

MAXIME GIROUARD

Critères d'évaluation des travaux d'étudiants en formation à l'enseignement

Thèse présentée à la
Faculté des études supérieures et postdoctorales
de l'Université d'Ottawa en vue de l'obtention
du grade de Maîtrise ès arts en éducation - M.A. (Éd.)

Faculté d'éducation,
Université d'Ottawa

© Maxime Girouard, Ottawa, Canada, 2011

Remerciements

L'exécution d'une étude ainsi que l'écriture d'une thèse ne s'accomplissent pas seules. Dans un premier temps, je voudrais remercier les participants de mon étude; malgré ma méthodologie peu commune et mon manque d'expérience avec la formation à l'enseignement, ils m'ont ouvert leurs portes et partagé leurs habitudes. Je vous remercie d'avoir permis à un étudiant de découvrir l'envers de la médaille; d'adopter le point de vue de l'évaluateur.

J'aimerais aussi remercier tous les membres de ma famille, qui m'ont toujours encouragé. Merci de m'avoir offert vos opinions, qui étaient beaucoup plus appréciées que vous ne le pensez. Un merci spécial à Jacinthe, ma femme, qui m'a soutenu de mille façons tout au long de cette aventure. Sans toi, ce projet n'aurait jamais pu voir sa fin.

Merci également aux membres de mon comité, Dany Laveault et Christine Suurtamm, vos commentaires m'ont permis de voir mon étude sous un autre œil et d'élargir ma compréhension de plusieurs concepts. Finalement, un immense merci à Marielle Simon, ma directrice de thèse. Merci de m'avoir toujours laissé la liberté que je désirais dans l'exécution de mon étude. Tu as su me conseiller et me diriger, peu importe les idées que je voulais discuter. Aussi, je serai toujours reconnaissant des multiples opportunités que tu m'as offertes. Merci mille fois.

Résumé

L'évaluation critériée a pris beaucoup d'expansion au cours des dernières années. Toutefois, les critères d'évaluation sont parfois employés à l'encontre des recommandations des experts. Cette étude a pour but d'explorer l'utilisation peu documentée de critères d'évaluation par les professeurs universitaires à la formation à l'enseignement. Une analyse qualitative basée sur des entrevues, des protocoles oraux et des analyses de documents a permis d'identifier les critères présents dans le processus d'évaluation de quatre professeurs. D'une part, des critères d'évaluation propres à chaque participant et d'autre part 7 critères présents chez 3 participants ou plus ont été identifiés ainsi qu'un seul critère implicite. Cependant, les critères d'évaluation identifiés ne se limitent pas tous à l'évaluation du rendement scolaire. La présence de documents facultaires a eu un impact sur les résultats. Cette étude a aussi décelé de nouvelles problématiques à explorer comme les influences dans l'utilisation des critères et l'appréciation des travaux eux-mêmes.

« Analyser qualitativement un matériau de recherche, c'est observer, percevoir, ressentir, comparer, nommer, juger, étiqueter, contraster, relier, ordonner, intégrer, vérifier; c'est tout à la fois découvrir et montrer que ceci est avant/après/avec cela, que ceci est plus important/évident/marqué que cela, que ceci est le contexte/l'explication/la conséquence de cela; c'est replacer un détail dans un ensemble, lier un sentiment à un objet, rapporter un événement à un contexte; c'est rassembler et articuler les éléments d'un portrait éclairant, juger une situation, dégager une interprétation, révéler une structure, construire ou valider une théorisation. » (Paillé et Mucchielli, 2008, p.48-49).

Table des matières

Remerciements.....	ii
Résumé.....	iii
Table des matières.....	v
Liste des tableaux.....	vii
Chapitre 1 – Introduction.....	1
Chapitre 2 – Recension des écrits.....	5
Planification de l'évaluation.....	5
Exécution du travail.....	12
Appréciation du travail.....	15
Utilisation des résultats.....	18
États des lieux.....	20
Chapitre 3 – Univers interprétatif.....	23
Posture épistémologique et perspective théorique.....	24
Concept de critère.....	26
Mise en relation des concepts.....	27
Questions de recherche.....	29
Chapitre 4 – Méthodologie.....	31
Choix théoriques.....	31
Contexte ontarien.....	32
Participants.....	34
Instrumentation.....	37
Procédure.....	40
Traitement et analyse des données.....	42
Traitement des données.....	43
Examen phénoménologique des données.....	44
Analyse par questionnement analytique.....	45
Rigueur méthodologique.....	52
Chapitre 5 – Résultats.....	57
Description des cas.....	57
Antoine.....	60
Béatrice.....	61
Charles.....	63
Denise.....	63
Présentation des résultats selon les trois questions.....	64
Chapitre 6 – Discussion.....	79
Retour sur les questions de recherche.....	79
Implications pratiques et théoriques.....	87
Limites.....	96

Chapitre 7 – Conclusion.....	99
Références.....	109
Annexe 1 - Texte de recrutement.....	110
Annexe 2 - Formulaire de consentement	111
Annexe 3 - Guide d’entrevue.....	113
Annexe 4 - Instructions du protocole oral.....	114
Annexe 5 - Exemples de protocoles oraux	115
Annexe 6 - Grille d’identification et d’analyse des critères	117

Liste des tableaux

Tableau 1 - Critères présents lors du processus d'évaluation d'Antoine	58
Tableau 2 - Critères présents lors du processus d'évaluation d'Antoine	66
Tableau 3 - Critères présents lors du processus d'évaluation de Béatrice	68
Tableau 4 - Critères présents lors du processus d'évaluation de Charles	70
Tableau 5 - Critères présents lors du processus d'évaluation de Denise	71
Tableau 6 - Critères présents chez trois participants ou plus.....	72
Tableau 7 - Critères externes au rendement scolaire	76

Chapitre 1 – Introduction

L'évaluation est une action très présente dans la vie quotidienne, qu'elle soit diagnostique, formative ou certificative. Elle joue aussi plusieurs rôles en éducation. Par exemple, Laurier, Tousignant et Morissette (2005) en nomment sept, dont les évaluations bilans ou de certification, qui sont utilisées fréquemment dans les salles de classe universitaires. En effet, les étudiants universitaires sont constamment évalués, souvent à plusieurs reprises à l'intérieur de chacun de leurs cours et rapportent d'importantes variations dans les évaluations des professeurs universitaires. En se référant autant au point de vue des étudiants universitaires qu'à celui des professeurs, Speck (1998) parle d'un mystère entourant les pratiques d'évaluation afin d'inciter au développement d'une compréhension accrue de ces dernières. Étant donné l'importance des résultats scolaires dans notre société, il est pertinent de chercher à mieux comprendre comment les professeurs universitaires procèdent lors de leurs évaluations des apprentissages des étudiants.

Plusieurs débats existent dans le domaine de l'évaluation et concernent entre autres la validité et la fidélité des interprétations à la suite d'une évaluation, les méthodes actuelles de collecte d'information en salle de classe ou la place de l'évaluation au sein de l'apprentissage. Cependant, ces thèmes demeurent souvent inaccessibles pour un des groupes concernés : les étudiants. Cette étude cherche à examiner les évaluations en salle de classe universitaire de façon à améliorer la compréhension du processus d'évaluation des apprentissages des étudiants et à contribuer aux connaissances générales sur l'évaluation universitaire.

Une piste potentielle pour une telle étude est l'évaluation critériée, si l'on considère que celle-ci est de plus en plus utilisée, autant en Ontario et en Amérique, (Aviles, 2001; Tillema et Smith, 2007) qu'en Europe (Rees et Sheard, 2004; Woolf, 2004; Webster, Pepper et Jenkins,

2000) et ailleurs dans le monde (Neil, Wadley et Phinn, 1999). Comme le décrit Aviles (2001), l'évaluation critériée permet de comparer les étudiants à un niveau de compétence désiré plutôt que de les comparer entre eux. Lorsque Wyatt-Smith et Castleton (2005) ont tenté de démystifier le rôle d'évaluateur, elles ont identifié six considérations ayant un impact sur les évaluations des travaux de leurs participants. Parmi ces considérations, qui varient de l'expérience des enseignants, à leurs connaissances pédagogiques et à l'observation des élèves, ce sont les critères d'évaluation et les standards qui représentent la facette du jugement la plus accessible. Le caractère accessible mentionné ici réfère au fait que ces informations devraient être mises à la disposition des étudiants (Case, 2007; McMillan, 2004; Ministère de l'Éducation de l'Ontario, 2000; Rust, Price, et O'Donovan, 2003; Sheaford et Sheaford, 2001; Tillema et Smith, 2007) et, par le fait même, être plus facilement accessibles aux chercheurs.

Les quelques études qui ont tenté d'identifier les critères utilisés par les professeurs universitaires discutent principalement de critères institutionnels ou départementaux mis à leur disposition (Webster et al., 2000; Woolf, 2004). Lorsque la question des critères d'évaluation développés par les professeurs universitaires est abordée, c'est souvent hors du contexte et hors du cadre d'un processus d'évaluation. De plus, les résultats de ces études témoignent d'une très grande variation dans les critères utilisés (Baume, Yorke et Coffey, 2004; Hand et Clewes, 2000; Tillema et Smith, 2007).

D'autre part, des études qui ont analysé les critères d'évaluation utilisés par des professeurs ont permis d'identifier des divergences entre les recommandations d'experts et les pratiques de certains enseignants et professeurs. Par exemple, certains professeurs ne se limitent pas aux critères d'évaluation qu'ils ont établis au préalable (Webster et al., 2000). Aussi, l'appréciation globale de la tâche par le professeur et sa connaissance antérieure des capacités de

l'étudiant influencent parfois son application de critères spécifiques (Baume et al., 2004), ayant alors un effet modérateur sur les jugements émis. Par ailleurs, l'utilisation de critères implicites, d'une pondération variable ou d'interprétations différentes des critères explicites engendre parfois une faiblesse au niveau de la fidélité des évaluations (Rust, O'Donovan et Price, 2005). De plus, divers partis concernés peuvent avoir différentes définitions de ce qui est inclus dans une note (Pilcher, 1994). En plus des critères d'évaluation du rendement scolaire, des professeurs se réfèrent parfois à des critères supplémentaires, tels que la participation et l'attitude envers le cours (Cross, Frary et Weber, 1993).

Cette étude porte sur les critères utilisés par les professeurs universitaires qui enseignent à la formation à l'enseignement, principalement en raison de leur influence sur les pratiques d'évaluation de futurs enseignants. Puisque peu de recherches étudient ce corpus spécifique, la recension comprend des études des pratiques de professeurs universitaires de même que d'enseignants du secondaire et primaire. Cette recherche vise d'abord à connaître quels sont les critères d'évaluation des travaux d'étudiants en contexte universitaire. Deux questions supplémentaires précisent davantage cette question et sont expliquées dans l'univers interprétatif, au chapitre trois. La nature des informations recherchées et la pénurie de connaissances actuelles sur le sujet justifient une étude exploratoire (De Ketele et Roegiers, 1993; Poupart et al., 1997). Un éventail de méthodes de collecte de données et d'analyse qualitative est employé afin de tirer un maximum d'information.

Cette étude comprend un chapitre sur la recension des écrits qui détaille les connaissances relatives à la sélection, à la présentation et à l'utilisation des critères selon les étapes du processus d'évaluation, décrites au début de cette même section. Ce chapitre présente les connaissances telles qu'acquises lors de la formulation de l'univers interprétatif de façon à

exposer adéquatement l'évolution de la conception de cette étude. Le chapitre de l'univers interprétatif énonce la posture épistémologique, la définition adoptée du concept de critère d'évaluation, une mise en relation des concepts ainsi que la présentation de la question de recherche précisée. Le chapitre de la méthodologie, basé sur les deux chapitres précédents, décrit le contexte, les participants, l'instrumentation, la procédure, ainsi que le traitement et l'analyse des données. Le chapitre des résultats est, quant à lui, structuré selon les questions de recherche traitées. Le chapitre discussion reprend les résultats et les interprète à la lumière des informations contenues dans la recension des écrits et l'univers interprétatif tout en les reliant à des études plus récentes.

Chapitre 2 – Recension des écrits

Afin de mieux comprendre la sélection, la communication et l'application des critères effectuées par les professeurs universitaires lors du processus d'évaluation, il importe de considérer chacune des étapes séparément. Il existe de multiples conceptions du processus d'évaluation dans la littérature (Laurier et al., 2005; McMillan, 2004; Moskal, 2000; Popham, 2000) et aucun consensus concernant une définition optimale. La définition du processus d'évaluation adoptée ici comporte quatre étapes généralement acceptées : la planification de l'évaluation, l'exécution du travail, l'appréciation du travail et l'utilisation des résultats. Cette division est utilisée à plusieurs reprises lors de cette étude, entre autres afin de structurer la recension des écrits ainsi que guider la collecte et l'analyse des données. La planification est vue comme l'étape où le professeur clarifie ses choix liés à l'évaluation tels que le sujet à évaluer, la fonction, le but, la méthode d'évaluation et les critères utilisés. L'exécution du travail comprend la communication des attentes et des exigences par le professeur, ainsi que l'achèvement par l'étudiant du travail exigé. L'appréciation est l'étape où le professeur compare le travail de l'étudiant aux critères d'évaluation et porte par la suite un jugement sur la valeur de celui-ci. Enfin, l'utilisation des résultats peut prendre plusieurs formes telles que la rétroaction qualitative ou la consignation d'une cote alpha numérique ou quantitative, ou servir à plusieurs fins comme la régulation ou la certification.

Planification de l'évaluation

C'est lors de la planification, première étape du processus d'évaluation, que les professeurs déterminent la fonction, la méthode et les détails de l'administration de l'évaluation. De plus, c'est à cette étape que les professeurs sélectionnent ou élaborent leurs critères d'évaluation. Cette section présente d'abord diverses définitions du concept de critère

d'évaluation. Par la suite, elle présente les études qui se sont penchées sur les critères d'évaluation institutionnels et départementaux ou ceux élaborés par les professeurs universitaires. D'autres études, présentées subséquemment, ont utilisé des méthodologies variées afin d'observer si les objets ciblés par les critères d'évaluation sont conformes aux recommandations des experts dans le domaine.

La conception d'un critère d'évaluation et de ses composantes varie selon les multiples définitions présentées dans les écrits. Dans un manuel sur les évaluations authentiques, Wiggins (1998) offre la définition suivante d'un critère « ... a way of defining success at meeting a target achievement or educational outcome » (p.128) et fournit une liste pertinente au choix des critères d'évaluation. Cette liste se divise en cinq catégories : (a) l'obtention de l'impact désiré (b) la qualité globale du travail (c) l'utilisation de méthodes appropriées (d) la validité du contenu et (e) le degré d'expertise. Bien que ces critères s'appliquent à divers contextes, la définition proposée par Wiggins semble limiter son concept de critère aux évaluations authentiques qui sont plus ou moins appliquées en contexte universitaire. Ces dernières réfèrent à l'évaluation de tâches qui s'apparentent à des problèmes de la vie courante, où les étudiants doivent utiliser leur jugement et puiser dans leurs habiletés et compétences afin de les résoudre (Scallon, 2000; Wiggins, 1998). Forgette-Giroux et Simon (2004) proposent une autre définition dans un article traitant des aspects problématiques de la notation qualitative. Selon elles, une distinction doit être faite entre les termes *critère* et *attributs*. Un critère est décrit comme étant une « ... caractéristique de la dimension d'une performance, d'un produit ou d'une réponse élaborée » (p.245) et les attributs comme étant « une qualité propre ou particulière d'un critère de performance, comme la pertinence (attribut) du choix de stratégie (critère) » (p.245). D'après cette définition, l'application d'un critère requiert l'identification d'un ou plusieurs attributs.

L'utilisation de critères et d'attributs précise et clarifie davantage le processus d'évaluation, mais les écrits s'y réfèrent rarement. Mertler (2001) et Tierney et Simon (2004) font référence à l'utilisation d'attributs dans leurs descriptions des composantes d'une échelle descriptive de travaux scolaires. Selon Mertler, l'attribut se trouve dans les niveaux de progression et précise les caractéristiques spécifiques et observables recherchées par rapport aux critères. Tierney et Simon suggèrent par ailleurs de définir les attributs parallèlement aux définitions des critères dans l'échelle descriptive, avant de les combiner dans les niveaux de progression. Selon elles, les attributs prennent la forme de mots de progression tels que *peu, quelque, la plupart et l'ensemble* en lien avec un critère spécifique comme *la présence d'éléments conventionnels dans une carte géographique* (Traduction libre, Tableau 1). D'autre part, dans un manuel sur l'évaluation en salle de classe, McMillan (2004) définit un critère comme étant « a clearly articulated and public descriptions of facets or dimensions of student performance that are used for judging the level of achievement » (p.32). Cette dernière définition se veut plus inclusive et s'applique à une plus grande étendue de situations d'évaluation, c'est-à-dire autant aux évaluations traditionnelles qu'aux évaluations authentiques.

Différentes études se sont penchées plus particulièrement sur les critères utilisés lors d'évaluations dans divers contextes universitaires. Une piste de recherche est l'étude de critères institutionnels et départementaux. Au Royaume-Uni, lors d'une recherche-action dans l'école des sciences sociales et l'école de droit à l'Oxford Brookes University, Webster et al. (2000) ont posé un regard sur les critères utilisés au cours de l'évaluation de dissertations d'étudiants du premier cycle. La dissertation sélectionnée est effectuée en fin de programme et doit représenter l'ensemble des études effectuées. Des sept départements étudiés, six avaient une liste de critères disponible. Le nombre de critères énoncés variait entre trois et 15 selon les départements. Neuf

critères étaient présents dans quatre départements ou plus, 13 critères étaient partagés par trois départements ou moins et huit critères étaient présents à une seule reprise. Les neuf critères les plus communs, rapportés par Webster et al. (2000) sont :

- « 1. Définit clairement le problème étudié et les questions à demander et à étudier.
2. Met en contexte en démontrant une compréhension de la perspective disciplinaire et la pertinence de l'étude pour la discipline.
3. Utilise et démontre une compréhension des méthodes de recherche appropriées.
4. Démontre la capacité d'analyser du matériel primaire et secondaire.
5. Fait preuve de pensée critique au sujet du problème, des idées préconçues et des opinions et valeurs rencontrées.
6. Tire des conclusions basées sur une compréhension des preuves et du matériel recueilli.
7. Discute et argumente clairement et avec cohérence (en choisissant une structure qui aide à faire ceci).
8. Écrit clairement et succinctement.
9. Cite correctement les sources et les met en référence [traduction libre] » (p.75).

Dans un contexte semblable, Woolf (2004) a mené une analyse de documents précisant les critères utilisés par sept départements d'histoire et quatre départements de gestion, lors de l'évaluation de dissertations de fin de premier cycle. L'auteur a observé une certaine conformité entre les établissements dans leurs choix de critères. Quatre critères de base se dégagent de sa recherche : « la capacité d'identifier et de rechercher un sujet, de développer une argumentation et de la soutenir avec les preuves pertinentes, de structurer ses idées et de les présenter de façon logique, lisible et précise [traduction libre] » (Woolf, 2004, p.485). Malgré un échantillon relativement petit, cette étude identifie plusieurs problèmes liés à l'utilisation de critères

d'évaluation dans ce contexte spécifique et propose des solutions. Par exemple, Woolf discute de la difficulté qu'ont les étudiants à s'approprier le vocabulaire lié aux évaluations et suggère de discuter les critères avec eux afin de clarifier le vocabulaire utilisé.

Dans une autre étude qualitative, Hand et Clewes (2000) proposent quant à eux d'identifier les critères d'évaluation utilisés par divers professeurs dans une école de commerce. À l'aide de discussions de groupe et d'un questionnaire, ces auteurs ont remarqué une grande diversité dans les critères d'évaluation utilisés, et ce, malgré une utilisation marquée de listes de critères déjà existantes par une majorité des professeurs questionnés. Aucune vérification n'est faite concernant la différence possible entre les données recueillies par cette approche méthodologique et les critères utilisés en contexte réel. Au cours de l'élaboration et de la mise en application d'une échelle descriptive d'appréciation des travaux universitaires, Simon et Forgette-Giroux (2001) réfèrent aux critères d'évaluation établis par diverses revues scientifiques lors de la soumission d'articles, tels la cohérence, la précision, l'envergure, la pertinence, la clarté et la force des arguments. La création de critères d'évaluation génériques, mais sujets à de multiples interprétations, s'est avérée une difficulté lors de la création de leur échelle.

Dans le but de calculer un accord interjuge sur la fiabilité d'évaluations de portfolios, Rees et Sheard (2004) ont identifié cinq critères utilisés dans une faculté de médecine du Royaume-Uni: (a) structure logique et cohérente (b) niveau de réflexion critique (c) niveau de développement, actuel ou futur, de compétences (d) utilisation de preuves écrites et (e) utilisation de sources pertinentes. Ceux-ci ont servi à évaluer 100 portfolios différents. Afin d'augmenter la fiabilité de l'évaluation, les auteurs proposent de définir quatre niveaux différents pour chacun des critères afin de former un continuum d'une performance exemplaire à

un échec. Cette façon de procéder permet aux auteurs de calculer leur accord interjuge et limite, par la même occasion, la variation possible par rapport à chaque critère. La méthodologie et le contexte limitent la pertinence de leurs résultats pour la présente étude, mais la précision accrue apportée aux critères mérite une attention particulière.

Située en contexte de formation à l'enseignement aux États-Unis, l'étude de Tillema et Smith (2007) cherche à identifier les critères utilisés lors de l'évaluation de portfolios par des professeurs de la didactique de l'anglais langue seconde. Puisqu'aucun critère n'était fourni au préalable aux professeurs participant, l'étude rapporte une très grande variation dans les critères utilisés et les commentaires donnés aux étudiants. Les auteurs ont toutefois identifié quatre critères semblant être présents lors de l'évaluation faite par leurs participants. En effet, les professeurs cherchaient à identifier si les étudiants faisaient preuve de : (a) progrès (b) compétence en enseignement (c) réflexions et connaissances du sujet étudié et (d) documentation de leur performance de façon claire et structurée.

Bien que les contextes spécifiques de ces études diffèrent beaucoup, certaines tendances ressortent. En fait, trois idées semblent revenir à plus d'une reprise : l'importance d'un travail clair et structuré, une argumentation soutenue par des preuves pertinentes, une réflexion critique au sujet du problème. Plusieurs études mentionnent également l'importance de démontrer un progrès. Un tel inventaire suffit si le seul intérêt est l'identification des critères d'évaluation. Cependant, une analyse plus approfondie des composantes des critères d'évaluation permet d'obtenir des informations additionnelles concernant la concordance des pratiques des professeurs aux recommandations des experts.

De multiples études se sont penchées sur les pratiques d'évaluation des enseignants et des professeurs en examinant les objets ciblés par les critères d'évaluation afin de mieux

comprendre les habiletés, les aptitudes et autres construits cernés par ces derniers. Dans cet ordre d'idées, plusieurs experts recommandent que le rendement scolaire soit le seul objet apprécié, mais remarquent que cela n'est pas nécessairement le cas (Allen, 2005; Brookhart, 1993; Cross, Frary et Weber, 1993; McMillan, 2001; Pilcher, 1994; Stiggins, Frisby, et Griswold, 1989). Dans une étude sur la composante d'une note, Pilcher (1994) définit le rendement scolaire comme étant « l'acquisition de connaissances et d'habiletés [traduction libre] » (p.70). Dans son dictionnaire de l'éducation, Legendre (1993) décrit pour sa part le rendement comme étant le « degré de réalisation des objectifs établis » (p.1109). Ces définitions suggèrent que des critères comme les traits personnels et les habitudes de travail se distinguent du rendement scolaire.

Lors de leur examen des pratiques d'évaluation de 15 enseignants du secondaire, Stiggins et al. (1989) ont été parmi les premiers à constater qu'en plus du rendement, les enseignants tiennent parfois compte de l'effort, des habiletés, de la motivation et des résultats aux évaluations formatives. Plus tard, grâce à un échantillon de près de 1 500 enseignants du secondaire, McMillan (2001) a confirmé les observations de Stiggins et al. et a ajouté à la liste la participation des étudiants, les habitudes de travail et le progrès démontré. Bien que les études mentionnées discutent des pratiques des enseignants du secondaire, ces observations s'appliquent également au milieu universitaire. Dans un article théorique discutant également des pratiques d'évaluation, Allen (2005) ajoute que l'utilisation de ces critères externes au rendement scolaire varie d'un professeur à l'autre. À la formation à l'enseignement, des professeurs ont mentionné, lors d'une autocritique de leurs évaluations, inclure la valeur apparente du portfolio (Tillema et Smith, 2007). Cross et al. (1993) ont constaté, grâce à un sondage mené auprès de 365 professeurs universitaires, que la moitié des participants

considérerait la participation et l'attitude envers le cours lors de l'appréciation de travaux comme des critères valables. L'incorporation de critères externes au rendement scolaire pose problème et complique l'interprétation de la note (Brookhart, 1993), puisque celle-ci n'indique pas si elle représente uniquement le rendement scolaire ou non.

Exécution du travail

À la deuxième étape du processus d'évaluation, l'étudiant doit utiliser les informations communiquées par le professeur afin d'accomplir les travaux ou les tâches d'évaluation qui s'imposent en tenant compte des critères qui leur sont spécifiques. La communication des critères peut parfois être une source d'incertitude dans l'exécution des travaux pour les étudiants, que ce soit en raison du vocabulaire choisi, du format de la présentation ou du poids accordé à chacun des critères.

Dans son article théorique, Speck (1998) explique que l'utilisation erronée des termes entourant le processus d'évaluation porte les étudiants à confusion. Il démontre clairement ce phénomène en exposant des situations où des termes distincts tels la notation, l'évaluation et la mesure sont utilisés par les professeurs comme s'ils étaient interchangeables. Selon lui, les variations dans le vocabulaire utilisé par les professeurs, qu'elles surviennent entre collègues ou chez un seul professeur, rendent les critères peu accessibles aux étudiants. Par ailleurs, Williams (2005) a remarqué dans son étude auprès d'étudiants de première année universitaire que la terminologie propre à un domaine d'études influence leur compréhension des critères et leur interprétation des résultats. Woolf (2004) met en garde contre une communication imprécise des critères en raison de la variation langagière d'un département à l'autre. Même à l'intérieur d'un département, les conceptions de l'évaluation des professeurs semblent varier selon leur domaine d'expertise spécifique (Higgins, 2000; Webster et al., 2000). Ces auteurs

s'entendent sur un fait : la simple présentation ne suffit pas à assurer la compréhension des critères par les étudiants.

