

Qu'est-ce qu'une bonne recherche?

Un rapport rédigé par Ethel Méndez

Introduction

L'évaluation de l'excellence en recherche suscite de plus en plus d'intérêt depuis quelques années. Les gouvernements du Royaume-Uni, de l'Australie et d'autres pays ont en effet entrepris d'examiner comment attribuer les fonds de recherche en fonction de la qualité des travaux de recherche produits. Toutefois, on se questionne depuis longtemps sur ce qui constitue une recherche de grande qualité ou sur ce qu'est l'excellence en recherche, non seulement parce que la réponse à cette question peut influencer sur l'octroi du financement, mais aussi en raison du rôle de la recherche dans la société. La recherche permet de comprendre le monde et de concevoir de nouvelles idées et applications pour celles-ci, d'en faire l'essai et de les prouver. Elle ouvre une porte sur l'inconnu et amène des solutions à de nombreux problèmes dans le monde. Compte tenu du rôle de la recherche, il est important qu'elle soit de grande qualité. Or, si la qualité élevée ou l'excellence en recherche s'avèrent souhaitables, qu'entend-on par ces notions ? À quoi reconnaît-on l'excellence en recherche ?

Le présent document fait le survol de la documentation sur l'évaluation de l'excellence en recherche et expose certaines des lacunes dans ce domaine. Il porte essentiellement sur ce qu'est l'excellence en recherche et sur les mécanismes pour l'évaluer. Les documents analysés permettent de constater que l'évaluation de l'excellence en recherche ne repose sur aucune définition unique, critères normalisés ou méthode particulière. En effet, de nombreuses définitions des termes « recherche » et « excellence » ont été recensées. De plus, les paramètres de qualité sur lesquels doit s'appuyer l'évaluation de la recherche ne font pas consensus et les méthodes d'évaluation de l'excellence en recherche (p. ex., examen par les pairs et analyse bibliométrique) suscitent de vastes débats. Le présent document ne vise pas à établir la définition ou l'approche à privilégier. Il expose plutôt l'éventail de raisonnements et d'idées provenant des documents analysés.

Il se compose de quatre grandes sections : (1) les définitions, (2) l'objet de l'évaluation de la recherche, (3) les méthodes d'évaluation de la recherche et (4) les principaux débats touchant l'évaluation de la recherche. La première section sur les définitions présente une analyse de certains des termes servant à décrire la recherche et met en lumière quelques-unes des difficultés que pose aux fins de l'évaluation l'écart entre les différentes définitions. La deuxième section présente les éléments de l'excellence en recherche les plus fréquemment mentionnés dans les documents analysés. La troisième section décrit des méthodes et des outils utilisés afin d'évaluer l'excellence en recherche. Bien que l'examen par les pairs et l'analyse bibliométrique constituent les principales méthodes d'évaluation de la recherche, ils ne sont pas abordés dans la troisième section du docu-

ment. Ces deux méthodes faisant l'objet de très nombreux ouvrages, elles figurent dans la quatrième section, consacrée aux principaux débats présentés dans les documents analysés. La quatrième section comporte également une sous-section sur les retombées de la recherche en tant que critère d'évaluation de l'excellence. En conclusion, le rapport présente certaines réflexions sur les lacunes de la documentation existante et les questions pouvant être analysées encore davantage.

Documents analysés et limites de la démarche

La présente analyse s'inscrit dans le cadre d'une étude d'envergure sur l'excellence en recherche menée par le Centre de recherches pour le développement international (CRDI). Bien que les concepts exposés doivent servir aux chercheurs et aux établissements de recherche soucieux de la façon d'évaluer la qualité de la recherche peu importe le domaine, l'analyse documentaire est essentiellement centrée sur la recherche pour le développement. Ainsi, plusieurs des exemples dont il est question sont tirés de ce domaine.

De plus, les documents analysés en profondeur ont été choisis à partir d'un vaste éventail de documents provenant de recherches sur les bases de données de revues scientifiques à comité de lecture, de Google et Google Scholar ainsi que de demandes précises adressées aux bibliothécaires et aux chercheurs ou présentées au moyen de listes de diffusion. Ces documents ont été sélectionnés à la suite d'un examen de leurs résumés. Des membres du personnel du CRDI ont également soumis des articles.

La grande majorité des sources proviennent de chercheurs des pays développés, notamment du Royaume-Uni, de l'Australie et des États-Unis. Seulement trois documents ont été rédigés par des chercheurs établis dans les pays du Sud. L'une des limites de l'étude tient au fait que la chercheuse n'était en mesure d'analyser que des documents rédigés en anglais ou en espagnol.

1. Définitions

Dans le cadre de tout processus d'évaluation, l'objet de l'évaluation doit être défini clairement. Il en va de même de l'évaluation de l'excellence en recherche. Ainsi, dans l'optique de l'évaluation de l'excellence en recherche, qu'entend-on par « recherche » ? En quoi les distinctions entre la recherche fondamentale, appliquée, novatrice et interdisciplinaire influent-elles sur l'évaluation de la qualité de la recherche ? Quelle est la différence entre la qualité et l'excellence en recherche ? Les paragraphes suivants fournissent des définitions répondant à ces questions.

