure » qui suit.

Enfin, les professeurs ont reçu un fichier en format Zip comportant un ensemble des 16 rubriques validées de l'AAC&U. Nous avons favorisé le recours à l'une de ces rubriques, mais quelques membres du corps professoral étaient réticents. Par exemple, une membre du corps professoral a demandé aux élèves de rédiger une réflexion à propos de l'apprentissage, et elle disposait d'une rubrique validée sur la réflexion dont elle souhaitait faire usage. Au bout du compte, nous avons estimé que le point de mire de l'étude était

le processus, plutôt qu'une rubrique en particulier, de sorte que nous avons autorisé ceux qui le souhaitent à recourir à la rubrique validée de leur choix.

Procédure

C'est Steve Joordens qui a choisi 14 participants, en raison partiellement du fait qu'ils lui ont répondu et ont établi un rapport avec lui. Chaque participant a reçu un protocole de recherche décrivant les étapes nécessaires afin que leur activité peerScholar soit conforme aux besoins en recherche. Au besoin, les participants recevaient de l'aide pour remplir les formulaires nécessaires (p. ex., le formulaire d'éthique) et en ce qui touche l'accès au logiciel nécessaire. Nous avons également proposé aux participants une aide quant à la mise en place de l'activité, mais tous ont préféré le faire eux-mêmes.

La technologie employée pour gérer leurs activités, peerScholar, comporte un certain nombre de paramètres qui permettent à l'enseignant de mettre en place son activité à sa guise. Globalement, nous avons autorisé les enseignants à faire comme ils l'entendaient, mais nous avons insisté afin qu'ils intègrent obligatoirement les éléments suivants :

- À la phase de création, la rubrique pertinente devait être jointe aux directives, de façon à ce que les élèves y aient accès en tout temps.
- À la phase d'évaluation, un message-texte de consentement a été ajouté. De façon précise, après avoir effectué l'évaluation de leurs pairs et avant de pouvoir soumettre le tout, une fenêtre contextuelle apparaissait et on y lisait ce qui suit : « La présente activité fait partie d'un projet de recherche entre plusieurs établissements d'enseignement, et nous aimerions intégrer vos données à notre recherche. Soyez assurés que vos données demeureront confidentielles : elles seront particulièrement dépersonnalisées et sécurisées. Afin d'utiliser vos données, il nous faut votre consentement [...] votre professeur ne sera pas informé de votre consentement ou non, et votre décision n'influera pas sur vos travaux de cours, de quelque façon que ce soit. Compte tenu de tout cela [...] » et sous ce message-texte, les élèves devaient choisir l'une des deux options à partir d'un menu déroulant : ou bien consentir à ce que leurs données soient utilisées, ou bien refuser.
- Enfin, à la phase de réflexion, après que les élèves ont vu et évalué la rétroaction reçue, mais avant qu'ils puissent soumettre leur réflexion, une autre fenêtre contextuelle est apparue : elle comportait les six questions qui figurent à l'annexe A. Les réponses à ces questions forment l'un des principaux ensembles de données d'intérêt dans la présente étude.

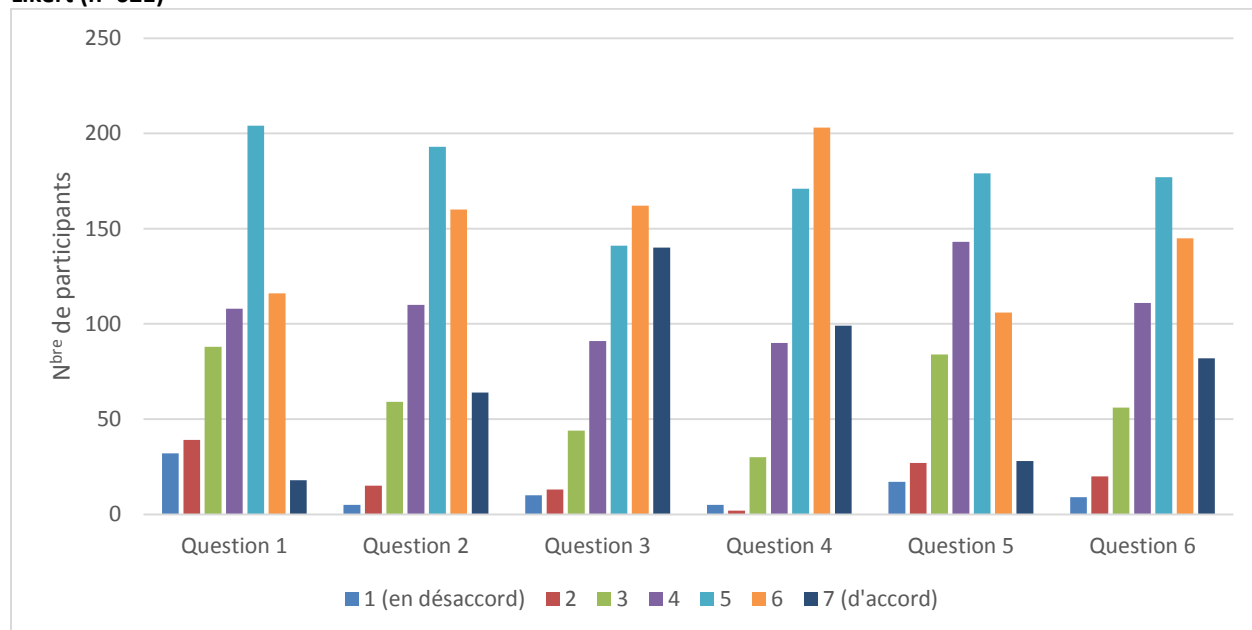
Une fois de plus, la plupart des enseignants ont suivi toute la procédure ci-dessus sans l'aide de l'équipe de recherche, sinon peu, mais nous avons dû dans certains cas examiner le tout et faire en sorte que tout soit

installé correctement. À cette étape-là, chacun a reçu l'autorisation d'aller de l'avant jusqu'à la fin de la session. Une fois arrivés à la fin de la session, nous avons de nouveau communiqué avec les enseignants participants pour leur demander de remplir un formulaire Google qui englobait le questionnaire à l'intention du corps professoral (voir l'annexe D). Les réponses à ces questions constituent le deuxième ensemble de données d'intérêt dans la présente étude.

