



Processus de coconstruction d'une grille critériée pour l'évaluation de productions écrites complexes à l'université

Elisabeth Martin, Catherine Lefrançois, Anne Guichard, Diane Tapp et
Louise Arsenault



Édition électronique

URL : <http://journals.openedition.org/ripes/1094>
ISSN : 2076-8427

Éditeur

Association internationale de pédagogie universitaire

Référence électronique

Elisabeth Martin, Catherine Lefrançois, Anne Guichard, Diane Tapp et Louise Arsenault, « Processus de coconstruction d'une grille critériée pour l'évaluation de productions écrites complexes à l'université », *Revue internationale de pédagogie de l'enseignement supérieur* [En ligne], 32-2 | 2016, mis en ligne le 20 juin 2016, consulté le 12 octobre 2018. URL : <http://journals.openedition.org/ripes/1094>

Ce document a été généré automatiquement le 12 octobre 2018.

Licence CC BY-NC-SA

Processus de coconstruction d'une grille critériée pour l'évaluation de productions écrites complexes à l'université

Elisabeth Martin, Catherine Lefrançois, Anne Guichard, Diane Tapp et Louise Arsenault

1. Introduction

- 1 Le champ de la mesure et de l'évaluation en contexte universitaire est en pleine effervescence, amenant les professeurs à revisiter leurs pratiques à l'égard de la détermination des objectifs d'apprentissage et de l'appréciation des productions étudiantes. Dans le contexte des études aux cycles supérieurs¹, l'évaluation des apprentissages présente des défis particuliers puisqu'elle repose davantage sur une appréciation subjective de l'aptitude à la réflexion critique, la capacité à mobiliser des compétences cognitives complexes et la maîtrise d'habiletés essentielles en recherche que sur une mesure des niveaux de connaissances acquises relativement à un objet, une discipline ou un champ d'études. Les écueils de l'évaluation se trouvent exacerbés, notamment pour les professeurs en début de carrière, souvent peu familiers avec la tâche de correction et les niveaux de performance attendus des étudiants à la maîtrise ou au doctorat dans les productions scientifiques qu'ils rédigent. Malgré les obstacles, les professeurs doivent néanmoins concevoir des grilles d'évaluation qui, à la fois, mesurent adéquatement les objectifs visés par les cours qu'ils offrent, mais permettent aussi d'apprécier le cheminement intellectuel des apprentis chercheurs qu'ils accompagnent.
- 2 Le projet relaté ici est né de la volonté de trois professeures en début de carrière de la Faculté des sciences infirmières (FSI) de l'Université Laval (Québec, Canada) qui, après avoir suivi une formation intensive en pédagogie universitaire, ont souhaité poursuivre

leurs réflexions et bonifier leurs pratiques d'évaluation des apprentissages. Un appui financier du Bureau de soutien à l'enseignement (BSE) de l'Université Laval a permis le développement d'un outil d'évaluation d'une production écrite complexe pour les programmes gradués de sciences infirmières et de santé communautaire de la FSI. L'objectif était de produire une grille d'évaluation dotée d'échelles descriptives, contenant des critères explicites assortis d'indicateurs observables permettant une mesure la plus valide et fiable possible de la compétence à rédiger des textes scientifiques tout en offrant une rétroaction de qualité aux étudiants.

- 3 Cet article présente le processus de développement, de mise à l'essai et de bonification de cette grille d'évaluation de la compétence des étudiants à rédiger des textes scientifiques en contexte d'études aux cycles supérieurs. Nous entendons ici par productions complexes des travaux écrits relativement longs et exhaustifs, prenant notamment la forme d'un travail scientifique, d'un travail recherche, d'un texte argumentatif, d'une dissertation ou d'un essai, qui visent à dresser un état des lieux ou à approfondir une problématique ou un sujet particulier. Après une brève recension des écrits sur les grilles d'évaluation, nous détaillerons les différentes étapes du processus de conception, présenterons les principaux critères et sous-critères d'évaluation retenus et proposerons, en guise de conclusion, quelques éléments de réflexion, sur les réussites et difficultés générées par la construction de l'outil.

2. Évaluation de productions scientifiques aux études aux cycles supérieurs

- 4 Tout processus d'évaluation exige d'emblée des enseignants qu'ils analysent « la 'preuve d'apprentissage' fournie par les étudiants et [qu'ils] attribue[nt] une valeur à cette 'preuve' » (Berthiaume, David & David, 2011, p. 2). Or, l'évaluation relève d'une pratique interprétative socialement située, faisant nécessairement intervenir la subjectivité individuelle dans le jugement posé sur les productions des étudiants (Read, Francis & Robson, 2005 ; Shay, 2005). En effet, la spécialisation disciplinaire, l'expérience et le degré d'investissement du professeur dans le sujet abordé sont autant de facteurs qui peuvent induire une variabilité dans l'appréciation d'un même travail par différents évaluateurs (Shay, 2005, 2008). À terme, pour les étudiants, ces fluctuations entraînent des difficultés à cerner les attentes et exigences propres à chaque enseignant.
- 5 Dans le contexte de l'enseignement universitaire aux cycles supérieurs, la nature même de l'exercice intellectuel demandé dans les travaux écrits du cursus de maîtrise et de doctorat complexifie la tâche d'évaluation. Axée essentiellement sur le développement de l'aptitude à la recherche scientifique et à la pensée critique (Kishbaugh *et al.*, 2012), la formation aux cycles supérieurs transcende la simple acquisition de connaissances. Elle tend plutôt vers le développement et la consolidation de capacités cognitives complexes liées à la résolution de problèmes, l'argumentation, l'analyse et l'appréciation de la validité et de la crédibilité de données scientifiques (Anderson, Krathwohl & Bloom, 2001 ; Dennison, Rosselli & Dempsey, 2014 ; Jonsson & Svingby, 2007). Cet ensemble d'habiletés est plus difficilement mesurable et requiert des professeurs qu'ils aillent au-delà de l'appréciation de la compréhension d'un phénomène ou de la richesse de l'information présentée afin de repérer, d'interpréter et d'évaluer les manifestations de ces habiletés cognitives complexes (Berthiaume *et al.*, 2011).

- 6 Cette évaluation concerne souvent des travaux écrits témoignant de l'intégration de concepts et de théories complexes et d'une analyse de la littérature scientifique, dans le but de développer un argumentaire apportant une réponse originale à un enjeu ou une problématique (Dennison *et al.*, 2014 ; Jonsson & Svingby, 2007 ; Prégent, 1990). Comme ces rendus ne peuvent être évalués de façon strictement binaire, sur la base d'une simple bonne ou mauvaise réponse, leur appréciation relève souvent de mécanismes implicites, et varie selon la tâche exigée et les manifestations des habiletés cognitives recherchées par les professeurs (Sadler, 2013). Le recours à des outils explicitant les critères et leur degré d'atteinte devient donc profitable pour soutenir l'enseignant dans cette tâche.