Encadré 1. L'exercice d'évaluation de la recherche (RAE) et le cadre d'excellence en recherche (REF) du Royaume-Uni

Le Higher Education Funding Council for England utilisait le RAE comme méthode d'évaluation de la qualité en recherche. Les résultats du RAE orientaient les décisions de financement des établissements d'enseignement supérieur. Cette méthode a été employée pour la première fois en 1986 et, pour la dernière fois, en 2008. Il s'agissait d'un système d'examen par les pairs portant sur trois dimensions de la recherche : l'originalité, la rigueur et l'importance. Les établissements d'enseignement supérieur étaient classés selon une échelle de cinq points. Ceux qui obtenaient le plus de points étaient désignés les meilleurs au monde au chapitre de l'originalité, de l'importance et de la rigueur (Higher Education Funding Council for England, 2011).

Certains ont critiqué le RAE, indiquant qu'il s'agissait d'une méthode non scientifique et subjective, coûteuse, exigeant beaucoup de temps, entraînant des biais et établissant des comparaisons entre des disciplines qui ne peuvent être comparées en raison des différences qui y existent dans la pondération des indicateurs de qualité (Kenna et Berche, 2011). Les chercheurs des universités, soit le public cible même de l'évaluation de cette méthode, étaient particulièrement critiques à l'endroit du RAE et des mesures d'incitation qui en découlaient de fait (publication, dans des revues scientifiques à comité de lecture, au détriment de la qualité de l'enseignement).

Le REF a été proposé comme solution de rechange au RAE. Le nouveau système, qui sera entièrement au point en 2014 et portera sur les résultats et les retombées de la recherche ainsi que sur l'environnement de recherche, a suscité beaucoup de critiques, car il met l'accent sur les paramètres et les retombées de la recherche.

Au moment de la rédaction du présent document, les critères provisoires sur lesquels s'appuie le REF laissent entendre que les comités d'examen décideront des indicateurs et de l'utilisation des paramètres. Cependant, le débat se poursuit sur les questions liées aux retombées de la recherche, à la façon de définir ce concept, de l'évaluer et de déterminer s'il faut en tenir compte.

Qualité ou excellence en recherche ?

Aucun terme ni aucune définition ne font l'unanimité pour définir ce qu'est une *bonne* recherche. Linda Graham, Robert Tijssen, Jonathan Grant, Philipp-Castian Brutscher, Susan Ella Kirk, Linda Butler et Steven Wooding soutiennent que la qualité et l'excellence, les termes les plus fréquemment employés pour désigner une bonne recherche, prennent un sens différent selon les personnes. En effet, la documentation démontre que les définitions de ces deux termes se recoupent quelque peu.

Un document du National Center for the Dissemination of Disability Research cite les descriptions données par divers auteurs à la qualité en recherche d'un point de vue presque strictement technique et méthodologique. Voici ce qu'il indique : « La *recherche de qualité* désigne le plus souvent le processus scientifique englobant tous les aspects de la conception d'une étude.

Cette notion s'applique notamment à l'évaluation faite de la correspondance entre les méthodes et les questions, la sélection des sujets, la mesure des résultats et la correction du biais systématique, du biais non systématique et des erreurs par déduction » [traduction libre] (2005, p. 2). »

Boaz et Ashby affirment que les « conceptualisations de la qualité en recherche doivent être poussées au-delà de la seule qualité méthodologique et traiter du “caractère pertinent” de la recherche » [traduction libre] (2003). Les auteurs traitent des dimensions de la « pertinence », c'est-à-dire la « pertinence des méthodes selon les objectifs de la recherche » et « la pertinence de la recherche en fonction de l'utilisation probable de ses constatations » [traduction libre] (p. 12). Ayant centré ses travaux sur la recherche pour le développement, Yule souligne aussi que l'utilité, l'accessibilité et la qualité des résultats selon les besoins des utilisateurs représentent d'importantes dimensions de la qualité en recherche (2010, p. 1).

Pour Yule, Boaz, Ashby et l'OCDE, l'utilité de la recherche constitue un élément important de la « qualité ». D'autres comme Grant et coll. estiment que la dimension de l'utilisation ou des retombées de la recherche est le critère permettant de distinguer la « qualité » de l'« excellence ». Ils n'assimilent pas les retombées à la qualité, mais voient les deux éléments comme formant ensemble le cadre d'excellence de la recherche (2010). Leurs travaux ont servi de fondement au nouveau cadre d'excellence en recherche (REF) du Royaume-Uni. (L'encadré 1 sur la page 3 présente des renseignements sur ce cadre.)