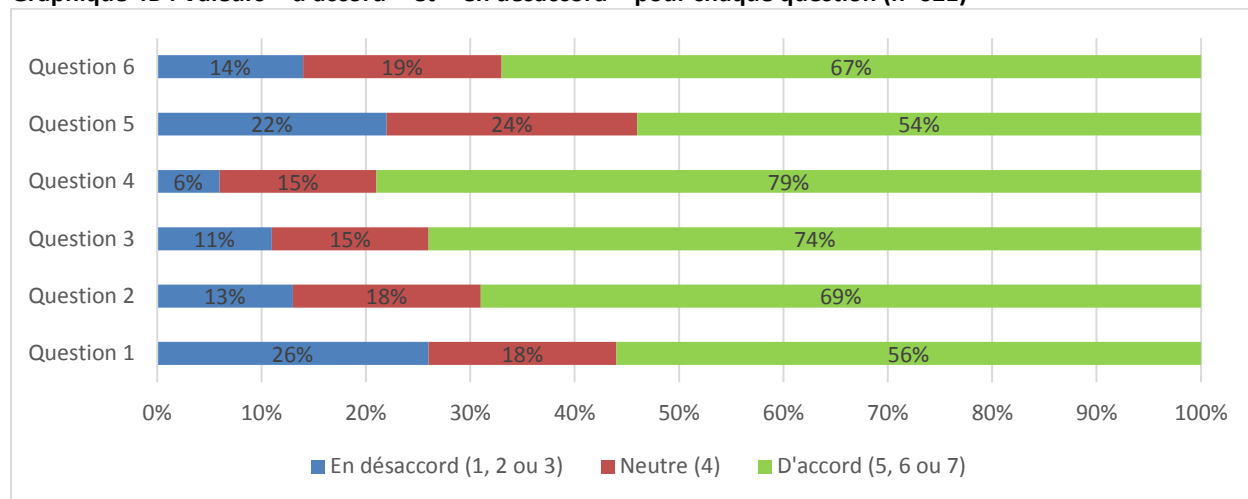
Résultats

Des 1 907 élèves qui ont rempli le sondage, seuls 32,6 % ont consenti à ce que leurs données servent à des fins de recherche. Cette valeur est inférieure à ce qui était attendu ou espéré, mais elle reflète peut-être la « distance » qui sépare chercheurs et participants lorsqu'une telle recherche se déroule dans plusieurs établissements d'enseignement. Les résultats du sondage en lien avec les élèves qui ont donné leur consentement se trouvent dans les graphiques 4A et 4B.

Graphique 4A : Fréquence des réponses de tous les participants dans chaque valeur de l'échelle à sept points de Likert (n=621)



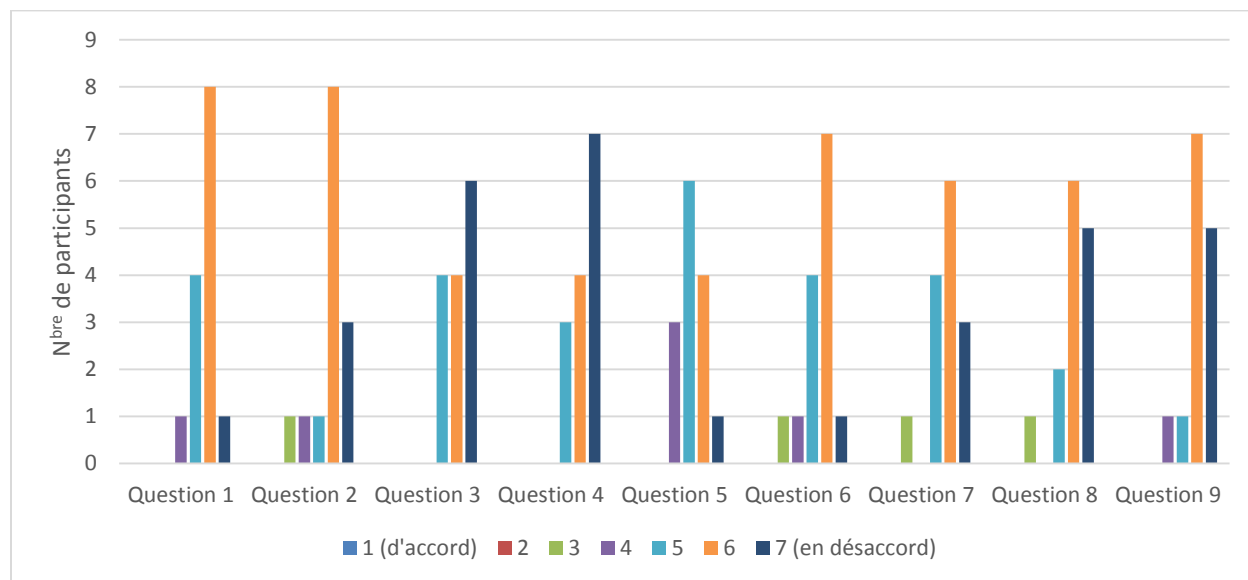
Graphique 4B : Valeurs « d'accord » et « en désaccord » pour chaque question (n=621)