2.1. Grilles d'évaluation critériées : apports pédagogiques et limites

- 7 Selon Pellegrino, Chudowsky et Glaser (2001), l'évaluation comporte une succession d'étapes : la clarification des apprentissages à juger, l'observation, soit le moyen mobilisé pour les attester et l'interprétation, c'est-à-dire la mise au point et le recours à des outils permettant de statuer sur le degré d'atteinte des objectifs fixés et de noter la performance en conséquence (Berthiaume *et al.*, 2011). Sur ce dernier volet, le recours aux grilles critériées s'avère une avenue qui gagne en popularité en contexte universitaire. En forçant l'explicitation de critères d'évaluation sur une base qualitative, de même que la détermination des niveaux de performance associés, de telles grilles peuvent diminuer les erreurs de mesure liées à une correction sur la base d'indications sous-entendues, réduisant ainsi la subjectivité du regard porté sur les productions étudiantes (Berthiaume *et al.*, 2011).
- 8 Malgré cet intérêt plutôt récent dans la pratique, les grilles critériées ont néanmoins fait leur apparition au début du 20^e siècle. Leur développement s'est inscrit dans le courant positiviste de quantification des apprentissages notamment à travers l'essor des tests psychométriques visant la réduction de la subjectivité et des différences entre évaluateurs (Broad, 2003 ; Diederich, French & Carlton, 1961 ; Turley & Gallagner, 2008). Intégrées plus formellement dans le champ de l'éducation dans les années 1960, leur popularité a culminé à la fin des années 90 (Broad, 2003 ; Turley & Gallagner, 2008). Au-delà de leur intérêt pour l'évaluation en elle-même, elles présentent un avantage pédagogique dans la mesure où elles facilitent le partage des attentes tout en permettant une rétroaction plus constructive à des fins d'apprentissage (Montgomery, 2002 ; Shepard, 2000).
- 9 Du point de vue des étudiants, ce dispositif pédagogique semble globalement apprécié (Gezie, Khaja, Chang, Adamek & Johnsen, 2012 ; Kishbaugh *et al.*, 2012). Les grilles critériées permettent effectivement de mieux saisir à la fois les objectifs et les exigences de la tâche (McKeown, 2011), en illustrant concrètement les caractéristiques d'un travail de qualité (Jonsson & Svingby, 2007 ; Reddy & Andrade, 2009). Ce faisant, elles constituent une pierre d'assise dans la communication des objectifs d'enseignement en cours de formation (Berthiaume *et al.*, 2011). Assorties d'explications claires du professeur sur leur contenu et leur mode d'emploi, les grilles peuvent servir d'aide à la rédaction et à l'auto-évaluation d'une production (Petkov & Petkova, 2006 ; Reitmeier, Svendsen & Vrchota, 2004) en permettant un repérage explicite et systématique des forces et faiblesses (Jonsson & Svingby, 2007 ; Smith, 2008). Au final, cette transparence accrue diminue l'incertitude (Bolton, 2006), au bénéfice d'une plus grande satisfaction à l'égard de la rétroaction reçue (Andrade & Du, 2007 ; Atkinson & Lim, 2013) et d'une amélioration

potentielle de la performance académique, bien que les connaissances scientifiques sur ce dernier point demeurent encore mitigées (Howell, 2011 ; Jonsson & Svingby, 2007 ; Reddy & Andrade, 2009).

- 10 Les professeurs y trouvent aussi des avantages. Le recours aux grilles critériées peut aider à diminuer la subjectivité et rehausser l'équité dans la correction (Jonsson & Svingby, 2007 ; Reddy, 2011 ; Reddy & Andrade, 2009 ; Timmerman, Strickland, Johnson & Payne, 2011). L'identification claire des niveaux attendus permet de mieux cadrer la performance des étudiants et ainsi de réduire l'ambivalence dans la correction, favorisant une plus grande confiance à l'égard des notes attribuées (Atkinson & Lim, 2013 ; Kishbaugh *et al.*, 2012 ; Stevens & Levi, 2005). Elles permettent aussi d'offrir une rétroaction plus concise et de diminuer le temps consacré à la correction (Kishbaugh *et al.*, 2012). En amont, dans la tâche de planification des évaluations, une grille critériée aide le professeur à clarifier les consignes d'un travail (Ruble, 2014) et même, en offrant la possibilité de poser un jugement plus valide sur des compétences complexes (Jonsson & Svingby, 2007), d'augmenter le niveau de difficulté des travaux (Andrade & Du, 2007).
- 11 Bien qu'elles aient leur place dans le contexte universitaire, les grilles critériées ne sont toutefois pas une panacée, ne serait-ce qu'en raison de l'investissement important requis du professeur qui développe un tel outil. Plusieurs auteurs identifient des écueils possibles dans leur conception notamment pour la clarté du langage (Moni, Beswick & Moni, 2005 ; Reddy & Andrade, 2009 ; Tierney & Simon, 2004) et la capacité à établir des échelles descriptives discriminantes, c'est-à-dire qui peuvent être distinguées de manière évidente et sans équivoque (Moskal & Leydens, 2000 ; Tierney & Simon, 2004). Paradoxalement, ce sont ces mêmes caractéristiques auxquelles se rattachent les critiques les plus souvent adressées aux grilles critériées, soit qu'elles limitent la créativité et la réflexion intellectuelle de par la spécification relativement détaillée des exigences (Atkinson & Lim, 2013 ; Bolton, 2006 ; Gezie *et al.*, 2012). Au demeurant, la valeur de la grille est hautement contingente de son utilisation : mettre à disposition une grille ne suffit pas, encore faut-il qu'elle soit introduite et expliquée, et que les étudiants soient accompagnés et guidés pour que cette dernière acquière un sens dans le contexte de l'exercice intellectuel demandé (Andrade & Du, 2007 ; Lovorn & Rezaei, 2011 ; Reddy, 2011 ; Reitmeier *et al.*, 2004 ; Rezaei & Lovorn, 2010).

3. Développement d'une grille d'évaluation critériée : retour sur le processus

3.1. Genèse du projet

- 12 Le projet d'élaboration d'une grille d'évaluation est né d'échanges informels entre trois professeures qui évoquaient leurs difficultés à définir et à énoncer des critères de correction qui soient, pour les étudiants, indicatifs des niveaux de performance attendus, mais qui puissent également leur permettre d'évaluer les compétences intellectuelles complexes construites au fil d'un cursus. Dans ce contexte, la participation à une formation intensive en pédagogie universitaire fut le catalyseur du projet. En posant les assises conceptuelles, cette formation a constitué une occasion de partages et de réflexions autour des enjeux de l'évaluation en milieu universitaire. Ce projet gagnait en pertinence dans le contexte du chantier de révision des programmes d'enseignement entrepris à la FSI.

- 13 C'est en conservant à l'esprit les limites inhérentes aux divers outils d'évaluation, mais également en reconnaissant leur potentiel appréciable pour guider et accompagner les professeurs dans la tâche de correction, qu'une équipe s'est formée et a conçu un projet de développement pédagogique, soutenu par la FSI, et financé par le BSE de l'Université Laval. L'équipe était constituée des trois professeures, issues d'horizons disciplinaires différents et ayant des champs de recherche variés. Deux expertes en pédagogie universitaire², dont une spécialisée en docimologie, ainsi qu'une étudiante de deuxième cycle issue du programme de maîtrise de santé communautaire s'y sont ajoutées.
- 14 Vu la double nature des programmes gradués relevant de la FSI, le souhait était de développer un outil transposable à la fois dans les programmes de sciences infirmières et de santé communautaire, dans lesquels les professeures impliquées enseignaient, mais qui, à priori, s'attachent à des objets différents. La grille produite devait donc être suffisamment générique et souple pour s'adapter à au caractère multidisciplinaire des divers programmes et aux exigences propres à chaque contexte d'enseignement. Plus spécifiquement, la production de la grille devait à terme viser : a) à clarifier les attentes des professeurs envers les étudiants gradués quant aux qualités recherchées dans une production écrite ; b) offrir une rétroaction aux étudiants gradués sur leur progression dans leurs apprentissages ; c) améliorer la justesse du travail de correction et l'équité dans l'appréciation des productions écrites.

3.2. Étapes d'élaboration de la grille

- 15 La littérature propose quelques pistes méthodologiques pour orienter la construction de grilles d'évaluation. Par exemple, Berthiaume *et al.* (2011), dans une synthèse des travaux de Stevens et Levi (2005), indiquent qu'il s'agit d'abord de spécifier et de détailler des critères d'évaluation qui soient congruents avec les objectifs d'apprentissage pour ensuite définir les niveaux de réussite, après quoi la sélection du type de grille à utiliser est effectuée. Dans une étape finale, la grille est testée, préférablement dans le cadre d'une évaluation formative, et ajustée en fonction de la rétroaction obtenue.
- 16 Bien qu'elle s'apparente à celle-ci, la démarche mobilisée ici rejoint davantage celle proposée par Reddy (2011) qui détaille huit étapes. Les activités en lien avec chacune des étapes décrites ont toutefois été légèrement adaptées aux spécificités du contexte, et ce, au fil du développement de l'outil, de la consultation des écrits recensés et des suggestions des conseillères en pédagogie. Le tableau 1 fait état des détails de cette démarche.