Robert Tijssen (2003) insiste sur la dimension de l'utilité de l'excellence en recherche et offre trois autres éléments permettant de différencier l'excellence et la qualité en recherche. En premier lieu, il explique que le terme « excellence » sous-entend l'idée de « dépassement par rapport à quelque chose ou à quelqu'un selon certains critères de qualité » [traduction libre] (p. 92), ce qui porte à croire que l'excellence est une caractéristique comparative. Selon ce point de vue, l'excellence ne peut être démontrée que si la recherche est comparée à une autre jugée excellente. En deuxième lieu, l'excellence en recherche s'appuie sur la « génération de nouvelles connaissances scientifiques et techniques de grande qualité » [traduction libre], ce qui fait référence à la justesse de la méthodologie de la recherche, mais aussi à son originalité. Enfin, Tijssen souligne la dimension de l'excellence rattachée à la publication et la commercialisation des produits de recherche.

Maureen O'Neil, ancienne présidente du CRDI, propose une autre façon d'entrevoir l'excellence en recherche dans le domaine de la recherche pour le développement. Voici ce qu'elle affirme : « Par excellence, on pourrait entendre une “recherche urgente présentant un défi”, ce qui signifie qu'elle doit se concentrer sur des enjeux faisant appel à une approche multidisciplinaire interpellant des réseaux internationaux provenant de différents chantiers de recherche et régimes politiques » (2002).

Dans l'ensemble, on peut avancer que les recoupements et les écarts dans les définitions de l'excellence et de la qualité en recherche témoignent de la croyance selon laquelle ces termes sont interchangeables. Néanmoins, le fait que certains auteurs comme Grant et coll. font une distinction claire entre ces termes indique qu'il faut apporter des précisions et poser des limites dans ce domaine. Le critère de comparaison fondé sur les retombées de la recherche est utile, et il semble compris par les intéressés qui connaissent le cadre d'excellence en recherche du Royaume-Uni. Toutefois, ce critère n'est pas largement accepté.

La présente analyse traite de l'excellence en recherche, interprétée d'une façon qui englobe les dimensions élargies de la recherche allant au-delà de sa rigueur méthodologique ou scientifique. Toutefois, comme la distinction entre la « qualité » et l'« excellence » n'est pas explicitée ni soulignée dans un grand nombre des documents analysés, et puisque l'objectif de l'analyse consiste à exposer l'éventail de points de vue relevés dans la documentation, aucun terme précis ne peut être privilégié dans le présent document. Celui-ci s'appuie plutôt sur une interprétation globale de la notion de qualité, et les termes « qualité » et « excellence » y sont employés de façon interchangeable.

B. Recherche fondamentale et appliquée

« Lorsque nous établissons les critères d'évaluation sur lesquels reposera notre obligation de rendre compte, nous devons reconnaître que cette démarche dépend de ce que nous croyons être le travail même de recherche » [traduction libre] (Hammersley, 2008, p. 754).

La principale différence entre la recherche fondamentale et la recherche appliquée tient à ce que la première vise l'enrichissement des connaissances et le développement de théories, tandis que l'autre a pour objectif la génération de constatations afin de résoudre un problème (Coryn, à paraître en 2013, p. 9). Stewart Donaldson et Christina Christie ajoutent que ces deux types de recherches présentent également des différences sur les plans de l'origine des questions à l'étude et des conditions dans lesquelles les travaux sont réalisés. Ils affirment que la recherche appliquée « est fondée sur un problème ou axée sur la solution et se déroule dans des conditions "réelles" et non dans l'environnement hautement contrôlé de laboratoires scientifiques traditionnels » [traduction libre] (2009, p. 2-3). Donald Stokes traite également du concept de la recherche inspirée par son utilisation, c'est-à-dire la recherche qui, à l'instar de la recherche fondamentale, vise l'enrichissement des connaissances et du savoir scientifique et qui, comme la recherche appliquée, présente des applications pratiques potentielles (Arnold, 2008, p. 2189).

Ces définitions ont de l'importance dans l'évaluation de l'excellence en recherche, car les éléments permettant d'évaluer la qualité de la recherche dépendront de l'objectif de la recherche. John Furlong et Alis Oancea, par exemple, soutiennent que les critères servant à évaluer la qualité de la recherche appliquée doivent intégrer les dimensions de l'application et de l'utilisation

de la recherche en plus de celle de la génération de connaissances, laquelle serait évaluée dans le cas de la recherche fondamentale (2005, p. 8).