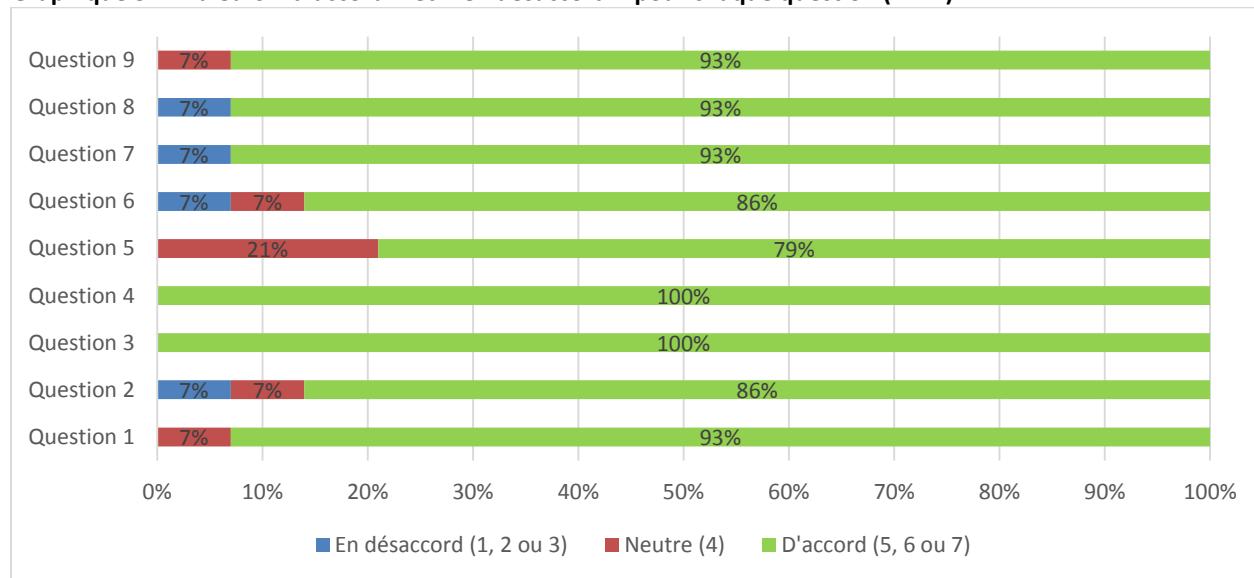
Malgré le faible taux de consentement, ces données reproduisent tous les phénomènes perçus à la Phase 1, lesquels semblent indiquer que les élèves qui passent par le processus perçoivent son utilité et souscrivent à son utilisation élargie. Dans la mesure où il existe des différences, celles-ci se manifestent essentiellement à l'échelle des taux de concordance légèrement inférieurs aux deux dernières questions. Autrement dit, les élèves étaient moins convaincus de l'importance de la mesure des compétences, bien qu'ils souscrivent encore au processus dans l'ensemble, et ils étaient moins convaincus que le processus par lequel ils étaient passés constituait un bon moyen de procéder, encore une fois malgré le fait qu'ils souscrivaient à cette prémisse en règle générale. Avec un peu de recul, le point crucial est le suivant : il n'y a pas de révolte et, de ce fait, les élèves souscrivent à cette démarche d'acquisition des compétences de même qu'à son utilisation élargie.

Si nous examinons maintenant le sondage auprès des membres du corps professoral, les 14 membres du corps professoral ont consenti à ce que leurs données soient utilisées et, par conséquent, les résultats de l'échantillon complet se trouvent en résumé aux graphiques 5A et 5B. Pour chaque question, nous avons également autorisé les commentaires d'ordre qualitatif et, pour des raisons de pertinence, chaque question fera l'objet d'une analyse détaillée.

Graphique 5A : Fréquence des réponses de tous dans chaque valeur de l'échelle à sept points de Likert (n=14)



Graphique 5B : Valeurs « d'accord » et « en désaccord » pour chaque question (n=14)



Les questions 1 et 2 se rapportaient essentiellement à la facilité d'utilisation de la solution technologique. L'activité était-elle facile à apprendre (Q1) et étaient-ils d'avis que les activités subséquentes seront rapides et faciles à mettre en place? Une note de 6 sur 7 correspondait à la réponse modale dans les deux cas, ce qui porte à croire que la technologie était, globalement, bien accueillie et qu'elle n'était pas perçue comme un obstacle d'envergure. Les commentaires d'ordre qualitatif s'inscrivaient essentiellement dans cette interprétation, quoique certains aient signalé des domaines où il y aurait encore matière à éclaircissement.