Tableau 1. Étapes de développement de la grille d'évaluation, adaptées de Reddy (2011)

Étapes	Activités
1. Identification des objectifs d'apprentissage et des critères	Recension des écrits et examen de la littérature scientifique Collecte et analyse de grilles d'évaluation utilisées par des professeurs de la FSI
2. Opérationnalisation des critères d'évaluation et développement d'un outil d'évaluation	Discussions et échanges au cours des neuf rencontres de l'équipe formée de professeurs, d'expertes en mesure et évaluation et d'une étudiante
3. Première expérimentation de l'outil	Pré-expérimentation de la grille d'évaluation dans deux cours sous la responsabilité de deux professeurs impliquées dans le projet
4. Obtention de rétroaction auprès des principaux utilisateurs attendus	Rencontre avec des professeurs Rencontre avec des étudiants
5. Révision de la grille en fonction de la rétroaction obtenue auprès des principaux utilisateurs et de la première expérimentation	Ajustement du contenu de la grille
6. Évaluation	Évaluation de la validité et de la fidélité
7. Mise à l'essai de la grille d'évaluation	Mise à l'essai par les professeurs impliquées dans le projet et par d'autres professeurs de la FSI et obtention de rétroaction auprès des étudiants de ces cours
8. Amélioration de la grille d'évaluation	Ajustement de la grille d'évaluation à partir des résultats obtenus lors de la mise à l'essai

- 17 Le présent article relate plus spécifiquement les trois premières étapes de cette démarche. Il apparaissait incontournable de décrire dans le détail le processus emprunté, d'autant plus que le travail de conception d'une grille d'évaluation relève généralement d'un travail solitaire de l'enseignant et non d'un processus de réflexion collective, ce qui constitue un trait original du projet. Néanmoins, les étapes subséquentes de développement seront aussi brièvement évoquées, afin de présenter les directions futures.

3.3. Première étape : identification des objectifs d'apprentissage et ébauche des critères

- 18 Selon Reddy (2011), outre la spécification de l'objet à évaluer, la première étape consiste à identifier les critères d'évaluation. Pour ce faire, la triangulation des sources, notamment le recours aux curriculums départementaux, à la littérature scientifique et la consultation d'experts, constitue une stratégie importante pour augmenter la validité de la grille d'évaluation. Pour notre part, le démarrage du projet s'est fait en deux temps, d'abord à travers une recension pour repérer les grilles d'évaluation publiées dans la littérature et, ensuite, par le biais d'une collecte intrafacultaire des outils d'évaluation utilisés par les collègues.

3.3.1. Recherche dans la littérature

- 19 Dans un premier temps, une recension des écrits a été réalisée afin de colliger les grilles d'évaluation des productions écrites aux études aux cycles supérieurs disponibles dans la littérature scientifique et pouvant servir d'inspiration ou de base de réflexion pour le

présent projet. Élaborée en collaboration avec des bibliothécaires spécialisées, la stratégie de recherche retenue prévoyait d'explorer les bases des données ERIC, Web of Science et CINAHL. Une première recherche s'est restreinte aux études aux cycles supérieurs, puis, au vu du faible nombre de résultats générés, une seconde recherche a étendu les mots-clés à l'enseignement universitaire plus largement, en incluant le premier cycle.

- 20 Sur près de 300 titres trouvés, 42 articles ont été retenus, dont 11 ont été repérés par le biais d'une recherche manuelle dans les bibliographies des publications identifiées. L'ensemble de ce corpus a ensuite été dépouillé. Les grilles incomplètes ou trop spécifiques à un domaine d'études ont été écartées, tout comme celles qui ne comptaient pas d'échelles descriptives, comme il s'agit d'une composante-clé des grilles critériées (Alter & Adkins, 2006 ; Haaga, 1993 ; Many, 1999 ; Smith, 2008 ; Topping, Smith, Swanson & Elliot, 2000).
- 21 À l'issue de ce premier examen approfondi, seulement neuf articles reproduisaient intégralement des grilles d'évaluation utilisées aux études aux cycles supérieurs. Deux grilles se sont montrées particulièrement intéressantes du fait qu'elles avaient fait l'objet d'un processus de validation rigoureux (Reddy, 2011 ; Sevian & Gonsalves, 2008). Toutefois, ayant été développées et formulées pour un type de production et une discipline spécifique, la plupart des critères y figurant relevaient d'une tâche propre à la discipline. Ces grilles cadraient donc plus ou moins bien avec la portée multidisciplinaire souhaitée pour l'outil à créer.
- 22 Au final, sept grilles d'évaluation de travaux longs (par exemple, dissertations, essais) ont été jugées pertinentes et retenues. Parmi celles-ci, seules deux grilles avaient été développées spécifiquement pour les études de 2^e et 3^e cycles (Rezaei & Lovorn, 2010 ; Truemper, 2004). Le tableau 2 synthétise les principales caractéristiques descriptives extraites de ces grilles et illustre notamment que le matériel sélectionné provenait d'une diversité de champs d'études, soit des sciences sociales (Gezie *et al.*, 2012 ; Pagano, Bernhardt, Reynolds, Williams & McCurrie, 2008 ; Rezaei & Lovorn, 2010 ; Rublee, 2014), des sciences administratives (Atkinson & Lim, 2013) et des sciences de la santé (Dennison *et al.*, 2014 ; Truemper, 2004).

Tableau 2. Caractéristiques des grilles retenues issues de la littérature

Auteurs	Discipline	Cycle d'études	Nombre de critères	Nombre d'échelons descriptifs	Critères utilisés ³
(Atkinson & Lim, 2013)	Affaires et commerce	1 ^{er}	10	5	<ul style="list-style-type: none"> - Démonstration de réflexion critique – analyse - Références (dans le texte et en bibliographie; sources et preuves qui soutiennent l'argumentaire) - Qualité de l'expression (orthographe, vocabulaire, grammaire, ton, objectif)
(Dennison <i>et al.</i> , 2014)	Sciences infirmières	1 ^{er}	13	4	<ul style="list-style-type: none"> - Respect des normes bibliographiques et de présentation matérielle - Organisation et structure - Corps du travail : contenu requis - Références
(Gezie <i>et al.</i> , 2012)	Travail social	1 ^{er}	5	4	<ul style="list-style-type: none"> - Décrire - Examiner - Articuler l'apprentissage - Utilisation d'un style bibliographique et format - Qualité générale de l'écriture
(Pagano <i>et al.</i> , 2008)	Arts libéraux et sciences sociales	1 ^{er}	5	3	<ul style="list-style-type: none"> - Réponse à la tâche - Engagement avec le(s) texte(s) - Développement - Organisation - Contrôle de la langue
(Rezaei & Lovorn, 2010)	Sciences sociales	2 ^e	5	4	<ul style="list-style-type: none"> - Organisation structurée et clarté - Compréhension et synthèse - Compréhension des objectifs et répercussions - Références et citations qui soutiennent l'argumentaire - Écriture (vocabulaire, orthographe, ponctuation)
(Rublee, 2014)	Sciences politiques	1 ^{er}	6	5	<ul style="list-style-type: none"> - Structure et organisation - Argumentation - Originalité analytique - Profondeur de la recherche et des références - Grammaire, syntaxe et ponctuation
(Truemper, 2004)	Sciences infirmières	2 ^e	8	4	<ul style="list-style-type: none"> - Transitions - Organisation - Contenu - Choix des mots - Références et respect des normes bibliographiques

Note³

- 23 L'exercice de recension des écrits a d'emblée permis de tirer un premier constat : la quasi-absence de littérature sur les modalités d'évaluation des productions écrites exigées dans le cursus académique aux études aux cycles supérieurs, alors que l'évaluation des mémoires, des thèses et des compétences cliniques présentent, pour leur part, des travaux somme toute exhaustifs. Néanmoins, les grilles collectées, bien qu'elles étaient exclusivement présentées en anglais, ont permis de dresser un portrait de certains des éléments les plus communément évalués et des échelles descriptives proposées.

3.3.2. Collecte et analyse de grilles des professeurs

- 24 Parallèlement à cette recension des écrits, les grilles d'évaluation utilisées par les professeurs de la FSI dans le cadre de leur enseignement aux différents programmes d'études aux cycles supérieurs ont été collectées et analysées. Lors d'une réunion facultaire, l'ensemble du corps professoral a été informé de la nature et des objectifs du projet de développement pédagogique et sollicité afin de partager, sur une base volontaire, ses outils d'évaluation avec l'équipe. À la suite d'un appel à tous diffusé par courriel, 52 grilles ont été collectées. La première étape de leur analyse a consisté à retenir celles conçues pour évaluer trois types de productions : les travaux longs de type dissertation, les commentaires de lecture et les présentations orales. Aux débuts de ce projet, le type exact de travaux visés par la ou les grilles à développer n'était pas clairement déterminé. Cependant, rapidement en cours de processus, les professeurs ont fixé leurs besoins quant à l'évaluation des productions écrites complexes telles les travaux longs scientifiques, les dissertations et les textes argumentatifs. Les grilles visant

l'appréciation de résumés de lecture ou d'exposés oraux ont donc été écartées. Celles dont le contenu était trop étroitement lié à un thème ou une tâche spécifique à accomplir ont aussi été exclues. Le nombre de grilles reçues, par type de travail évalué et cycle d'études, retenues pour analyse est indiqué au tableau 3.