C. Recherche novatrice et nouvelles disciplines

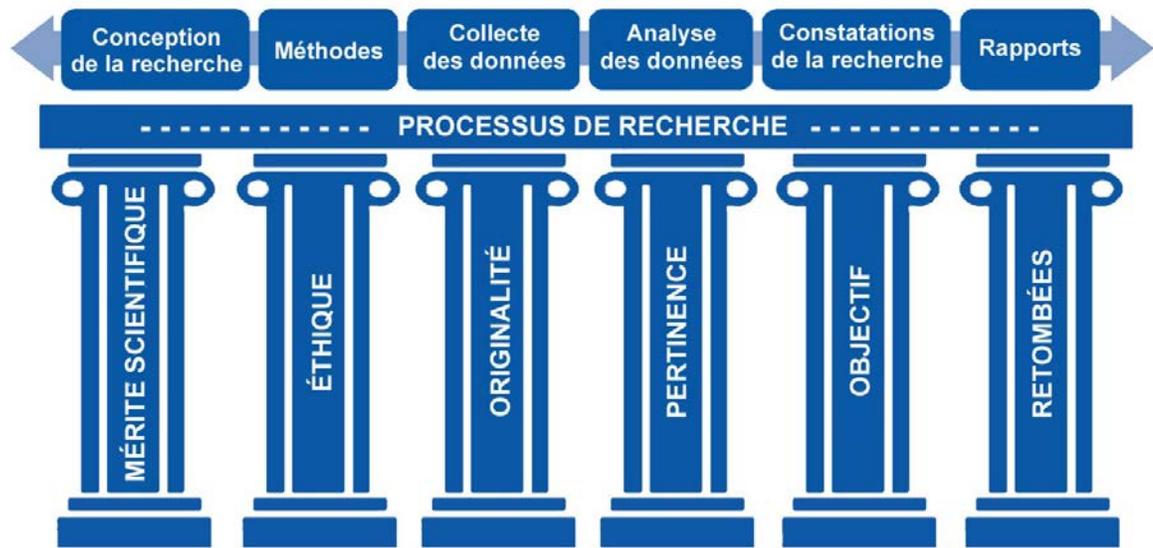
« C'est sans grande hésitation que je vous signale mon refus de recevoir quelque document de votre part. Je crois cependant qu'après avoir réexaminé le sujet, vous conviendrez que je n'ai d'autre choix. Les sujets que vous proposez en vue d'une série d'essais sur les mathématiques et la métaphysique sont si pointus qu'aucun des abonnés de notre revue n'est probablement en mesure de les comprendre » [traduction libre] (Rodrik, 2011).

Ces phrases sont tirées du billet de blogue de Danni Rodrick intitulé « Lettre de refus que j'aimerais recevoir d'une revue scientifique un jour » [traduction libre]. Il s'agit d'un extrait d'une lettre envoyée en 1821 à Charles Babbage, reconnu comme le « père d'un ordinateur », par *The Edinburgh Journal of Science*, tel que cité dans le livre *The Information: A History, a Theory, a Flood* de James Gleick. Bien que les propos de Rodrik fassent sourire, ils soulèvent une question importante au sujet de l'évaluation de la recherche novatrice. « On peut considérer que la recherche novatrice vise à combler soit des lacunes relativement importantes sur le plan théorique (par un raisonnement déductif comportant de nombreuses étapes), soit des lacunes relativement importantes sur le plan de la collecte de données expérimentales (par la mise au point de techniques tout à fait nouvelles) » [traduction libre] (Andras, 2011, p. 94).

Peter Andras s'appuie sur l'idée de science révolutionnaire de Kuhn, qui entraîne généralement un changement dans les hypothèses de base ou un changement de paradigme (Kuhn, 1962). Comme la science révolutionnaire ne s'appuie pas sur une riche tradition théorique, comme le fait la science normale, les nouvelles découvertes scientifiques ou constatations de la science révolutionnaire risquent d'être rejetées en raison de leur caractère innovateur. La communauté scientifique accorde souvent peu de crédibilité à ces découvertes, ce qui retarde leur reconnaissance, leur publication et leur application. Cette reconnaissance tardive a elle-même des répercussions sur l'évaluation de la recherche, surtout lorsque l'excellence est mesurée en fonction du nombre de publications. Andras explique ce qui suit : « En ce qui concerne la présentation des résultats de recherche dans les revues scientifiques de prestige, les travaux de la science normale sont beaucoup plus susceptibles d'être publiés, car il est bien plus probable que la communauté scientifique accepte d'emblée les résultats de ces travaux que de ceux de la science révolutionnaire, qui peuvent semer la controverse » [traduction libre] (2011, p. 97).

De même, l'Organisation de coopération et de développement économiques (OCDE) indique que « des problèmes particuliers se posent dans l'évaluation des [...] nouvelles disciplines, lesquelles sont souvent mal représentées dans les analyses bibliométriques actuelles et par les mesures qualitatives fondées sur le jugement des pairs » [traduction libre] (1997, p. 9). Lorsque

Diagramme 1. Éléments conceptuels courants de l'excellence en recherche au fil du processus de recherche



les travaux de recherche novatrice sont soumis à des examens par les pairs, il est difficile de trouver des personnes qualifiées pour en évaluer la qualité (Wooding et Grant, 2003, p. 25).

D. Recherche interdisciplinaire

Veronica Boix Mansilla, Irwin Feller et Howard Gardner décrivent la recherche interdisciplinaire comme « une forme de questionnement qui fait appel aux connaissances et aux modes de pensée propres à au moins deux disciplines ou domaines d'études établis afin d'en arriver à une percée cognitive ou pratique » [traduction libre] (2006, p. 70). Craig Stephen et Ibrahim Daibes ajoutent que, bien qu'il soit important de générer de nouvelles connaissances de pointe, dans des domaines interdisciplinaires comme la recherche mondiale en santé, l'étendue de la recherche est au moins aussi importante (sinon plus) que son degré de spécialisation (2010, p. 4). L'évaluation de la qualité de la recherche interdisciplinaire doit prendre en compte de tels éléments.