La question 3 avait trait aux perceptions de l'apprentissage : on demandait aux membres du corps professoral s'ils étaient d'avis que les élèves avaient beaucoup appris, plus qu'ils ne l'auraient fait sans l'intervention de l'évaluation par les pairs. Les membres du corps professoral étaient d'accord avec cet énoncé, la valeur modale correspondant à 7 sur 7. De toute évidence, les membres du corps professoral perçoivent l'utilité pédagogique du processus. L'un d'eux a affirmé ce qui suit : « Certains élèves ont affirmé que c'était la chose la plus importante qu'ils avaient apprise. »

La question 4 portait plus directement sur la démarche à l'étude : on a demandé aux membres du corps professoral si, à leur sens, le fait d'allier l'évaluation par les pairs à la mise en application de rubriques valides semblait un bon moyen d'enseigner les compétences transférables. Ici encore, tous étaient d'accord avec cet énoncé et, une fois de plus, la réponse modale s'est établie à 7 sur 7.

À la question 5, on a demandé aux membres du corps professoral s'ils croyaient que les élèves avaient aimé l'expérience. Rappelons que l'apprentissage actif exige des élèves qu'ils consacrent du temps à leur apprentissage et qu'ils y investissent des efforts considérables. Par conséquent, peut-être que les élèves allaient reculer face à l'apprentissage actif, puis faire part de leur mécontentement aux membres du corps professoral. Heureusement, nous n'avons perçu aucune donnée probante d'un tel recul dont les membres du corps professoral auraient pris connaissance. En réalité, 79 % des membres du corps professoral estimaient que leurs élèves avaient aimé l'activité, pendant que les autres sentaient que les élèves étaient neutres face à celle-ci.

La question 6 allie deux enjeux fondamentaux : Les membres du corps professoral étaient-ils d'avis que ce processus constituait une expérience d'apprentissage efficace, tout en nécessitant peu de temps et d'effort de leur part? Douze membres du corps professoral sur 14 (86 %) étaient d'accord avec cette question, la réponse modale se fixant à 6 sur 7. Certes, le processus exige temps et efforts des membres du corps professoral afin de mettre en place ce processus et de le gérer. Le taux d'adhésion élevé chez les membres du corps professoral s'est révélé encourageant à cet égard.

La question 7 portait sur l'extensibilité de cette démarche à « n'importe quel contexte pédagogique et n'importe quel niveau d'études ». Malgré l'aspect étendu de cette question, tous les membres du corps professoral, sauf un, étaient d'accord, la valeur modale s'établissant à 6 sur 7. Le membre du corps professoral qui n'était pas d'accord a ajouté un commentaire révélant qu'il soupçonnait que cette démarche risque de ne pas fonctionner au niveau d'éducation primaire.

La question 8 prêtait davantage attention à l'utilisation continue : les membres du corps professoral étaient-ils suffisamment impressionnés à un point tel qu'ils continueraient de recourir à ce processus? Cet aspect est parfois qualifié d'« adhérence », les démarches adhérentes ayant un plus fort potentiel d'échelonnement et de croissance. Une fois de plus, tous les membres du corps professoral, sauf un, étaient d'accord avec l'idée qu'ils allaient continuer de recourir à ce processus, la valeur modale se fixant à 6 sur 7 dans l'échelle d'accord.

Enfin, à la question 9, on leur demandait s'ils aimeraient voir l'adoption de cette démarche dans un plus grand nombre de cours, de façon à donner aux élèves davantage de pratique entre les contextes variés. Tel serait évidemment l'objectif si le recours à ce processus était échelonné au sein de l'établissement d'enseignement. Une fois de plus, tous les membres du corps professoral, sauf un, étaient d'accord avec l'idée de voir cette démarche dans davantage de cours, ce qui montre encore une fois leur soutien général à la démarche que leurs élèves et eux ont vécu directement.

Une fois combinées, les données des élèves et des membres du corps professoral se révèlent très positives en ce qui touche le potentiel de cette démarche nouvellement proposée en vue de l'acquisition et de la mesure en bonne et due forme des compétences transférables. Ni le corps professoral, ni les élèves ne montrent des signes de réticence. Ces derniers semblent plutôt souscrire à la démarche et à son utilité, et ils sont disposés à voir celle-ci utilisée plus vastement. Fait crucial, l'acquiescement provient des enseignants et des élèves ayant vécu le processus dans une vaste gamme de contextes pédagogiques.

Analyse générale

L'auteur Joordens (2018) fait valoir que la démarche actuelle dans l'enseignement supérieur de n'enseigner qu'officieusement les compétences transférables de base, comme la pensée critique et créative, de même que la communication claire et efficace, doit être officialisé en raison de l'importance que prennent ces compétences au sein d'un milieu de travail dynamique. Il propose ce qui peut être la première démarche officielle quant à l'acquisition de la mesure des compétences qui peuvent être saisies par une rubrique. La démarche, qui allie l'évaluation par les pairs par voie technologique à la mise en application raisonnée de rubriques validées, pourrait en théorie s'appliquer à n'importe quelle discipline et à n'importe quel niveau.

Le potentiel de cette démarche d'influer sur la réussite des élèves à l'avenir semble très élevé, mais seulement si, bien sûr, le processus décrit est aussi extensible que proposé.

La recherche décrite dans le présent document a pour objet de mettre à l'essai plusieurs aspects de la démarche qui sont pertinents à son extensibilité. D'abord, la démarche part du principe qu'il est possible de se fier aux moyennes tirées de la mise en application des rubriques fondées sur les pairs à titre de base valide de mesures de la compétence en question. Ensuite, pour qu'une technologie éducative soit extensible, elle doit être facile à utiliser, son utilité par rapport à l'apprentissage doit être admise, et elle ne doit susciter d'indignation ni chez les élèves, ni chez les membres du corps professoral qui s'en servent. Si ceux qui aiment le processus perçoivent également son utilité et souhaiteraient voir l'élargissement de son utilisation, c'est encore mieux. Chacun de ces enjeux sera traité progressivement dans les sections suivantes.