Tableau 3. Description des grilles collectées auprès des professeurs

Type de travaux	Maîtrise	Doctorat	Maîtrise et doctorat	Non spécifié	Reçues	Retenues
Travaux longs de recherche	5	2	3	5	15	10
Commentaires de lecture	2	2	0	0	4	4
Présentations orales	2	2	5	5	16	8
Autres travaux	4	5	3	3	17	0

- 25 Notons que malgré la grande qualité des grilles recueillies, la plupart cadraient plus ou moins avec les objectifs du projet, soit en raison du caractère très spécifique des critères, souvent calqués sur le contenu recherché du travail exigé, ou encore par l'absence d'échelles descriptives détaillées. D'ailleurs, sur les dix grilles d'évaluation des travaux longs de recherche retenues, seulement deux présentaient un certain nombre d'échelons comportant des énoncés descriptifs illustrant les différents degrés d'atteinte des critères. En somme, si plusieurs comportaient les éléments de contenu classiques attendus d'une grille de correction, notamment la pondération relative des critères, la performance observable attendue dans les travaux n'était pas explicitée.
- 26 Une fois les grilles sélectionnées, une analyse de contenu a permis d'extraire les critères généraux d'évaluation et la manière dont ils étaient opérationnalisés en fonction du type de travail exigé et sur la base d'une classification par thèmes communs. En raison des objectifs du projet, l'analyse s'est centrée sur dix grilles évaluant des travaux longs, ce qui a permis de dégager quatre principaux critères d'évaluation : 1) la présentation générale, incluant des éléments de forme, la structure et la qualité de l'écriture ; 2) la capacité analytique, particulièrement la réflexivité, le jugement critique et la rigueur ; 3) la mobilisation de la littérature scientifique, et 4) la compréhension. L'ensemble des résultats a été colligé dans un document synthèse auquel l'équipe s'est référé périodiquement pour nourrir sa réflexion sur l'outil d'évaluation à créer.
- 27 Notons que ce travail de collecte et d'analyse s'est avéré fort utile pour jeter les assises de la validité de contenu de la grille élaborée en équipe, et ce, sous deux angles. D'abord, il a mis en lumière à plus large échelle, à travers une grande diversité de cours et la nature variée des travaux exigés, les éléments fondamentaux constituant la base de l'évaluation sommative. Ensuite, il a mis au jour le langage généralement employé dans les différents champs disciplinaires couverts au sein de la FSI pour articuler les critères de correction. Les grilles reproduites dans la littérature étant exclusivement en anglais, disposer d'outils développés en français fournissait une aide supplémentaire pour la formulation des critères, des sous-critères et des échelles descriptives, en plus de contribuer à augmenter la validité contextuelle de l'outil. Au final, à défaut de générer des grilles « clé en main », l'examen des écrits scientifiques et l'analyse des grilles utilisées à la FSI ont permis de baliser le processus de développement du nouvel outil. En effet, le document synthèse produit, consignait les divers critères en usage et leur mode d'opérationnalisation, a servi de source d'inspiration quant aux éléments de contenu à intégrer, la manière de les agencer et de les formuler.

3.4. Deuxième étape : développement des critères et des échelles descriptives

- 28 Sur cette base, l'équipe s'est réunie pour entreprendre la deuxième étape, soit la conception et l'élaboration de sa propre grille, une opération qui selon Reddy (2011) implique d'identifier les critères et de formuler des échelles descriptives pour préciser les performances attendues des étudiants. Tôt dans le processus, la direction des rencontres a été assumée par la docimologue qui a instauré un mode de réflexion collectif fortement guidé par l'induction. En début de processus, il a délibérément été choisi de laisser le matériel empirique recueilli à distance pour laisser émerger spontanément et sans a priori les idées et les besoins quant aux critères d'évaluation et aux indicateurs à intégrer. Les premières rencontres ont surtout servi à échanger sur les qualités d'un travail universitaire aux études aux cycles supérieurs en fonction des perceptions des professeures et de l'étudiante graduée sur la base de leurs expériences respectives en lien avec la réalisation et la notation de telles productions. Ce premier remue-méninges se voulait un exercice d'explicitation et de mise-en-mots d'éventuels critères d'évaluation, sans chercher à structurer d'emblée la réflexion.
- 29 Les discussions se sont ensuite rapidement tournées vers la formulation des premiers énoncés d'échelles descriptives, un exercice poursuivi au fil des rencontres ultérieures. Toujours sous la direction de la docimologue, les professeures et l'étudiante illustraient simultanément, à l'aide d'exemples vécus et d'images, les caractéristiques d'une production écrite perçue exemplaire et, par opposition, d'une de piètre qualité ou déficiente, l'idée étant de faire émerger des distinctions claires par l'utilisation forcée du contraste.
- 30 Ces premiers allers-retours ont permis de clarifier les grandes composantes de la grille, c'est-à-dire les critères généraux de correction et le nombre d'échelons, représentant les niveaux de performance, qui fut fixé à quatre au départ. Ce chiffre, proposé par l'experte en mesure et évaluation en pédagogie, correspond aux nombres d'échelons normalement suggérés dans la littérature, soit entre trois et cinq (Reddy & Andrade, 2009). Sur la base des échanges, les premiers critères généraux ont été dégagés : la structure du texte, la recherche, le raisonnement, la synthèse, le développement, le respect des consignes, le respect de normes de présentation, la compréhension, la qualité de la rédaction, l'originalité analytique et la rigueur. Au fil des rencontres, ces critères et ces indicateurs ont été définis, raffinés et opérationnalisés, puis assemblés dans la mouture finale de la grille. L'essentiel du travail réalisé pendant les rencontres renvoyait à l'opérationnalisation de ces critères en niveaux de performance à travers la formulation d'énoncés qualitatifs descriptifs. Vu l'hétérogénéité de l'équipe rassemblée, cette étape a suscité de vives discussions au regard de la signification de chacun des critères, échanges qui ont permis de mettre au jour les nuances et les points de désaccord sur le sens attribué à certains concepts. Des discussions exhaustives ont parfois été nécessaires pour parvenir à un consensus sur la signification de chaque critère et le cristalliser en un énoncé clair et porteur de sens pour les étudiants et les professeures. À cet égard, la taxonomie de Bloom (Anderson *et al.*, 2001) s'est avérée utile pour encadrer le processus. Malgré les nombreuses transformations et reformulations au fil de cette démarche itérative, la plupart des critères énoncés au départ figurent toujours dans la grille finale et en forment les principales sections.

- 31 En somme, le contenu préliminaire de la grille a nécessité neuf rencontres d'une demi-journée pour être mis au point, lesquelles ont été entrecoupées de travail individuel de la part des membres de l'équipe pour en clarifier le contenu. Ainsi, la grille s'est peaufinée sur la base d'un travail inductif de coconstruction, au gré des allers et retours entre les discussions en équipe et le travail individuel, en s'inspirant toutefois sporadiquement de la recension des écrits, des grilles de collègues et d'autres outils pédagogiques, particulièrement les 88 clefs de Thibaudeau (2009), dans le but de préciser certaines idées ou la terminologie employée. La grille était d'emblée bonifiée et ajustée lorsque des lacunes étaient mises en évidence lors des discussions.