L'évaluation de la qualité de la recherche reposant sur diverses disciplines est également problématique, car les critères d'excellence ne concordent pas nécessairement d'une discipline à l'autre et présentent parfois même des contradictions. À titre d'exemple, Annette Boaz et Deborah Ashby expliquent que les paradigmes épistémologiques positivistes et constructivistes qui sous-tendent les disciplines peuvent compliquer l'évaluation de la recherche interdisciplinaire : « Les débats méthodologiques en sciences naturelles sont axés sur la recherche de la "vérité" et l'élimination du biais. En sciences sociales, l'existence de la vérité objective est souvent contestée, et on considère généralement le biais comme étant une dimension du savoir, qu'il faut souligner plutôt qu'éliminer » [traduction libre] (2003, p. 9).

Ces positions distinctes sur les méthodes, l'objectivité et la vérité produisent des points de vue différents sur les critères d'excellence fondés sur les méthodes, ce qui peut entraîner des divergences dans l'analyse de la recherche interdisciplinaire. Les essais comparatifs sur échantillons aléatoires en clinique sont par exemple souvent considérés comme étant la norme de référence en matière d'excellence en recherche dans le domaine de la santé. Cependant, l'utilisation de cette méthode en sciences sociales est de plus en plus contestée, particulièrement dans les domaines interdisciplinaires caractérisant le développement international comme la gouvernance et les technologies de l'information et de la communication.

Enfin, la sélection d'évaluateurs qualifiés et la mise à profit de leur « expertise concordante » durant l'évaluation de la recherche interdisciplinaire représentent une lourde tâche. Comme dans le cas de la recherche novatrice, la difficulté tient au fait que peu de personnes possèdent la combinaison juste de connaissances interdisciplinaires pour évaluer certains types de travaux de recherche. Ainsi, l'évaluation de la qualité de la recherche interdisciplinaire se limite souvent au cadre de certaines disciplines, ce qui pose problème, tel que mentionné plus haut (Boix et coll., 2006, p. 70). (Boix et coll. proposent de faire appel à des comités d'examen constitués stratégiquement de façon à représenter de multiples perspectives disciplinaires ainsi qu'à des « interprètes » pouvant combler les lacunes épistémiques parmi les experts. Les auteurs suggèrent également d'intégrer au sein des comités des représentants d'universités ou des agents de programme qui peuvent relever les innovations et contrer les tendances conservatrices des comités d'examen (2006, p. 70-71).)

2. L'objet de l'évaluation de la recherche : les cadres d'excellence

Le présent document a déjà traité de certaines définitions fondamentales et souvent controversées de l'évaluation de la recherche. Toutefois, abstraction faite des débats que ces définitions suscitent, quels sont les critères sur lesquels doit s'appuyer l'excellence en recherche ?

Pratiquement tous les livres et les manuels sur la recherche ou les méthodes comportent un ou plusieurs ensembles de dimensions servant à évaluer la qualité. Les documents analysés présentent au moins 30 ensembles de critères dont la longueur, le degré de précision et l'approche sur laquelle ils sont fondés varient. L'examen de ces 30 ensembles a permis de relever des critères conceptuels et spécifiques fréquemment utilisés dans l'évaluation de l'excellence en recherche. Les éléments conceptuels – l'objectif, la pertinence, l'originalité, le mérite scientifique, l'éthique et les retombées – sont abordés dans la section qui suit et illustrés au diagramme 1.

Les critères plus spécifiques liés à ces éléments conceptuels qui soulignent le processus de recherche se trouvent dans l'annexe 1.

Encadré 2. Classement des critères de qualité en recherche

Un rapport commandé par la Lavenham Suffolk Social Policy Association et le Joint Universities Council Social Policy Committee du Royaume-Uni résume les conclusions de divers ateliers, durant lesquels plus de 250 chercheurs et directeurs de recherche ont analysé et classé 35 critères de qualité en recherche. Voici certaines des conclusions :

- Les critères les plus élevés dans le classement étaient ceux de la « facilité d'accès aux résultats de recherche par les groupes cibles pertinents » et de la « conception adéquate en fonction des questions ». Le critère de la publication a été classé en fin de liste.
- Parmi les méthodes, les entrevues exhaustives ont été jugées comme présentant la qualité la plus élevée, tandis que les essais comparatifs sur échantillons aléatoires, habituellement considérés comme la méthode par excellence en recherche, ont été associés à une faible qualité.
- Sur le plan de la recherche quantitative, les critères de la validité et de la fiabilité ont été classés en tête de liste. Au nombre des autres critères figuraient la possibilité de reproduire la recherche, la généralisabilité, la rigueur, la transparence de la méthodologie, la cohérence, la formulation claire et précise des questions et la mise à l'essai appropriée des hypothèses.
- Dans le cas de la recherche qualitative, le critère de la validité a été jugé important par rapport à ceux de la fiabilité, de la possibilité de reproduire la recherche et de la généralisabilité. Une certaine importance a été rattachée à la fiabilité et à la corroboration, de même qu'à la précision et à la transparence, à la pertinence et à la participation des utilisateurs ainsi qu'à la réflexivité.
- Quant aux méthodes combinées de recherche, la majorité des chercheurs n'étaient pas en faveur de l'utilisation des mêmes critères pour les éléments quantitatifs et qualitatifs.
- Les participants ont proposé d'autres critères qui englobent la pertinence par rapport aux questions de recherche, la transparence, la nécessité d'intégrer les constatations des méthodes combinées de recherche et la justification des méthodes combinées (Becker, Bryman et Sempik, 2006).