Certaines études discutent de diverses façons de présenter les critères d'évaluation afin de les rendre plus accessibles aux étudiants. Selon Sadler (2002), un choix de critères clairement définis permet tout de même plusieurs interprétations de ce qui est attendu au niveau de la *qualité* du travail. Cette notion de qualité, difficile à cerner, complique la compréhension des attentes par les étudiants. À ce sujet, Sambell, Miller et Hodgson (2002) ont observé l'avantage de présenter et de discuter des attentes avec les étudiants plutôt que de simplement les mentionner. En accord avec cette constatation, O'Donovan, Price et Rust (2004) suggèrent que les critères eux-mêmes soient toujours soumis à une interprétation de la part des professeurs et étudiants. Afin de remédier à tout écart d'interprétation entre les deux partis, ils proposent que les professeurs discutent des critères d'évaluation au moment de la distribution des tâches d'évaluation, dans le but d'amener les étudiants à participer au processus d'élaboration des critères. Cette discussion permet aux étudiants de mieux comprendre l'appréciation de leurs travaux. Bloxham et West (2007) ont observé, lors d'une étude visant une meilleure compréhension des critères d'évaluation, que deux tiers des étudiants considèrent les explications verbales des critères comme une aide dans l'amélioration de leur performance. Cet exercice sert à clarifier ce qui est attendu plutôt qu'à tenter d'obtenir des informations additionnelles (Bloxham et West, 2007, p.84). Similairement, lors d'études de cas distinctes témoignant des problèmes encourus par les professeurs, Sadler (2002) et Sambell et al. (2002) ont constaté que l'utilisation de copies types permet aux étudiants de mieux saisir ce qui est attendu d'eux. Selon Simon et Forgette-Giroux (2001), l'utilisation de l'échelle descriptive d'appréciation lors d'évaluations formatives aide à éliminer les écarts entre l'interprétation des

étudiants et des professeurs. D'autre part, dans une étude visant à connaître la valeur psychométrique d'évaluations authentiques en sciences sociales et en sciences naturelles au secondaire, Stokking et al. (2004) ont observé plusieurs inconsistances au cœur des évaluations. Plus particulièrement, la description de la tâche, ses buts et les critères d'évaluation diffèrent sur le plan des attentes. Ils ajoutent que les enseignants formulent parfois plus de buts à l'évaluation qu'il n'est possible d'évaluer à l'aide de critères.

Nombre de chercheurs, praticiens et experts s'accordent sur l'importance de présenter clairement les critères d'évaluation (McMillan, 2004; Ministère de l'Éducation de l'Ontario, 2000; Sheaford et Sheaford, 2001; Stokking et al., 2004; Tillema et Smith, 2007; Woolf, 2004). En effet, une connaissance au préalable des critères d'évaluation a une incidence bénéfique sur la qualité des apprentissages effectués (Case, 2007; Rust et al. 2003). Toutefois, Webster et al. (2000) expliquent qu'à l'intérieur de facultés de sciences sociales et de droit en Europe où une liste de critères d'évaluation est fournie aux étudiants, les professeurs n'utilisent parfois qu'une partie de la liste de critères sans spécifier lesquels, ou encore ajoutent des critères implicites à ceux-ci sans les préciser. Sheaford et Sheaford (2001) mentionnent comme exemples de critères implicites, l'aptitude à écrire, la lisibilité du texte, la qualité de la grammaire et la syntaxe. Bien que ces critères puissent être pertinents lors d'un cours de rédaction, s'ils n'ont pas été clairement communiqués préalablement à l'exécution de la tâche, ils ne devraient pas servir de critères du rendement scolaire lors de l'appréciation. Baume et al. (2004) mettent en garde que l'utilisation de critères autres que ceux fournis explicitement peut faire varier grandement la note attribuée.

D'autre part, la présentation explicite de la pondération des critères est aussi importante que celle des critères. Hand et Clewes (2000) ont observé de grandes variations dans les

pondérations des critères utilisés par les professeurs d'une école de commerce au Royaume-Uni. Webster et al. (2000) ont, quant à eux, remarqué que certains professeurs mentionnent des critères d'évaluation sans préciser les poids relatifs de chacun et leur accordent pourtant une valeur différente. La question de la pondération des critères dépend en partie de l'approche utilisée par les enseignants et les professeurs. En effet, l'utilisation d'une approche analytique des critères, où la performance à chaque critère est jugée individuellement, peut donner lieu à des situations problématiques. Comme l'explique Dyrud (2003) dans une revue professionnelle, si un étudiant remet un texte respectant la majorité des critères, mais aucunement pertinent, l'étudiant obtiendrait tout de même un « A » si tous les critères ont une pondération égale. La solution adoptée par cette professeure est d'utiliser une approche holistique lors de l'évaluation. Bien que l'évaluation holistique soit aussi informée par des critères d'évaluation, l'apport relatif de chacun d'eux n'est pas clairement défini et l'appréciation porte sur le travail dans son ensemble (Mertler, 2001). Cette pratique n'est cependant pas la solution à tous les problèmes. Comme l'explique Mertler, « the advantage to the use of analytic rubrics is quite substantial. The degree of feedback offered to students-and to teachers-is significant. Students receive specific feedback on their performance with respect to each of the individual scoring criteria » (¶ 4). Baume et al. (2004) ont observé que des professeurs combinent parfois ces deux types d'évaluation en considérant particulièrement certains critères tout en se basant majoritairement sur une évaluation holistique. Une fois encore, cette pratique complique la compréhension et l'utilisation des critères d'évaluation de la part des étudiants.

Appréciation du travail

L'appréciation des travaux, troisième étape du processus d'évaluation, se fait en comparant les productions des étudiants aux critères d'évaluation retenus afin de porter un

jugement de valeur. Comme le dit Tremblay (1998), « l'acte d'évaluation a comme finalité de poser un jugement, de se prononcer sur la façon dont les attentes concernant cette réalité sont satisfaites ou non » (p.40). Le jugement professionnel constitue un champ d'études en soi, mais comme le mentionne Laveault (2005), il est difficilement défini. Dans un article théorique cherchant à définir le jugement professionnel des enseignants lors d'évaluations, Laveault explique la nécessité de considérer les valeurs et les croyances personnelles, les compétences, les stéréotypes et autres influences dans l'étude du jugement comme acte professionnel. Ce concept peut ainsi englober toutes les décisions que doivent prendre les professeurs durant le processus d'évaluation et plus encore. L'étude de la subjectivité inhérente aux jugements des enseignants n'est pas récente. Par exemple, Weiss discute dès 1992 de l'objectivité impossible en évaluation en raison du contexte social dans lequel cette activité s'effectue. Bressoux et Pansu (2003) présentent pour leur part l'évaluation comme un acte social, inévitablement influencé tant par le contexte culturel que par les caractéristiques individuelles des élèves. C'est en raison de la grande complexité de l'étude du jugement professionnel que cette étude adopte une définition plus restrictive. Le jugement, comme discuté ici, se limite à la prise de décisions quant à l'application des critères à un travail précis (Simon et Forgette-Giroux, 2001) et exclut les considérations sociales, culturelles et contextuelles qui informent parfois les jugements.

Dans leur étude concernant l'appréciation de portfolios, Baume et al. (2004) ont noté une différence entre les jugements émis lors des cours où les portfolios ont été recueillis et ceux émis après ces cours. Les participants de l'étude mentionnent avoir modifié leur façon de noter les portfolios, car leur jugement n'était pas influencé par la connaissance de l'étudiant en question. Ces auteurs expliquent que les professeurs ont tendance à porter un jugement global sur les travaux aux dépens des critères individuels. Par ailleurs, lors de l'appréciation, les professeurs

se voient obligés de mettre beaucoup d'effort à identifier les éléments recherchés lorsque ces derniers ne sont pas clairement définis au préalable (Baume et al, 2004). Green et Emerson (2007) ont identifié des *projections* de la part de professeurs, ce qui consiste à puiser dans leurs propres connaissances afin d'identifier des éléments dans les travaux des étudiants qui ne sont pas réellement présents. Ces pratiques ont pour effet d'augmenter la possibilité de se référer à des critères implicites ou externes au rendement scolaire lors de l'appréciation. Wyatt-Smith et Castle (2005) ont aussi observé que c'est lors de la formulation d'un jugement que des variables autres que le rendement scolaire peuvent être considérées, telles la motivation de l'étudiant, l'effort perçu et des traits de personnalité. Comme mentionnés auparavant, ces objets d'évaluation sont tous externes au rendement scolaire et ne devraient pas faire l'objet d'évaluations (Allen, 2005; Brookhart, 1993; Cross et al.,1993; McMillan, 2001; Pilcher, 1994; Stiggins et al., 1989). Par ailleurs, Wyatt-Smith et Castle (2005) expliquent que ces mêmes enseignantes ont davantage tendance à se limiter aux critères d'évaluation et à leurs expériences en enseignement lorsqu'elles ne connaissent pas les élèves évalués. Tillema et Smith (2007) ont demandé à des professeurs à la formation à l'enseignement d'évaluer des portfolios d'étudiants inconnus, sans avoir accès aux critères d'évaluation spécifiés lors de l'évaluation initiale. Ils ont observé que les participants portaient des jugements très diversifiés et n'offraient que peu d'explications de ces jugements ou des critères considérés. Ces participants ont aussi affirmé être influencés par l'apparence générale du portfolio.

Tout comme l'utilisation de copies types aide à la présentation des critères et attentes (Sadler, 2002; Sambell et al., 2002), certaines pratiques peuvent faciliter l'appréciation. Scallon (2004) explique dans son manuel sur l'évaluation des compétences que l'utilisation de divers instruments, tels que des échelles descriptives, peut faciliter l'appréciation. Il justifie cette

recommandation entre autres grâce à « un degré élevé de concordance des évaluations menées par différentes personnes » et à « la qualité du *feed-back* adressé aux élèves » (p.181). Dans leur description des aspects positifs de l'utilisation d'échelles descriptives, Simon et Forgette-Giroux (2001) mentionnent que l'évaluation, en lien avec chacun des critères, se fait selon des niveaux de performance, ce qui permet à l'étudiant de comparer sa performance à une performance optimale. Ces échelles peuvent être développées afin de présenter des critères propres à une tâche ou génériques à l'ensemble du cours ou même d'un programme (Forgette-Giroux et Simon, 2004), ces deux options étant chacune plus appropriées que l'autre selon les situations. Tierney et Simon (2004) expliquent elles aussi que les échelles augmentent l'uniformité de l'application des critères, et ce, grâce à l'utilisation de mots de progression dans les niveaux de progression qui varient en quantité, en fréquence ou en intensité. Une échelle descriptive offre la possibilité d'augmenter la fidélité de l'application de critères d'évaluation. Cependant, la création d'un tel outil nécessite souvent une formation supplémentaire en évaluation. Bien qu'une augmentation soit notée dans l'utilisation de cet instrument (Tierney et Simon, 2004), l'occurrence exacte de son utilisation est difficile à estimer.

Utilisation des résultats

Tel que présenté plus tôt, l'utilisation des résultats peut prendre plusieurs formes. L'attribution de résultats permet de communiquer, grâce à une note, la qualité de la performance d'un étudiant, ce qui constitue la dernière étape du processus d'évaluation. Green et Emerson (2007) définissent la notation comme étant « le processus par lequel le travail se voit assigner un code, normalement de A à F ou un pourcentage, qui représente la qualité intégrale du travail [traduction libre] » (p.497). Les notes servent à communiquer des informations à plusieurs partis intéressés (Allen, 2005) et risquent d'incorporer des critères implicites et

externes au rendement scolaire selon les divers besoins ou clientèles. Dans une étude concernant des élèves du primaire et du secondaire, Brookhart (1994) explique qu'un élève faible est parfois noté sur son effort en plus de son rendement scolaire, mais qu'un élève fort sera noté uniquement sur son rendement scolaire. Ces variations compliquent l'interprétation des résultats communiqués, mais l'occurrence exacte de ces pratiques est inconnue aussi bien au primaire, au secondaire qu'à l'université.

Lorsqu'elle est utilisée dans un but de régulation, la rétroaction est une utilisation informelle des résultats de l'évaluation (Laurier et al. 2005). Malgré la grande importance de la rétroaction dans l'apprentissage, il existe toujours un manque de recherche à ce sujet (Weaver, 2006). Les pratiques optimales en termes de rétroaction sont difficiles à identifier. Guénette (2007) a effectué une méta-analyse sur la question et présente pourquoi l'étude de l'efficacité de types de rétroaction doit prendre en compte un nombre phénoménal de variables. Suite à un regard sur plus de vingt ans de recherche, peu de conclusions peuvent être tirées en raison de la grande variation entre les études qui présentent souvent des résultats contradictoires. Shepard (2000) suggère que ce manque de consensus soit dû à un changement de paradigme en éducation qui rend les études passées difficilement comparables. Par ailleurs, Crisp (2007) a observé grâce à l'administration consécutive de deux travaux semblables, peu de différences entre le rendement aux deux évaluations, malgré une rétroaction décrivant les points à améliorer. Cependant, plusieurs variables ayant possiblement un effet sur les résultats ne sont pas discutées, par exemple, la mesure dans laquelle les travaux sont réellement équivalents, s'il y a eu une augmentation dans les attentes entre les deux travaux ou si le correcteur vise une distribution normale malgré l'utilisation d'évaluations critériées.

Weaver (2006) et Woolf (2004) proposent tous deux que l'utilité de la rétroaction dépende de son lien aux critères d'évaluation. Case (2007) ajoute que les étudiants démontrent une meilleure compréhension des attentes, des critères et une approche plus centrée sur l'utilisation des niveaux cognitifs supérieurs lorsqu'ils reçoivent une rétroaction constructive et positive plutôt que sommaire. Higgins (2000) offre cependant la même mise en garde qu'au sujet de la présentation des critères : bien que la rétroaction soit considérée comme utile par les étudiants, ces derniers doivent faire attention aux variations de vocabulaire. Le vocabulaire varie entre différentes facultés, mais aussi à l'intérieur des facultés. Il explique aussi que les étudiants provenant de contextes différents interprètent la rétroaction différemment.

États des lieux

Plusieurs études présentées ci-dessus adoptent des conceptions, des perspectives ou des méthodologies pertinentes à l'étude de la problématique visée. Afin d'identifier les critères d'évaluation utilisés au niveau des études postsecondaires, Woolf (2004) a procédé à une étude de documents de sept départements d'histoire et quatre départements de gestion. Cette étude énonce les critères répertoriés, mais ceux-ci représentent les suggestions de départements et non les critères établis par les professeurs. Webster et al. (2000) ont, eux aussi, d'abord procédé à une étude de documents afin d'identifier les critères promus par sept départements différents. Les grandes différences recensées les ont motivés à développer une nouvelle méthodologie qui, à l'aide de questionnaires et d'entrevues semi-dirigées, permettait de voir l'utilisation de critères tout au long du processus d'évaluation. Hand et Clewes (2000) recherchent eux aussi les critères utilisés, dans une école de commerce cette fois, mais ils adoptent des questionnaires et des discussions de groupe comme méthodes de collecte de données. Aucune méthode n'est utilisée afin de confirmer les critères mentionnés par les participants. De leur côté, Baume et al.

(2004) ont tenté d'étudier l'ensemble du processus d'évaluation. Pour ce faire, ils ont adopté une méthodologie où les professeurs évaluaient des portfolios d'étudiants auxquels ils n'avaient pas enseigné et devaient décrire lors de leur appréciation les jugements qu'ils portaient sur le travail. Ils ont tenté par la suite d'expliquer les différences entre les résultats obtenus et les appréciations originales des travaux. Tillema et Smith (2007) adoptent une méthodologie mixte qui comporte des questionnaires, des entrevues et des évaluations de portfolios, afin de connaître les critères d'évaluation utilisés par les professeurs universitaires qui enseignent à la formation à l'enseignement. Ces auteurs ont rapporté l'utilisation de très peu de critères explicites et encore moins de critères communs employés par les professeurs. Cependant, certaines limites de l'étude doivent être considérées lors de l'interprétation de ces résultats, entre autres, les portfolios évalués ne provenaient pas d'un cours enseigné par les participants et aucun critère n'était fourni aux participants. Wyatt-Smith et Castleton (2005) ont utilisé des protocoles oraux lors de l'appréciation de travaux en contexte et hors contexte afin de présenter les différences présentes dans leur jugement. Cette étude examine cependant la formation du jugement professionnel plutôt que l'utilisation de critères d'évaluation.

Il est donc possible de constater que malgré les nombreuses études traitant des critères d'évaluation des professeurs universitaires, il est difficile de dresser un portrait clair des pratiques courantes. Les chercheurs discutent rarement des critères d'évaluation réellement utilisés par les professeurs universitaires (Webster et al., 2000; Woolf, 2000). De plus, de grandes variations existent dans les résultats des études, souvent influencées par des contextes ou des méthodologies difficilement comparables (Hand et Clewes, 2000; Rees et Sheard, 2004; Simon et Forgette-Giroux, 2001). Finalement, les études qui examinent les critères d'évaluation des professeurs universitaires le font souvent hors du processus d'évaluation (Tillema et Smith,

2007). La définition même d'un critère varie d'une recherche à l'autre (Forgette-Giroux et Simon, 2004; McMillan, 2004; Simon et Forgette-Giroux, 2001; Wiggins, 1998). La présente étude pose donc un regard sur la création ou la sélection et l'utilisation des critères d'évaluation par les professeurs universitaires tout au long du processus d'évaluation, de la planification à la rétroaction. Étant donné le manque de consensus, il est particulièrement important de définir le concept de critère d'évaluation et de le mettre en relation avec d'autres concepts pertinents à cette étude.

Chapitre 3 – Univers interprétatif

Les choix théoriques servant à encadrer la question de recherche influencent l'ensemble de l'étude et méritent d'être présentés en détail. Un cadre théorique est souvent utilisé afin de décrire la présentation d'une théorie précise, sélectionnée afin de situer l'étude. Dans leur manuel d'analyse qualitative, Paillé et Mucchielli (2008) émettent une mise en garde :

« L'apriorisme théorique trop étroitement défini tient peu compte de la richesse des recherches plus exploratoires et des terrains où s'exerce la capacité créatrice du chercheur » (p.76). Selon ces auteurs, le choix d'une théorie ou d'un amalgame de concepts définis préalablement dans un cadre conceptuel limite l'univers interprétatif disponible au chercheur. Celui-ci « est constitué de référents très divers, actualisés en cours d'analyse à un rythme, à une ampleur et selon des modalités difficiles à prévoir » (Paillé et Mucchielli, 2008, p.74).

Ainsi donc, Paillé et Mucchielli proposent qu'une description de l'univers interprétatif remplace le cadre théorique ou le cadre conceptuel. L'inventaire des référents qui constituent l'univers interprétatif devient ainsi un regroupement des ressources à utiliser « peuplé de diverses constellations : des savoirs formels mais aussi des savoirs tacites; des constats issus de lectures ou trouvant leur origine dans la vie professionnelle; des postulats de nature épistémologique; des préjugés; etc. » (p.77). L'écriture de ces référents sert alors de conscientisation, pour le chercheur, de l'univers interprétatif initial du développement de l'étude.

Cette vision du rapport au théorique est adoptée dans cette étude puisque, comme le disent Paillé et Mucchielli (2008), « la théorie devrait nous guider dans l'observation du réel et non constituer en soi un objet de connaissance fermé ou de développement indépendant » (p.80). Ce chapitre se veut donc une présentation des référents constituant l'univers interprétatif à

l'étape de la planification de cette étude. Le choix de la posture épistémologique et de la perspective théorique et les répercussions envisagées de ce choix sont présentés en premier lieu. Par la suite, j'expose ma vision d'un critère d'évaluation et une mise en relation des concepts qui la constituent. Je termine cette section par une présentation détaillée des questions de recherche.

Posture épistémologique et perspective théorique

La posture épistémologique témoigne de la vision du rapport au savoir employée par le chercheur. Cette épistémologie fait partie de son univers interprétatif et influence ses décisions concernant autant les choix méthodologiques que l'analyse des données. La posture épistémologique adoptée dans cette étude est le constructivisme. Par ailleurs, au sein de cette posture, la perspective proposée par la théorie cognitiviste renseigne de façon utile sur l'acquisition et l'utilisation des connaissances. La perspective cognitiviste est la plus appropriée pour cette étude puisque, en éducation, elle propose plusieurs explications autant des pratiques d'enseignement que de celles d'évaluation. Elle tente, entre autres, de comprendre les processus de traitement de l'information, toujours omniprésents dans ce domaine. L'apport de cette perspective à l'étude est double; elle guide les interactions et la collecte de données auprès des professeurs participant, mais elle offre aussi une vision des rôles distincts du professeur et de l'apprenant dans l'apprentissage.

Bien que le choix d'une posture épistémologique doive guider le chercheur dans la conduite de sa recherche, l'application de la perspective cognitiviste au milieu éducationnel renseigne aussi sur la collecte de données dans cette étude. Comme le décrit Tardif (1997), cette perspective place l'apprenant au centre de ses apprentissages, car « L'apprentissage est un processus actif et constructif. La personne qui apprend ne reste pas passive devant ce qui lui est

présenté. Au contraire, ..., elle fait une sélection. Ce processus de sélection implique la création de règles » (p.34). Le professeur a donc un rôle important à jouer lui aussi; il doit favoriser l'intégration des connaissances déclaratives, procédurales et conditionnelles à la mémoire à long terme (Laurier et al., 2005) en prenant en considération les processus cognitifs des étudiants et en favorisant le développement de leur gestion de celle-ci par la métacognition (Tardif, 1997). L'apprenant est donc responsable de son apprentissage, à condition d'avoir les outils nécessaires. En ce sens, le cognitivisme suggère de communiquer aux étudiants les stratégies adéquates afin de réaliser les tâches demandées (Tardif, 1997, p.44), lesquelles devraient être « [...] complètes, complexes et signifiantes » (Tardif, 1993, p.4).

Ces recommandations sont actualisées par des pratiques en enseignement telles l'utilisation d'évaluations authentiques et le développement de l'auto-évaluation chez les étudiants. L'auto-évaluation est un outil d'enseignement très utile qui permet aux étudiants de prendre la responsabilité de leurs apprentissages (Orsmond, Merry et Reiling, 1997). Les critères d'évaluation sont un outil favorisant la réflexion en plus d'être la pierre angulaire de l'auto-évaluation. Puisque cette perspective favorise la réflexion au sujet des pratiques d'enseignement et d'évaluation, elle réitère l'importance pour les étudiants d'avoir accès aux informations concernant l'évaluation de leurs travaux.

Par ailleurs, l'adoption d'une perspective cognitiviste permet aussi de mieux comprendre le traitement de l'information effectué par les professeurs lors du processus d'évaluation. Elle renseigne le chercheur sur la manière dont les professeurs sélectionnent et appliquent les critères tout au long du processus d'évaluation. Concrètement, cette perspective propose le contexte de situation problème comme étant le plus approprié à la collecte d'informations nécessaires à l'étude des réflexions entourant le choix des critères des professeurs (Tardif, 1993, 1997).

L'utilisation d'un contexte semblable à celui rencontré par les professeurs dans l'exercice de leur profession est donc nécessaire afin de recueillir les données recherchées. De plus, l'apport de la perspective cognitive permet de percevoir les critères d'évaluation comme une composante du processus d'évaluation autant pour les professeurs que pour les étudiants.

Plus concrètement, la posture épistémologique constructiviste, surtout informée par une perspective cognitive permet de définir l'enseignement et l'évaluation en salle de classe. Contrairement à l'ancien paradigme, basé sur l'associationnisme et le béhaviorisme, l'apprentissage n'est pas perçu comme une transmission de connaissances segmentées, mais plutôt comme la construction de savoirs complexes (Shepard, 2000). Dans une perspective cognitive, les étudiants sont responsables autant de leurs apprentissages que de leurs processus de pensée concernant ces apprentissages. Comme le mentionne Shepard (2000), « intelligent thought involves self-monitoring and awareness about when and how to use skills » (p.6). Dans ce contexte, l'évaluation formative, l'auto-évaluation et la rétroaction sont des outils essentiels de l'apprentissage qui dépendent d'une connaissance des critères d'évaluation. En ce sens, la présentation des critères d'évaluation est une condition nécessaire à une évaluation juste et équitable (Shepard, p.12-13). En somme, l'évaluation, dans une perspective cognitive, joue un rôle central dans la construction du savoir et l'utilisation de stratégies cognitives adaptées, en offrant la possibilité à l'étudiant d'être responsable de ses apprentissages.

Concept de critère

La définition de McMillan (2004), citée dans la recension, exprime clairement la conception de critère désirée pour cette étude. Certains termes sont cependant discutés ici afin de modifier cette définition de manière à ce qu'elle représente davantage tous les concepts traités. Par exemple, la notion d'attribut apporte une précision accrue aux critères (Forgette-

Giroux et Simon, 2004; Mertler, 2001; Tierney et Simon, 2004). Un attribut est une qualité (Forgette-Giroux et Simon) qui varie sur une échelle, soit d'intensité, de quantité ou de fréquence (Tierney et Simon) et précise l'objet d'évaluation. Cet objet peut représenter uniquement le rendement scolaire, tel que recommandé par les experts (Allen, 2005; Brookhart, 1993; Cross et al. 1993; McMillan, 2001; Pilcher, 1994; Stiggins et al., 1989), ou inclure des critères externes tels l'effort, l'amélioration, la participation et l'attitude. Bien que la définition de McMillan (2004) mentionne le besoin de rendre les critères publics, celle-ci ne réfère pas nécessairement à leur caractère explicite. En effet, les critères doivent être présentés explicitement aux étudiants, c'est-à-dire, communiqués par écrit dans le syllabus ou dans tout document remis aux étudiants, afin que ces derniers puissent les prendre en considération dans l'accomplissement de la tâche (Case, 2007; McMillan, 2004; Ministère de l'Éducation de l'Ontario, 2000; Rust, Price, et O'Donovan, 2003; Sheaford et Sheaford, 2001; Tillema et Smith, 2007, Woolf, 2000). Le caractère explicite des critères doit être appliqué tant à leurs attributs qu'à leurs objets d'évaluation. En amalgamant ces concepts à la définition sélectionnée, un critère d'évaluation est considéré dans la présente étude comme étant *une description explicite de facettes ou de dimensions d'une performance, qui comporte un ou plusieurs attributs précisant un objet d'évaluation clairement articulé, utilisé afin d'apprécier le rendement scolaire d'un étudiant.*

Mise en relation des concepts

Les concepts présentés jusqu'ici illustrent la complexité du processus d'évaluation auquel font face les professeurs universitaires. Je propose une interprétation des relations entre ces concepts, fondée sur les connaissances recensées dans la littérature. Cette section met en relation les concepts de méthode et de processus d'évaluation, d'attribut et d'objet d'évaluation,

de critère explicite ou implicite et de critère du rendement scolaire ou externe à celui-ci. Elle fournit également les bases nécessaires à la collecte et à l'analyse des données. Par ailleurs, cette étude porte uniquement sur des travaux universitaires à réponse élaborée, ce qui comprend tous les travaux où l'étudiant doit produire un ou plusieurs textes afin de répondre aux exigences du cours. Ces travaux prennent différentes formes telles que les examens à développement, les dissertations, les critiques, les journaux, les portfolios et les examens de rédaction en classe ou à la maison.