3.4.1. Mise en forme finale et pondération des critères

- 32 Au terme des neuf rencontres, l'étape de construction de la grille a été jugée bouclée. Un dernier regard a été posé par tous les membres de l'équipe sur l'outil construit pour identifier les redondances et les inconsistances dans le vocabulaire. Une fois ces derniers détails remaniés, la grille a été soumise à la docimologue afin de réaliser un dernier élagage et de procéder à la mise en page finale. Le travail de la docimologue à cette étape a été particulièrement précieux puisqu'il a permis de ramener à une page recto-verso la version initiale de la grille qui comportait six pages, une opération essentielle pour rendre l'outil convivial et plus simple à utiliser.
- 33 Ainsi, cette première mouture de grille d'évaluation complète se décline sous huit critères, eux-mêmes définis en sous-critères. Au terme du processus, ceux-ci constituent, selon l'équipe, les éléments essentiels à évaluer dans le contexte des études aux cycles supérieurs dans les programmes de la FSI. Ces critères sont ici présentés, assortis de leurs principaux sous-critères.
- *Objectifs du travail : respect de la tâche et de ses exigences.* Ce critère porte essentiellement sur la tâche accomplir, c'est-à-dire si le travail répond à la question posée ou correspond à la tâche exigée dans la forme attendue.
 - *Respect des normes de présentation telles qu'énoncées dans les consignes.* Ce critère renvoie à la présentation matérielle du travail : utilisation d'un système de référence constant entre le texte et la bibliographie, police d'écriture, pagination, mise en page, etc.
 - *Qualité de la langue et de la rédaction.* La qualité du français, tant au plan de la grammaire que de style, est ici abordée. Au-delà de la syntaxe, de l'orthographe et de la grammaire, ce critère comporte aussi une composante sur la clarté de l'expression ainsi que le ton adopté dans le travail.
 - *Repérage et mobilisation de la littérature grise et scientifique.* Il s'agit ici d'évaluer la qualité et la pertinence des références et leur utilisation pour soutenir l'argumentaire dans un texte.
 - *Argumentation et logique.* Ce critère évalue le cheminement logique, ou le fil conducteur, du texte dans son ensemble ainsi qu'à l'intérieur des paragraphes, par exemple la facilité à repérer les idées et la logique de leur développement.
 - *Compréhension et capacité de synthèse.* Ce critère évalue la démonstration de la compréhension du sujet d'étude, notamment à travers la délimitation du sujet à l'étude, ainsi que la synthèse de l'objet d'étude, c'est-à-dire la concision ainsi que la capacité à présenter et distinguer les idées principales et accessoires.
 - *Jugement critique.* Ce critère se déploie en trois étapes. D'abord, il évalue comment le travail compare et contraste des idées, ou fait dialoguer les auteurs. Ensuite, il évalue la capacité à questionner, saisir et nuancer la problématique à l'étude et la littérature scientifique qui soutient l'argumentaire. Enfin, il évalue la capacité de l'étudiant à tirer un sens nouveau à

partir de son argumentaire et à se positionner à l'égard des idées et des critiques des auteurs évoqués dans le texte.

- *Structure du travail.* Ce critère évalue la structure du texte et de ses parties, soit la structure de l'introduction, du développement et de la conclusion. Il évalue aussi l'importance relative du développement par rapport à ces parties ainsi que la présence et la qualité des titres et sous-titres dans le texte.

34 Ces critères et ces sous-critères sont accompagnés d'énoncés qualitatifs descriptifs déclinés en quatre échelons distincts. Ces échelles descriptives explicitent les différents niveaux d'atteinte relatifs à chaque critère et sous-critère. Par exemple, le critère « argumentation et logique » se décline en un sous-critère intitulé « cheminement logique de l'argumentaire dans l'ensemble du texte » qui lui renvoie à quatre échelles descriptives distinctes.

- Niveau a : Les arguments sont bien identifiés, organisés de manière cohérente et avec logique (liens de déduction, d'induction et d'analogie).
- Niveau b : Les arguments sont bien identifiés, mais à de rares moments, les idées manquent légèrement de cohérence.
- Niveau c : Plus rarement, les arguments sont identifiés et organisés d'une manière logique et cohérente ou des contradictions s'observent.
- Niveau d : On ne perçoit pas les liens entre les idées énoncées ou les arguments défendus.

35 L'ensemble des critères et des sous-critères a été opérationnalisé de la sorte. Dans la mise en forme finale, il est possible de cocher chaque élément dans les échelles descriptives afin d'offrir une rétroaction plus ciblée aux étudiants.

36 Quant à la pondération et la répartition des points, en raison de l'échelle descriptive à quatre niveaux retenue, un simple pointage sur quatre points a été testé. Ainsi, pour chaque sous-critère, l'étudiant recevait quatre points s'il se classait au niveau A ou aucun point si sa performance se situait à l'autre bout du spectre, au niveau D. Les niveaux B et C valaient trois points et deux points, respectivement. L'attribution d'une pondération exprimée en pourcentage pour chacun des critères permettait de multiplier la note attribuée sur quatre pour obtenir une note finale cumulative sur cent points. Comme la pondération relative attribuée à chaque critère était laissée à la discrétion des professeures, cette méthode apparaissait comme la plus appropriée dans le contexte.

3.5. Troisième étape : première mise à l'essai et révision finale

37 Une fois la première version finalisée, la grille a été mise à l'essai. Selon Reddy (2011), cette première expérimentation peut prendre la forme d'un projet pilote ou d'obtention de rétroaction auprès des principaux utilisateurs, soit des professeurs et des étudiants. Pour ce faire, deux professeures de l'équipe l'ont testée dans les cours dont elles étaient titulaires au trimestre d'hiver 2015, soit un cours de 2^e cycle en sciences infirmières et un de 3^e cycle en santé communautaire, offrant ainsi l'opportunité de la mettre à l'essai dans des contextes d'enseignement variés. Toutefois, vu les petites cohortes d'étudiants, le nombre de travaux corrigés à l'aide de la grille ne s'est élevé qu'à neuf, mais l'utilisation de l'outil dans une situation d'enseignement concrète a permis de mettre à l'épreuve la première mouture.

38 Cette première expérimentation par les professeures a été jugée positive, notamment en ce qui a trait à la réduction du temps de correction. Malgré les imperfections du système

de notation, l'évaluation à l'aide des échelles descriptives fournissait un solide cadre sur lequel juger la qualité des travaux et appuyer le jugement sur la note finale attribuée. Les professeures se sentaient plus confiantes envers les notes remises aux étudiants et la rétroaction fournie puisque la grille permettait clairement d'identifier la plupart des points forts et des aspects à améliorer dans les productions écrites.

- 39 Cette première évaluation à l'aide de la grille a aussi permis de mettre en évidence un certain nombre de failles, notamment autour du recoupement de certains sous-critères, redondances qui furent dès lors corrigées. Des imprécisions dans le vocabulaire ou le sens donné aux énoncés dans les échelles descriptives ont aussi pu être clarifiés. Néanmoins, deux éléments plus problématiques ont été mis en lumière à travers cet exercice. Le premier renvoie à l'exigence de transposition de la grille dans de multiples contextes d'enseignement. Une telle visée laisse nécessairement peu de place pour spécifier finement les éléments de contenu à retrouver ou les connaissances à maîtriser dans le contexte d'un cours particulier, ces éléments devant être ajoutés a posteriori de manière à « personnaliser » la grille générique en fonction des exigences propres à un travail particulier. Bien que ces ajouts à la pièce soient nécessaires, ils demeurent quelque peu noyés au travers de l'ensemble et s'y intègrent parfois mal.
- 40 Le second renvoie à la méthode de pointage mise à l'essai qui s'est avérée difficile à opérationnaliser et à calculer pour l'obtention de la note finale, surtout avec une échelle à quatre niveaux qui se prête moins bien à une répartition du pointage sur la base de multiples de cinq ou de dix. De plus, le système de notation par la dominante n'a pas donné les résultats escomptés. Les professeures devaient évaluer quel échelon de l'échelle correspondait le mieux à ce qui était observé dans le travail de l'étudiant, à partir d'une appréciation globale de la tendance prépondérante ressortant de leur performance sur chacun des sous-critères, une pratique qui induisait des aberrations. Par exemple, pour le critère de la qualité de la langue et de la rédaction, l'étudiant pouvait se voir attribuer la totalité des quatre points parce que son travail rencontrait complètement les exigences (niveau A) sur deux sous-critères, soit la justesse du ton et à la clarté du propos, et ce, en dépit d'une performance jugée médiocre (niveau D) sur le troisième sous-critère (orthographe, syntaxe et grammaire) et pour laquelle il ne se trouvait finalement pas pénalisé.
- 41 De plus, la méthode de pointage semblait aussi produire un effet discriminant sur les notes finales en rehaussant les productions d'excellent niveau et en ramenant à la baisse la notation des travaux de moins bonne qualité. La répartition des points entre les quatre niveaux a donc dû être ajustée en cours de route. À ce titre, la réalisation de simulations de différents scénarios de ventilation de pointages et des discussions à ce propos lors de deux rencontres additionnelles ont permis de revoir le système de notation.
- 42 À la suite de cette première mise à l'essai, les prochaines étapes amèneront l'équipe à valider l'outil auprès d'autres enseignants, et à documenter davantage son utilité, notamment son apport pédagogique pour les étudiants, et ce, à travers des questionnaires et des entretiens réalisés avec des utilisateurs réels et potentiels. Cette rétroaction permettra subséquentement d'apporter quelques ajustements, de s'assurer que le langage est accessible et que les critères et les échelles descriptives ont été adéquatement opérationnalisés, deux indices de qualité fondamentaux pour des grilles critériées. En outre, le langage est un des aspects les plus complexes dans l'élaboration de telles grilles, d'où la pertinence d'un processus incluant les étudiants afin de valider la compréhension (Moni *et al.*, 2005 ; Tierney & Simon, 2004).