A. Éléments conceptuels de l'excellence en recherche

Objectif. La recherche doit comporter un objectif clair sous-tendant les questions de recherche (Aagaard-Hansen et Svedin, 2009). Aagaard-Hansen et Svedin font un lien entre l'objectif et la définition précise du problème, l'adoption d'une terminologie rigoureuse, la cohérence et le savoir tiré des documents pertinents (2009). Les partisans de la recherche-action, par exemple, soutiennent que l'objectif consiste à générer des processus de transformation. Les éléments servant à concevoir la recherche, soit les questions, l'approche, les méthodes et les outils, sont ensuite choisis selon l'objectif établi (1999, tel que cité dans Groundwater-Smith et Mockler, 2007).

Pertinence. Les critères de qualité définis par l'Association des universités des Pays-Bas présentent la pertinence de deux points de vue : (1) la pertinence scientifique ou l'importance de la recherche pour l'avancement de la discipline et (2) la pertinence sociétale, soit les retombées de la recherche sur les plans sociétal et technologique (Organisation de coopération et de développement économiques, 1997). Boaz et Ashby traitent également de la pertinence en discutant de l'intérêt que présente la recherche pour les décisions stratégiques et dans la pratique : « [...] la mesure dans laquelle la recherche répond aux besoins des principaux intéressés constitue une dimension importante de la qualité » [traduction libre] (p. 12). Selon les critères établis par la RAND Corporation, la recherche doit de plus « être convaincante, utile et pertinente pour les intéressés et les responsables des décisions » (2011). (Les notions de pertinence, de retombées et d'objectif sont étroitement liées, mais elles sont traitées séparément dans les documents analysés. C'est pourquoi elles sont présentées séparément dans le présent document.)

Originalité. Dans une étude réalisée par Saul Becker, Alan Bryman et Joe Sempik, un groupe de 250 chercheurs et directeurs de recherche du Royaume-Uni ont convenu que « l'originalité tient à la façon nouvelle ou différente d'examiner des idées, des données ou des problèmes existants, beaucoup plus qu'à la génération de nouvelles données ou méthodes. L'originalité reposait aussi sur le développement de nouvelles vues et de nouveaux concepts théoriques et pratiques » [traduction libre] (2006, p. 12). Wooding et Grant soulignent également que de « définir le programme de recherche en formulant de nouvelles questions de recherche et en faisant évoluer un domaine » [traduction libre] représente l'une des principales caractéristiques de la recherche de grande qualité (2003, p. 14).

Mérite scientifique. Les critères de qualité englobent les éléments de justesse ou de rigueur des méthodes de recherche, les différentes formes de validité (ces termes varient selon le type de recherche) et l'interprétation logique des données. Plusieurs de ces éléments figurent parmi les critères spécifiques de la première annexe. Certains auteurs ont rattaché des critères précis à certaines méthodes, par exemple l'assignation occultée au hasard à des groupes de traitement dans le cadre d'essais comparatifs sur des échantillons aléatoires (Boaz et Ashby, 2003) et la capacité d'agir sans subir les effets des pathologies psychologiques et organisationnelles dans les enquêtes collaboratives (Martí et Villasante, 2009).

Éthique. Susan Groundwater-Smith et Nicole Mockler font valoir l'idée selon laquelle « l'éthique représente le premier critère de qualité dans la recherche par les praticiens » [traduction libre] (2007, p. 204). Elles considèrent l'éthique comme une « orientation donnée à la recherche qui anime profondément les personnes travaillant dans le domaine et suscitent grandement leur intérêt » [traduction libre] et proposent des lignes directrices relatives à l'éthique dans la recherche menée par les praticiens (p. 206) :

- La recherche doit respecter des protocoles et des processus éthiques.
- Les méthodes de recherche doivent être transparentes.

- La recherche doit être axée sur la collaboration.
- L'objectif et l'application de la recherche doivent mener à une transformation.
- La recherche doit pouvoir se justifier auprès de la communauté de pratique.

Retombées. Les personnes qui estiment que les retombées de la recherche constituent un élément de qualité soutiennent que la recherche doit être utile pour être jugée de qualité élevée. L'évaluation de la recherche doit donc aller au-delà des résultats attendus et porter sur les conséquences de la recherche, son influence ou les changements qui y sont rattachés. De nombreux débats concernent la façon de définir et de mesurer les retombées de la recherche et même la question de savoir si celles-ci représentent une dimension de l'excellence en recherche. Les principaux arguments des parties qui débattent de ces questions sont traités plus loin dans le document.