Validité des notes par rubrique attribuées par les pairs

Il convient de souligner que l'auteur Joordens (2018) formule deux affirmations à propos du processus qu'il met en relief. D'abord, il affirme qu'il s'agit d'une démarche efficace en vue de l'acquisition de compétences transférables par l'ajout d'une sorte de « formation à l'oreille » [Woody (2012)] à l'expérience vécue par les élèves pendant qu'ils apprennent une nouvelle compétence. Par exemple, la proposition selon laquelle si les élèves apprennent, disons, la pensée critique, au sein du processus décrit, tout l'apprentissage qui se déroulerait par le processus habituel de peerScholar [Joordens et al. (2014)] s'en trouverait alors bonifié, alors que les élèves recherchent et notent l'ampleur de la pensée critique présente dans les travaux de leurs pairs. Cette amélioration dans l'acquisition des compétences n'a pas fait l'objet d'un examen direct dans le présent document, mais elle semble censée, du moins à première vue.

En deuxième lieu, l'auteur Joordens (2018) avance que la démarche pourrait également servir de mesure de n'importe quelle compétence donnée dans le contexte de l'évaluation. De façon précise, il affirme que si le travail d'un élève donné est évalué en fonction, disons, de la pensée critique par cinq pairs ou plus de cet élève, les cinq notes ou plus pourraient être mises en moyenne, et cette moyenne peut procurer des mesures valides de la compétence, tout comme les moyennes par les pairs se sont préalablement révélées des mesures valides de la qualité globale des travaux [Cho et al. (2006); Paré et Joordens (2008)].

La capacité de mesure joue un rôle crucial dans l'adoption et l'extensibilité d'une démarche officielle en vue de l'acquisition des compétences au sein des établissements d'enseignement supérieur [Liu, Bridgeman et Adler (2012)]. Ces établissements se servent de mesures pour donner forme à l'apprentissage et lui attribuer des crédits, ce qui constitue, dans une certaine mesure, un rôle capital que les établissements d'enseignement supérieur jouent au sein de la société. Par conséquent, l'affirmation de l'auteur Joordens

(2018) comme quoi la démarche proposée peut fournir facilement de telles mesures dans le contexte des travaux ordinaires constitue une composante fondamentale qui donne à la démarche de telles retombées potentielles élevées.

Les notes par rubrique fondées sur les moyennes par les pairs procurent-elles des mesures valides des compétences? Cette affirmation était le point de mire de la Phase 1 et s'est avérée en quelque sorte, mais pas complètement, du moins en fonction de sa mise en œuvre ici. De façon précise, la capacité des moyennes attribuées par les pairs à prédire de façon relative et absolue des notes données par les assistants à l'enseignement et les experts a fait l'objet d'une évaluation. Les notes attribuées par les pairs ont permis de bien prédire les notes relatives, ce qui signifie que les compositions jugées relativement fortes (ou faibles) par les pairs étaient également celles que les AE et les experts estimaient relativement fortes (ou faibles). Toutefois, lorsque venait le temps de faire des prédictions absolues, à savoir déterminer la note réelle qu'un adjoint à l'enseignement ou à l'expert décernerait à une composition donnée, les moyennes décernées par les pairs n'ont pas été concluantes. Cet échec dans l'exactitude des notes sur le plan absolu est attribuable en grande partie au fait que les pairs n'ont pas manqué de donner des notes supérieures à celles des évaluateurs davantage experts.

Avant de nous attarder sur l'échec quant à la formulation de prédictions exactes sur le plan absolu, il peut être sage de contextualiser dans un premier temps les données quant aux mérites de la démarche proposée. Dans certaines démarches de mesure des compétences transférables, comme le test de la pensée critique Cornell, les élèves doivent ouvrir une session dans un site Web puis consacrer passablement de temps à répondre à une vaste gamme de questions [Ennis (1993)]. Si un établissement d'enseignement souhaite mesurer les retombées de l'apprentissage sur la pensée critique au moyen de cette démarche, il doit alors mettre en place des séances de mise à l'essai, disons au début de la première année et à la fin de la quatrième année, puis inciter un échantillon d'élèves à procéder à l'évaluation. Qui plus est, cette méthode ne permettrait qu'une mesure de la pensée critique. Des tests comme celui-ci n'existent pas pour chaque compétence transférable et, si tel était le cas, les exigences logistiques de gestion de cette démarche seraient alors nécessaires pour chaque compétence à laquelle l'établissement d'enseignement attache de l'importance. Une démarche nettement supérieure consisterait à trouver un moyen de mesurer les compétences relatives aux travaux que les élèves exécutent habituellement dans le cadre de l'éducation reçue [Simper, Frank, Scott et J. Kaupp (2018)], ce que permet le processus recommandé par l'auteur Joordens (2018). En outre, parce que la démarche permet de mesurer n'importe quelle compétence qui correspond à la rubrique mise en œuvre, elle procure un moyen de mesurer n'importe quelle compétence dans n'importe quel contexte à n'importe quel stade dans le cheminement universitaire des élèves.

Il importe également de souligner que la capacité de faire avec exactitude des prédictions sur le plan relatif a quelque chose de réjouissant. Les pairs réussissent à discerner la présence (ou l'absence) d'une compétence, puis ils font part de leurs réflexions aux élèves quant à la vigueur de leurs compétences par rapport à leur cohorte. Lorsque cet élément est pris en considération à la lumière des avantages logistiques décrits ci-dessus, il est clair que des progrès dans l'acquisition et la mesure de telles compétences sont réalisés.