Les relations sont nombreuses entre les concepts spécifiés. Selon la définition adoptée, un critère est composé d'un ou plusieurs attributs (Forgette-Giroux et Simon, 2004; Mertler, 2004; Tierney et Simon, 2004) et il porte sur un seul objet d'évaluation. Le rôle des attributs est de préciser les qualités recherchées de l'objet d'évaluation. Ce dernier peut représenter le rendement scolaire, ou y être externe (Allen, 2005). Par exemple, l'exactitude (attribut) des liens avec une théorie (objet) est un critère complet du rendement scolaire, mais l'envergure (attribut) du progrès démontré (objet) est un critère complet externe au rendement scolaire.

D'autre part, un critère peut être explicite ou implicite en fonction de la façon dont il est présenté (Case, 2007; Rust, et al., 2003; Sheaford et Sheaford, 2001). Le caractère explicite d'un critère est influencé par l'étape du processus d'évaluation à laquelle il est ajouté à l'ensemble des critères d'évaluation. Selon la conception du processus d'évaluation adoptée, les critères d'évaluation devraient être exprimés clairement, et préférablement par écrit, par le professeur dès la planification, afin d'être explicites à toutes les étapes subséquentes. Toutefois, un critère pourrait devenir explicite s'il est ajouté lors de l'exécution de la tâche, à la suite d'une discussion entre le professeur et les étudiants. Un critère qui n'a pas été mentionné clairement lors de la planification, ou ajouté explicitement lors de l'exécution de la tâche, ne peut être rendu

explicite lors de l'appréciation ou de l'utilisation des résultats, car le caractère explicite fait référence à la capacité de l'élève de considérer le critère avant la production et l'appréciation de son travail.

Le caractère implicite ou explicite d'un critère peut s'appliquer différemment à l'attribut et à l'objet qui le composent. De plus, même si un objet et un attribut sont nécessaires afin de former un critère complet, ces concepts peuvent apparaître de façon indépendante. La présence d'un attribut ou d'un objet seul permet donc de deviner qu'une partie du critère est implicite. Par exemple, un professeur qui mentionne évaluer la profondeur infère cet attribut à un objet d'évaluation tel que la réflexion.

Si cette conception d'un critère d'évaluation exprime ce qui peut être considéré comme un critère optimal, elle permet aussi de faire ressortir les caractéristiques manquantes. Par exemple, un critère externe fait référence à un objet autre que le rendement scolaire tel l'efficacité des habitudes de travail. Cette façon d'effectuer une discrimination entre les critères est reprise par la suite afin de développer une question de recherche et d'y ajouter deux questions plus précises.

Questions de recherche

Les professeurs universitaires ont plusieurs responsabilités lors du processus d'évaluation. Il arrive cependant qu'ils éprouvent de la difficulté à cerner leurs attentes et leurs critères d'évaluation et à les communiquer clairement (O'Donovan et al., 2004; Speck, 1998; Williams, 2005; Woolf, 2004). Cette imprécision des critères résulte dans une difficulté à saisir exactement quelles informations constituent la note (Allen, 2005; Brookhart, 1994; Pilcher, 1994; Sadler, 2003). De plus, les professeurs mesurent parfois d'autres variables que le rendement scolaire (Allen, 2005; Brookhart, 1993; Cross, et al., 1993; McMillan, 2001; Pilcher,

1994; Stiggins, et al., 1989). D'autre part, des critères implicites s'ajoutent parfois lors de l'appréciation de travaux universitaires (Case, 2007; Rust, et al., 2003; Sheaford et Sheaford, 2001). Les professeurs et enseignants basent leurs pratiques d'évaluation principalement sur leur expérience personnelle (Allen, 2005; Speck, 1998; Tremblay, 1998) en raison de formations et ressources limitées (Allen, 2005; Brookhart, 1993). Les pratiques d'évaluations des professeurs universitaires qui enseignent à la formation à l'enseignement jouent un rôle exemplaire dans le développement professionnel des enseignants (Allen, 2005; Speck, 1998). En réponse à ces constatations, il semble pertinent de se questionner sur les pratiques d'évaluation des professeurs universitaires, particulièrement concernant les critères utilisés. Cette étude propose donc de répondre à la question suivante :

1. *Quels critères certains professeurs universitaires d'un programme de formation à l'enseignement utilisent-ils pour évaluer des travaux de leurs étudiants?*

Deux questions additionnelles servent à approfondir les résultats de la première question, en plus de guider la méthodologie exploratoire. Ces questions sont en lien direct avec les composantes de la définition d'un critère, présentées dans la mise en relation des concepts, et issues de la recension des écrits. Cette étude examinera donc :

2. *Dans quelle mesure ces critères sont-ils communiqués explicitement?*
3. *Sur quels objets d'évaluation ces critères portent-ils, si autres que le rendement scolaire?*

Chapitre 4 – Méthodologie

Ce chapitre détaille la méthodologie développée afin de répondre aux questions de recherche énoncées précédemment. Il est structuré de façon à ce que la présentation des choix théoriques structurant la méthodologie précède la description détaillée du contexte, des participants de l'étude, de l'instrumentation et de la procédure adoptée par la suite. La description du traitement et de l'analyse des données et de la rigueur méthodologique de cette étude complète ce chapitre.

Choix théoriques

Cette étude exploratoire tente d'éclaircir la nature et les composantes des critères utilisés par les professeurs universitaires en contexte de formation à l'enseignement lors des diverses étapes du processus d'évaluation. Selon De Ketele et Roegiers (1993), une étude exploratoire permet de « faire l'inventaire des variables susceptibles de jouer ... et donc de bien cerner la problématique de l'objet d'étude » (p.102) et de faire émerger de nouvelles hypothèses. Elle permet aussi de « déterminer les impasses et les blocages susceptibles d'entraver un projet de recherche à grande échelle » (Poupart, 1997, p.88). En optant pour une telle étude, une grande quantité d'informations pertinentes peut être recueillie à partir d'un échantillon de faible taille. Dans cette optique, une étude de cas multiples permet l'étude complète et détaillée d'un sujet précis dans un contexte naturel, ce qui est en lien avec la posture épistémologique adoptée. Selon Karsenti et Savoie-Zajc (2004), l'étude de cas multiples permet à la fois de constater l'interaction de plusieurs variables ainsi que de faire ressortir les ressemblances et les particularités des cas. Ils ajoutent que celle-ci permet l'utilisation de techniques variées de collecte d'information, telles le recueil de documents, les entrevues, les protocoles oraux et autres.

Puisque cette étude vise à explorer les critères d'évaluation employés lors du processus d'évaluation par les professeurs universitaires, elle doit considérer plusieurs informations diversifiées dans un contexte précis tout en tenant compte des différences présentes entre les participants. Pour ce faire, la collecte de données est divisée selon le processus d'évaluation et fait appel à des entrevues, au recueil de documents écrits et à des protocoles oraux afin d'accéder aux données nécessaires. Le traitement et l'analyse des données qui suit continuent de tisser des liens entre les questions de recherche, le processus d'évaluation et les choix théoriques décrits ici. Les répercussions de ces choix sont traitées en détail dans les sections suivantes.

Contexte ontarien

L'éducation étant de juridiction provinciale au Canada, chaque province est libre de mettre en place un système qui réponde à ses besoins spécifiques. À ce sujet, l'Ontario a adopté un système éducatif dédoublé afin de servir une clientèle anglophone et une clientèle francophone. Selon Simbagoye (2007), les communautés francophones en Ontario sont caractérisées par des « disparités sur le plan des effectifs, l'hétérogénéité linguistique et culturelle et la dispersion géographique des écoles » (p.135). Il ajoute que « dans un tel contexte, la transmission de la langue et de la culture françaises fait partie du mandat des écoles de langue française » (p.135).

Cette séparation a été implantée lors d'une réforme de grande envergure qui a lieu au sein du système éducatif ontarien depuis les dernières années. En 1993, le gouvernement provincial a demandé à la Commission royale sur l'éducation (CRE) de réaliser un examen du système d'éducation, puis de formuler des recommandations (Simbagoye, 2007). En réponse à cette demande, la Commission d'amélioration de l'éducation (CAE) a été créée afin de mettre en

place les modifications nécessaires dans le but d'implanter une partie des recommandations. Ces changements incluent, entre autres, la réduction du nombre d'années du programme secondaire, la réforme du curriculum, la restructuration des conseils scolaires, la réforme du système de financement, l'instauration d'un bulletin scolaire uniforme, la création de l'Office de la qualité et de la responsabilité en éducation (OQRE) et de l'Ordre des enseignantes et des enseignants de l'Ontario (CAE, 2000; Simbagoye, 2007). C'est lors de la restructuration des conseils scolaires que la création des nouveaux conseils scolaires distincts au plan linguistique a eu lieu. Par ailleurs, la mise en place d'évaluations critériées a coïncidé avec la mise en place des nouveaux programmes d'études en Ontario.

À la suite de ces changements, la CAE a formulé à son tour quatre piliers dans lesquels le ministère de l'Éducation de l'Ontario (MEO) devrait investir afin de continuer la réforme du système éducatif entamée. Parmi ses recommandations, la CAE propose comme deuxième pilier, le « Perfectionnement professionnel, formation et leadership » (CAE, 2000, p.11). Dans cette optique, le MEO a entrepris le développement du perfectionnement professionnel des enseignants, entre autres, par un financement accru des universités (Simbagoye, 2007). La relation entre les facultés d'éducation, le MEO et les acteurs concernés par la formation à l'enseignement est complexe et traite d'aspects financiers, légaux et curriculaires. Cependant, les obligations engendrées par cette relation n'enfreignent pas la liberté académique des professeurs universitaires quant aux évaluations en salle de classe. Cette étude n'élaborera donc pas cette relation de façon plus approfondie.

La formation universitaire initiale des enseignants dure un an et comprend deux stages pratiques. Simbagoye (2007) qualifie cette formation d'« unique en son genre » (p.155) et se questionne sur la valeur d'une formation d'une telle durée. La CAE, quant à elle, explique que

« personne ne commence à exercer une profession doté de toutes les connaissances et compétences dont il a besoin durant toute sa carrière » (CAE, 2000, p.11) et prône une formation continue pour les enseignants de l'Ontario, laquelle est offerte, en partie, par les universités responsables de la formation initiale.

Des discussions avec les participants de la présente étude ont permis de dégager une vision de la formation qu'ont certains professeurs de la faculté d'éducation où s'est déroulée l'étude. Le portrait qui se dresse est celui d'une formation exigeante en raison de la charge de travail demandée, compte tenu du temps alloué. Au niveau des évaluations, les professeurs mentionnent un souci de préparer adéquatement les futurs enseignants en utilisant des méthodes d'évaluation semblables à celles que ces derniers utiliseront à leur tour en salle de classe. De ce fait, l'utilisation d'échelles descriptives semble répandue dans cette faculté. Bien que certains participants ressentent un manque de formation par rapport à l'utilisation de ces instruments, il existe un partage de connaissances et d'échelles entre collègues. Ceux-ci consultent les professeurs donnant des cours similaires et parfois les ressources du MEO lorsqu'ils recherchent des informations concernant l'évaluation. Au cours de la discussion concernant les ressources utilisées, aucun participant n'a mentionné l'existence de ressources universitaires ou facultaires en matière d'évaluation.

Participants

La sélection des participants a été basée sur une série de critères soutenus par des études recensées. Les participants devaient être professeurs universitaires à temps plein dans une faculté d'éducation. Ce critère est nécessaire puisque les pratiques d'enseignement et d'évaluation des professeurs varient entre facultés (Cross et al., 1993). Ces professeurs devaient aussi enseigner au moins un cours à la formation à l'enseignement durant le semestre précédant

l'étude. Selon Goubeaud et Yan (2004), les professeurs à la formation à l'enseignement utilisent une plus grande variété d'évaluations que les professeurs des autres facultés. Par ailleurs, Kelly (2002) a observé des différences significatives entre les méthodes d'évaluation utilisées par les professeurs à temps plein et ceux à temps partiel, les professeurs à temps plein utilisant plus de travaux à réponse élaborée. De plus, ces professeurs forment de futurs enseignants et les préparent à évaluer à leur tour. Allen (2005) explique que la formation en évaluation est souvent limitée et facultative pour les professeurs universitaires. Allen (2005), Speck (1998) et Tremblay (1998) sont en accord qu'un manque de formation en évaluation pousse les professeurs et enseignants à évaluer selon les méthodes avec lesquelles ils ont eux-mêmes été évalués. Par ailleurs, les professeurs dont le domaine d'expertise est l'évaluation ont été exclus dans le but de refléter le niveau de connaissances typique des professeurs à la formation à l'enseignement.

En plus de répondre à ces critères de sélection, les participants devaient effectuer eux-mêmes l'appréciation des travaux sélectionnés afin d'avoir l'application la plus juste de leur conception des critères. Afin de participer à l'étude, ces professeurs devaient également avoir accès et choisir un travail à réponse élaborée produit dans un cours qu'ils enseignent. Le travail à réponse élaborée, tel que décrit dans la mise en relation des concepts, a été choisi comme modalité d'évaluation en raison de sa popularité au niveau universitaire (Kelly, 2002). Dans une étude visant une meilleure compréhension de l'évaluation des tâches complexes, Shay (2005) attribue l'utilisation accrue de travaux à réponse élaborée au fait qu'ils requièrent l'utilisation de plusieurs concepts et théories ainsi qu'une application des méthodes de recherche et de rédaction. Plusieurs études (Condon et Kelly-Riley 2004; McMillan, 2001; Sheaford et Sheaford, 2001; Woolf, 2004) mentionnent la préférence des professeurs pour les travaux à

réponse élaborée, puisqu'ils considèrent que ces travaux font appel aux niveaux cognitifs supérieurs des étudiants. Selon Struyven, Dochy et Janssens (2005), ces travaux sont considérés comme étant plus exigeants par les étudiants universitaires que les autres méthodes d'évaluation, car ils les amènent à fournir un effort accru et à chercher une compréhension réelle du sujet. Les participants devaient par ailleurs être prêts à exprimer leur raisonnement lors d'une entrevue et lors de l'appréciation du travail, ainsi qu'accepter de partager les documents écrits qui présentent et expliquent leurs critères d'évaluation.

L'échantillon de cette étude est composé de quatre professeurs universitaires, deux hommes et deux femmes, enseignant à la formation à l'enseignement, ce qui assure la faisabilité de cette étude tout en permettant de mettre en place une étude de cas multiples. Compte tenu de la taille de la population professorale à la faculté d'éducation sélectionnée, dont les participants représentent le tiers des professeurs répondant aux critères de sélection, certaines caractéristiques des participants ne peuvent être présentées afin de conserver l'anonymat. De plus, aucune section des documents n'est reproduite dans l'ensemble ou en partie et aucune citation verbatim n'est présente dans ce document, toujours dans le but de préserver l'anonymat.

L'expérience des participants en enseignement à la formation initiale varie de 10 cours à près de 30 ans d'enseignement. Ces professeurs enseignent à de futurs enseignants aux niveaux primaire, intermédiaire et secondaire. Les domaines d'expertise des participants sont diversifiés et représentent une bonne partie de l'éventail des sujets présentés aux étudiants. Cependant, aucun professeur n'a comme expertise la mesure et l'évaluation. Le niveau de formation en évaluation des participants varie des cours de base en évaluation donnés à la formation à l'enseignement à plusieurs cours au niveau de la maîtrise et du doctorat. Seul un participant a

mentionné avoir assisté à un cours de perfectionnement au sujet de l'évaluation après l'obtention d'un poste de professeur.

Des informations additionnelles ont été recueillies concernant les pratiques d'évaluation des participants. Tel que spécifié dans les critères de sélection, tous les participants utilisent des travaux à réponse élaborée, souvent en combinaison avec des présentations, des analyses critiques, des travaux synthèses, des entrées de journal ou des examens. Tous les participants disent présenter le travail et les critères d'évaluation au début du semestre grâce au syllabus. La source d'information concernant l'évaluation la plus souvent mentionnée par les participants est les collègues qui donnent des cours équivalents, mais certains consultent aussi les ressources du MEO ainsi que d'autres livres et outils sur le sujet. Aucun participant n'a mentionné avoir consulté des documents produits par la faculté ou par l'université en matière d'évaluation.

Instrumentation

Étant donné sa nature exploratoire, des instruments de collecte de données et d'analyse ont été développés spécifiquement pour cette étude. La collecte de données suit approximativement les étapes du processus d'évaluation. À l'étape de la planification de l'évaluation, la collecte de données a nécessité la création de questions d'entrevue (Annexe 3). Concernant l'exécution du travail, la collecte des données de cette étape se résume à récupérer l'ensemble des documents contenant des informations pertinentes à l'étude des critères. Lors d'une courte rencontre avec les participants, tous les documents servant à communiquer aux étudiants la tâche à accomplir, les compétences visées et les critères d'évaluation ont été demandés. Ces documents comprennent entre autres des syllabus du cours, des descriptions de la tâche d'évaluation, l'échelle descriptive et autres documents fournis aux étudiants afin de préciser le travail à accomplir. L'étape de l'appréciation du travail a nécessité la création

d'instructions (Annexe 4) et d'exemples de protocole oral (Annexe 5). Finalement, à l'étape de l'utilisation des résultats, les documents comprenant tous les commentaires écrits offerts en rétroaction et qui renseignent l'étudiant sur sa performance ont été recueillis. À la suite de la collecte de données, l'identification et l'analyse des critères ont été faites à l'aide d'une grille développée à cet effet (Annexe 6).

Les entrevues permettent de recueillir des informations contextuelles précises au sujet des cas étudiés quant à leur formation, leur expérience ainsi que leurs pratiques d'évaluation. Les questions d'entrevue (Annexe 3) sont inspirées de recherches sur les pratiques d'évaluation d'enseignants et de professeurs (Gaeddert, 2001; McKenna, 2003; Shanahan, 2001) et de pratiques additionnelles identifiées dans la recension des écrits. L'entrevue est composée de huit questions distinctes, réparties en deux catégories. Les trois premières questions portent sur la formation générale des participants ainsi que leur formation en évaluation et servent principalement à la sélection des participants. Les cinq autres questions traitent plus spécifiquement des pratiques d'évaluation des participants dans leurs cours. Ces questions concernent, entre autres, les méthodes d'évaluation, le déroulement du processus d'évaluation et les critères d'évaluation. L'utilisation d'entrevues de type semi-dirigées permet aux participants de discuter librement et d'apporter des éléments qu'ils considèrent importants à la discussion (Karsenti et Savoie-Zajc, 2004). Étant donné que les pratiques des participants ne sont pas connues, cet instrument offre également la possibilité d'avoir accès à des données contextuelles imprévues en plus d'obtenir les informations recherchées. L'accès à de telles données est discuté dans la section limites de cette étude.

Le deuxième instrument de collecte de données a pour but de renseigner sur les critères d'évaluation utilisés par les professeurs lors de l'étape de l'appréciation du travail. Cet

instrument, le protocole oral (De Serres, 2001) ou protocole « Think Aloud » (Ericsson et Simon, 1980), permet de recueillir des informations situées dans la mémoire à court terme et n'implique donc pas une réflexion additionnelle de la part des participants (Ericsson et Simon, 1984), laquelle pourrait venir influencer les données. En lien avec la posture épistémologique décrite dans l'univers interprétatif, cet instrument permet de concevoir l'appréciation comme un processus cognitif qui se construit autour des critères d'évaluation. Le protocole oral donne accès au processus de sélection de l'information et à l'application des critères en temps réel. Cet instrument est donc utilisé afin de déceler les différences possibles entre les critères d'évaluation communiqués aux étudiants et ceux utilisés lors de l'appréciation. Van Someren, Barnard et Sanberg (1994) mentionnent l'utilisation du protocole oral comme méthode de collecte de données depuis les années 30. Ils expliquent que les participants sont souvent plus habitués à réaliser une tâche qu'à la décrire et qu'il est fort possible que leurs descriptions, lorsqu'effectuées sans accomplir la tâche, soient incomplètes ou incorrectes (Section 1.1). Il existe plusieurs variations de cette méthode; cette recherche utilise le protocole oral dans un contexte simulé, ce qui permet de mieux contrôler le déroulement, mais peut donner des résultats différents des pratiques réelles (Aitken et Mardegan, 2000).

Afin d'avoir une application uniforme d'une rencontre à l'autre et d'obtenir les réactions désirées par les participants, les explications et les instructions (annexe 4) sont rédigées avec une attention particulière au choix du vocabulaire (Ericsson et Simon, 1984). Les explications présentent d'abord les deux exemples de protocole oral transcrit (Annexe 5). Ces exemples montrent deux stratégies cognitives différentes : la création d'une formule mathématique par le premier participant et une répétition d'essais erreurs par le deuxième participant. Ces exemples sont utilisés afin de démontrer aux participants comment un problème peut engendrer diverses

réflexions, et ce, pour aboutir à la même réponse. L'annexe 4 présente aussi deux problèmes, utilisés afin de préparer et d'habituer les participants à exprimer leur pensée à voix haute.

La grille d'identification et d'analyse des critères (Annexe 6), qui reprend la définition de critère et la mise en relation des concepts présentés dans l'univers interprétatif, sert à repérer les critères explicites ou implicites présents dans les documents fournis par les professeurs, les verbatim et la rétroaction sur les travaux appréciés par les participants. Des définitions des possibilités d'attributs, d'objets ou de critères implicites sont incluses dans cette grille et accompagnées d'exemples. Les qualités techniques de ces instruments sont discutées dans la section rigueur méthodologique, après l'explication de la procédure et des techniques d'analyse des données prévues.

Procédure

Afin de procéder au recrutement des participants, une recherche sur le site Internet de la faculté sélectionnée a permis d'identifier les professeurs admissibles. En raison du nombre limité de participants désiré, le recrutement a eu lieu à une seule université qui offre la formation à l'enseignement et compte plusieurs professeurs à temps plein à la faculté d'éducation. Les professeurs à temps plein ont leurs propres locaux, ce qui a permis d'effectuer l'enregistrement de leurs réflexions dans des lieux privés. Cette étude a eu lieu lors du semestre d'été en raison de retards imprévus.

Les professeurs répondant au critère de sélection ont d'abord été approchés afin de sonder leur intérêt à participer à cette étude. Les professeurs admissibles ont été contactés par courriel afin de leur présenter l'étude et de prendre rendez-vous. Les professeurs intéressés ont reçu le formulaire de consentement (Annexe 2). Une lecture du formulaire a permis de clarifier

ce qui était attendu des participants. Par la suite, les professeurs participant m'ont remis une copie signée du formulaire et ont conservé la deuxième copie.

Le premier participant à avoir démontré un intérêt envers l'étude a été sélectionné afin d'effectuer une étude pilote de la méthodologie. À la suite de la collecte de données, aucune modification n'a été jugée nécessaire à la méthodologie employée. Les résultats de ce participant ont été conservés et inclus dans l'analyse, avec son consentement. La méthodologie utilisée lors de l'étude pilote a donc été appliquée uniformément à l'ensemble des participants.

La collecte de données a suivi les étapes du processus d'évaluation tel que présenté précédemment. L'étape de la planification de l'évaluation sert principalement au professeur afin qu'il clarifie ses choix liés à l'évaluation tels que le sujet à évaluer, la fonction, le but, la méthode d'évaluation et les critères utilisés. Une entrevue (Annexe 3) avec les professeurs a permis d'ouvrir un dialogue et de mieux connaître leurs pratiques d'évaluation entourant les critères. Ces entrevues ont eu une durée variant entre 15 et 45 minutes et ont été conduites directement après la remise des documents. Ces entrevues ont été enregistrées, avec le consentement des participants, afin de faciliter le traitement subséquent des données. En lien avec l'étape de l'exécution du travail, tous les documents pertinents à la présentation des critères ont été demandés aux participants, par exemple les syllabus des cours, les descriptions des tâches d'évaluation et les échelles descriptives. Les participants ont ensuite pris part à un protocole oral, où ils devaient exprimer leurs pensées à voix haute lors de l'appréciation d'un travail à réponse élaborée de leur choix. Avant d'amorcer le protocole oral, cet instrument leur a été présenté (Annexe 4) à l'aide d'un exemple de transcription d'un protocole oral (Annexe 5) traduit de van Someren, Barnard et Sanberg, (1994). Les participants ont alors eu à prendre part à leur tour à un exercice préparatoire de réflexion à voix haute (Annexe 4), mais aucun des

participants n'a ressenti le besoin de prendre part au deuxième exercice qui avait été prévu (Annexe 4) avant d'entreprendre le protocole oral lors de l'appréciation du travail sélectionné. Le déroulement des protocoles oraux a eu lieu avec un nombre restreint d'interventions de ma part, limitées à demander aux participants de continuer d'exprimer leurs réflexions de façon verbale en cas d'arrêts de 30 secondes consécutives, conformément à l'utilisation de cet instrument (Aitken et Mardegan, 2000; De Serres, 2001; Ericsson et Simon, 1984). La durée des protocoles oraux a varié entre 15 et 30 minutes. Des enregistrements audio ont été utilisés afin d'effectuer des transcriptions verbatim par la suite. La dernière étape consistait à recueillir les commentaires émis sur les travaux appréciés par les participants afin d'identifier les critères utilisés lors de la rétroaction. Puisque les travaux utilisés pour cette étude avaient déjà été appréciés dans le cadre d'un cours à la formation à l'enseignement, la rétroaction fournie aux étudiants a été recopiée par les participants afin de m'en remettre une copie anonyme lors de notre rencontre. De cette façon, les commentaires écrits n'ont pas pu être influencés par leur participation à cette étude, puisqu'ils avaient été écrits avant celle-ci. Les quatre participants ont complété l'ensemble de la collecte de données lors d'une seule visite. Le traitement des données a commencé immédiatement à la suite de la collecte de données.