B. La pondération des critères de l'excellence en recherche

Différents auteurs ont accordé de l'importance à divers critères de qualité (voir l'encadré 2 pour une étude sur le classement des critères de qualité). Certains écarts dans l'importance accordée aux critères par les auteurs s'expliquent par les différentes unités d'analyse. (Les documents analysés comportent les unités d'évaluation suivantes : résultats; individus; équipes de recherche; laboratoires et établissements comme les universités; discipline scientifique; programmes gouvernementaux et organismes de financement; effectif de chercheurs du pays et autres (OCDE, 1997; Rons, De Bruyn et Cornelis, 2008; Hammersley, 2008).) Laudel et Glaser constatent qu'il faut définir ce qui compte comme un résultat si des indicateurs fondés sur des résultats seront utilisés dans l'évaluation (Laudel & Glaser, 2006). Par exemple, le REF, dans lequel les établissements d'enseignement supérieur du Royaume-Uni représentent l'unité d'analyse, traite des éléments de l'« environnement de recherche » tels les doctorats décernés, les fonds de recherche et les fonds de recherche en nature. Il ne serait pas logique de tenir compte de ces éléments dans d'autres méthodes d'évaluation prenant par exemple comme unité d'analyse les résultats de la recherche. Toutefois, dans le cas du REF, l'examen portant strictement sur les résultats de recherche pourrait inciter les chercheurs à mettre l'accent sur la publication, ce qui nuirait à l'enseignement, à l'encadrement et aux autres activités universitaires requises au niveau institutionnel.

3. Les méthodes d'évaluation de l'excellence en recherche

L'examen par les pairs ou par un comité d'experts et l'analyse bibliométrique constituent les méthodes les plus fréquemment utilisées pour évaluer les travaux de recherche. Ces deux méthodes ont soulevé de nombreuses critiques et sont examinées en détail à la section V du présent document. Toutefois, malgré la prédominance de l'examen par les pairs et de l'analyse bibliométrique, plusieurs personnes font l'essai et la promotion d'autres méthodes d'évalua-

ses désavantages, la méthode ne peut être facilement adaptée en cas de changements. De plus, elle peut perpétuer les inégalités, incite à la complaisance, ne favorise pas l'amélioration, tient strictement compte du passé et manque de crédibilité (Wooding et Grant, 2003, p. 24).

Autoévaluation : Wooding et Grant décrivent cette méthode comme un système où « les établissements, les départements ou les personnes s'évaluent eux-mêmes » [traduction libre] (p. 20). Des chercheurs et des administrateurs de la recherche du Royaume-Uni se sont montrés critiques envers cette méthode, affirmant qu'elle manque de crédibilité, constitue une démarche lourde, est subjective, entraîne des biais, permet la manipulation des données et la surévaluation, n'offre aucune comparaison et est trop centrée vers l'intérieur. Toutefois, cette méthode présente les avantages d'être simple, peu coûteuse et souple, de tenir compte des réalités locales, de favoriser l'apprentissage, la réflexion et la responsabilité et de mettre à contribution les chercheurs (Wooding et Grant, 2003, p. 27).

Analyse sémantique informatisée : Le projet EERQI, du 7^e programme-cadre, de l'Union européenne porte sur la possibilité d'utiliser l'analyse sémantique informatisée pour relever dans les documents les principaux indicateurs sur lesquels fonder l'évaluation de la qualité (Bridges, 2009, p. 507). Bridges indique que, même si l'objectif de cette démarche est de dégager les principaux indicateurs dans les documents, une méthode informatisée ne peut convenir, car l'analyse ainsi menée n'offre pas un point de vue éclairé par les connaissances et l'expérience de spécialistes.

Analyse comparative (système fondé sur des mesures) : Le Fonds monétaire international (FMI) a réalisé une évaluation de la qualité de ses travaux de recherche en 2011. Cette évaluation tenait compte de la fréquence des citations et des publications et a été menée dans huit établissements semblables responsables des politiques. Ceux-ci ont été classés selon la qualité de leurs travaux de recherche. L'évaluation présente les limites de l'analyse comparative en tant que « méthode qui continue d'entraîner des biais probables liés aux barèmes différentiels, à l'hétérogénéité et aux missions différentes des huit établissements » [traduction libre] (Aizenman, Edison, Leony et Sun, 2011, p. 19). (L'évaluation permet de conclure que le nombre d'auteurs qui se citent eux-mêmes est élevé au FMI, plus que dans les autres établissements étudiés, ce qui peut fausser les perceptions sur les retombées de la recherche. Étant donné sa taille, le FMI compte relativement peu de publications.)