Mais au bout du compte, le processus comporterait nettement plus de retombées si les moyennes par les pairs permettaient de mesurer, avec exactitude et sur le plan absolu, les compétences en question. De ce point de vue, les données actuelles procurent un point de départ en vue de travaux supplémentaires. Nous mettons en relief deux démarches évidentes afin de rehausser ces notes de mesure.

La première consiste à prêter attention à la formation des élèves avant de leur demander d'évaluer leurs pairs. L'organisme AAC&U a créé un programme de formation et il en est actuellement à se pencher sur des moyens par lesquels offrir cette formation en ligne [Terrel Rhodes, communication personnelle]. Peut-être que les élèves devraient être tenus de suivre cette formation dans le cadre d'un cours. Voilà une idée qui est loin d'être absurde, car les élèves qui suivent une formation axée, disons, sur la nature de la pensée critique de même que l'aspect qu'elle prend lorsqu'elle est présente, vivraient une expérience scolaire fantastique. Nous pourrions, à tout le moins, imaginer un cours de courte durée qui pourrait consister en un module en ligne dans lequel on pourrait jeter les bases du processus de mise en application des rubriques de façon plus intensive et peut-être plus pertinente que la courte vidéo de formation que nous avons utilisée ici. De plus amples recherches pourraient permettre d'évaluer différentes démarches quant à la formation ainsi que leurs effets sur la validité en vue de déterminer une marche à suivre optimale.

Quoi qu'il en soit, même au moyen d'une excellente formation d'entrée de jeu, il demeure probable que l'expérience de mise en application des rubriques débouche sur des notes inférieures au fur et à mesure que les évaluateurs se mettent à percevoir un peaufinement progressif du discernement dans le travail [MacLean et al. (2010)]. Il s'agit peut-être tout simplement d'une étape du processus d'apprentissage qui ne peut jamais être éliminée. Si tel est le cas, une autre bonne démarche pourrait alors consister à évaluer la mesure dans laquelle le biais est présent, pour ensuite le neutraliser sur le plan statistique. À titre d'exemple, un enseignant pourrait disposer d'un ensemble de compositions ayant trait à certaines activités notées par des experts. Il pourrait ensuite intégrer ces activités aux compositions réelles des élèves, puis demander aux élèves de noter également ces compositions. La différence entre les notes moyennes attribuées par les élèves et celles attribuées par les experts pourrait servir à évaluer l'ampleur du biais présent dans les notes attribuées par les élèves, et ce biais pourrait être soustrait des notes réelles. Bien qu'elle puisse exiger du temps et des efforts supplémentaires, cette démarche permettrait non seulement

de neutraliser les notes gonflées, mais de connaître la mesure dans laquelle le biais est présent, et la façon dont ce biais évolue au fil du temps.

En conclusion, notre évaluation de l'affirmation de l'auteur Joordens (2018) comme quoi la démarche proposée peut fournir des mesures des compétences transférables de base dans le contexte des travaux habituels en classe est globalement positive. Conformément à l'évaluation, les moyennes par les pairs donnent de fortes notes relatives des compétences en question et, moyennant un peu de suivi, nous avons toutes les raisons de croire que de fortes notes sur le plan absolu peuvent également être obtenues, peut-être au moyen d'une formation davantage approfondie combinée au recours à des compositions préalablement notées afin d'évaluer le biais et de le neutraliser.

S'en serviront-ils?

La réussite d'une technologie éducative, quelle qu'elle soit, dépend évidemment de son efficacité et de sa pertinence relativement au but recherché par l'établissement d'enseignement, mais il dépend également de facteurs beaucoup plus liés à la dimension humaine. Les membres du personnel de chaque centre de formation et d'apprentissage le savent bien : lorsque vient le temps d'inciter les membres du corps professoral à envisager une nouvelle technologie ou démarche, la facilité d'utilisation constitue systématiquement un souci fondamental à leurs yeux [Bates (2000)]. En effet, les membres du corps professoral ne souhaitent tout simplement pas chambarder de fond en comble ce qu'ils font déjà, tout particulièrement de façon constante. Cependant, si une nouvelle technologie ou démarche est censée rehausser la pédagogie, elle peut être mise en œuvre sans grande perturbation, et son utilisation sera alors beaucoup plus susceptible de connaître une extension au sein de l'établissement d'enseignement.

Ce qui est également pertinent, c'est la réaction des élèves face à la nouvelle démarche. La plupart des nouvelles démarches qui sont défendues font intervenir l'apprentissage actif sous une certaine forme et, au cœur de celui-ci, les élèves sont tenus d'exécuter ce qu'un grand nombre d'entre eux considèrent comme du travail [Michael (2006)]. Le simple fait que quelque chose soit différent risque déjà de susciter un malaise [Reis, Maniaci, Caprariello, Eastwick et Finkel (2011)]. La nécessité supplémentaire de faire quelque chose suscite également des problèmes de gestion du temps et exige un effort. Par conséquent, il n'y a rien de surprenant à voir des élèves résister parfois face à de nouvelles possibilités d'apprentissage actif, même parmi ceux qui peuvent améliorer leur réussite à l'avenir. Bien entendu, cette résistance des élèves peut se traduire par des évaluations de cours à la baisse, et étant donné que les évaluations des cours contribuent en forte partie aux décisions relatives à l'avancement et aux augmentations salariales, l'extensibilité d'une technologie est également sensible à la réaction des élèves [Simpson et Siguaw (2000)].