Traitement et analyse des données

Étant donné la nature qualitative et exploratoire de cette étude, le processus de traitement et d'analyse des données présente une certaine complexité. Cette section détaille comment les données des quatre étapes de la collecte de données des quatre participants ont été transcrites, codifiées et regroupées. Deux techniques d'analyse ont été utilisées à la suite du traitement des données : un examen phénoménologique des données et une analyse par questionnement analytique. Ces techniques sont expliquées, puis leur application est décrite de façon détaillée.

Traitement des données

Une grande variété de données a été recueillie lors de la collecte, autant de façon écrite que de façon verbale. Un traitement complexe a eu lieu avant que débute l'analyse. D'abord, les documents papier reçus de la part des participants ont été numérisés, c'est-à-dire les syllabus ou les sections de syllabus, la rétroaction, ainsi que les explications additionnelles obtenues lors de l'entrevue. Par la suite, les enregistrements des entrevues et des protocoles oraux des quatre participants ont été transcrits de façon verbatim. Lors du traitement des données, les noms des participants ont été remplacés par les lettres A, B, C et D. Afin de conserver leur anonymat et de simplifier le traitement, les noms fictifs Antoine, Béatrice, Charles et Denise sont utilisés pour représenter ces participants. Subséquemment aux transcriptions des entrevues et des protocoles oraux, les participants ont eu l'occasion de vérifier les verbatim liés à leur participation. Les corrections apportées aux verbatim étaient peu nombreuses et habituellement liées à des passages où les enregistrements étaient moins clairs, donc attribuables à un problème technique.

La codification des protocoles oraux a débuté après que ceux-ci ont été transcrits et validés. Cette étape permet d'analyser tous les documents de façon égale (De Serres, 2001; van Someren et al., 1994). Le protocole oral doit être codé afin de fournir des résultats comparables à ceux des autres collectes de données. La codification des transcriptions verbatim commence par la segmentation de l'enregistrement selon les pauses, les phrases et les idées, tel que présenté dans l'exemple de protocole oral (Annexe 5). Lors de l'analyse, les segments peuvent alors être regroupés en épisodes qui réfèrent à des éléments uniques dans le modèle cognitif utilisé. Plus précisément dans cette étude, les épisodes représentent une référence à un critère utilisé lors de l'appréciation des travaux.

Par la suite, les documents recueillis lors de la collecte de données ont été regroupés afin de faciliter l'analyse. Les syllabus, les annexes de syllabus et les échelles descriptives remises par les participants constituent les documents écrits. Les transcriptions et les questions des entrevues ont été regroupées avec les notes écrites par les participants durant ces entrevues. Les protocoles oraux forment une autre catégorie de documents. Finalement, les commentaires écrits en rétroaction ont été ajoutés dans une quatrième catégorie avec les créations d'échelles descriptives utilisées lors de l'appréciation comme sources de rétroaction.

Examen phénoménologique des données

« Notre esprit est en réalité beaucoup plus confirmatoire qu'explorateur »

(Paillé et Mucchielli, 2008, p.87).

L'examen phénoménologique des données est une première étape, suggérée par Paillé et Mucchielli (2008), à toutes études qualitatives et ne se limite pas nécessairement aux études phénoménologiques. Son but est de tirer le maximum d'information des données, c'est-à-dire des témoignages des participants, avant de les analyser avec des techniques spécifiques.

« Concrètement, dans la situation d'analyse qualitative, il s'agit d'un appel à se tourner et à retourner vers les témoignages et les observations avec la volonté d'être instruits par eux avant de se les approprier » (Paillé et Mucchielli, 2008, p.86). Cet examen des données est possible par la lecture et la relecture des entretiens constituant les données à analyser. Dans ces lectures, le chercheur doit mettre de côté ses préconceptions et ses catégories a priori et plutôt définir les thèmes émergents. Après ces lectures à répétition, le chercheur consulte les notes accumulées et les utilise afin de voir le récit qui s'en dégage.

Cette technique d'analyse est utilisée à deux fins dans cette étude; premièrement afin de dégager une vision du contexte propre à l'échantillon tel que présenté par les participants ainsi

que de développer un portrait des cas à l'étude, exposé dans la section résultats. Une lecture des transcriptions d'entrevues, des syllabus et des documents connexes ainsi que des protocoles oraux a permis de faire ressortir certaines pratiques et préoccupations au sujet des évaluations présentes chez les participants. Les résultats de cet examen phénoménologique des données ont donc été présentés précédemment, dans la description du contexte et dans la présentation de l'échantillon.

À la suite de cet exercice, les informations recueillies lors de l'analyse des entrevues ont été regroupées selon les questions et ont été ajoutées à ces données. Pour ce faire, les transcriptions et les questions d'entrevues ont été importées dans le logiciel Nvivo afin de poser un regard approfondi sur ces données. Ce logiciel permet de traiter une grande quantité de documents texte, de créer des regroupements d'extraits, d'ajouter des mémos et de présenter les données de façon à faciliter l'analyse. Ces nouvelles informations, mises en liens avec celles observées précédemment dans l'examen phénoménologique, ont servi à la description des cas dans le chapitre suivant, c'est-à-dire les résultats.

Analyse par questionnement analytique

L'analyse par questionnement analytique (Paillé et Mucchielli, 2008) a été choisie comme technique d'analyse principale. Cette technique consiste à procéder à l'analyse des données en essayant de répondre à des questions précises plutôt que de formuler des catégories ou d'identifier des thèmes présélectionnés. Les réponses à ces questions permettent de formuler de nouvelles questions qui améliorent à leur tour la compréhension du phénomène étudié. Les données ont été utilisées afin de répondre aux trois questions de recherche exprimées dans l'univers interprétatif. Ces questions se prêtent à l'analyse par questionnement analytique, puisque la deuxième et la troisième question précisent davantage les réponses obtenues à la

première question. Les questions de recherche ont donc été utilisées afin de structurer l'analyse ainsi que l'écriture de sections de plusieurs chapitres.

Tout comme à l'étape de la collecte de données, la procédure adoptée pour l'analyse a d'abord été effectuée avec les données du participant à l'étude pilote dans le but d'en vérifier la qualité du fonctionnement. Les étapes du processus d'évaluation, décrites dans la recension des écrits, ont de nouveau été utilisées afin de guider les étapes à suivre lors de l'analyse des données pour la première question. L'analyse de la première question de l'étude pilote a été effectuée par deux évaluateurs séparément et les résultats obtenus ont ensuite été comparés. La personne sélectionnée afin d'agir comme deuxième évaluateur est un praticien dans le domaine de l'éducation avec une spécialisation autre que la mesure et l'évaluation. Des instructions lui ont été fournies, avant de débiter, concernant la procédure à suivre lors de l'identification et l'analyse des critères et de l'utilisation de l'instrument prévu à cette fin (Annexe 6). La procédure suivie par les évaluateurs est décrite dans la section suivante, qui discute de l'analyse de la première question.

Un accord interjuge a été calculé à chaque étape où les évaluateurs ont travaillé séparément. Hartman (1977) propose trois étapes à la sélection d'une mesure de l'accord interjuge : le choix de l'unité de mesure, l'étendue de temps ou de texte et le choix de la méthode de calcul. Dans cette étude, l'unité de mesure demeure les critères d'évaluation identifiés et l'étendue est délimitée par l'ensemble des documents pertinents à une étape du processus d'évaluation. Les méthodes de calcul d'accord interjuge existantes sont réparties en pourcentages d'accord ou en indices corrélationnels (Hartman, 1977). Étant donné que l'unité de mesure exclut les sections qui ne sont pas des critères, donc l'accord de non-occurrence, et

qu'aucun ordre n'est donné aux critères d'évaluation, les méthodes de calcul disponibles sont limitées.

La méthode utilisée afin de calculer l'accord interjuge est le pourcentage d'accord d'occurrence, c'est-à-dire $O\% = A / (A + B + C) * 100$, où A est le nombre d'accords sur la présence de critères et B et C représentent l'identification d'un critère uniquement par l'un ou l'autre des évaluateurs (House, House et Campbell, 1981). Cette méthode n'inclut pas de correction pour la possibilité d'accord dû à la chance. Cette caractéristique n'est toutefois pas considérée comme une limite dans cette étude en raison de l'unité de mesure choisie. Selon House et al., le pourcentage d'accord d'occurrence offre comme avantages la facilité de son calcul et de son interprétation. Ils ajoutent qu'un pourcentage d'accord supérieur à 70 est nécessaire, supérieur à 80 est suffisant et supérieur à 90 est bien (Traduction libre, p.46). Les accords obtenus sont présentés de concert avec la description de l'analyse à chacune des étapes du processus d'évaluation. Ces exercices étaient suivis d'une discussion des opinions divergentes au besoin.

Afin de répondre à la première question, *quels critères certains professeurs universitaires d'un programme de formation à l'enseignement utilisent-ils pour évaluer des travaux de leurs étudiants*, l'analyse a débuté par la lecture des documents, tels les syllabus et les échelles descriptives des participants, dans le but de recenser les critères présents à l'aide de la grille d'identification et d'analyse des critères (Annexe 6). Cette grille présente la définition d'un critère d'évaluation et reprend la mise en relation des concepts de façon à préciser si un critère est composé d'attributs et d'un objet mesurant le rendement scolaire ou non et s'il est exprimé explicitement. Dès le début de l'identification des critères, le deuxième évaluateur a relevé la difficulté d'identifier certains critères lorsque ceux-ci comportent des attributs ou des

objets implicites ou peu évidents. L'ajout d'une définition minimale, c'est-à-dire le minimum nécessaire afin que des instructions soient considérées comme un critère d'évaluation, a permis de résoudre ce problème. Cette définition, qui décrit un critère d'évaluation comme étant *une description de facettes ou dimensions d'une performance utilisée afin d'apprécier le rendement d'un étudiant*, a été ajoutée à l'annexe 6. De ce fait, il est plus simple d'identifier le format du travail comme étant un objet d'évaluation ayant un attribut implicite à l'aide de la nouvelle définition plus inclusive. Cet ajout a précédé l'identification réelle des critères. Par conséquent, uniquement la nouvelle version de l'annexe 6 a été utilisée dans cette étude. Par ailleurs, seuls les critères pertinents aux travaux sélectionnés par les participants ont été retenus. En ce sens, une seule exclusion a eu lieu. La ponctualité est possiblement utilisée afin d'apprécier le rendement d'un étudiant lors du calcul de la note finale, mais ne réfère pas à un travail précis et n'est donc pas retenue comme critère d'évaluation.

Les évaluateurs ont dressé individuellement une liste des critères identifiés ainsi que les justifications de leurs choix. Ces justifications ont pris la forme d'une section de texte dans le syllabus ou d'une rangée d'une échelle descriptive qui représentent la définition minimale. Les sections de texte propres à un seul critère d'évaluation ont varié de quelques mots à une ou deux phrases. Elles ont parfois été regroupées et parfois été évoquées à plus d'une reprise dans le syllabus. Une rangée d'une échelle descriptive représente ici une suite horizontale de cellules dans un tableau, en opposition à une suite verticale nommée colonne et qui représente souvent dans une échelle descriptive un niveau de progression. Puisqu'un seul critère devrait être présent par rangée (Tierney et Simon, 2004), il est difficile d'identifier un terme mieux adapté pour expliquer ces regroupements de critères. Des exemples de tels regroupements sont présentés au début du chapitre des résultats.

Tel qu'expliqué plus tôt, le calcul de l'accord interjuge entre les évaluateurs a pris en considération les noms des critères ainsi que les éléments inclus dans les justifications des critères recensés. Un accord interjuge de 100% a été calculé lors de cette étape à l'aide de la formule du pourcentage d'accord d'occurrence [$100\% = 11 / (11 + 0 + 0) * 100$], puisque 11 critères ont été identifiés par les deux évaluateurs et qu'aucun critère additionnel n'a été identifié par l'un ou l'autre des évaluateurs. Les variations dans le choix des noms des critères, tels *Pertinence des concepts* par un évaluateur et *Pertinence des éléments du programme disciplinaire* par l'autre évaluateur n'ont pas été incluses dans le calcul de l'accord interjuge lorsque ces variations du vocabulaire étaient présentes dans les documents du participant. Dans ce cas, les termes *concepts* et *éléments du programme disciplinaire* étaient utilisés afin de désigner un seul concept. Par la suite, les listes de critères des deux évaluateurs ont été combinées en une seule.

Le processus suivi afin de nommer les critères a tenté d'être le plus fidèle possible aux noms donnés par les participants, tout en incluant un maximum d'information. Afin de poursuivre avec l'exemple précédent, *les éléments du programme disciplinaire* offre davantage d'information que *les concepts*. Par conséquent, le critère est nommé *Pertinence des éléments du programme disciplinaire*. Par ailleurs, les noms précisés dans le chapitre des résultats représentent la meilleure estimation d'un critère complet, c'est-à-dire une estimation des attributs ou objets implicites inclus dans un critère précis. Dans les cas où les attributs ou l'objet ne pouvaient être identifiés, une recherche d'indices pointant vers ces derniers a été entreprise. Dans l'échelle descriptive, une recherche pour les variations d'attributs ou d'objets a aussi été effectuée dans les niveaux de progression mentionnés. Par ailleurs, puisqu'un seul objet d'évaluation est possible par critère, lorsqu'un ou plusieurs attributs étaient appliqués

simultanément à plusieurs objets d'évaluation, cette situation a été traitée comme plusieurs critères distincts.

Une fois les critères d'évaluation identifiés dans les syllabus et autres documents de référence, ils ont été utilisés afin de repérer les critères présents dans les protocoles oraux. De nouveau, deux évaluateurs ont tenté d'identifier les critères utilisés lors de l'appréciation, à l'aide de la liste de critères dressée à l'étape précédente, tout en étant conscients de l'apparition possible de nouveaux critères. Les évaluateurs ont identifié tous les critères présents dans le syllabus d'Antoine en plus d'une référence à un critère additionnel. Lors de cette étape, les deux évaluateurs ont donc obtenu encore une fois un accord interjuge de 100%, basé sur 12 critères cette fois. Aucune modification n'a été apportée à la liste de critères préparée à l'étape précédente, mis à part l'ajout d'informations précisant si les critères étaient présents ou non à l'étape du protocole oral. Le critère additionnel identifié par les évaluateurs dans le protocole oral est présenté dans le chapitre des résultats.

Les évaluateurs ont répété par la suite la même procédure avec les données reliées à la rétroaction fournie à l'étudiant. De nouveau, la liste des critères identifiés aux étapes précédentes a été utilisée conjointement à la grille d'identification et d'analyse des critères. Encore une fois, un accord interjuge de 100% a été calculé en fonction des 11 critères identifiés au départ et aucun nouveau critère n'a été ajouté. Lors des trois étapes, un total de 34 critères (11+12+11) ont été identifiés par les deux évaluateurs et aucun critère n'a été identifié par un seul évaluateur pour un accord interjuge de 100% pour le participant de l'étude pilote. De tels accords interjuges sont souhaitables dans une étude qualitative, mais peu communs. Une explication possible de ces accords est la présence d'une échelle descriptive qui présente

explicitement des critères d'évaluation. L'effet de la présence d'échelles descriptives dans cette étude est élaboré lors de la discussion des résultats au chapitre six.

Une fois l'analyse pilote terminée, ces étapes ont été reprises avec les données des trois autres participants. Les critères identifiés pour les quatre participants sont présentés dans le chapitre résultats. Les critères imprécis ou implicites, les changements dans les attributs ou les objets d'évaluation dans les niveaux de progression et autres problèmes encourus lors du choix des noms des critères y sont aussi présentés.

La procédure d'analyse de la deuxième question, *dans quelle mesure ces critères sont-ils communiqués explicitement*, diffère de celle de la première. Puisque les attributs et les objets d'évaluation ont été recensés lors de la première question, cette question part des résultats obtenus afin de préciser dans quelle mesure ils sont explicites. La mesure mentionnée ici n'est pas une quantification, mais plutôt un inventaire et une description des instances où un attribut, un objet d'évaluation ou un critère entier est implicite. Pour ce faire, une comparaison a été effectuée entre le critère tel qu'il se retrouve dans les étapes du processus d'évaluation et le critère identifié lors de l'analyse pour la première question.

De façon similaire à la question précédente, la troisième question, *sur quels objets d'évaluations ces critères portent-ils, si autres que le rendement scolaire*, précise les résultats obtenus à la première question. Un regard est désormais porté sur les aspects sur lesquels les professeurs évaluent les travaux des étudiants autres que le rendement scolaire. Au départ, cette tâche était perçue comme un simple inventaire, puisque ces informations avaient été identifiées lors de la première analyse. Cependant, la grille d'identification et d'analyse des critères d'évaluation (Annexe 6) n'a pas permis d'établir les critères qui ne traitent pas du rendement scolaire. D'abord, un retour à la définition du concept de rendement scolaire a été nécessaire

afin de répondre adéquatement à la troisième question. Par la suite, une comparaison entre les critères recensés à la première question et les objectifs du cours établis par les participants a été effectuée. Les raisons qui ont amené ces changements sont présentées dans le chapitre suivant, avec les résultats.

Selon Paillé et Mucchielli (2008), la recherche qualitative évolue tout au long de son déroulement. L'analyse par questionnaire analytique traite des trois questions présentées dans *l'univers interprétatif*. Ces questions, sélectionnées afin de fournir une direction à l'analyse, permettent à leur tour d'identifier de nouvelles pistes de questionnement. Cette étude étant menée dans le cadre d'une maîtrise, elle ne peut poursuivre toutes les avenues identifiées. Néanmoins, plusieurs pistes de recherches futures sont discutées à la suite de la présentation des résultats.

Rigueur méthodologique

Toute recherche scientifique se doit de fournir des preuves de la rigueur méthodologique employée durant sa réalisation. Ces preuves permettent au lecteur de juger la valeur de la recherche selon des critères précis. Les critères méthodologiques de cette étude sont ceux suggérés pour une étude qualitative par Karsenti et Savoie-Zajc (2004), c'est-à-dire la crédibilité, la transférabilité, la fiabilité et la confirmation. Chacun de ces critères est défini préalablement à l'énumération des preuves qui y sont associées.

Le critère de crédibilité réfère à « la plausibilité de l'interprétation du phénomène étudié » (Karsenti et Savoie-Zajc, 2004, p.142). Une façon d'augmenter la crédibilité de l'étude est l'utilisation de techniques de triangulation. Karsenti et Savoie-Zajc (2004) ainsi que Patton (1990) suggèrent qu'une triangulation des méthodes de collecte de données permette de faire ressortir divers éléments de réponse selon les méthodes utilisées. Karsenti et Savoie-Zajc (2004)

ajoutent que « plusieurs modes de collecte de données compensent pour les limites de chacun pris individuellement » (p.146). La collecte des documents de référence, du protocole oral et de la rétroaction écrite, fournit trois sources différentes à la recherche des critères d'évaluation utilisés par les participants. En plus d'utiliser de multiples sources de données, une partie de l'analyse a été effectuée par deux évaluateurs indépendants. En effet, les données des documents de référence, du protocole oral et de la rétroaction du premier participant ont toutes été analysées par deux évaluateurs. Un accord interjuge a été calculé à chacune de ces étapes, puis comparé au taux d'accord de 0,70 établi avant de débiter (House et al., 1981). Dans les trois cas, aucune discussion n'a été nécessaire afin de concorder les différences d'opinions puisque l'accord interjuge calculé était de 100%. Une troisième preuve de crédibilité de l'étude vient de l'entrevue avec les participants, de l'étude phénoménologique des données et de la recherche documentaire, toutes effectuées dans le but de développer une compréhension fine des dynamiques du contexte étudié (Karsenti et Savoie-Zajc, 2004, p.143). De plus, un retour au participant, sous la forme de la confirmation des verbatim des entrevues et des protocoles oraux, apporte une autre preuve de la crédibilité des données, nécessaire à la crédibilité des interprétations.

Le deuxième critère de rigueur méthodologique de Karsenti et Savoie-Zajc (2004), la transférabilité, est un critère qui réfère à la possibilité d'adapter les résultats de l'étude à un autre contexte. Cette étude étant effectuée dans un contexte particulier, la description de celui-ci a été alimentée grâce à une entrevue avec les participants, à une étude phénoménologique des données et à de la recherche documentaire, tel que mentionné ci-haut. De plus, la section participants décrit et justifie les critères de sélection. Cependant, les précautions prises afin de conserver l'anonymat des participants limitent la transférabilité des résultats. D'autre part, le

critère de transférabilité de Karsenti et Savoie-Zajc s'apparente à deux critères proposés par De Ketele et Roegiers (1993), soit la généralisabilité des résultats et la répétabilité de la démarche. Afin de permettre une reproduction des démarches suivies, l'ensemble de la méthodologie est documenté avec souci du détail, que ce soit la procédure, le traitement ou l'analyse des données. Également, les instruments de collecte de données et d'analyse sont mis en annexe et sont à la disposition du lecteur. De plus, un exemple composite d'échelle descriptive est présenté dans la section résultats afin de faciliter la répétabilité de l'analyse. Par ailleurs, selon De Ketele et Roegiers, la généralisabilité des résultats, qui est limitée entre autres par l'omission de données afin de préserver l'anonymat, n'est pas une visée des études exploratoires.

Le troisième critère proposé par Karsenti et Savoie-Zajc (2004) est celui de la fiabilité. Ils définissent ce critère selon la « cohérence entre les questions posées au début de la recherche, l'évolution qu'elles ont subie, la documentation de cette évolution et les résultats de la recherche » (p.143). L'adoption d'un univers interprétatif en remplacement d'un cadre théorique ou conceptuel, permet un processus de pensée où les changements qui surgissent lors de l'étude sont intégrés (Paillé et Mucchielli, 2008). De tels ajouts sont exposés dans la méthodologie, dans la présentation des résultats ainsi que dans la discussion. De plus, une étude exploratoire favorise aussi l'inclusion de nouvelles informations non anticipées. L'évolution de l'étude est ainsi intégrée dans la description de celle-ci. Par ailleurs, l'analyse par questionnement analytique, technique principale d'analyse de cette étude, favorise le développement de la compréhension par la formulation de questions de recherche. Tel que mentionné précédemment, les réponses à ces questions donnent lieu au développement de nouvelles questions plutôt qu'à la remise en question de ces premières. Les questions de recherche ont donc un rôle central tout au long de l'analyse et servent également à structurer la

présentation des résultats et la discussion qui s'en suit. Outre l'importance donnée aux questions de recherche dans l'étude, l'utilisation des étapes du processus d'évaluation, décrites dans la recension des écrits, est présente dans l'ensemble de l'étude. Les étapes du processus d'évaluation aident à structurer la recension et à définir l'ordre des opérations de collecte et d'analyse des données. Toujours dans un souci de documenter les changements potentiels à la méthodologie au cours de l'étude, certaines mesures ont été mises en place durant son exécution. Premièrement, comme décrit auparavant dans la section procédure, un des professeurs avait accepté de participer à une étude pilote de la méthodologie. Puisqu'aucune modification ne s'est avérée nécessaire, les données ont été incluses au corpus, tel que convenu avec le participant. Par la suite, les données de ce même participant ont été utilisées afin d'effectuer une étude pilote du traitement et de l'analyse des données. Cet exercice a permis d'identifier une difficulté lors de l'utilisation de la grille d'identification et d'analyse des critères (annexe 6), entraînant un ajout à l'instrument afin d'augmenter sa précision. Par ailleurs, l'existence de certaines politiques et documents de référence découverts grâce aux commentaires du comité responsable d'évaluer cette étude ont été inclus dans le chapitre résultats et les chapitres subséquents.

Karsenti et Savoie-Zajc (2004) proposent le critère de confirmation comme quatrième et dernier critère. Selon eux, ce critère « renvoie au processus d'objectivation pendant et après la recherche » (p.144). Des preuves de rigueur en lien avec ce critère peuvent prendre la forme de justifications des instruments de collecte de données, de clarification des approches d'analyse appliquées ainsi que la cohérence entre les deux (Karsenti et Savoie-Zajc, 2004, p.144). La sélection des instruments de collecte de données est basée sur le processus d'évaluation; un instrument est présent pour chacune des étapes où les professeurs doivent utiliser les critères

d'évaluation. Lors de la planification et de l'utilisation des résultats, sous la forme d'une rétroaction à l'étudiant, une analyse de traces écrites est suffisante afin d'accéder aux informations recherchées. Cependant, selon la perspective cognitive, l'utilisation d'un protocole oral au cours de l'appréciation est plus appropriée puisqu'elle permet d'avoir accès aux processus cognitifs en jeu lors de l'application des critères. Les études exploratoires qualitatives reposent beaucoup sur la description du contexte, c'est pourquoi une entrevue semi-dirigée a été ajoutée aux instruments de collecte de données. Tel que mentionné plus haut, l'analyse par questionnement analytique part des questions de recherche afin de structurer l'interprétation des données. Elle permet aussi de combiner les informations provenant de divers instruments de collecte de données et de les interroger à répétition afin de répondre aux multiples questions de recherche.

Afin de clore cette section, il est important de mentionner que malgré toutes les preuves de rigueur méthodologique qu'une étude peut amasser, celle-ci continue de présenter des limites. Il est tout aussi important de savoir reconnaître les limites d'une étude que de pouvoir en juger la rigueur. Une section expliquant certaines des limites de cette étude est présente dans le sixième chapitre, à la suite de la discussion.

Chapitre 5 – Résultats

Étant donné le grand nombre de données recueillies par cette étude, une structure définie facilite leur présentation. Conformément à l'analyse par questionnement analytique, la présentation des résultats est organisée selon les trois questions de recherche. Cependant, cette section est précédée d'une description du contexte commun à l'ensemble des participants et de celui propre à chacun des cas étudiés, à partir des résultats de l'examen phénoménologique des données. Les informations communes concernent majoritairement la présentation des critères aux étudiants. Les informations particulières à chacun des participants sont présentées selon leur place dans le processus d'évaluation.