4. Discussion : retour réflexif sur le processus de construction d'une grille d'évaluation

- 43 Au terme de ce processus réflexif et participatif de 12 mois, l'équipe a pu identifier un certain nombre de défis à l'égard de la conception d'une grille d'évaluation pour des productions écrites en contexte d'études aux cycles supérieurs que nous évoquerons ici en guise de discussion.

4.1. Investissement considérable pour créer un sens commun

- 44 D'entrée de jeu, la démarche a requis un investissement en temps plus important qu'anticipé au départ. À ce chapitre, le soutien de la docimologue s'est avéré essentiel pour optimiser l'efficacité de la démarche et limiter les errements. Si l'identification des critères globaux d'évaluation s'est faite relativement aisément, les énergies ont surtout été mobilisées dans le travail d'explicitation et d'opérationnalisation des critères et des sous-critères pour les traduire dans des échelles descriptives détaillant distinctement les divers niveaux de performance. En effet, les échelles descriptives, au-delà de leur formulation factuelle et observable, doivent être les plus mutuellement exclusives possible, de manière à marquer une division claire entre chacun des niveaux (Moskal & Leydens, 2000 ; Tierney & Simon, 2004). Dans ce contexte, les difficultés ont surtout porté sur la formulation de critères et de sous-critères mutuellement exclusifs, c'est-à-dire qui ne se recoupent pas et qui ne soient évalués qu'à un seul endroit dans la grille. Au final, compte tenu de la complexité et des interconnexions entre les habiletés cognitives évaluées dans le contexte des études aux cycles supérieurs, les chevauchements demeurent inévitables, mais lorsque suffisamment décortiqués et explicités, ils peuvent être considérablement réduits.
- 45 Plusieurs discussions animées ont également porté sur la classification des sous-critères, leur rattachement logique aux critères plus généraux et leur disposition à travers différentes catégories pour éviter les redondances. Bref, le consensus sur une arborescence finale qui ait un sens pour tous les membres de l'équipe a été difficile à obtenir. À cet effet, les écueils rencontrés reflètent l'hypothèse de Shay (2005) posant l'évaluation comme une pratique socialement ancrée. Les manifestations de ces habiletés cognitives, le reflet de leur importance relative à travers les critères et les échelles descriptives et la logique de leur classification sont tout autant d'objets de discussions grandement teintées par la spécialisation disciplinaire des professeures et les expériences des membres de l'équipe (Shay, 2005 ; Thibaudeau, 2009). Si la diversité de l'équipe réunie et le temps dédié à cette démarche réitérent le caractère intrinsèquement construit de l'évaluation des productions écrites complexes aux 2^e et 3^e cycles universitaires, il s'agit néanmoins d'un investissement nécessaire pour assurer la validité de l'outil développé et sa pertinence contextuelle.

4.2. À la recherche d'un outil convivial

- 46 La contrepartie manifeste à la fois d'une démarche itérative de coconstruction, et d'un travail approfondi et délibéré d'explicitation a été de déboucher sur une première version complète de la grille d'une longueur démesurée (11 pages) qui correspondait plus ou

moins aux attentes formulées à l'aube du démarrage du projet. Si le processus pour aboutir à cette ébauche a été ardu, l'élagage subséquent s'est avéré encore plus difficile. En effet, bien que très raffinée dans sa forme initiale, la longueur de la grille contraignait sensiblement sa convivialité pour les professeurs et compromettait son utilité pour la tâche de correction, noyant aussi les étudiants dans des détails superflus et risquant davantage de semer la confusion que de les aider à saisir les exigences. Cette opération « réductionniste » a été un moment charnière du processus qui a questionné l'équipe quant à sa capacité d'arriver à une version allégée, utile pour la tâche d'évaluation, tout en questionnant la pertinence même du recours à des grilles critériées dans le contexte des études aux cycles supérieurs. Confrontée à cet obstacle majeur, l'équipe a confié cette première tâche d'élagage à la docimologue qui a procédé à un remaniement, en éliminant les répétitions, en formulant plus efficacement les échelles descriptives et en maximisant la mise en page.

- 47 Rétrospectivement, les écueils rencontrés en cours de processus ont sensibilisé l'équipe à la confusion que peuvent vivre les étudiants confrontés à des critères de correction tacites et à la nécessité de les formuler le plus clairement possible pour qu'une grille d'évaluation devienne réellement un outil pédagogique pour les orienter dans la réalisation de leurs productions écrites. Malgré les mises en garde énoncées dans la littérature à cet effet, les difficultés à l'égard du langage, de la clarté, d'une certaine univocité des critères et des échelles descriptives afin de maintenir l'apport pédagogique de l'outil développé ne sauraient suffisamment être soulignées (Reddy, 2011 ; Tierney & Simon, 2004). Du même souffle, la démarche souligne la complexité à mettre en exergue des critères, des sous-critères et des échelles descriptives qui soient congruents avec les habiletés cognitives à développer aux études aux cycles supérieurs, sans multiplier indûment les nuances qui alourdissent l'outil et, au final, peuvent lui faire perdre son sens et sa visée première. Comme le soulignent Berthiaume *et al.* (2011) en relatant leur propre expérience d'élaboration : « Le langage descriptif de la grille, parce qu'il porte sur des phénomènes qualitatifs [...] et non sur des phénomènes mesurables [...] évoque des degrés clairs mais dont les seuils sont très flous [...]. Il y a un flou que rien ne résorbe dans l'élaboration de la grille elle-même. » (p. 7)
- 48 Malgré un certain consensus quant aux compétences visées aux cycles supérieurs et le recours à une taxonomie des habiletés cognitives complexes, la difficulté de quantifier leur degré d'atteinte conjuguée à un usage commun et peu défini de l'« esprit critique » posent des défis additionnels à l'explicitation de celles-ci (Johnson & Blair, 2009 ; Seldomridge & Walsh, 2006). En somme, le point d'équilibre entre le détail et l'essentiel, sans toutefois compromettre la visée même de clarification, est ardu à trouver. La capacité d'évoquer en mots des habiletés cognitives complexes demeure limitée et le simple recours à un outil ne résout pas entièrement ce problème.

4.3. Défi de la portée transversale

- 49 L'objectif initial de l'équipe était de concevoir une grille d'évaluation suffisamment générique pour être transférable et adaptable dans plusieurs contextes. Cette visée était justifiée, rappelons-le, par la composition même de l'équipe de professeures évoluant dans des contextes d'enseignement variés et par la volonté de jeter des ponts entre les programmes gradués de sciences infirmières et de santé communautaire au sein de la FSI. En règle générale, l'ébauche de critères d'évaluation se fait en concordance avec les

objectifs pédagogiques visés par un cours, et non par ceux d'un programme, encore moins par plusieurs. Or, créer une grille critériée plus universelle qui chevauche plusieurs cours et plusieurs programmes, en plus de couvrir le 2^e et le 3^e cycle, ne s'est pas avéré une mince tâche. Entre les programmes et leurs cursus pédagogiques respectifs, les contenus sont forts variés et les apprentissages visés distincts. Néanmoins, un objectif commun demeure, celui de la consolidation de l'aptitude à la recherche scientifique et la maîtrise de la méthodologie intellectuelle et du travail de rédaction qui la sous-tend. C'est autour de cette tangente que l'équipe a voulu asseoir la construction de la grille, en articulant le développement des critères, des sous-critères et des échelles descriptives autour de la compétence à rédiger des textes scientifiques aux études aux cycles supérieurs plutôt que sur des contenus ou des connaissances disciplinaires à acquérir. Par conséquent, en raison de son caractère généraliste, la grille d'évaluation devra vraisemblablement être chaque fois adaptée selon son contexte d'utilisation, colorée selon la nature de la production écrite exigée et des objectifs pédagogiques visés par le cours.