Compte tenu de cela, l'adhésion des membres du corps professoral et des élèves détermine la mesure dans laquelle les nouvelles innovations pédagogiques seront extensibles à l'échelle des établissements d'enseignement. Lorsque l'innovation nécessite le recours à une nouvelle technologie, la facilité d'utilisation de cette technologie est alors cruciale, tout comme les perceptions de l'apprentissage et les perceptions de l'acceptation chez les élèves. Si une innovation peut permettre d'apaiser toutes ces craintes, les possibilités que son usage se répande sont alors élevées.

L'auteur Joordens a pris fait et cause pour un processus dans ses travaux en 2018 : il a avancé que la démarche peut fonctionner dans n'importe quelle discipline à n'importe quel niveau, et, selon la rubrique employée, elle peut servir dans l'acquisition et la mesure de n'importe quelle compétence transférable. Ce sont là de fortes affirmations qui, si elles sont avérées, portent à croire qu'il ne s'agit pas d'un processus ciblé, mais plutôt d'une démarche qui peut comporter de vastes retombées sur l'évaluation des compétences transférables au sein de divers établissements d'enseignement.

La Phase 2 correspond à une tentative d'évaluation directe de telles affirmations d'extensibilité, essentiellement en demandant aux membres du corps professoral d'en faire l'essai dans leur classe, quel que soit celle-ci et à quelque niveau d'enseignement que ce soit. Après en avoir fait l'essai, les membres du corps professoral et les élèves ont rempli des questionnaires afin d'évaluer les enjeux mis en relief précédemment, y compris l'utilité perçue du processus, la facilité de mise en œuvre, le caractère convivial ainsi que l'adhésion à l'idée selon laquelle le processus est effectivement aussi extensible que proposé.

La démarche a été mise à l'essai dans plusieurs contextes variés, allant des cours de français oral en première année à un cours en éducation au niveau de la maîtrise, en passant par des cours de droit en troisième année. En règle générale, les membres du corps professoral ont fait une évaluation très positive du potentiel de la démarche. Ce qui est peut-être le plus pertinent, presque tous ont convenu qu'ils allaient continuer de recourir à la démarche et que celle-ci devrait être employée à une plus vaste échelle. Fait crucial, les élèves ont également réagi de façon positive dans l'ensemble à l'expérience, malgré le fait que la démarche nécessitait du temps et des efforts considérables de leur part. Ils ont également souscrit à l'affirmation selon laquelle la démarche devrait être employée dans un plus grand nombre de leurs cours et, ce qui a été constaté à répétition, ils attachent de l'importance à la possibilité d'apprendre de leurs pairs et en leur compagnie [p. ex., Topping (1998)].

Bien sûr, la difficulté que présente toute nouvelle démarche ou technologie est d'être utilisée en premier lieu, mais d'après les données actuelles, il semble que la démarche décrite par l'auteur Joordens (2018), une fois mise à l'essai, suscite l'« adhérence ». Sachant que ceux qui en ont fait l'essai y adhèrent peut constituer en soi un facteur important pour amener les autres à en faire l'essai, et les données relatées ici dans

l'ensemble sont très encourageantes en ce qui concerne la réaction des membres du corps professoral et des élèves face à la démarche proposée.

Conclusions globales

Par le passé, les établissements d'enseignement supérieur ont porté attention à l'enseignement officiel de l'information, tandis que l'enseignement des compétences transférables de base se révèle officieux. À titre d'exemple, on demande aux élèves de prendre position dans une dissertation, ou de poser des questions pendant un cours magistral. Le mouvement en faveur d'une démarche plus officielle dans l'acquisition des compétences nécessite une démarche qui peut s'insérer dans nos contextes universitaires caractérisés par des compressions, et dans laquelle les compétences peuvent être mesurées avec exactitude, de façon à permettre la reconnaissance de l'apprentissage. Nous avons fait allusion à ces difficultés précédemment en ce qui touche le « Comment? » (pouvons-nous acquérir les compétences efficacement) ainsi que le « Comment savons-nous? » (ce que nous faisons comporte-t-il des retombées).

L'auteur Joordens (2018) a proposé un processus qui, avance-t-il, apporte une réponse concrète et fondamentalement extensible aux deux questions. Dans le présent document, nous avons évalué plusieurs facteurs en lien avec cette affirmation. Notre conclusion fondée sur nos constatations est la suivante : la démarche proposée comporte un potentiel des plus élevés. En ce qui touche le volet mesure des éléments, il reste encore du travail à exécuter afin de fournir des mesures qui se situent au niveau approprié de la validité (c.-à-d. exemptes de tout biais), mais la voie à suivre pour ce faire est claire, et nous avons toutes les raisons de croire que la démarche nous mènera là où nous devons aller. En ce qui touche le volet acquisition des éléments, l'essentiel du processus proposé a déjà révélé des retombées sur les compétences transférables de base [Joordens et al. (2014)]. Les bonifications à cette base devraient rehausser encore l'apprentissage, pendant que les réactions des membres du corps professoral et des élèves portent à croire que la démarche peut servir judicieusement dans une vaste gamme de contextes pédagogiques.