Description des cas

Dans un premier temps, tous les cours à la formation à l'enseignement étant de 39 heures sur une période de neuf semaines, les étudiants ont donc environ quatre heures et demie de cours par semaine. Tous les participants ont remis un syllabus d'en moyenne 15 pages qui comporte, entre autres, une échelle descriptive liée au travail sélectionné. Ces échelles sont généralement présentées sur une seule page, quoique celle d'Antoine soit d'environ une page et demie. Le tableau 1 est un exemple fictif d'une échelle descriptive, grandement inspiré par les échelles des participants. Ce cas composite sert à illustrer certaines observations et à préciser les étapes de l'analyse tout en conservant l'anonymat des participants.

Cet exemple composite d'une échelle descriptive permet d'illustrer la structure générale des échelles descriptives employées par les participants. Le tableau 1 présente une échelle descriptive qui compte sept rangées, dont trois sont omises afin de faciliter la présentation, tandis que celles des participants varient de cinq à neuf rangées. Le nombre de colonnes présent

dans les échelles des participants est fidèlement représenté par une première colonne, nommée critères, suivie de colonnes subséquentes qui représentent les niveaux de progression.

Tableau 1
Exemple composite d'échelle descriptive

Critères	Bien C et C+	Très bien B et B+	Excellent A- et A	Exceptionnel A+
Pertinence	Le sujet est peu pertinent à la discipline	Le sujet est assez pertinent à la discipline	Le sujet est très pertinent à la discipline	Le projet reflète un travail modèle en ce qui concerne la pertinence du sujet à la discipline
...
Précision et clarté des concepts développés	Le travail démontre une utilisation plutôt imprécise des concepts et de la terminologie du domaine	Le travail démontre une certaine maîtrise des concepts et de la terminologie du domaine	Le travail démontre une utilisation exacte des concepts et de la terminologie du domaine	Le texte démontre une utilisation claire concise et exacte des concepts et de la terminologie du domaine
...
Utilisation de la langue	L'utilisation de la langue est peu rigoureuse	L'utilisation de la langue est assez rigoureuse	L'utilisation de la langue est très rigoureuse	L'utilisation de la langue est exemplaire
...
La présentation orale	Le discours est parfois adapté à la situation. La façon de s'exprimer manque de clarté. Le recours aux notes est fréquent, ce qui influence la fluidité.	Le discours est assez souvent adapté à la situation. La façon de s'exprimer est assez claire. Le recours aux notes est occasionnel, sans toutefois nuire à la fluidité.	Le discours est généralement adapté à la situation. La façon de s'exprimer est claire. Les notes sont utilisées à bon escient.	Le discours est continuellement adapté à la situation. La façon de s'exprimer est exemplaire. L'utilisation des notes n'est pas nécessaire.

Le tableau 1 illustre aussi des pratiques observées lors de l'identification des critères. D'abord, la colonne *critère* ne présente pas uniquement des critères complets. La *pertinence* est un attribut tandis que *l'utilisation de la langue* et *la présentation orale* sont des objets d'évaluation. Les colonnes des niveaux de progression présentent souvent les attributs et objets manquants dans la colonne des critères. Ces informations aident à identifier des critères complets, comme *la pertinence du sujet à la discipline* ou *la rigueur de l'utilisation de la langue*.

Un autre exemple illustré dans le tableau 1 est la présence de critères d'évaluation multiples dans une seule rangée de l'échelle. La définition d'un critère d'évaluation, présentée dans l'univers interprétatif, précise qu'un critère peut avoir plus d'un attribut, mais un seul objet d'évaluation. Dans la rangée *précision et clarté des concepts développés*, les attributs sont *la précision* et *la clarté* et l'objet d'évaluation est *les concepts développés*. Cependant, un deuxième objet d'évaluation, *la terminologie du domaine* est ajoutée dans tous les niveaux de progression. La rangée *présentation orale* démontre une autre façon d'inclure des critères d'évaluation multiples. Le critère présenté dans la première colonne sert en fait de regroupement pour trois critères complets dans les niveaux de progression, *l'adaptation du discours à la situation*, *la clarté du message exprimé* et *la fréquence du recours aux notes*.

D'autre part, les niveaux de progression présentent parfois une variation dans le vocabulaire utilisé comme mots de progression. Par exemple, dans la rangée *précision et clarté des concepts développés*, se trouve la progression (1) utilisation plutôt imprécise, (2) une certaine maîtrise, (3) utilisation exacte et (4) utilisation claire concise et exacte. Ces variations sont particulièrement présentes dans le dernier niveau de progression où le terme exemplaire remplace le ou les attributs, comme il est possible de voir dans la rangée *utilisation de la langue*.

Cet exemple composite n'est pas une analyse complète des données, mais sert à illustrer diverses pratiques observées lors de l'identification des critères, dans les échelles descriptives des participants. Les résultats de chacun des participants sont présentés séparément à la suite de la description détaillée des cas.

Antoine

Afin de communiquer les détails relatifs à l'évaluation du travail apprécié, Antoine a remis le syllabus du cours. Ce cours obligatoire porte sur une discipline d'enseignement à l'élémentaire. Ce syllabus décrit les différents travaux qui sont requis au cours du semestre. Il inclut aussi des instructions générales sur l'exécution et l'évaluation des travaux de même que des instructions propres à chacun des travaux. Comme mentionné, une échelle descriptive dédiée au travail apprécié fait aussi partie du syllabus. Par ailleurs, les compétences visées ainsi que les *normes d'exercice de la profession enseignante* sont présentées dès les premières pages du syllabus. Enfin, l'échelle de notation de l'université est présentée en annexe, avec un accent mis sur le caractère exceptionnel du A+ .

Le travail apprécié par Antoine requiert des étudiants qu'ils planifient, créent et adaptent un certain nombre de leçons liant la didactique du cours à une deuxième didactique au choix. Il compte pour plus du tiers de la note finale et est à remettre la semaine avant la fin du semestre. Antoine a mentionné lors de l'entrevue l'utilisation de copies types afin d'aider les étudiants à comprendre ce qui est attendu d'eux lors de l'exécution de la tâche. Un exemple de copie type n'a pas été demandé dans le cadre de cette étude, puisque les copies types sont des instruments qui aident à la compréhension des critères déjà présentés.

Selon le syllabus, il s'agit d'un travail pouvant être effectué en équipe de 2 ou 3 étudiants. Cependant, lors du protocole oral, Antoine réfère à un seul étudiant en tant qu'auteur

du travail. Antoine a procédé à plusieurs lectures du travail lors de l'appréciation. Toutes ces lectures remplissent, selon lui, des buts différents tels prendre connaissance du travail, vérifier la présence d'éléments importants, juger de la qualité de la langue écrite et finalement juger de la réussite à chacun des critères. Antoine a mentionné, lors du protocole oral, utiliser l'échelle descriptive durant l'appréciation.

Concernant la rétroaction, Antoine a utilisé l'échelle descriptive afin de surligner le niveau de progression exact atteint par l'étudiant. Cette pratique permet, en théorie, de fournir de la rétroaction propre à chaque critère. D'autre part, un commentaire global est aussi offert par écrit aux étudiants sous l'échelle descriptive. Ce commentaire traite du travail dans son ensemble en relevant les points forts et les points à améliorer.

Béatrice

Béatrice a partagé les sections du syllabus qui concernent le travail sélectionné aux fins d'évaluation. La première de ces sections propose une description de l'ensemble des travaux du cours, de leur échéance et de leur pondération. Le travail sélectionné compte pour le quart de la note finale. Une deuxième section présente les buts du travail, de même que des instructions détaillées de la marche à suivre. Ces instructions incluent certaines descriptions de ce qui est attendu, surtout au sujet de la structure du travail, mais constituent majoritairement des suggestions aux étudiants, par exemple au niveau de l'analyse à effectuer. Le dernier extrait du syllabus est l'échelle descriptive liée à ce travail. Béatrice a également mentionné l'utilisation de copies types en plus de la présentation de l'échelle descriptive aux étudiants.

Comme dans le cas d'Antoine, ce cours porte sur l'enseignement d'une didactique à l'élémentaire et est obligatoire. Le travail demandé consiste à démontrer et à analyser une stratégie d'enseignement en lien avec la didactique du cours. Une composante écrite et une

composante orale forment ce travail, toutes deux incluses dans l'échelle descriptive fournie. Ces présentations orales ont lieu tout au long du semestre et la composante écrite est un résumé des informations présentées oralement. Par ailleurs, ce travail est exécuté en groupe, mais une seule appréciation du travail a lieu et la même note est attribuée aux membres de l'équipe.

Béatrice a procédé lors du protocole oral à une lecture détaillée du travail. Tout au long de cette étape, Béatrice a émis verbalement des commentaires en lien avec les critères d'évaluation mentionnés dans l'échelle descriptive qu'elle a consultée en parallèle. Ces commentaires se sont avérés dans plusieurs cas similaires à la rétroaction écrite sur le travail lors de l'appréciation originale du travail. Les critiques émises envers la stratégie d'enseignement sélectionnée ne semblent pas avoir joué un rôle dans l'appréciation de la performance des étudiants. Elle a terminé l'appréciation en surlignant dans l'échelle descriptive les niveaux de progression appropriés pour chacune des descriptions composant ses critères.

Concernant la rétroaction, Béatrice a fourni des informations aux étudiants de trois façons. En plus d'utiliser l'échelle descriptive où elle a surligné chaque commentaire qui s'appliquait au travail apprécié, elle a écrit des commentaires tout au long du travail. Béatrice a de plus rédigé un court texte au verso de l'échelle descriptive expliquant un à un les critères réussis et les points à améliorer. La rétroaction offerte de ces trois façons est constante concernant les critères soulevés, autant lorsque la performance est adéquate que lorsqu'elle est à améliorer. Lors de l'entrevue, Béatrice a mentionné l'importance de fournir une rétroaction rapide aux étudiants et dit remettre dans la majorité des cas les travaux appréciés une semaine après la présentation.

Charles

Deux documents représentent l'étape de la planification de Charles, le syllabus du cours et l'échelle descriptive présentée séparément. Le cours enseigné par Charles est aussi un cours obligatoire, offert aux étudiants des cycles intermédiaire et supérieur. Le syllabus présente une description des travaux à accomplir par l'étudiant, des compétences visées et de la pondération des travaux. Le syllabus de Charles mentionne aussi l'importance des normes d'exercice de la profession. L'échelle descriptive fournie s'applique à l'ensemble des travaux.

Les travaux demandés sont regroupés en un portfolio, composé de cinq entrées de journaux, de cinq analyses critiques et de trois fiches synthèses. Le portfolio compte pour plus de la moitié de la note finale du cours. Lors de l'entrevue, Charles a expliqué que les étudiants ont l'option de soumettre leur portfolio lors d'une évaluation formative afin de recevoir de la rétroaction. Cependant, le nombre d'étudiants qui utilisent cette option n'a pas été adressé. Il a aussi confirmé l'utilisation de l'échelle descriptive lors de l'appréciation sommative du travail.

Lors du protocole oral, Charles a effectué un retour sur le portfolio d'un étudiant et s'est limité à discuter de l'appréciation originale plutôt que d'y procéder à nouveau. Charles a qualifié le portfolio sélectionné d'exemplaire, car celui-ci répondait exactement aux objectifs du cours. Durant ce retour, Charles n'a pas consulté l'échelle descriptive ni n'a précisé si celle-ci est utilisée dans le but de fournir de la rétroaction. Concernant la rétroaction, Charles a écrit un court commentaire général à la fin du portfolio.

Denise

Denise a aussi fourni le syllabus du cours d'où provient le travail apprécié. Ce cours concerne l'enseignement d'une didactique au cycle supérieur. Le syllabus comprend une section présentant de façon générale les travaux demandés et la pondération respective de

chacun d'eux. Une section du syllabus concernant les évaluations mentionne certains critères, le barème de notation de l'Université et le caractère exceptionnel du « A+ ». Les trois annexes du syllabus décrivent les travaux mentionnés précédemment et sont accompagnées d'échelles descriptives propres à chaque travail.

L'échelle descriptive retenue aux fins de l'étude s'applique à un travail composé de quatre tâches de réflexion d'environ une page, comptant au total pour le quart de la note finale. Ces tâches sont effectuées individuellement puis commentées par les pairs. Une seule de ces tâches a été sélectionnée pour cette étude. Le protocole oral a permis d'apprendre que Denise a consulté l'échelle descriptive lors de l'appréciation. En terme de rétroaction, Denise a écrit quelques commentaires sur le travail de l'étudiant. Si l'échelle descriptive est remise aux étudiants, que ce soit après chaque évaluation ou suite à l'achèvement du travail, ceci n'a pas été précisé.

La présentation des cas permet de dresser un portrait des pratiques d'évaluation, mais aussi de constater l'utilisation d'échelles descriptives par l'ensemble des participants. Cette dernière constatation est exceptionnelle, puisque ni la recension des écrits, ni l'expérience de l'auteur n'avaient permis d'anticiper une utilisation aussi répandue de cet outil d'évaluation. La suite des résultats présente les critères d'évaluation identifiés dans les processus d'évaluation des participants afin de répondre à la première question de recherche.

Présentation des résultats selon les trois questions

L'analyse des données d'Antoine en réponse à la première question de recherche a révélé plusieurs critères d'évaluation. Ceux-ci sont présentés dans le tableau 2 de façon à dégager les attributs, les objets d'évaluation ainsi que leur présence lors des étapes du processus d'évaluation. Plusieurs raisons ont compliqué l'identification des critères d'évaluation d'Antoine.

Premièrement, dans une échelle descriptive, chaque rangée représente normalement un critère distinct, ce qui permet de porter un jugement sur chacun d'eux. Tel que mentionné précédemment, une rangée est une suite horizontale de cellules dans un tableau. Dans son échelle, Antoine utilise plusieurs rangées qui comportent deux objets d'évaluation. Les huit premiers critères présentés dans le tableau 2 représentent en fait les quatre premières rangées de son échelle, ayant toutes des objets multiples. Par exemple, Antoine regroupe les *éléments théoriques et pratiques* sous un seul critère dans son échelle descriptive, mais les traite comme des critères séparés lors de son appréciation dans le cadre du protocole oral. Une autre observation lors de l'identification des critères a été la variation du vocabulaire utilisé afin de désigner les critères d'évaluation. Qu'il s'agisse de l'utilisation de synonymes entre les niveaux de progression de l'échelle ou à l'intérieur du syllabus, Antoine utilise parfois plus d'un terme afin de désigner un seul objet d'évaluation, tels les *assises*, les *éléments* ou les *concepts théoriques*. D'autres fois, les variations sont seulement présentes dans la colonne représentant le A+. Par exemple, l'ajout de l'attribut *rigueur* afin de désigner l'évaluation de *l'utilisation de la langue* ou la *mise en application des éléments pratiques* en contraste aux *liens aux éléments pratiques* pour tous les autres niveaux. Cette dernière information est importante en raison du caractère exceptionnel donné au A+ par plusieurs participants.

Tableau 2

Critères présents lors du processus d'évaluation d'Antoine

Attributs	Objets	Présence
Profondeur de la mise en application des éléments pratiques	S, E, P, R
Profondeur de la mise en application des éléments théoriques	S, E, P, R
Précision et clarté des concepts développés	S, E, P, R
Précision et clarté de la terminologie du domaine	E, R
Rigueur de la préparation des activités	S, E, P, R
Rigueur de l'organisation des activités	S, E, R
Cohérence des activités	S, E, P, R
Logique de l'enchaînement des activités	S, E, P, R
Pertinence des éléments du programme disciplinaire	S, E, P, R
Adaptation de la séquence à un autre niveau scolaire	S, E, P, R
Pertinence des liens avec d'autres matières	S, E, P, R
Rigueur dans l'utilisation de la langue	S, E, P, R
Originalité et créativité du travail	S, E, P, R
Respect du format du travail	S, P
Ponctualité de la remise des travaux	S

Notes : S = Syllabus; E = Échelle descriptive; P = Protocole oral; R = Rétroaction.

Le tableau 3 présente les critères d'évaluation de Béatrice. L'échelle de Béatrice est divisée en six rangées composées des multiples descriptions de facettes de la performance recherchées et ce sont ces dernières qui ont été retenues comme étant les critères d'évaluation. Les critères qui composent une rangée présentent une certaine uniformité. Par exemple, cinq critères dans une seule rangée de l'échelle descriptive réfèrent à la présentation orale. Puisque les rangées représentent une idée générale, certains critères semblables se retrouvent dans deux rangées différentes. Par exemple, le critère *clarté du message exprimé* qui est dans la rangée présentation orale est semblable au critère *précision et clarté de la formulation des idées*, qui lui, est dans une rangée qui réfère à la composante écrite. Par ailleurs, des variations de vocabulaire sont aussi présentes dans les documents de Béatrice. Elle aussi utilise des synonymes afin de préciser un attribut, tel que pour le critère visant la *clarté du message exprimé* où elle mentionne le *volume*, la *fluidité*, le *ton*, le *contact visuel* et la *prosodie* en tant qu'éléments observés. De plus, certains termes sont utilisés à répétition, par exemple, l'attribut *efficacité* est mentionné par rapport à six critères distincts et ceux *précision* et *clarté* le sont pour trois critères chacun. D'autre part, dans la colonne représentant le A+, les attributs *précision* et *concision* ainsi que l'attribut *originalité* sont ajoutés à divers critères. De façon similaire, les attributs des critères *clarté du message exprimé* et *rigueur de l'application des conventions linguistiques*, autant à l'oral qu'à l'écrit, sont remplacés par le terme *exemplaire*, toujours dans la colonne du A+.

Tableau 3

Critères présents lors du processus d'évaluation de Béatrice

Attributs	Objets	Présence
Envergure de la compréhension	E, P, R
Efficacité de la soutenance du jugement	E, P, R
Profondeur du sujet traité	E, P, R
... recherchés et soutenus	Les liens entre la théorie et la pratique sont ...	E, P, R
Clarté et cohérence de l'organisation	E, P, R
Pertinence du support pédagogique	E, P, R
Efficacité de la gestion du temps alloué	E, P, R
Clarté du message exprimé	E, P, R
Fréquence du recours aux notes	E, P, R
Adaptation du discours à la situation	E, P, R
Originalité et efficacité des moyens utilisés pour susciter l'intérêt	E, P, R
Précision et clarté de la formulation des idées	E, P, R
Précision et clarté de la transmission des idées	E, P, R
Efficacité des techniques d'écriture	E, P, R
Précision et concision du développement des éléments demandés	E, P, R
Description de la stratégie	S
Analyse de la stratégie	S
Rigueur de l'application des conventions linguistiques à l'oral	E, P, R
Rigueur de l'application des conventions linguistiques à l'écrit	E, P, R

Notes : S = Syllabus; E = Échelle descriptive; P = Protocole oral; R = Rétroaction.

La source principale dans l'identification des critères d'évaluation de Charles, présentés dans le tableau 4, a été l'échelle descriptive, mais un critère a aussi été révélé lors du protocole oral. Une fois encore, plusieurs raisons ont nui à la clarté des critères d'évaluation. Premièrement, comme l'ont fait Antoine et Béatrice, Charles a regroupé plus d'un critère d'évaluation dans une seule rangée de l'échelle descriptive. Le critère *précision et clarté de la terminologie, des concepts et des principes à l'étude* regroupe trois objets d'évaluation et est donc présenté comme trois critères dans le tableau 4. Par ailleurs, l'échelle présente une liste des attributs recherchés et les objets d'évaluation sont présents dans les niveaux de progression. Le vocabulaire est plus constant que pour les deux participants précédents, mais on remarque toujours l'utilisation de nouveaux attributs et l'ajout de synonymes, particulièrement dans la colonne représentant le A+. Par exemple, dans la rangée représentant le critère *innovation des idées et des approches*, on retrouve les attributs *créatif, original, différent, varié* et, de façon péjorative, *limité aux contenus étudiés en classe*. Comme décrit dans la méthodologie, le choix des noms des critères a tenté d'être fidèle aux termes employés par les participants tout en conservant une certaine concision.

Tableau 4

Critères présents lors du processus d'évaluation de Charles

Attributs	Objets	Présence
Pertinence des éléments du projet à la compétence visée	E, P, R
Profondeur de l'analyse des composantes du sujet	E, P, R
Envergure des composantes du sujet	E, P, R
Précision et clarté de la terminologie	E
Précision et clarté des concepts à l'étude	E
Précision et clarté des principes à l'étude	E
Cohérence des divers éléments du projet	E
Autonomie de la production du projet	E
Rigueur dans l'utilisation de la langue	S,E
Innovation des idées ou des approches	E
Respect du format du travail	P

Note : S = Syllabus; E = Échelle descriptive; P = Protocole oral; R = Rétroaction.

Le tableau 5 énumère les critères de Denise. Une fois de plus, l'échelle descriptive est la source principale des critères, avec deux critères provenant uniquement du syllabus. Denise a mentionné lors du protocole oral l'importance particulière de deux des critères d'évaluation en lien avec le travail sélectionné. Une seule partie de l'échelle a donc été utilisée lors de l'appréciation. Chaque rangée de l'échelle représente une compétence visée par la tâche demandée. Dans ce cas, les deuxième, troisième et quatrième critères du tableau 5 sont liés à la tâche appréciée. L'échelle descriptive de Denise comporte aussi des variations de vocabulaire dans les niveaux de progression. Par exemple, pour le critère concernant les *interventions aux discussions sur les lectures*, les attributs *abondance* et *fréquence* sont interchangeables au mot de progression *peu*. Autre variation, à l'exception du critère *pertinence et richesse de*

l'argumentation, les attributs de tous les critères sont remplacés par *exemplaire* ou *exceptionnel* au niveau A+. Quant au critère *pertinence et richesse de l'argumentation*, la *pertinence* recherchée à tous les niveaux de progression est remplacée par la *richesse*.

Tableau 5

Critères présents lors du processus d'évaluation de Denise

Attributs	Objets	Présence
Pertinence et fréquence des interventions aux discussions sur les lectures	E
Précision du résumé des idées de l'auteur	E, P
Enchaînement et constance des idées énoncées	E, P
Pertinence et richesse de l'argumentation	E, P
Participation aux ateliers et aux activités du cours	E
Degré d'engagement dans son rapport au savoir didactique	E
Rigueur dans l'utilisation de la langue	S, E, P
Ponctualité de la remise des travaux	S
Respect du format du travail	S

Note : S = Syllabus; E = Échelle descriptive; P = Protocole oral

L'étude des critères des quatre participants a révélé à la fois des variations et des ressemblances concernant les critères d'évaluation. Par exemple, malgré l'utilisation commune d'échelles descriptives, celles-ci comportent plusieurs différences au sujet des informations présentées. Plusieurs variations identifiées dans les critères peuvent être attribuées en partie à l'utilisation d'échelles descriptives, telles que l'inconstance dans le choix des noms des critères ou dans les mots de progression, ainsi que le regroupement de critères dans une seule rangée. D'autre part, le nombre de critères recensés varie grandement d'un participant à l'autre. Calculer le nombre de critères utilisés s'avère cependant plus complexe qu'il ne le semble

préalablement, en raison des regroupements d'attributs et d'objets d'évaluation. Concernant les critères recensés, des similarités ont tout de même été observées. En comparant les attributs des critères et en faisant une approximation de correspondance quant aux objets d'évaluation, tout en démontrant un peu de latitude par rapport à l'utilisation de synonymes et de critères propres aux travaux et aux didactiques, de grandes ressemblances surgissent. De ce fait, le tableau 6 présente sept critères mentionnés par un minimum de trois participants. Les répercussions de ces observations sont discutées dans le chapitre suivant.

Tableau 6

Critères présents chez trois participants ou plus

Attributs	Objets	Présence
Pertinence d'éléments spécifiques dans le travail	A, B, C, D
Cohérence des diverses sections du travail	A, B, C, D
Profondeur du sujet traité	A, B, C
Précision et/ou clarté des idées	A, B, C, D
Rigueur dans l'utilisation de la langue	A, B, C, D
Originalité, créativité et/ou innovation du travail	A, B, C
Respect du format du travail	A, C, D

Note : A = Antoine, B = Béatrice, C = Charles, D = Denise

L'identification des critères d'évaluation utilisés par les participants permet maintenant de répondre à la deuxième question. La définition présentée dans l'annexe 6 concernant le caractère explicite d'un critère ne nécessite que la présence de l'attribut et de l'objet lors de l'étape de la préparation de l'évaluation. En ce sens, un seul critère complètement implicite a été trouvé ainsi qu'un seul attribut implicite. Charles a été le seul à mentionner un critère lors du protocole oral qui n'était ni présent dans son échelle descriptive ni dans le syllabus : *le respect du format du travail*. Au sujet de l'attribut implicite, c'est lors du protocole oral qu'Antoine a

mentionné rechercher la *profondeur de la mise en application des éléments pratiques* et de la *mise en application des éléments théoriques*.

Un deuxième critère implicite a été trouvé grâce au protocole oral d'Antoine. Cependant, ce critère n'a pas été inclus dans le tableau 2, puisqu'il n'est pas dirigé vers le travail sélectionné, mais plutôt vers un critère d'évaluation mentionné précédemment. Antoine ajuste son jugement en lien au critère *originalité et créativité du travail* en se basant sur l'*effort fourni par l'étudiant*. L'utilisation de ce critère semble donc être dépendante de la performance de l'étudiant à un autre critère.

La grille d'identification et d'analyse des critères d'évaluation (Annexe 6), bien que suffisamment précise afin de répondre aux deux premières questions de recherche, n'a pas permis d'identifier les critères qui ne traitent pas du rendement scolaire. En conséquence, la définition du rendement scolaire a dû être revue. Le rendement scolaire est présenté par Pilcher (1994) comme étant l'acquisition de connaissances ou d'habiletés, ce qui constituait la définition de départ. Par contre, Legendre (1993) et Pilcher présentent aussi le rendement comme étant le degré de réalisation des objectifs établis. De plus, d'évaluer uniquement le rendement scolaire lié aux objectifs établis permet à tous d'interpréter facilement la note (Brookhart, 1993). Dans ce sens, plusieurs connaissances ou habiletés sont utiles dans une panoplie de situations, mais la notation devrait se limiter à celles qui sont nécessaires à la réalisation des objectifs. En considérant ces nouvelles informations, le rendement scolaire est défini comme l'acquisition de connaissances ou d'habiletés qui sont nécessaires à la réalisation des objectifs du cours. Ce qui est considéré comme étant du rendement scolaire dépend donc du cours.