4.4. Reflet de la progression des compétences

- 50 La grille d'évaluation créée visait à être appliquée aux 2^e et 3^e cycles universitaires, soit à la maîtrise et au doctorat, niveaux entre lesquels est prévue une progression dans les compétences intellectuelles, notamment du point de vue des habiletés de recherche. Comment alors refléter les distinctions dans les niveaux de performance attendus dans une seule et même grille ? Cette préoccupation a constamment refait surface dans les discussions et les échanges, mais les quatre niveaux d'échelles descriptives reflétés dans l'outil mis au point, offrent la possibilité de positionner le curseur de réussite différemment, selon que l'évaluation concerne la maîtrise ou le doctorat ou encore de pondérer différemment chacun des critères de correction en fonction de la réalité du niveau d'études et des objectifs qu'il poursuit.
- 51 Finalement, compte tenu des parcours différenciés et de la nature diversifiée des productions écrites exigées des étudiants, l'équipe a parfois eu du mal à concilier les exigences personnelles propres à chaque professeure dans une grille d'évaluation consolidée. Comme le mentionnent à juste titre Romainville (2011) : « Surtout dans l'enseignement supérieur, la multiplicité des points de vue en ce qui concerne les acquis des étudiants constitue une richesse, qui pourrait être mise à mal par un trop grand souci de standardiser les pratiques sous prétexte de les rendre objectives. » (p. 6)
- 52 Dans cette lignée, nous avons souhaité tableur sur l'hétérogénéité de l'équipe, en dépit des visions différenciées au regard des habiletés complexes à évaluer. En effet, le caractère multidisciplinaire apportait une validité supplémentaire à la démarche et constituait un gage de la versatilité de l'outil, au-delà d'un cours, d'un programme ou d'une discipline.

5. Conclusion

- 53 Cet article a détaillé la démarche entreprise dans un projet de conception d'une grille d'évaluation des productions écrites aux études aux cycles supérieurs. Dans un contexte de révision des programmes de la FSI, il a permis de jeter des ponts et de nourrir le dialogue entre les acteurs engagés dans les programmes gradués de santé communautaire et de sciences infirmières. Son caractère innovant se situe notamment dans la démarche mobilisée, soit celle d'associer des professeurs et une étudiante, à l'intérieur d'une même

équipe, dans une réflexion conjointe sur les attendus et la coconstruction de critères de correction. De plus, l'assise scientifique du processus mobilisé permet d'aller au-delà des démarches individuelles souvent proposées dans les livres de pédagogie et de suggérer une façon de faire reflétant les dernières avancées dans ce champ d'études. En détaillant cette démarche et ses résultats ainsi qu'en discutant des difficultés rencontrées, cet article a permis de nourrir la réflexion sur l'évaluation de la compétence à rédiger des textes scientifiques dans la formation aux cycles supérieurs, un domaine encore peu exploré.

- 54 Les étapes suivantes du projet permettront d'étoffer davantage l'outil construit, notamment en ce qui a trait aux échelles descriptives, et d'alimenter le désir de coconstruction qui est au cœur de cette démarche. Une phase de validation permettra également d'obtenir davantage de rétroaction auprès d'étudiants de cycles supérieurs et de professeurs. Au-delà de valider leur compréhension de l'outil, cette étape permettra de collecter leurs réactions sur la grille et leur appréciation, après quoi une mise à l'essai plus étendue est prévue en conformité avec la démarche présentée précédemment (tableau 1). Ainsi, la grille d'évaluation sera utilisée par d'autres professeurs dans leurs cours, permettant de recueillir leurs commentaires sur leur expérience de même qu'une rétroaction d'un nombre plus important d'étudiants. Ces différents résultats permettront d'améliorer la grille pour en venir à une version finalisée ainsi qu'à la rédaction d'un guide d'accompagnement pour soutenir les professeurs et les étudiants dans son utilisation. Le partage de ces outils permettra d'étoffer davantage la réflexion ici engagée.

BIBLIOGRAPHIE

- Alter, C. & Adkins, C. (2006). Assessing student writing proficiency in graduate schools of social work. *Journal of Social Work Education, 42*(2), 337-354.
- Anderson, L. W., Krathwohl, D. R. & Bloom, B. S. (2001). *A taxonomy for learning, teaching, and assessing: a revision of Bloom's Taxonomy of educational objectives*. New York, NY: Longman.
- Andrade, H. & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education, 32*(2), 159-181.
- Atkinson, D. & Lim, S. L. (2013). Improving assessment processes in higher education: student and teacher perceptions of the effectiveness of a rubric embedded in a LMS. *Australasian Journal of Educational Technology, 29*(5), 651-666.
- Berthiaume, D., David, J. & David, T. (2011). Réduire la subjectivité lors de l'évaluation des apprentissages à l'aide d'une grille critériée : repères théoriques et applications à un enseignement interdisciplinaire. *Revue internationale de pédagogie de l'enseignement supérieur, 27*(2). Repéré à <http://ripes.revues.org/524>.
- Bolton, F. C. (2006). Rubrics and adult learners: andragogy and assessment. *Assessment Update, 18* (3), 5-6.

- Broad, B. (2003). *What we really value: beyond rubrics in teaching and assessing writing*. Logan, UT: USU Press Publications.
- Dennison, R., Rosselli, J. & Dempsey, A. (2014). *Evaluation beyond exams in nursing education: designing assignments and evaluating with rubrics*. New York, NY: Springer Publishing Company-LLC.
- Diederich, P., French, J. & Carlton, S. (1961). Factors in judgments of writing ability. *ETS Research Bulletin Series*, 1961(2), i-93.
- Gezie, A., Khaja, K., Chang, V. N., Adamek, M. E. & Johnsen, M. B. (2012). Rubrics as a tool for learning and assessment: what do baccalaureate students think? *Journal of Teaching in Social Work*, 32(4), 421-437. doi:10.1080/08841233.2012.705240
- Haaga, D. A. F. (1993). Peer review of term papers in graduate psychology courses. *Teaching of Psychology*, 20(1), 28-32.
- Howell, R. J. (2011). Exploring the impact of grading rubrics on academic performance: findings from a quasi-experimental, pre-post evaluation. *Journal on Excellence in College Teaching*, 22(2), 31-49.
- Johnson, R. L. & Blair, J. A. (2009). Teaching the dog's breakfast: some dangers and how to deal with them. *APA Newsletters: Newsletter on teaching philosophy*, 9(1), 2-5.
- Jonsson, A. & Svingby, G. (2007). The use of scoring rubrics: reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130-144.
- Kishbaugh, T. L. S., Cessna, S., Horst, S. J., Leaman, L., Flanagan, T., Neufeld, D. G. & Siderhurst, M. (2012). Measuring beyond content: a rubric bank for assessing skills in authentic research assignments in the sciences. *Chemistry Education Research and Practice*, 13(3), 268-276.
- Lovorn, M. & Rezaei, A. R. (2011). Assessing the assessment: rubrics training for pre-service and new in-service teachers. *Practical Assessment, Research & Evaluation*, 16(16).
- Many, J. (1999). Assessment and instruction in a graduate literacy education class: reflecting what I'm learning in what I do. *Journal of adolescent and adult literacy*, 42(7), 566-579.
- McKeown, R. (2011). Using rubrics to assess student knowledge related to sustainability: a practitioner's view. *Journal of Education for Sustainable Development*, 5(1), 61-74.
- Moni, R. W., Beswick, E. & Moni, K. B. (2005). Using student feedback to construct an assessment rubric for a concept map in physiology. *Advances in Physiology Education*, 29(4), 197-203.
- Montgomery, K. (2002). Authentic tasks and rubrics: going beyond traditional assessments in college teaching. *College Teaching*, 50(1), 34-39.
- Moskal, B. M. & Leydens, J. A. (2000). Scoring rubric development: validity and reliability. *Practical Assessment, Research & Evaluation*, 7(10).
- Pagano, N., Bernhardt, S. A., Reynolds, D., Williams, M. & McCurrie, M. K. (2008). An inter-institutional model for college writing assessment. *College Composition and Communication*, 60(2), 285-320.
- Pellegrino, J. W., Chudowsky, N. & Glaser, R. (éd.) (2001). *Knowing what students know: the science and design of educational assessment*. Washington, DC: National Academy Press.
- Petkov, D. & Petkova, O. (2006). Development of scoring rubrics for IS projects as an assessment tool. *Issues in Informing Science and Information Technology*, 3, 499-510.
- Prégent, R. (1990). *La préparation d'un cours*. Montréal : Éditions de l'École polytechnique.