Bibliographie

- American Association of Colleges & Universities (AAC&U) (s. d.). Current VALUE Research.
<https://www.aacu.org/current-value-research>
- Bates, A. W. (2000). *Managing Technological Change: Strategies for College and University Leaders*. The Jossey-Bass Higher and Adult Education Series. San Francisco, Jossey-Bass.
- Bean, J. C. (2011). *Engaging Ideas: The Professor's Guide to Integrating Writing, Critical Thinking, and Active Learning in the Classroom*, John Wiley & Sons.
- Bennett, R. E. (2011). « Formative Assessment: A Critical Review », dans *Assessment in Education: Principles, Policy & Practice*, vol. 18 n° 1, p. 5-25.
- Boud, D. (2013). *Enhancing Learning through Self-assessment*, Routledge.
- Cho, K., Schunn, C. D. et R.W. Wilson (2006). « Validity and Reliability of Scaffolded Peer Assessment of Writing from Instructor and Student Perspectives », dans *Journal of Educational Psychology*, vol. 98 n° 4, p. 891–901.
- Collimore, L. M., Paré, D. E. et S. Joordens (2015). « SWDYT: So What Do You Think? Canadian Students' Attitudes about peerScholar, an Online Peer-assessment Tool », dans *Learning Environments Research*, vol. 18 n° 1, p. 33-45.
- Dochy, F. J. R. C., Segers, M. et D. Sluijsmans (1999). « The Use of Self-, Peer and Co-assessment in Higher Education: A review », dans *Studies in Higher education*, vol. 24 n° 3, p. 331-350.
- Ennis, R. H. (1993). « Critical Thinking Assessment », dans *Theory into practice*, vol. 32 n° 3, p. 179-186.
- Finley, A. P. (2011). « How Reliable are the VALUE Rubrics? », dans *Peer Review*, vol. 13 n° 4/1, p. 31.
- Hanrahan, S. J. et G. Isaacs (2001). « Assessing Self- and Peer-Assessment: The Students' Views », dans *Higher Education Research & Development*, vol. 20 n° 1, p. 53-70, DOI: 10.1080/07294360123776
- Joordens, S. (2018). « Developing and Measuring Learning Outcomes at Scale: The Promise of Peer Assessment », dans F. Deller, J. Pichette et E. K. Watkins (éd.), *Driving Academic Quality: Lessons from Ontario's Skills Assessment Projects* (p. 13-27), Toronto, Conseil ontarien de la qualité de l'enseignement supérieur.
- Joordens, S., Paré, D. et L. M. Collimore (2014). *Programme d'entraînement pour les résultats d'apprentissage : Méthode d'établissement et d'évaluation des résultats d'apprentissage fondé sur les travaux scolaires*, Toronto, Conseil ontarien de la qualité de l'enseignement supérieur.

- Liu, O. L., Bridgeman, B. et R. M. Adler (2012). « Measuring Learning Outcomes in Higher Education: Motivation Matters », dans *Educational Researcher*, vol. 41 n° 9, p. 352-362.
- Liu, N. F. et D. Carless (2006). « Peer Feedback: the Learning Element of Peer Assessment », dans *Teaching in Higher Education*, vol. 11 n° 3, p. 279-290.
- MacLean, K. A., Ferrer, E., Aichele, S. R., Bridwell, D. A., Zanesco, A. P., Jacobs, T. L. [...] et B.A. Wallace (2010). « Intensive Meditation Training Improves Perceptual Discrimination and Sustained Attention », dans *Psychological science*, vol. 21 n° 6, p. 829-839.
- Michael, J. (2006). « Where's the Evidence that Active Learning Works? », dans *Advances in Physiology Education*, vol. 30 n° 4, p. 159-167.
- Nicol, D. J. et D. Macfarlane-Dick (2006). « Formative Assessment and Self-regulated Learning: A Model and Seven Principles of Good Feedback Practice », dans *Studies in Higher Education*, vol. 31 n° 2, p. 199-218.
- Paré, D. E. et S. Joordens (2008). « Peering into Large Lectures: Examining Peer and Expert Mark Agreement Using peerScholar, an Online Peer Assessment Tool », dans *Journal of Computer Assisted Learning*, vol. 24 n° 6, p. 526-540.
- Reis, H. T., Maniaci, M. R., Caprariello, P. A., Eastwick, P. W. et E. J. Finkel (2011). « Familiarity does Indeed Promote Attraction in Live Interaction », dans *Journal of Personality and Social Psychology*, vol. 101 n° 3, p. 557.
- Rhodes, T. L. (2016). « The VALUE of Assessment: Transforming the Culture of Learning », dans *Change: The Magazine of Higher Learning*, vol. 48 n° 5, p. 36-43.
- Simper, N., Frank, B., Scott, J. et J. Kaupp (2018). *Learning Outcomes Assessment and Program Improvement at Queen's University*, Toronto, Conseil ontarien de la qualité de l'enseignement supérieur.
- Simpson, P. M. et J. A. Siguaw (2000). « Student Evaluations of Teaching: An Exploratory Study of the Faculty Response », dans *Journal of Marketing Education*, vol. 22 n° 3, p. 199-213.
- Topping, K. (1998). « Peer Assessment between Students in Colleges and Universities », dans *Review of Educational Research*, vol. 68 n° 3, p. 249-276.
- Tseng, S. C. et C. C. Tsai (2007). « On-line Peer Assessment and the Role of the Peer Feedback: A Study of High School Computer Course », dans *Computers & Education*, vol. 49 n° 4, p. 1161-1174.
- Woody, R. H. (2012). « Playing by Ear: Foundation or Frill? », dans *Music Educators Journal*, vol. 99 n° 2, p. 82-88.



Conseil ontarien
de la qualité de
l'enseignement
supérieur

Un organisme du gouvernement de l'Ontario