Il a donc été nécessaire de déterminer les objectifs établis par les participants afin de les comparer aux critères d'évaluation recensés. Antoine et Denise ont inclus une liste de

compétences visées par leurs cours, tandis que Béatrice et Charles ont mentionné des objectifs pour le travail ou le cours. Dans un souci de conserver l’anonymat des participants, les objectifs et compétences précis visés dans les divers cours ne sont pas présentés ici. Quels qu’ils soient, tous les participants cherchent à développer la compétence en enseignement des futurs enseignants, particulièrement en lien avec une discipline spécifique. Une fois les compétences identifiées, une concordance entre celles-ci et les critères possiblement externes au rendement scolaire était suffisante pour considérer que le critère évalue le rendement scolaire. Plusieurs exemples sont présents dans les paragraphes qui suivent.

Selon la définition du rendement scolaire présentée, *la qualité de la langue écrite* ne semble pas, à première vue, être nécessaire à l’évaluation de la capacité d’enseigner une didactique à l’élémentaire. Cependant, Antoine l’a incluse dans ses compétences visées, ce qui aide à justifier sa présence en tant que critère d’évaluation. *Le respect du format du travail* et *la ponctualité de la remise des travaux* peuvent être des qualités désirées chez les futurs enseignants, mais ne sont pas en lien avec les compétences visées d’Antoine. Concernant le critère d’évaluation identifié lors du protocole oral, *l’effort de l’étudiant*, il ne renseigne pas sur le degré de réalisation des objectifs et fausse même ces dernières informations en modifiant la note censée représenter le degré de compétence réel.

Le cas de Béatrice est plus complexe puisque peu d’informations sont connues concernant les objectifs du cours. Au premier coup d’œil, plusieurs critères semblent ne pas être en lien avec le rendement scolaire, puisqu’ils réfèrent à la capacité des étudiants de faire une présentation ou à la qualité de la langue écrite. Cependant, Béatrice a mentionné lors de l’entrevue que la capacité de présenter des informations, et ce, dans un langage adéquat, était une compétence nécessaire à tout enseignant. De plus, certaines parties du syllabus n’ont pas

été remises lors de la collecte de documents puisqu'elles ne présentaient pas un lien direct avec les critères d'évaluation du travail. Tout de même, le critère *envergure de la compréhension* demeure injustifié quant aux objectifs connus. Ce critère vise comme objet d'évaluation la compréhension de l'étudiant plutôt qu'une composante du travail. Tout comme l'effort évalué par Antoine, la compréhension ne renseigne pas sur le degré de réalisation des objectifs. De plus, ce critère semble vouloir évaluer à deux reprises les mêmes informations puisqu'une mauvaise compréhension aura assurément une influence sur d'autres critères tels que la *profondeur du sujet traité* ou la *précision et clarté des idées formulées et transmises*.

Charles inclut lui aussi la *qualité de la langue* dans ses critères d'évaluation et mentionne cette habileté comme une compétence visée par son cours. Parmi les autres critères, Charles mentionne l'*autonomie de la production du projet*. Ce critère n'est ni une habileté, ni une connaissance et s'apparente plutôt à une habitude de travail ou à un trait personnel. De plus, ce critère ne renseigne pas sur le degré de réalisation des objectifs du cours. Par ailleurs, Charles inclut, tout comme Antoine, le *respect du format du travail* dans sa liste de critères.

Denise inclut également la *qualité de la langue* parmi ses critères d'évaluation, mais sans mentionner celle-ci dans les compétences visées. Elle inclut aussi la *ponctualité de la remise des travaux* et le *respect du format du travail*, deux critères déjà identifiés chez d'autres participants. La liste de critères d'évaluation de Denise inclut aussi les critères : *la pertinence et la fréquence des interventions aux discussions sur les lectures, la participation aux ateliers et aux activités du cours* ainsi que *le degré d'engagement dans son rapport au savoir didactique*. Ces critères réfèrent tous de façon différente à la participation au cours, aux activités et à l'apprentissage. Une fois encore, ces critères ne sont pas en lien avec la réalisation des objectifs

établis et semblent être plus près d’habitudes de travail que d’habiletés ou de connaissances nécessaires à une compétence visée.

En posant un regard sur l’ensemble des participants, il est possible de remarquer que tous ont inclus d’une façon ou d’une autre la qualité de la langue dans leurs critères d’évaluation. Ce critère n’est généralement pas un critère du rendement scolaire, puisqu’il n’affecte pas directement la capacité de réussite des objectifs. Cependant, puisque tous les participants ont mentionné ce critère et que la majorité l’inclut dans leur description des compétences nécessaires au cours, ce critère semble important à la formation à l’enseignement. Dans ce contexte, ce critère est considéré comme pertinent au rendement scolaire. Cela dit, le tableau 7 présente les autres critères externes au rendement scolaire. Plus d’une mention du *respect du format du travail* a été identifiée, de même que pour la *punctualité de la remise des travaux*. Autre observation, tous les participants ont inclus au moins un critère qui est davantage en lien avec les traits personnels qu’avec les habiletés nécessaires à la réussite des objectifs établis.

Tableau 7

Critères externes au rendement scolaire

Attributs	Objets	Présence
Respect du format du travail	A, C, D
Ponctualité de la remise des travaux	A, D
Envergure de la compréhension	B
Autonomie de la production du projet	C
Pertinence et fréquence des interventions aux discussions sur les lectures	D
Participation aux ateliers et aux activités du cours	D
Degré d’engagement dans son rapport au savoir didactique	D

Note : A = Antoine, B = Béatrice, C = Charles, D = Denise

Tel que mentionné plus tôt, Paillé et Mucchielli (2008) proposent que les changements et nouvelles informations concernant les études qualitatives soient incorporés à l'intérieur même de celles-ci. En ce sens, une discussion avec le comité responsable de l'évaluation de cette thèse a mis à jour l'existence de documents ayant une influence sur les résultats présentés ici. Certains règlements, politiques et directives concernant l'évaluation à la faculté ainsi qu'à l'université où se déroule l'étude guident les pratiques d'évaluation des participants. Comme le propose l'analyse par questionnement analytique (Paillé et Mucchielli), ces nouvelles informations incitent à se questionner sur les pratiques d'évaluation exigées par ces documents.

L'absence de mention de politique facultaire par les participants durant l'ensemble de la collecte de données, ainsi qu'une recherche infructueuse sur le site Internet de la faculté en question avait fait croire en l'absence de politique facultaire. Au total, trois documents ont été identifiés avec l'aide du comité. Le premier document est une série de règlements scolaires publiée sur le site de l'université où se déroule l'étude. Ce document est important puisque le barème de notation officiel de l'université y est disponible et présente le caractère exceptionnel du A+. Le deuxième document est la politique d'évaluation de la faculté concernée. Ce document semble avoir une circulation limitée. La politique d'évaluation compte six consignes et mentionne par exemple que la présence au cours est obligatoire, mais ne doit pas être évaluée. Il y est aussi inscrit que les présentations orales doivent être accompagnées de preuves écrites, comme l'a demandé Béatrice. Par ailleurs, cette politique énonce de nouveau le caractère exceptionnel du A+ qui doit, par le fait même, être attribué de façon exceptionnelle.

Le troisième document est une note de service qui est distribuée par courriel aux professeurs qui donnent un cours à un semestre précis. Cette note discute de directives à inclure dans le syllabus. Ce document apporte des explications à plusieurs pratiques observées. Par

exemple, cette note de service demande de présenter les normes de la profession qui sont en lien avec les compétences visées par le cours, le nombre de travaux demandés et de préciser que l'assiduité et la ponctualité sont des qualités requises d'un enseignant. Conformément à ces directives, les professeurs devraient aussi présenter le caractère exceptionnel du A+ tel que mentionné dans la politique d'évaluation facultaire. Cependant, l'information la plus importante pour cette étude est l'exigence d'énoncer clairement les modalités d'évaluation ainsi que les critères d'évaluation à l'intérieur du syllabus. De plus, ces critères doivent obligatoirement inclure la qualité de la langue. Tous ces documents ne sont pas reproduits dans cette étude afin de ne pas révéler l'université où l'étude a eu lieu ainsi que pour des raisons de droit d'auteur.

L'analyse des données a donc permis de présenter des points communs et des différences dans les pratiques d'évaluation. Tous les participants utilisent des échelles descriptives comme outil d'évaluation. Toutefois, certaines échelles comptent deux fois plus de critères que d'autres. Tous utilisent certains critères communs, présentés dans le tableau 6, à la suite de quoi ils ajoutent des critères propres au travail ou à la didactique. La présence des critères lors du déroulement du processus d'évaluation varie aussi grandement d'un participant à l'autre. Cependant, l'échelle descriptive est dans tous les cas la source de critères la plus complète disponible aux étudiants. Les critères d'évaluation implicites sont peu nombreux, le seul critère entièrement implicite répertorié étant l'évaluation du *respect du format du travail* chez Charles. En comparaison, le nombre de critères qui ne traite pas du rendement scolaire est important puisque tous les participants en utilisent au moins un et certains en utilisent jusqu'à cinq, comme le démontre le tableau 7. Par ailleurs, les politiques et directives relatives à l'évaluation soulèvent autant de questions qu'elles n'apportent de réponses concernant les résultats. Le prochain chapitre tente de fournir une explication possible à ces résultats.

Chapitre 6 - Discussion

L'objectif de cette étude est d'acquérir une meilleure compréhension des critères d'évaluation utilisés par les professeurs universitaires dans le contexte de la formation à l'enseignement. Les résultats obtenus aux questions de recherche, à la suite de la collecte et de l'analyse des données, offrent un aperçu des critères utilisés par ce corpus. Ce chapitre fait d'abord un retour sur les questions de recherche. Cette section se poursuit par une discussion des implications pratiques et théoriques de cette étude. Finalement, ce chapitre se termine par une réflexion concernant les limites de l'étude, les pratiques mises en place afin de minimiser leur impact et quelques recommandations pour les études futures.

Retour sur les questions de recherche

L'identification des critères d'évaluation utilisés par certains professeurs à la formation à l'enseignement a fait surgir plusieurs informations d'intérêt. De façon générale, similairement à plusieurs études recensées (Baume et al., 2004; Hand et Clewes, 2000; Tillema et Smith, 2007), les critères d'évaluation des participants ont plusieurs ressemblances notables ainsi que des variations importantes. Dans un premier temps, lors de la communication des critères de l'évaluation, tous les participants présentent un syllabus qui comporte des informations sur les travaux demandés et les échelles descriptives utilisées par la suite (Case, 2007; Rust et al. 2003). Cette pratique est obligatoire selon les directives facultaires. Antoine et Béatrice mentionnent l'utilisation de copies types afin de mieux communiquer leurs critères (Sadler, 2002; Sambell et al., 2002). Charles, quant à lui, offre aux étudiants la possibilité de soumettre leur portfolio à des fins d'évaluation formative dans le but d'ajuster l'interprétation des critères (Simon et Forgette-Giroux, 2001). Lors de l'appréciation, mis à part le critère implicite employé par Charles, les participants se sont limités aux critères d'évaluation présentés

explicitement aux étudiants (McMillan, 2004; Ministère de l'Éducation de l'Ontario, 2000; Roegiers, 2004; Sheaford et Sheaford, 2001; Stokking et al., 2004; Tillema et Smith, 2007). Finalement, concernant la rétroaction, Antoine, Béatrice et Charles ont offert aux étudiants une rétroaction en lien avec tous ou certains critères d'évaluation (Weaver, 2006; Woolf, 2004).

Tous les participants ont regroupé à un certain moment plus d'un objet d'évaluation dans une seule rangée de leurs échelles descriptives. Ce regroupement, qui a pour effet d'augmenter le nombre de critères réellement présents lors de l'appréciation, nuit à la clarté de la présentation des critères. En fait, Antoine présente neuf critères dans son échelle descriptive, mais à la suite de l'analyse, le compte augmente à 15 en raison d'objets regroupés. Béatrice présente six regroupements, ce qui se traduit par 19 critères d'évaluation, dont certains pourraient potentiellement être à nouveau rassemblés. Le même commentaire s'applique pour Charles : 11 critères d'évaluation sont présents à la suite de l'analyse, mais les trois critères ayant la précision et la clarté comme attributs demeurent très similaires. Finalement, neuf critères ont été trouvés lors de l'analyse des documents de Denise. Antoine, Charles et Denise présentent donc un nombre de critères semblable à ceux observés par Webster et al. (2000) dans les établissements étudiés. Cependant, comme le mentionnent Laurier et al. (2005) concernant le nombre de critères employés, « Il faut éviter de les multiplier, car il y a des limites à ce qu'un évaluateur peut observer adéquatement » (p.117). Le nombre de critères optimal est discuté plus en détail dans une section future en lien avec le concept d'interdépendance des critères.

Il est important de considérer que, cette étude étant exploratoire, les comparaisons possibles avec les études recensées sont limitées par leurs contextes disparates. La recension des écrits mentionne des études proposant des critères d'évaluation utilisés par des professeurs

universitaires. Cependant, il est difficile de savoir si les compétences nécessaires pour un baccalauréat en sciences sociales ou en droit (Webster et al., 2000), en histoire, en gestion (Woolf, 2004) ou en commerce (Hand et Clewes, 2000) sont les mêmes que celles requises à la formation à l'enseignement. Le portrait est similaire quant à la variation des critères. L'étude de Tillema et Smith (2007) est la seule qui se situe dans le contexte de la formation à l'enseignement, mais les participants y ont apprécié un portfolio anonyme, sans aucun critère de référence. Malgré cela, le tableau 6 reprend les sept critères les plus communs qui présentent des similitudes avec certaines listes de critères recensées. En particulier, l'étude de Simon et Forgette-Giroux (2001), qui s'inspire des critères d'évaluation de revues scientifiques, rapporte, elle aussi, la *pertinence*, la *cohérence*, la *profondeur*, la *précision*, la *clarté*, et la *rigueur* comme attributs importants. Deux critères présents dans le tableau 6 ne sont pas mentionnés par Simon et Forgette-Giroux, soient l'*originalité du travail* et le *respect du format du travail*. Ces deux critères sont souvent tenus pour acquis dans la rédaction d'articles scientifiques. Par contre, les critères présents dans le tableau 6 réfèrent généralement à l'écriture d'un travail précis, plutôt qu'à une compétence unique, ce qui est plus semblable à l'étude de Webster et al. Il est possible, en raison de la ressemblance entre les compétences visées dans les divers cours, que ces critères réfèrent à des compétences transversales de la formation à l'enseignement (Forgette-Giroux et Simon, 2004). Les raisons précises expliquant ces ressemblances ne peuvent être identifiées avec le devis méthodologique employé par cette étude. Cependant, les influences possibles dans la création ou l'adaptation des échelles descriptives sont traitées dans les implications pratiques et théoriques.

D'un autre point de vue, les différences recensées dans les critères d'évaluation des participants sont majoritairement attribuables aux didactiques étudiées et aux travaux

demandés. Par exemple, Béatrice inclut plusieurs critères qui réfèrent à la présentation orale demandée, une composante qui n'est pas reprise par les autres participants. De son côté, Denise inclut plusieurs critères en lien avec la participation au cours, par exemple la *pertinence et la fréquence des interventions aux discussions sur les lectures*. Charles a par contre opté pour une échelle descriptive plus générique (Forgette-Giroux et Simon, 2004), possiblement liée à l'utilisation d'un portfolio qui regroupe naturellement plusieurs travaux diversifiés. L'utilisation de critères propres au cours, aux travaux demandés ou aux compétences visées est une explication possible des grandes variations systématiquement identifiées lors d'études des critères d'évaluation (Baume et al., 2004; Hand et Clewes, 2000; Tillema et Smith, 2007).

L'identification des critères lors de la première question a par la suite permis de répondre à la deuxième question, *dans quelle mesure ces critères sont-ils communiqués explicitement?* Un seul critère implicite a été identifié dans les données des quatre participants, le critère *respect du format du travail* évalué par Charles. Ce résultat est contraire aux observations de Tillema et Smith (2007), la seule étude recensée qui étudie les critères d'évaluation des professeurs à la formation à l'enseignement, qui a identifié peu de critères explicites. Cette différence est attribuable à l'utilisation d'un portfolio hors du contexte d'un processus d'évaluation par ces auteurs et à l'obligation de présenter explicitement les critères d'évaluation à la faculté où s'est déroulée l'étude. Par ailleurs, la présence de ce critère ne semble pas être liée à un cas frontière (Cross et al., 1993), puisque le portfolio évalué par Charles est qualifié d'exemplaire. Outre ce critère, Antoine révèle un attribut implicite à un critère présent dans le syllabus et l'échelle descriptive. Le caractère implicite de ce critère semble cependant lié à un manque de clarté dans son écriture.

D'autres parts, Antoine mentionne l'effort lors du protocole oral, servant à modifier le jugement émis concernant le critère *originalité et créativité du travail*, déjà présent dans son échelle descriptive. L'utilisation de l'effort comme critère d'évaluation externe au rendement scolaire n'est pas une pratique récente (Brookhart, 1994; Stiggins et al., 1989; McMillan, 2001), sans toutefois être commun aux autres cas étudiés. Cet ajout semble être utilisé afin d'ajuster la note de façon à minimiser l'impact de la performance sur le critère auquel il réfère. La faiblesse de la performance générale de l'étudiant offre une explication possible. Brookhart (1993) suggère que les étudiants moyens ou forts soient notés selon leur performance tandis que les étudiants sous la moyenne reçoivent un traitement plus clément si ce dernier peut être justifié (p.140). La force des étudiants qui ont produit les travaux utilisés lors des appréciations n'a pas fait partie de la collecte de données. Il est cependant possible de conclure que Charles a choisi un étudiant très fort et que Béatrice a choisi des étudiants moyens. L'impact des caractéristiques particulières des étudiants est discuté dans les implications pratiques et théoriques de cette étude.

Par ailleurs, certaines pratiques nuisent à la clarté de la présentation des critères (Case, 2007; Rust et al. 2003). Par exemple, lors de la présentation du travail, Antoine, Charles et Denise ont tous les trois écrit des critères soit uniquement dans le syllabus ou dans leurs échelles descriptives. Cette séparation n'est pas clairement justifiée (Stokking et al., 2004) et les critères *respect du format du travail et ponctualité de la remise des travaux* ne se retrouvent à aucune reprise dans les échelles descriptives des participants. Une explication possible est l'idée proposée par Roegiers (2004) de critères d'exclusion, c'est-à-dire « des critères tels que, à eux seuls, ils peuvent entraîner la non-maîtrise d'une compétence, et dès lors amener à prononcer l'échec de l'élève » (p.195). En ce sens, ces critères entreraient en jeu avant l'utilisation de la

grille. Les données recueillies dans le cadre de cette étude ne permettent toutefois pas de confirmer ou d'infirmier cette hypothèse.

D'autre part, la variation du vocabulaire dans le domaine des évaluations est un thème récurrent dans les écrits (Higgins, 2000; Speck, 1998; Webster et al., 2000; Williams, 2005; Woolf, 2004). Les critères qui sont présentés dans le syllabus et dans l'échelle descriptive utilisent parfois des synonymes plutôt qu'un vocabulaire constant. C'est particulièrement le cas d'Antoine. La variation de vocabulaire est aussi présente dans les niveaux de progression des échelles descriptives des participants, souvent par l'ajout d'attributs ou d'objets d'évaluation. Plus précisément, le caractère exceptionnel du A+ est mentionné dans le syllabus d'Antoine et Denise et des critères ayant des exigences différentes à ce niveau sont utilisés par tous les participants. Bien que les règlements, politiques et directives universitaires et facultaires mentionnent unanimement l'importance du caractère exceptionnel du A+, ceci ne se traduit pas nécessairement par l'ajout d'attributs ou d'objets dans ce niveau de progression. Tierney et Simon (2004) prônent plutôt l'utilisation d'attributs constants dans les niveaux de progression d'une échelle descriptive. Par ailleurs, Shepard (2000) mentionne à ce sujet que le constructivisme favorise une présentation des critères d'excellence si transparente que les étudiants peuvent évaluer leurs travaux comme le feraient leurs professeurs (p.11).

Tel que mentionné au chapitre précédent, un retour à la méthodologie a été nécessaire afin de répondre à la question *sur quels objets d'évaluation ces critères portent-ils si autres que le rendement scolaire?* La définition du rendement scolaire adoptée au chapitre 5 se résume à *l'acquisition de connaissances et d'habiletés nécessaires à la réalisation des objectifs établis.* D'évaluer uniquement le rendement scolaire clarifie l'objectif de l'étudiant, les attentes des enseignants et la signification de la note (Stiggins et al., 1989, p.6). Toutefois, l'évaluation de

compétences complique la définition du rendement scolaire, puisque le développement de compétences complexes implique plus d'habiletés et de connaissances que la réalisation d'objectifs précis. En ce sens, « la compétence se définit comme un savoir-faire qui mobilise un certain nombre de ressources pour faire face à un ensemble de situations » (Laurier et al., 2005, p.10).

Une des difficultés principales de définir le rendement scolaire est que celui-ci dépend du cours où a lieu l'évaluation (Gerard, 2008). Les compétences sélectionnées par les participants doivent être en lien avec l'objectif du cours afin qu'il soit facile de savoir, de façon générale, quelles habiletés et connaissances sont incluses dans la note (Brookhart, 1993). D'ailleurs, afin d'améliorer la validité de la notation, Greville (2009) propose que les critères soient en lien avec les compétences visées par les cours, lesquelles seraient également en lien avec les apprentissages visés par les programmes. En ce sens, les politiques et directives facultaires sont un premier pas vers des évaluations plus uniformes et des compétences mieux définies puisque certaines valeurs et compétences sont valorisées pour le programme entier.

D'autre part, bien que la concordance des critères au rendement scolaire soit jugée selon les compétences établies par les participants ou par la faculté, la question de l'interprétation de la note par les différents acteurs demeure. Il est important de prendre en considération qu'a posteriori, le rendement scolaire est précisément ce qui est inclus dans la note. Cette façon de concevoir le rendement scolaire incite à remettre en doute ce que l'on y inclut a priori. À ce sujet, Brookhart (1993) réitère l'importance d'inclure uniquement le rendement scolaire dans la note, afin qu'il soit clair que celle-ci représente les capacités de l'étudiant.

Plusieurs des critères externes au rendement scolaire semblent se référer à des habitudes de travail, entre autres, la *ponctualité de la remise des travaux*, l'*autonomie de la production du*

projet, la pertinence et la fréquence des interventions aux discussions sur les lectures et la participation aux ateliers et aux activités du cours. Un regard aux bulletins utilisés en Ontario révèle que des habiletés semblables sont évaluées du primaire au secondaire. On y retrouve entre autres *l'utilisation du français oral, la remise des travaux et des devoirs, l'autonomie au travail et la participation en classe* (Ministère de l'Éducation et de la Formation, 1998; MÉO, 1999). Cependant, il est important de préciser que ces habiletés sont évaluées de façon externe au rendement aux didactiques, c'est-à-dire que le rendement scolaire et ces habiletés reçoivent chacun une note distincte. Cette idée n'est pas récente puisque déjà en 1994, Brookhart proposait d'inclure une note représentant uniquement le rendement scolaire et une autre qui représenterait l'effort et les habitudes de travail de l'élève. Ces critères peuvent donc être justifiés. Le problème réside dans le fait qu'une seule note est attribuée à l'université où se déroule l'étude, ce qui ne permet pas de communiquer clairement les habiletés considérées. Par exemple, les directives facultaires exigent que l'évaluation de la langue soit incluse dans la notation. Cependant, comment un futur employeur peut-il savoir si la note d'un enseignant représente sa compétence en enseignement ou la qualité de son français écrit? Dans ce contexte, la note devrait donc représenter uniquement le rendement scolaire (Brookhart, 1993).

Dans un autre ordre d'idées, les critères *autonomie de la production du projet, envergure de la compréhension et engagement dans son rapport au savoir didactique* sont eux aussi semblables entre eux. Ces critères tentent d'évaluer directement des concepts non mesurables. Le désir des professeurs d'évaluer les niveaux cognitifs supérieurs et autres concepts abstraits est bien connu (Condon et Kelly-Riley, 2004; McMillan, 2001; Sheaford et Sheaford, 2001; Woolf, 2004). Par contre, cette volonté ne se traduit pas par une évaluation directe de processus cognitifs, tel que le suggèrent ces critères. Laurier et al. (2005) expriment bien cette distinction :

« ... l'apprentissage que l'on veut évaluer est un changement interne, invisible, qui échappe à l'observation directe » (p.13) et « En ce sens, les apprentissages ne sont jamais observables ni mesurables. Seules leurs manifestations le sont » (p.7). Ce sont donc des occurrences précises de construits complexes qu'il est possible d'évaluer (Crocker et Algina, 1986). Ceci ne signifie pas que l'évaluation de niveaux cognitifs supérieurs n'est pas possible. En réalité, certains outils sont disponibles afin de faciliter l'évaluation de tels construits, comme la taxonomie des objectifs cognitifs de Bloom par exemple (Laveault et Grégoire, 2002).

Par ailleurs, le critère *créativité du travail*, bien qu'il soit présenté explicitement et qu'il renseigne sur le rendement scolaire des étudiants, est tout de même problématique. Walker et Gleave (2008) ont étudié la compréhension d'étudiants de ce terme et ont observé que malgré une familiarité avec le concept, les étudiants comprennent mal ce critère en plus d'avoir de la difficulté à l'appliquer dans leurs travaux. Les auteurs proposent qu'en raison des problèmes encourus avec ce critère, il soit préférable de ne pas l'utiliser. Fait intéressant, c'est en lien avec ce critère, potentiellement parce qu'il n'avait pas été compris, qu'Antoine évalue l'effort de l'étudiant plutôt que *l'originalité et la créativité du travail*. Certains auteurs expriment un doute relativement à la capacité d'évaluer la créativité à l'aide d'échelles descriptives (Chapman et Inman, 2009; Cooper et Gargan, 2009; Hand et Clewes, 2000; Mertler, 2001; Norton, 2004). Par contre, en incitant la métacognition des étudiants à l'aide de discussions des critères et d'auto-évaluations favorisant la compréhension, l'inclusion de ce critère est justifiée (Forgette-Giroux et Simon, 2004).