- Read, B., Francis, B. & Robson, J. (2005). Gender, « bias », assessment and feedback: analyzing the written assessment of undergraduate history essays. *Assessment and Evaluation in Higher Education*, 30(3), 241-260.
- Reddy, M. Y. (2011). Design and development of rubrics to improve assessment outcomes: a pilot study in a master's level business program in India. *Quality Assurance in Education: An International Perspective*, 19(1), 84-104.
- Reddy, M. Y. & Andrade, H. (2009). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435-448. doi:10.1080/02602930902862859
- Reitmeier, C. C., Svendsen, L. K. & Vrchota, D. A. (2004). Improving oral communication skills of students in food science courses. *Journal of Food Science Education*, 3(2), 15-20.
- Rezaei, A. R. & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing*, 15(1), 18-39. doi :10.1016/j.asw.2010.01.003
- Romainville, M. (2011). Objectivité versus subjectivité dans l'évaluation des acquis des étudiants. *Revue internationale de pédagogie de l'enseignement supérieur*, 27(2). Repéré à <http://ripes.revues.org/499>.
- Ruble, M. R. (2014). Rubrics in the political science classroom: packing a serious analytical punch. *PS: Political Science & Politics*, 47(01), 199-203. doi:10.1017/S1049096513001704
- Sadler, R. (2013). Assuring academic achievement standards: from moderation to calibration. *Assessment in Education: Principles, Policy & Practice*, 20(1), 5-19.
- Seldomridge, L. A. & Walsh, C. M. (2006). Measuring critical thinking in graduate education: what do we know? *Nurse Educator*, 31(3), 132-137.
- Sevian, H. & Gonsalves, L. (2008). Analysing how scientists explain their research: a rubric for measuring the effectiveness of scientific explanations. *International Journal of Science Education*, 30, 1441-1467.
- Shay, S. (2005). The assessment of complex tasks: a double reading. *Studies in Higher Education*, 30(6), 663-679.
- Shay, S. (2008). Researching assessment as social practice: implications for research methodology. *International Journal of Educational Research*, 47(3), 159-167.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.
- Smith, L. J. (2008). Grading written projects: what approaches do students find most helpful? *Journal of Education for Business*, 83(6), 325-330.
- Stevens, D. D. & Levi, A. J. (2005). *Introduction to rubrics: an assessment tool to save grading time, convey effective feedback and promote student learning*. Sterling, VA : Stylus.
- Thibaudeau, V. (2009). 88 clefs pour identifier dans un texte un problème de logique ou d'expression de la pensée. Repéré à : <http://www.fp.ulaval.ca/88clefs/francais/liste.html> (consulté le 28 janvier 2016).
- Tierney, R. & 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), 1-10.
- Timmerman, B. E. C., Strickland, D. C., Johnson, R. L. & Payne, J. R. (2011). Development of a « universal » rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assessment & Evaluation in Higher Education*, 36(5), 509-547.

- Topping, K. J., Smith, E. F., Swanson, I. & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment & Evaluation in Higher Education*, 25(2), 149-169.
- Truemper, C. M. (2004). Educational innovations. Using scoring rubrics to facilitate assessment and evaluation of graduate-level nursing students. *Journal of Nursing Education*, 43(12), 562-564.
- Turley, E. & Gallagner, C. (2008). On the « uses » of rubrics: reframing the great rubric debate. *The English Journal*, 97(4), 87-92.

NOTES

1. Les études aux cycles supérieurs font ici référence aux études de niveau universitaire dans des programmes de niveau maîtrise (2^e cycle) et de doctorat (3^e cycle).
2. L'une des deux expertes en pédagogie, Amélie-Véronique Dubé, a dû, après avoir participé à la conception du projet et aux premiers échanges sur le contenu de la grille, se retirer du projet. Nous souhaitons toutefois la remercier chaleureusement pour sa contribution.
3. Les critères reproduits ici sont l'ensemble de ceux non spécifiques à la tâche demandée. Comme les grilles recensées étaient exclusivement disponibles en anglais, le libellé français des critères est une traduction libre.

RÉSUMÉS

Axée sur le développement de l'aptitude à la recherche scientifique et à la pensée critique, la formation universitaire aux cycles supérieurs transcende la simple acquisition de connaissances propres à un objet, une discipline ou un champ d'études. La nature complexe des apprentissages visés par ce cursus, notamment la compétence à la rédaction de textes scientifiques, pose de nombreux défis pour leur évaluation. Parmi les outils disponibles, les grilles critériées ont certes connu un essor, mais peu d'intérêt leur est accordé dans les écrits pédagogiques dans le contexte des études aux cycles supérieurs. Cet article présente le processus de développement, de mise à l'essai et de bonification d'une grille critériée coconstruite par une équipe de professeures, une étudiante et des conseillères pédagogiques de l'Université Laval (Québec, Canada) visant à évaluer la compétence à rédiger des productions scientifiques complexes aux 2^e et 3^e cycles universitaires. À la suite d'une recension des écrits et de l'exploration des outils jusqu'alors employés au sein de leur faculté, une démarche inductive et itérative a permis d'élaborer une grille d'évaluation comprenant huit critères subdivisés en sous-critères, tous accompagnés d'échelles descriptives qualitatives. Alors que les premières expérimentations de l'outil appuient son utilité et sa pertinence, les écueils ayant émergé au cours de la démarche sont discutés en regard des apports et limites de ce type de grille dans le contexte des études aux cycles supérieurs. Au-delà de la création de l'outil en soi, le caractère innovant du projet se situe dans le processus de réflexion collective qui l'a sous-tendu, s'éloignant ainsi du mode de production individuel habituellement emprunté pour la construction de grilles d'évaluation.

Graduate studies go way beyond the simple acquisition of knowledge, aiming among other things at developing critical thinking as well as research and scientific writing skills. Evaluating the latter poses real challenges. Rubrics have long been used to assess writing skills but have not yet

made their way into the evaluation practices in the context of graduate studies. This article presents a co-construction process conducted by a team of professors, students and educational counsellors from Université Laval (Quebec, Canada). It led to the development and testing of a rubrics-based evaluation tool to assess scientific writing skills among master's and doctoral students. Following a literature review and the exploration of tools previously used in one university faculty, an inductive and iterative approach helped develop an evaluation grid including eight criteria divided into sub-criteria, each being accompanied by a qualitative descriptive scale. While the first experiments of the tool support its usefulness and relevance, the pitfalls that emerged from the process are discussed in light of the contributions and limits of such tools in the context of graduate studies. Beyond the creation of the tool itself, the innovative nature of the project was also related to the process of collective thinking that was used, going beyond the individual production method usually followed for the construction of such tools.

INDEX

Mots-clés : compétence, cycles supérieurs, études universitaires, évaluation des apprentissages, grilles critériées, rédaction, rubrics

AUTEURS

ELISABETH MARTIN

Université Laval, Québec, Canada
elisabeth.martin@fsi.ulaval.ca

CATHERINE LEFRANÇOIS

Université Laval, Québec, Canada
catherine.lefrancois.2@ulaval.ca

ANNE GUICHARD

Université Laval, Québec, Canada
anne.guichard@fsi.ulaval.ca

DIANE TAPP

Université Laval, Québec, Canada
diane.tapp@fsi.ulaval.ca

LOUISE ARSENAULT

Université Laval, Québec, Canada
louise.arsenault@dGPC.ulaval.ca