Implications pratiques et théoriques

Cette étude avait comme objectif de poser un premier regard sur les critères d'évaluation utilisés par certains professeurs universitaires à la formation à l'enseignement. Cependant, en

plus des résultats discutés dans les sections précédentes, cette étude a relevé plusieurs pistes d'intérêt, et ce, pour des raisons variées. Tel que mentionné dans la méthodologie, une étude exploratoire ainsi que l'analyse par questionnaire analytique favorisent le développement de questions de recherche basées sur les résultats aux questions précédentes, afin de continuer d'approfondir les connaissances obtenues. Cette étude étant menée dans le cadre d'une maîtrise, elle ne peut donc pas poursuivre toutes les pistes identifiées. Dans une optique de faisabilité, les trois questions présentées dans l'univers interprétatif ont été répondues et cette section présente des thèmes identifiés parmi les résultats qu'il pourrait être pertinent de développer. En ce sens, les thèmes abordés ici ont la possibilité d'informer autant les professeurs que les étudiants sur les pratiques d'évaluation universitaires et de les conscientiser aux incohérences existant entre les recommandations des experts et les pratiques réelles.

D'entrée de jeu, il est pertinent de se questionner sur l'apport du contexte aux résultats. La formation à l'enseignement est décrite comme un programme à la fois court et exigeant, tant par les auteurs (Simbagoye, 2007) que par les participants. Pour cette raison, Charles mentionne accepter une négociation à la baisse du nombre de tâches demandées dans son cours. Malgré ces restrictions de temps, des travaux à réponse élaborée sont tout de même demandés aux étudiants. Les travaux sélectionnés par les participants avaient des exigences différentes et les critères reflétaient ces différences. Par exemple, Béatrice est la seule à inclure une présentation orale dans le travail demandé, un choix qu'elle justifie lors de l'entrevue en invoquant l'importance de cette compétence pour un enseignant. D'autre part, Denise a mentionné à la suite du protocole oral un manque d'uniformité dans les évaluations à la faculté. Afin de faciliter la compréhension des critères par les étudiants, l'université où se situe l'étude pourrait étudier la pertinence d'adopter une liste de critères institutionnels (Webster et al., 2000; Woolf,

2004). Les directives facultaires concernant l'évaluation sont un premier pas dans l'adoption de pratiques plus uniformes. Un groupe de discussion constitué de professeurs universitaires pourrait développer et piloter une échelle descriptive commune (Greville, 2009). Une telle échelle pourrait porter sur des compétences « transversales telles que le processus d'écriture, la résolution de problèmes, la création artistique, l'investigation scientifique, la recherche expérimentale et le processus d'évaluation » (Forgette-Giroux et Simon, 2004, p.243). Étant donnée la présence d'habiletés à développer, semblables à celles utilisées par le MEO, dans les critères d'évaluation des participants, il serait pertinent de considérer l'ajout d'une notation distincte de celle du rendement scolaire (Brookhart, 1994). Par exemple, des notes ou cotes séparées pourraient représenter la qualité de la langue, laquelle est valorisée dans les directives facultaires. Les résultats d'une mise à l'essai de ce genre pourraient par la suite être adoptés par d'autres facultés d'éducation.

Un thème peu adressé dans cette étude, mais qui a été soulevé par certains participants, est la formation des professeurs concernant l'évaluation des apprentissages. Comme l'explique Speck (1998), les professeurs universitaires ont généralement peu de formation concernant l'appréciation de travaux à l'aide de critères d'évaluation. À titre d'exemple, lors de l'entrevue, Béatrice a expliqué comprendre les réticences de certains étudiants face aux échelles descriptives, ayant elle-même eu à s'adapter à cet outil d'évaluation lors de son arrivée à la faculté en question. Un des buts de la formation à l'enseignement est de préparer les enseignants à évaluer, mais la formation en évaluation des enseignants est aussi souvent limitée (Allen, 2005; Simbagoye, 2007; Stokking et al., 2004). En ce sens, Antoine, Béatrice, Charles et Denise ont tous mentionné, comme raison d'utiliser des échelles descriptives, le désir de préparer adéquatement les étudiants. Cette pensée est conforme à plusieurs études (Allen, 2005;

Speck, 1998; Tremblay, 1998) qui mentionnent le réflexe chez les professeurs et les enseignants d'utiliser les méthodes d'évaluation avec lesquelles ils ont eux-mêmes été évalués. D'ailleurs, Reynolds-Keefer (2010), qui a procédé à l'interrogation de 45 étudiants à la formation à l'enseignement, confirme que l'utilisation d'échelles à l'université influence les pratiques des futurs enseignants. L'auteur a observé que plusieurs des étudiants adoptaient une vision plus positive des échelles descriptives suite à leur utilisation dans un cours. La crainte principale articulée concernant l'évaluation à l'aide d'échelles descriptives était liée au manque de formation des étudiants dans la création et l'utilisation de cet outil (Reynolds-Keefer, 2010, p.5). Il semble donc pertinent de mieux connaître le niveau de formation des professeurs afin d'offrir des formations supplémentaires basées sur les compétences manquantes. Une étude future pourrait évaluer les pratiques autant des professeurs que celles des enseignants dans la création et l'utilisation de diverses méthodes d'évaluation ou spécifiquement l'utilisation d'échelles descriptives.

Quoi qu'il en soit, l'utilisation d'échelles descriptives d'évaluation par les participants semble avoir eu un effet sur les résultats de cette étude. Dans un premier temps, la meilleure source de critères d'évaluation parmi les données recueillies est constamment l'échelle descriptive. Malgré les directives facultaires qui exigent une présentation claire des critères d'évaluation, l'utilisation d'échelles descriptives n'est pas une obligation. Par contre, certains critères, tels le *respect du format du travail*, sont systématiquement exclus des échelles descriptives des participants. Les raisons qui expliquent ces distinctions ne sont pas connues (Stokking et al., 2004), mais l'utilisation de critères d'exclusion est une explication possible (Roegiers, 2004). Par ailleurs, même si l'utilisation d'échelles descriptives par les participants aide à l'articulation ainsi qu'à une application uniforme des critères d'évaluation (Scallon, 2004;

Tierney et Simon, 2004), les ambiguïtés du vocabulaire et de l'expression écrite sont très présentes. Étonnamment, même les types de variations du vocabulaire sont nombreux et diversifiés entre les participants. Certaines ambiguïtés sont liées à l'utilisation de synonymes dans les mots de progression ou à des différences présentes entre l'échelle et le syllabus. L'ajout ou la différenciation des attributs, des objets ou des critères d'évaluation entiers entre les niveaux de progression est une autre source de variation du vocabulaire. Tierney et Simon (2004) discutent d'importantes répercussions des variations de vocabulaire dans les niveaux de progression. Selon elles, celles-ci affectent grandement l'utilité des échelles descriptives et la validité de leurs résultats.

La présence d'un certain nombre de critères comparables dans ces échelles soulève des questions. Notamment, quelles sont les influences externes sur les professeurs lors de la création ou la sélection de critères d'évaluation? Plusieurs éléments de réponse ont été rencontrés au cours de cette étude. Tous les participants ont répondu dans l'entrevue que leur référence en matière d'évaluation était des collègues qui donnent un cours similaire. Dans une étude sur le jugement professionnel en évaluation, Lafortune et Allal (2008) ont confirmé ce recours aux collègues. Ces consultations touchent le choix des critères d'évaluation ou l'élaboration d'échelles descriptives. Par ailleurs, Béatrice mentionne un partage d'échelles descriptives entre collègues. D'autre part, les échelles d'Antoine et de Charles réfèrent à un même document qui s'avère être un syllabus contenant des exemples d'échelles descriptives, mis à la disposition des professeurs. Malgré toutes ces discussions des influences avec les participants, aucun n'a mentionné les règlements, les politiques ou les directives identifiés par la suite. Une nouvelle étude pourrait examiner dans quelle mesure ces documents sont connus des professeurs universitaires. Une autre influence possible est le ministère de l'Éducation, qui est

une source particulièrement importante, du fait que cette étude se déroule à la formation à l'enseignement. Une fois encore, tous les participants ont mentionné l'utilisation d'échelles descriptives afin de préparer les étudiants à leur emploi futur. Plusieurs critères externes au rendement scolaire sont aussi retrouvés dans les habiletés à développer du ministère (Ministère de l'Éducation et de la Formation, 1998; MEO, 1999). Une nouvelle étude pourrait identifier, à l'aide de questionnaires, quelles sont les influences externes sur les professeurs universitaires dans la création ou la sélection de critères d'évaluation. Étant donné la simplicité de ce type d'instrument de collecte de données, des questionnaires pourraient être distribués auprès d'un grand nombre de professeurs, potentiellement dans plusieurs facultés ou universités. Une table de discussion invitant des professeurs universitaires à échanger sur ce thème aurait également le potentiel de faire ressortir des influences non connues. Par ailleurs, une étude pourrait explorer les liens présents entre les associations et ordres professionnels, le MEO et les diverses facultés d'éducation de la province. Une meilleure compréhension des rapports entre ces organismes pourrait être un premier pas dans l'étude des compétences essentielles à la profession.

L'évaluation à l'aide d'échelles descriptives est vue comme une répétition de jugements émis afin d'identifier la performance d'un étudiant à chacun des critères (Mertler, 2001; Simon et Forgette-Giroux, 2001; Woolf, 2004). Cependant, plusieurs études ont constaté que les professeurs ont tendance à porter un jugement holistique qu'ils adaptent ensuite à l'échelle descriptive (Baume et al., 2004; Sadler, 2009; Tillema et Smith, 2007). Cette pratique est clairement observée dans les protocoles oraux d'Antoine et de Béatrice et semble aussi être présente chez Charles et Denise. À ce sujet, Sadler propose que le recours à un jugement holistique ainsi qu'à un jugement analytique lors d'une seule évaluation puisse faire ressortir une inadéquation entre les jugements obtenus. Il a observé trois réponses à une telle situation :

les professeurs peuvent décider d'ignorer le jugement holistique lorsqu'il ne s'adapte pas aux critères, d'ajuster le jugement de chacun des critères afin de diminuer l'inadéquation ou bien, si un critère d'évaluation précis est passablement identifiable, d'ajouter ce dernier à la note de façon implicite. Il semble possible que certaines pratiques d'évaluation identifiées au cours de cette étude soient dues au désir de répondre à l'inadéquation présente entre les jugements holistiques et analytiques, telles que l'ajout : (a) de critères hors de l'échelle descriptive (b) de synonymes (c) d'attributs ou d'objets dans les niveaux de progression (d) d'un grand nombre de critères et (e) de critères externes au rendement scolaire. Toutes ces pratiques ont comme effet de rendre l'échelle descriptive plus inclusive en donnant une certaine latitude dans l'interprétation des critères d'évaluation et nuisent, par la même occasion, à la présentation claire de ces critères. Sadler vient à la conclusion que cette inadéquation est inhérente à l'utilisation d'échelles descriptives. Plusieurs indices recueillis au cours de la présente étude tendent à appuyer cette hypothèse. Une étude des jugements des professeurs universitaires pourrait utiliser des protocoles oraux afin de faire la comparaison de jugements analytique et holistique d'un même travail par divers participants. Les différences de notation, le choix d'une pondération ainsi que les critères d'évaluation mentionnés lors des appréciations seraient des informations complémentaires à cette étude.

Certains concepts présentés dans l'univers interprétatif sont ici remis en question. Comme le mentionnent Paillé et Mucchielli (2008), la conception d'une étude, particulièrement d'une étude exploratoire, évolue tout au long de son déroulement. Concernant le nombre d'objets par critère d'évaluation, le critère *la précision et la clarté de la terminologie, des concepts et des principes à l'étude* de Charles pousse à réflexion. Selon la définition de critère adoptée dans cette étude, ce critère doit être divisé tel que présenté dans le tableau 4. Cette

façon de procéder est justifiée à plus d'une reprise, par exemple lorsque Antoine tente d'inclure *la mise en application des éléments pratiques et théoriques* dans un seul critère. La différence entre ces exemples est l'interdépendance des critères. Le regroupement théorique d'objets d'évaluation a comme répercussion que le jugement d'un critère n'est pas entièrement indépendant du jugement d'un autre critère (Gerard, 2008; Roegiers, 2004; Sadler, 2009).

L'interdépendance de critères est possible soit par le regroupement de plusieurs critères en un seul, soit par l'énumération de critères distincts, qui s'avèrent par la suite hautement corrélés dans les résultats. Par exemple, dans le cas d'Antoine, il semble y avoir une interdépendance entre le critère explicite *originalité et créativité du travail* et le critère implicite *effort*. Donc, plutôt que de définir un nombre précis d'objets par critères, ne serait-il pas plus pertinent de se questionner sur l'interdépendance de ces objets? En ce sens, la réflexion peut aussi s'étendre sur le nombre d'attributs inclus dans un seul critère. Parallèlement, tel que proposé dans le paragraphe précédent, l'ajout d'attributs et d'objets est peut-être lié au nombre total de critères d'évaluation utilisés. En ce sens, Roegiers et Gerard proposent de limiter le nombre de critères à trois critères minimaux et un ou deux critères de perfectionnement où « Les critères minimaux ... déterminent la réussite, c'est-à-dire la maîtrise des compétences. Les critères de perfectionnement sont ... non strictement indispensables, qui situent les productions des élèves entre une production tout juste satisfaisante et une production excellente » (Roegiers, 2004, p.192). Cette façon de procéder est une alternative à l'ajout d'attributs ou d'objets afin de représenter le caractère exceptionnel du A+. D'ailleurs, un nombre restreint de critères minimise le risque d'interdépendance des critères et garantit une note plus juste (Gerard, 2008; Roegiers, 2004). Une étude pourrait d'ailleurs tenter d'éclairer le concept d'interdépendance des critères (Gerard, 2008; Roegiers, 2004; Sadler, 2009). Par exemple, une analyse factorielle

exploratoire pourrait faire ressortir un certain nombre de critères d'évaluation distincts ou peu corrélés.

De multiples possibilités de modification du devis adopté, afin d'élargir les informations recueillies concernant les critères d'évaluation, ont été identifiées lors du déroulement de cette étude. Dans un premier temps, une étude semblable qui se concentrerait sur un seul professeur, mais qui inclurait les travaux de tous les étudiants pourrait informer sur la variation dans l'articulation et l'application des critères en fonction de la force respective des étudiants dans le cours. La force des étudiants semble avoir joué un rôle dans cette étude, mais l'impact réel est inconnu. Une étude alternative serait d'observer l'ensemble d'un cours afin de voir si les mêmes critères d'évaluation sont utilisés pour tous les travaux et si des critères additionnels sont ajoutés lors du calcul de la note. Les échelles descriptives d'évaluation de tous les participants incluaient des habiletés semblables à celles mesurées par le MÉO dans les échelles descriptives, mais la considération de telles habiletés au niveau du cours reste à démontrer. Une autre étude comparable pourrait élargir les informations recherchées lors de la collecte de données. Afin d'incorporer toutes les utilisations potentielles de critères, l'observation d'évaluations formatives et sommatives, d'auto-évaluations et de discussions des critères permettrait d'identifier la présence ou non des différences recherchées. Bien sûr, une étude à plus grande échelle permettrait de dresser un inventaire complet des critères d'évaluation et de constater la fréquence d'utilisation de critères précis par faculté, par université ou à un niveau de scolarisation précis. Aussi, la présence de politiques et directives facultaires donne beaucoup d'intérêt à une étude des différences et ressemblances entre les pratiques de diverses facultés. Par ailleurs, l'exercice de création d'une liste de critères ou d'une échelle descriptive avec des

étudiants permettrait de comparer les compétences importantes pour ce groupe et leurs professeurs.

Limites

Bien que cette étude ait été planifiée et exécutée avec un souci de rigueur méthodologique, elle comporte certaines limites comme toute autre recherche. Une limite importante à considérer lors de toute étude et particulièrement lors d'études interprétatives est le biais du chercheur. Au début de l'étude, mon expérience d'étudiant ainsi que la recension des écrits suggéraient que les critères seraient peu structurés. Dès le début de la collecte de données, l'observation de l'utilisation d'échelles descriptives par l'ensemble des participants a eu comme effet de déstabiliser mes préconceptions et de permettre une étude vraiment exploratoire. L'examen phénoménologique a aussi favorisé une ouverture à être instruit par les données avant d'entamer l'analyse et est recommandé pour toute recherche qualitative.

L'utilisation exceptionnelle d'échelles descriptives par l'ensemble des participants est aussi à considérer. L'usage d'échelles influence la communication et l'utilisation des critères ainsi que la rétroaction. Cependant, le choix des critères d'évaluation doit être fait par les participants, peu importe s'ils emploient une échelle descriptive ou non. Pour cette raison, la méthodologie d'analyse prévue a été appliquée plutôt que de développer une nouvelle méthodologie orientée vers l'analyse des échelles descriptives. D'ailleurs, l'identification des critères d'évaluation n'était pas toujours limitée aux échelles descriptives des participants.

L'envergure de cette étude a aussi posé certaines limites. Malgré la collecte d'un nombre important de données, il n'était pas possible de tout inclure ou de tout prévoir. Dans un premier temps, l'utilisation d'une entrevue n'a pas suffi à être informé par les participants de l'existence de documents très importants. Par ailleurs, la décision d'opter pour une

méthodologie inclusive des diverses pratiques d'évaluation des professeurs fait en sorte que le type de travaux à réponse élaborée ou la force des étudiants peuvent avoir une influence difficilement mesurable avec la méthodologie employée. En ce sens, le retour sur la définition du rendement scolaire fait en sorte que davantage d'importance aurait dû être donnée aux compétences visées lors de la collecte de données. Par ailleurs, aucune information n'a été demandée concernant les formules utilisées afin de générer une note à la suite des jugements portés sur les critères d'évaluation. La décision de ne pas demander ces informations a peut-être eu une incidence sur l'absence de mention des règlements, politiques et directives universitaires et facultaires de la part des participants. D'autre part, les professeurs ont pu donner des informations verbales lors du cours qui ne sont pas incluses dans les documents écrits. Des critères implicites pourraient s'appliquer à un sous-groupe d'étudiants, tel que les cas frontière, sans être utilisés lors de l'appréciation de tous les travaux. À ce sujet, il aurait été préférable de procéder à l'étude pilote de la collecte de données jusqu'à l'obtention de résultats finals afin de prendre en considération les informations additionnelles à recueillir auprès des autres participants.

Une autre limite de l'étude est due à la variation du vocabulaire propre aux évaluations, observée lors d'échanges avec les participants. La question du langage a eu un impact sur la compréhension des questions d'entrevues. Cependant, l'utilisation d'entrevues semi-structurées a fait en sorte que les thèmes recherchés étaient souvent adressés par les participants dans le cadre d'une autre question. D'autre part, les ambiguïtés du vocabulaire et de l'expression écrite ont aussi compliqué le choix des noms de critères tel que décrit dans la méthodologie, puisqu'un choix devait être fait entre les synonymes présents dans les documents et les transcriptions analysées. Ces choix étaient basés sur le vocabulaire le plus utilisé par les participants ainsi que

dans un but de conformité entre les participants, lorsque pertinent. De demander aux participants d'expliquer leur compréhension de leurs propres critères aurait possiblement été une façon de limiter cette ambiguïté dans le vocabulaire.

L'utilisation d'un protocole oral dans la collecte de données soulevait des craintes quant à la qualité des informations recueillies. Dans cette étude, le protocole oral a été utilisé dans un contexte simulé. Aitken et Mardegan (2000) expliquent que le recours à un contexte simulé rend plus difficile de juger la correspondance des résultats obtenus avec un contexte naturel. De plus, Charles a mentionné lors de cet exercice que l'appréciation dans un contexte simulé lui semblait artificielle. Cependant, la perspective cognitiviste suggère qu'un tel contexte de situation problème soit particulièrement pertinent dans l'étude des réflexions (Tardif, 1993, 1997). De plus, malgré la réticence de Charles, son protocole oral a permis d'identifier un critère implicite, un des buts principaux de cet instrument. L'utilisation d'exemples et d'exercices afin de préparer les participants s'est avérée une étape essentielle, qu'il aurait été bénéfique dans certains cas d'élaborer davantage et qui mérite de faire partie de toute utilisation de protocoles oraux.

Chapitre 7 – Conclusion

Cette étude avait comme objectif de développer les connaissances concernant les critères d'évaluation des professeurs universitaires qui enseignent à la formation à l'enseignement. Les écrits présentés tout au long de cette étude sont très hétéroclites puisque la littérature dédiée précisément aux évaluations des professeurs universitaires continue d'être rare. De ce fait, les résultats présentés ici permettront d'élargir cette littérature. Par ailleurs, les contextes spécifiques des études recensées complexifient la comparaison aux résultats. Tout de même, la recension a permis de développer une compréhension des concepts clés qui à son tour a permis l'étude des critères présentés ici.

Le contexte de la formation à l'enseignement a été sélectionné puisque le partage de connaissances concernant l'évaluation semble concerner autant les professeurs que les étudiants et potentiellement leurs futurs élèves. La méthodologie mise en place a permis en grande partie de recueillir les informations désirées ainsi que des informations additionnelles, même non anticipées par instants. L'utilisation d'un examen phénoménologique des données a fait ressortir un portrait de la formation à l'enseignement en plus d'instruire sur les participants de l'étude. L'utilisation d'entrevues semi-structurées a été bénéfique puisqu'elles ont facilité la communication malgré différentes conceptions des termes liés à l'évaluation. D'autre part, le protocole oral a donné accès à des informations uniques et a permis l'identification d'un critère d'évaluation implicite. Par contre, la méthodologie mise en place n'a pas permis d'identifier les politiques et directives facultaires ainsi que les règlements universitaires comme étant des documents essentiels à cette analyse. Mis à part cette omission, le devis méthodologique s'est avéré adapté à la collecte des données voulues.

Les résultats obtenus offrent un premier regard sur les critères employés dans le cadre d'un processus d'évaluation. En effet, un certain nombre de critères communs ont été identifiés, dont la *pertinence d'éléments spécifiques dans le travail*, la *cohérence des diverses sections du travail*, la *précision et/ou clarté des idées* et la *rigueur dans l'utilisation de la langue* qui sont présents dans les listes de critères des quatre participants. Ce dernier critère est exigé par les directives facultaires. Les divergences dans les critères employés sont souvent attribuables aux didactiques ou aux travaux précis demandés aux étudiants. Contrairement à ce qui a été observé dans la littérature, très peu de critères implicites ont été identifiés, ce qui est dû aux directives facultaires. Concernant les critères externes au rendement scolaire, la majorité des critères répertoriés présentaient de grandes ressemblances avec les habiletés à développer utilisées par le ministère d'Éducation de l'Ontario.

La collecte de données a révélé une utilisation répandue d'échelles descriptives chez les participants. L'utilisation de cet outil est justifiée par les professeurs par le désir de préparer les futurs enseignants à les utiliser, ce qui semble être fondé. La présence de cet outil peut avoir eu une incidence sur les résultats, en favorisant par exemple la présentation explicite des critères d'évaluation. Cependant, l'utilisation d'échelles descriptives n'est pas toujours adéquate, par exemple en raison des grandes variations dans le choix des attributs et objets d'évaluation. L'utilisation d'échelles descriptives par les professeurs universitaires mérite d'être mieux étudiée.

En fait, la nature exploratoire de cette étude, combinée avec l'analyse par questionnement analytique, a fait ressortir plusieurs pistes de recherches futures. Que ce soit au niveau de la formation des professeurs, de l'utilisation d'échelles descriptives ou bien de la modification du devis méthodologique afin d'acquérir des informations additionnelles, un

manque de connaissances est toujours présent entourant les évaluations des professeurs universitaires.

Plusieurs pratiques ont été mises en place tout au long de cette étude afin d'assurer la rigueur méthodologique. D'autre part, une recherche peut difficilement adresser toutes les failles possibles et c'est pourquoi une discussion des limites de cette étude est aussi présente. En terminant, cette étude espère avoir réussi à contribuer aux connaissances concernant les critères d'évaluation et à faire place à des questionnements futurs sur le sujet. Les professeurs autant que leurs étudiants gagnent à développer une meilleure compréhension des évaluations en contexte universitaire.

Références

- Aitken, L. M. et Marddegan, K. J. (2000). "Thinking Aloud": Data collection in the natural setting. *Western Journal of Nursing Research*, 22(7), 841-853.
- Allen, J. D., (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House*, 78(5), 218-223.
- Aviles, C. B. (2001). Grading with norm-referenced or criterion-referenced measurements: to curve or not to curve, that is the question. *Social Work Education*, 20(5), 603-608.
- Baume, D., Yorke, M. et Coffey, M. (2004). What is happening when we assess, and how can we use our understanding of this to improve assessment? *Assessment & Evaluation in higher Education*, 29(4), 451-477.
- Bloxham, S. et West, A. (2007). Learning to write in higher education: Students' perceptions of an intervention in developing understanding of assessment criteria. *Teaching in Higher Education*, 12(1), 77-89.
- Bressoux, P. et Pansu, P. (2003). *Quand les enseignants jugent leurs élèves*. Paris, Fr : Les presses universitaires de France.
- Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement*, 30(2), 123-142.
- Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education*, 7(4), 279-301.
- Case, S. (2007). Reconfiguring and realigning the assessment feedback processes for an undergraduate criminology degree. *Assessment & Evaluation in Higher Education*, 32(2), 285-299.
- Chapman, V. G. et Inman, M. D. (2009). A Conundrum: Rubrics or Creativity/Metacognitive Development? *Educational Horizons*, 87(3), 198-202.
- Commission d'amélioration de l'éducation (2000). *La voie de l'avenir : Rapport sur l'amélioration du rendement des élèves : Cinquième et dernier rapport*, Ontario, Imprimeur de la Reine pour l'Ontario, Disponible : www.edu.gov.on.ca.
- Condon, W. et Kelly-Riley, D. (2004). Assessing and teaching what we value: The relationship between college-level writing and critical thinking abilities. *Assessing Writing*, 9, 56-75.
- Cooper, B. S. et Gargan, A. (2009). Rubrics in Education : Old Term, New Meaning. *Phi Delta Kappan*, 91(1), 54-55.

- Crisp, B. R. (2007). Is it worth the effort? How feedback influences students' subsequent submission of assessable work, *Assessment & Evaluation in Higher Education*, 32(2), 571-581.
- Crocker, L. et Algina, J. (1986). *Introduction to classical and modern test theory*. New York : Holt, Rinehart and Winston.
- Cross, L., Frary, R. et Weber, L. J. (1993). College grading. *College Teaching*, 41(4), 143.
Disponible : Academic Search Premier database :
<https://proxy.bib.uottawa.ca/login?url=http://search.epnet.com/login.aspx?direct=true&db=aph&an=9706113839>.
- De Ketele, J-M. et Roegiers, X. (1993). *Méthodologie du recueil d'information : Fondement des méthodes d'observation, de questionnaires, d'interviews et d'étude de documents*. Bruxelles, Belgique : De Boeck.
- De Serres, L. (2001). Déroulement et apport de la technique du protocole oral chez le lecteur adulte. *Mesure et évaluation en éducation*, 24(1), 1-26.
- Dyrud, M. A. (2003). Preserving sanity by simplifying grading, *Business Communication Quarterly*, 66(1), 78-85.
- Ericsson, K. A. et Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87(3) 215-251.
- Ericsson, K. A. et Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. London : MIT Press.
- Forgette-Giroux, R. et Simon, M. (2004). Échelle descriptive d'appréciation du rendement de l'élève : Acquis et problématiques. *McGill Journal of Education*, 39(3), 241-259.
- Gaeddert, B., K. (2001). Improving graduate theological instruction: A qualitative description of professors' attitudes toward the use of classroom assessment techniques. *Dissertation Abstracts International*, 62-05A (UMI No. AAI3014247).
- Gerard, F.-M. (2008). *Évaluer des compétences : Guide pratique*. Bruxelles, Belgique : De Boeck.
- Goubeaud, K. et Yan, W. (2004). Teacher educators' teaching methods, assessment, and grading: A comparison of higher education faculty's instructional practices. *The Teacher Educator*, 40(1), 1-16.

- Green, K. H. et Emerson, A. (2007). A new framework for grading. *Assessment & Evaluation in Higher Education*, 32(4), 495-511.
- Greville, E. C. (2009). A Rose by Any Other Name : Grading and Assessment. *Assessment Update ; Progress, Trends, and Practices in Higher Education*. 21(5), 1-13.
- Guénette, D. (2007). Is feedback pedagogically correct? Research design issues in studies of feedback on writing. *Journal of Second Language Writing*, 16, 40-53.
- Hand, L. et Clewes, D. (2000). Marking the difference : an investigation of the criteria used for assessing undergraduate dissertation in business school. *Assessment & Evaluation in Higher Education*, 25(1), 5-21.
- Karsenti, T. et Savoie-Zajc, L. (2004). *La recherche en éducation : étapes et approches*. Sherbooke, Qc : Éditions du CRP.
- Hartmann, D.P. (1977). Considerations in the choice of interobserver reliability estimates. *Journal of Applied Behavior Analysis*, 10, 103-116.
- Higgins, R. (2000). 'Be more critical!': rethinking assessment feedback, paper presented at the *British Educational Research Association Conference*, Cardiff University, 7-10. September.
- House, A.E., House, B.J. et Campbell, M.B. (1981). Measures of interobserver agreement: Calculation formulas and distribution effects. *Journal of Behavioral Assessment*, 3(1), 37-57.
- Kelly, K. B. (2002). A comparative study of the assessment practices utilized by full-time and adjunct instructors in preservice teacher education courses. *Dissertation Abstracts International*, 63-04A (UMI No. AAI3050037).
- Lafortune, L. et Allal, L. (2008). *Jugement professionnel en évaluation : Pratiques enseignantes au Québec et à Genève*. Québec, Qc : Presses de l'Université du Québec.
- Laurier, M. D., Tousignant, R. et Morissette, D. (2005). *Les principes de la mesure et de l'évaluation des apprentissages*. Montréal, Qc : Gaëtan Morin Éditeur.
- Laveault, D. et Grégoire, J. (2002). *Introduction aux théories des tests : En psychologie et en science de l'éducation*. Bruxelles, Belgique : De Boeck.
- Laveault, D. (2005). Le jugement professionnel de l'enseignant : quel impact sur l'acte d'évaluer? *Mesure et évaluation en éducation*, 28(2), 93-114.
- Legendre, R. (1993). *Dictionnaire actuel de l'éducation* (2^e ed.). Montréal : Guérin.

- McKenna, B., W. (2003). Classroom assessment: What teachers already know and what they will need to benefit student learning. *Dissertation Abstracts International*, 64-10A (UMI No. AAI3107591).
- McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practice. *Educational Measurement: Issues & Practice*, 20(1), 20-32.
- McMillan, J. H. (2004). Chapter 3: Establishing high quality classroom assessments. Dans J. H. McMillan (Éd.), *Classroom Assessment: Principles and Practice for Effective Teaching* (3nd ed.). Needham Heights, MA: Allyn & Bacon.
- Ministère de l'éducation et de la formation. (1998). *Guide d'utilisation du bulletin scolaire de l'Ontario de la 1er à la 8 année*. Toronto, ON.: Ministère de l'éducation.
- Ministère de l'Éducation de l'Ontario. (1999). *Guide du bulletin scolaire de l'Ontario de la 9e à la 12e année*. Toronto, ON.: Ministère de l'éducation.
- Ministère de l'Éducation de l'Ontario. (2000). *Le curriculum de l'Ontario de la 9e à la 12e année : Planification des programmes et évaluation*. Toronto : Imprimeur de la Reine pour l'Ontario.
- Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation*, 7(25). Disponible : <http://pareonline.net/getvn.asp?v=7&n=25>.
- Moskal, B. M. (2000). An assessment model for the mathematics classroom. *Mathematics Teaching in the Middle School*. 6(3), 192-194.
- Neil, D. T., Wadley, D. A. et Phinn, S. R. (1999). A generic framework for assessment of undergraduate essays. *Journal of Geography in Higher Education*, 23(3), 303-325.
- Norton, L. (2004). Using assessment criteria as learning: A case study in psychology. *Assessment & Evaluation in Higher Education*, 29(6), 687-702.
- O'Donovan, B. Price, M. et Rust, C. (2004). Know what I mean? Enhancing student understanding of assessment standards and criteria. *Teaching in Higher Education*, 9(3), 325-335.
- Orsmond, P., Merry, S. et Reiling, K. (1997). A study in self assessment : tutor and students' perceptions of performance criteria. *Assessment & Evaluation in Higher Education*. 22(4), 357-368.

- Paillé, P. et Mucchielli, A. (2008). *L'analyse qualitative en sciences humaines et sociales* (2^e éditions). Paris : Armand Colin.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2^e éditions). Californie : SAGE Publication Inc.
- Pilcher, J. K. (1994). The value-driven meaning of grades. *Educational Assessment*, 2(1), 69-88.
- Popham, W. J. (2000). *Modern educational measurement: Practical guidelines for educational leaders* (3^e ed.). Needham Heights, MA: Allyn & Bacon.
- Poupart, J., Deslauriers, J.-P., Groulx, L.-H., Laperrière, A., Mayer R. et Pires, A. P. (1997). *La recherche qualitative : Enjeux épistémologiques et méthodologiques*. Boucherville, QC : Gaëtan Morin éditeur.
- Rees, C. et Sheard, C. (2004). The reliability of assessment criteria for undergraduate medical students' communication skills portfolios: the Nottingham experience. *Medical Education*, 38, 138-144.
- Reynolds-Keefer, L. (2010). Rubric-referenced assessment in teacher preparation : An opportunity to learn by using. *Practical Assessment, Research & Evaluation*, 15(8).
Disponible : <http://pareonline.net/getvn.asp?v=15&n=8>.
- Roegiers, X. (2004). *L'école et l'évaluation : Des situations pour évaluer les compétences des élèves*. Bruxelles, Belgique : De Boeck.
- Rust, C., O'Donovan, B. et Price, M. (2005). A social constructivist assessment process model: how the research literature shows us this could be best practice. *Assessment & Evaluation in Higher Education*, 30(3), 231-240.
- Rust, C., Price, M. et O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education*, 28(2), 147-164.
- Sadler, D. R. (2002). Ah!... So that's 'Quality'. Dans Schwatz, P. et Webb, G. (Eds.), *Assessment : Case studies, experience and practice from higher education* (pp. 130-136). London : Kogan Page.
- Sadler, D. R. (2003). *How criteria-based grading misses the point*. Presentation to the effective teaching and learning conference, Griffith University, Southbank campus.
- Sadler, R. D. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, 34(2), 159-179.

- Sambell, K., Miller, S. et Hodgson, S. (2002). Let's get the assessment to drive the learning. Dans Schwatz, P. et Webb, G. (Eds.), *Assessment: Case studies, experience and practice from higher education* (pp. 137-143). London : Kogan Page.
- Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétences*. Saint-Laurent, Québec : ERPI.
- Scallon, G. (2000). *L'évaluation des apprentissages: Vocabulaire : Appréciation authentique*. Consulté le 10 janvier 2011, sur le site de l'Université Laval, Département des fondements et pratiques en éducation, Disponible : http://www.fse.ulaval.ca/gerard.scallon/vocabulaire/vocabul_page.htm.
- Shanahan, J. E., Jr. (2001). Teachers and the process of assessing students in the classroom: A study of classroom assessment practices. *Dissertation Abstracts International*, 62-04A (UMI No. AAI3012572).
- Shay, S. (2005). The assessment of complex tasks: a double reading. *Studies in Higher Education*, 30(6), 663-679.
- Sheaford, B. W. et Sheaford, B. A. (2001). Grading. Dans D. D. Royce (Ed.), *Teaching tips for college and university instructors: A practical guide* (pp.183-211), Needham Heights, MA : Allyn & Bacon : A Pearson Education Company.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 1-14.
- Simbagoie, A. (2007). Chapitre 7 : La réforme curriculaire dans les écoles de langue française de l'Ontario. Dans L. Lafortune, M. Ettayebi et P. Jonnaert (Eds.), *Observer les réformes en éducation* (pp. 131-168). Québec : Presses de l'Université du Québec.
- Simon, M. et Forgette-Giroux, R. (2001). A rubric for scoring postsecondary academic skills. *Practical Assessment, Research, & Evaluation*, 7(18), Disponible : <http://pareonline.net/getvn.asp?v=7&n=18>.
- Speck, B. W. (1998). Unveiling some of the mystery of professional judgment in classroom assessment. Dans Anderson, R. S. et Speck, B. W. (Eds.), *Changing the way we grade student performance: Classroom assessment and the new learning paradigm. New Directions for Teaching and Learning : New Learning Paradigm*, 74, 17-31.
- Stiggins, R. J., Frisby, D. A. et Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement : Issues and Practice*, 8, 5-14.

- Stokking, K., Van der Schaaf, M., Jaspers, J. et Erkens, G. (2004). Teachers' assessment of students' research skills. *British Educational Research Journal*, 30(1), 93-116.
- Struyven, K., Dochy, F. et Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education*, 30(4), 325-341.
- Tardif, J. (1993). L'évaluation dans le paradigme constructivisme. Dans R. Hivon (Éd.). *L'évaluation des apprentissages*. Sherbrooke: Édition du CRP.
- Tardif, J. (1997). *Pour un enseignement stratégique : L'apport de la psychologie cognitive* (2^e ed.). Montréal, Québec : Les Éditions Logiques.
- Tierney, R et Simon, M. (2004). What's still wrong with rubrics: focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research & Evaluation*, 9(2). Disponible : <http://PAREonline.net/getvn.asp?v=9&n=2>.
- Tillema, H. et Smith, K. (2007). Portfolio appraisal : In search of criteria. *Teaching and Teacher Education*. 23, 442-456.
- Tremblay, M-C. (1998). Élaboration d'une grille d'auto-évaluation : instrument de réflexion sur son enseignement. *Mesure et évaluation en éducation*, 21(2), 37-63.
- van Someren, M. W., Barnard, Y. F. et Sanberg, J. A. C. (1994). *The think aloud method: A practical guide to modeling cognitive processes*. Academic Press, London.
- Walker, C. et Gleaves, A. (2009). An exploration of students' perceptions and understandings of creativity as an assessment criterion in undergraduate-level studies within higher education. *Irish Educational Studies*. 27(1), 41-54.
- Weiss, J. (1992). L'enseignant au « cœur froid » ou l'objectivité en éducation. *Mesure et évaluation en éducation*, 14(4), 19-31.
- Webster, F., Pepper, D. et Jenkins, A. (2000). Assessing the undergraduate dissertation. *Assessment & Evaluation in Higher Education*, 25(1), 71-80.
- Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco, CA: Jossey-Bass Publishers.
- Williams, K. (2005). Lecturer and first year student (mis)understanding of assessment task verbs: 'mind the gap'. *Teaching in Higher Education*, 10(2), 157-173.
- Weaver, M. R. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education*, 31(3), 379-394.

- Woolf, H. (2004). Assessment criteria: Reflections on current practices. *Assessment & Evaluation in Higher Education*, 29(4), 479-493.
- Wyatt-Smith, C. et Castleton, G. (2005). Examining how teachers judge student writing: an Australian case study, *Journal of Curriculum studies*, 37(2), 131-154.

Annexe 1 - Texte de recrutement

Madame/Monsieur,

Vous êtes invité à prendre part à une étude menée par Maxime Girouard, étudiant à la maîtrise ès arts en enseignement, apprentissage et évaluation. Le but de cette étude est d'acquérir une meilleure compréhension des critères auxquels les professeurs universitaires font référence tout au long du processus d'évaluation du travail d'un étudiant. Cette recherche est supervisée par Marielle Simon, Ph.D. et a reçu l'approbation du comité de thèse et du comité d'éthique.

Procédure : Si vous acceptez de participer à cette étude, vous serez appelé à contribuer à quatre reprises. Dans un premier temps, vous devrez fournir une copie de tous les documents qui ont été remis aux étudiants afin de les guider dans l'exécution d'un travail à réponse élaborée spécifique. Par la suite, vous devrez prendre part à une courte entrevue traitant de vos expériences de travail et de vos pratiques d'enseignement. Cette entrevue aura une durée d'environ 30 minutes et la rencontre pourra avoir lieu dans un local de votre choix ou bien dans un local réservé à cet effet. Ensuite, vous devrez participer à un protocole oral, qui consiste à verbaliser vos raisonnements à voix haute, lors de l'appréciation du travail à réponse élaborée de l'un de vos étudiants. Ce protocole vous sera préalablement expliqué plus en détails, à l'aide d'exemples et de problèmes préparatoires. Cette session durera environ 2 heures et une copie du verbatim produit suite à l'enregistrement vous sera remise afin d'être validée. Une fois encore, cette rencontre peut avoir lieu dans un local de votre choix ou bien dans un local réservé à cet effet. Finalement, une copie de la rétroaction inscrite sur le travail apprécié devra être remise à l'étudiant responsable de cette étude.

Bénéfice : En participant à cette étude, vous allez contribuer à l'avancement des connaissances dans le domaine des évaluations au niveau universitaire. Vous allez de plus avoir la chance de partager vos croyances et pratiques en matière d'évaluation.

Participation volontaire : La participation à cette étude se fait sur une base volontaire et vous avez le droit de cesser votre participation en tout temps et/ou de refuser de répondre à toute question sans subir de conséquences négatives.

Anonymat et confidentialité : Toute l'information obtenue à l'intérieur de l'étude demeurera confidentielle. Les données seront conservées de façon électronique avec des mots de passe limitant l'accès à l'étudiant et à sa superviseure. Tous les efforts seront faits afin de conserver l'anonymat des participants. Les données qui pourraient permettre d'identifier un participant seront amalgamées et des pseudonymes seront utilisés.

Contact : Toutes questions concernant cette étude peuvent être adressées à Maxime Girouard, l'étudiant responsable, ou à Marielle Simon. Pour tout renseignement sur les aspects éthiques de cette recherche, vous pouvez vous adresser au Responsable de l'éthique en recherche.

Étude effectuée par :
Maxime Girouard,
Étudiant à la maîtrise
Faculté d'éducation
Université d'Ottawa

Sous la supervision de :
Marielle Simon, Ph. D.
Professeure titulaire
Faculté d'Éducation
Université d'Ottawa

Annexe 2 - Formulaire de consentement

Madame/Monsieur,

Vous êtes invité à prendre part à une étude menée par Maxime Girouard, étudiant à la maîtrise ès arts en enseignement, apprentissage et évaluation. Le but de cette étude est d'acquérir une meilleure compréhension des critères auxquels les professeurs universitaires font référence tout au long du processus d'évaluation du travail d'un étudiant. Cette recherche est supervisée par Marielle Simon, Ph.D. et a reçu l'approbation du comité de thèse et du comité d'éthique.

Procédure : Si vous acceptez de participer à cette étude, vous serez appelé à contribuer à quatre reprises. Dans un premier temps, vous devrez fournir une copie de tous les documents qui ont été remis aux étudiants afin de les guider dans l'exécution d'un travail à réponse élaborée spécifique. Par la suite, vous devrez prendre part à une courte entrevue traitant de vos expériences de travail et de vos pratiques d'enseignement. Cette entrevue aura une durée d'environ 30 minutes et la rencontre pourra avoir lieu dans un local de votre choix ou bien dans un local réservé à cet effet. Ensuite, vous devrez participer à un protocole oral, qui consiste à verbaliser vos raisonnements à voix haute, lors de l'appréciation du travail à réponse élaborée de l'un de vos étudiants. Ce protocole vous sera préalablement expliqué plus en détails, à l'aide d'exemples et de problèmes préparatoires. Cette session durera environ 2 heures et une copie du verbatim produit suite à l'enregistrement vous sera remise afin d'être validée. Une fois encore, cette rencontre peut avoir lieu dans un local de votre choix ou bien dans un local réservé à cet effet. Finalement, une copie de la rétroaction inscrite sur le travail apprécié devra être remise à l'étudiant responsable de cette étude.

Bénéfice : En participant à cette étude, vous allez contribuer à l'avancement des connaissances dans le domaine des évaluations au niveau universitaire. Vous allez de plus avoir la chance de partager vos croyances et pratiques en matière d'évaluation.

Participation volontaire : La participation à cette étude se fait sur une base volontaire et vous avez le droit de cesser votre participation en tout temps et/ou de refuser de répondre à toute question sans subir de conséquences négatives.

Anonymat et confidentialité : Toute l'information obtenue à l'intérieur de l'étude demeurera confidentielle. Les données seront conservées de façon électronique avec des mots de passe limitant l'accès à l'étudiant et à sa superviseure. Tous les efforts seront faits afin de conserver l'anonymat des participants. Les données qui pourraient permettre d'identifier un participant seront amalgamées et des pseudonymes seront utilisés.

Contact : Toutes questions concernant cette étude peuvent être adressées à Maxime Girouard, l'étudiant responsable, ou à Marielle Simon. Pour tout renseignement sur les aspects éthiques de cette recherche, vous pouvez vous adresser au Responsable de l'éthique en recherche.

Si vous êtes intéressé à être mis au courant lors de la publication des résultats, veuillez inscrire votre intérêt sur ce formulaire ou en faire part à l'étudiant de façon verbale.

Veillez conserver une copie de ce document dans vos dossiers.

Signature : _____

Nom : _____

Date : _____

J'aimerais être informé lors de la publication de la thèse

Étude effectuée par :
Maxime Girouard,
Étudiant à la maîtrise
Faculté d'éducation
Université d'Ottawa

Sous la supervision de :
Marielle Simon, Ph. D.
Professeure titulaire
Faculté d'Éducation
Université d'Ottawa

Annexe 3 - Guide d'entrevue

Questions générales

1. Dans quel domaine avez-vous fait vos études ?
2. Avez-vous déjà suivi des formations en évaluation ?
3. Combien de cours avez-vous enseigné à la formation à l'enseignement ?

Questions sur l'évaluation

1. Quelles méthodes d'évaluation utilisez-vous le plus souvent ?
2. Comment décidez-vous ce que vous évaluez dans votre classe ?
3. Pouvez-vous me décrire votre processus d'évaluation pour cette évaluation ?
4. Comment choisissez-vous les critères d'évaluation ?
5. Comment présentez-vous les tâches et les critères d'évaluation aux étudiants ?

Annexe 4 - Instructions du protocole oral

Explications

Dans le but de vous familiariser avec le protocole oral, je vais d'abord vous montrer un exemple de problème et les résultats de deux étudiants afin de vous donner une idée de ce qui est recherché. Par la suite, un problème à résoudre vous sera présenté avec les instructions nécessaires pour vous y préparer. Cet exercice servira à vous pratiquer à exprimer vos pensées lors de réflexions. Un deuxième exercice est prévu si vous voulez un essai supplémentaire. Après ces exercices, s'il n'y a pas de questions, vous allez pouvoir débiter l'appréciation du travail que vous avez choisi.

Exercices préparatoires

Exercice 1 :

Déterminer la ligne suivante :

1
11
21
1211
111221
..... ?

Exercice 2 :

Mathieu, Anne et George ont eu un total de 23 points lors de leur test de mathématiques. Si George a 3 points de moins que Mathieu et que Anne a deux points de plus que George, combien de points ont-ils chacun ?

Instructions

C'est maintenant à votre tour. Lors de l'appréciation du travail que vous avez choisi, je vais vous demander de penser à voix haute. Je veux que vous me disiez tout ce qui vous vient à l'esprit, peu importe si cela vous semble important ou pertinent. Vous faites peut-être cet exercice lorsque vous êtes seul et tentez de résoudre un problème. Ne planifiez pas quoi dire ou parler après la pensée, mais laissez plutôt votre pensée parler comme si vous pensiez vraiment à haute voix. Ne tentez pas d'expliquer votre pensée à personne, prétendez plutôt que vous êtes seul. Ne parlez pas du problème mais solutionnez-le.

Annexe 5 - Exemples de protocoles oraux

Un père, une mère et leur fils sont âgés de 80 ans ensemble. Le père est deux fois plus vieux que son fils. La mère a le même âge que le père. Quel âge a le fils ? (van Someren et. al.,1994 : ma traduction)

Élève 1 :

- 1 : Un père, une mère et leur fils sont âgés de 80 ans ensemble
- 2 : Le père est deux fois plus vieux que son fils
- 3 : La mère est aussi vieille que le père
- 4 : Quel âge a le fils
- 5 : Ça a l'air compliqué
- 6 : Bon, commençons
- 7 : Je vais les nommer P, M et F
- 8 : P plus M plus F donne 80
- 9 : P est 2 fois F
- 10 : Et M est égal à P
- 11 : Qu'est-ce que ça nous donne ?
- 12 : Trois équations et trois inconnus
- 13 : Donc, F ...
- 14 : 2 fois P plus F donne 80
- 15 : Donc 4 fois F plus F donne 80
- 16 : Donc 5 fois F donne 80
- 17 : F a 16
- 18 : Oui, c'est possible
- 19 : Donc le père et la mère ont 80 moins 16
- 20 : 64
- 21 : Donc, ... 32

Élève 2 :

- 1 : Un père, une mère et leur fils sont âgés de 80 ans ensemble
- 2 : Comment est-ce possible ?
- 3 : Si un tel père a 30 ans et la mère aussi
- 4 : Alors le fils a 20 ans
- 5 : Non, c'est impossible
- 6 : Si tu as 30 ans tu ne peux pas avoir un fils de 20 ans
- 7 : Donc ils devraient être plus vieux
- 8 : Plus ou moins 35, environ
- 9 : Voyons voir
- 10 : Le père est deux fois plus vieux que son fils
- 11 : Donc s'il a 35 et le fils 17
- 12 : Non, c'est impossible
- 13 : 36 et 18
- 14 : Donc la mère a ...
- 15 : 36 plus 18 donne 54
- 16 : 26 ...

17 : Ça pourrait être possible
18 : Non, sinon elle aurait eu un enfant à l'âge de 9 ans
19 : Ah, non
20 : Non, le père devrait, la mère devrait être plus vieille
21 : Par exemple 30
22 : Mais alors je n'aurai pas 80
23 : 80 moins 30, 50
24 : Alors le père devrait être près de 35 et le fils 18
25 : Quelque chose du genre
26 : Voyons voir, où suis-je rendu ?
27 : Le père est deux fois ...
28 : La mère a le même âge que le père
29 : Oh la la !
30 : Ma mère, bien pas ma mère
31 : Mais ma mère avait 30 et mon père près de 35
32 : C'est impossible
33 : Si je mets les deux à 33
34 : Alors j'ai 66 ensemble
35 : Alors le fils est âgé de ... 24
36 : Non, c'est impossible
37 : Je ne comprends plus rien
38 : 66, ..., 80
39 : Non, attends, le fils a 14
40 : Presque, ... Les parents sont trop vieux
41 : 32, 32, 64, 16, Oui
42 : Le fils a 16 ans et les parents 32, ensemble 80.

Annexe 6 - Grille d'identification et d'analyse des critères

Définition : Un critère est une description explicite de facettes ou dimensions d'une performance, qui comporte un ou plusieurs attributs précisant un objet d'évaluation clairement articulé, utilisée afin d'apprécier le rendement scolaire d'un étudiant.

Définition minimale: Un critère est une description de facettes ou dimensions d'une performance utilisée afin d'apprécier le rendement d'un étudiant.

Attributs				
Définition	Exemples			
Un attribut est « une qualité propre ou particulière d'un critère de performance » (Forgette-Giroux et Simon, 2004, p.245) qui varie selon une échelle d'intensité, de fréquence ou de quantité.	Cohérence Efficience Pertinence Profondeur Clarté Efficacité Précision	Envergure Rigueur Exactitude Originalité Créativité Authenticité	Attrait Intérêt Utilité Organisation Méthodique Coopération	Collaboration Présentation Concision Logique Exhaustivité Structure

Objets d'évaluation		
Définitions	Exemples	
Un critère du rendement scolaire représente l'acquisition de connaissances et d'habiletés (Pilcher, 1994).	Travail Texte Présentation Sujet traité Choix de stratégie Affirmation	Méthode de travail Structure du texte Idée Information Liens avec la théorie Argumentation
Un critère externe au rendement scolaire ne renseigne pas sur la capacité de l'élève à répondre à la compétence visée.	Effort Motivation Participation Progrès démontré	Attitude envers le cours Habitudes de travail Traits personnels Résultats aux évaluations formatives

Critères implicites	
Définitions	Exemples
Un critère est implicite s'il n'est pas présenté de façon explicite lors de la planification, mais utilisé lors d'une étape future.	Lors de l'appréciation, le professeur commente le manque d'originalité du sujet traité sans que l'originalité du sujet traité ne soit identifiée en tant que critère lors de la planification.
Un attribut est implicite si seul l'objet d'évaluation est exprimé explicitement lors de la planification.	Un professeur inscrit évaluation des liens avec la théorie couverte dans un cours sans exprimer s'il recherche la pertinence, la clarté ou l'exhaustivité des liens.
Un objet d'évaluation est implicite si seul l'attribut est exprimé explicitement lors de la planification.	Un professeur mentionne la pertinence comme critère d'évaluation sans préciser si l'objet est le travail entier, l'argumentation, les sources utilisées ou les sous-titres choisis.