Tijssen (2003) se penche également sur l'utilisation des tableaux de bord, qu'il définit comme un « mode d'analyse comparative fondé sur des ensembles structurés de données quantitatives pertinentes portant sur divers aspects de l'excellence en recherche [...] dont le but est précisément d'élargir systématiquement (sic) l'éventail d'indicateurs quantitatifs en vue d'un examen et d'une comparaison plus poussés, lesquels sont menés à des fins d'analyse et de représentation » [traduction libre] (p. 96).

Étude de cas : Dans une étude sur l'histoire de l'évaluation de la recherche, la RAND Corporation (2009) a constaté que la méthode de l'étude de cas a été la plus fréquemment employée en recherche pour évaluer l'apport des travaux de recherche à la société (Marjanovic, Hanney et Wooding, 2009, p. 17). Grant et coll. décrivent l'étude de cas comme une méthode qui s'appuie sur la présentation des retombées d'une recherche à partir de données probantes quantitatives et qualitatives. L'évaluation ainsi menée repose donc sur l'opinion d'experts. Le cadre de la qualité en recherche de l'Australie a été conçu comme un système d'évaluation des retombées fondé sur l'étude de cas. Il n'a toutefois pas été mis en place en raison du changement de gouvernement de l'Australie (Grant, Brutscher, Kirk, Butler et Wooding, 2010, p. 7-20).

Indicateurs : Grant et coll. examinent le système de cotation des retombées établi par la RAND Corporation pour l'Arthritis Research Campaign (ARC). Il s'agit d'un système fondé sur des indicateurs qui comprend 187 questions dont la réponse est soit « oui » soit « non » et qui vise à recueillir des données sur les subventions de recherche octroyées par l'ARC. Cette démarche permet de connaître les retombées des travaux de recherche réalisés et met en évidence celles qui sont les plus vastes, par exemple l'amélioration des capacités en recherche. Cette méthode n'exige pas beaucoup de temps. Toutefois, les réponses se limitant à « oui » ou « non », il est difficile de recueillir divers points de vue et des explications permettant de comprendre les constatations (Grant et coll., 2010, p. 21-31).

Méthodes combinées : L'approche d'évaluation de la recherche en contexte des Pays-Bas (ERiC) est une combinaison de méthodes comprenant l'autoévaluation, la collecte de données selon le secteur, la mise en correspondance des résultats et la consultation auprès des principaux intéressés au moyen d'entrevues et de sondages (une forme d'examen par les pairs). Une grande charge de travail est associée à ce type d'évaluation, mais les nombreuses sources de données et étapes nécessaires permettent de bien évaluer la qualité et de fournir des éléments dont les établissements peuvent tirer des leçons (Grant et coll., 2010, p. 47-58).

Délibération et recherche d'un consensus : Le National Center for the Dissemination of Disability Research (2005) a formulé les observations suivantes sur l'importance de la délibération et de la recherche d'un consensus pour évaluer la qualité en recherche : « La recherche d'un consensus parmi les chercheurs est l'un des modes d'évaluation de la qualité les plus respectés. Parmi les stratégies employées pour atteindre un consensus figurent la présentation d'énoncés de position, la tenue de conférences, l'examen par les pairs et l'examen systématique » [traduction libre] (2005).

Bridges (2009) ainsi que Stephen et Daibes (2010) soulignent également l'importance du consensus et proposent une forme d'examen par les pairs dans le cadre duquel de multiples intéressés provenant de nombreuses disciplines décident des critères (s'il y a lieu). Selon ce modèle, les processus d'examen et de décision au sujet de la qualité présentent potentiellement plus de dynamisme que celui de l'examen de documents pur et simple par un comité, qui n'en-

Bridges (2009) ainsi que Stephen et Daibes (2010) soulignent également l'importance du consensus et proposent une forme d'examen par les pairs dans le cadre duquel de multiples intéressés provenant de nombreuses disciplines décident des critères (s'il y a lieu). Selon ce modèle, les processus d'examen et de décision au sujet de la qualité présentent potentiellement plus de dynamisme que celui de l'examen de documents pur et simple par un comité, qui n'entraîne aucune interaction entre les experts. S'inspirant du livre *Love's Knowledge* de Martha Nussbaum, Bridges indique que la souplesse, la capacité d'adaptation et l'ouverture sont essentielles à la délibération efficace.

4. Principaux débats touchant l'évaluation de l'excellence en recherché

Compte tenu des pressions qui s'exercent sur les fonds de recherche et de l'importance accordée à la gestion fondée sur les résultats ainsi qu'à l'utilisation efficace des ressources, particulièrement dans le secteur de l'éducation au Royaume-Uni et en Australie, une attention accrue est portée à la qualité des travaux de recherche et, donc, à la façon de l'évaluer. Cette situation a déclenché de nombreux débats sur l'évaluation de l'excellence en recherche. La présente section traite de trois de ces débats, soit ceux qui portent sur (1) les retombées de la recherche comme dimension de l'excellence ainsi que (2) l'examen par les pairs ou par un comité d'experts et (3) l'analyse bibliométrique en tant que méthodes d'évaluation de l'excellence en recherche.

A. Le débat sur les retombées

L'évaluation des retombées en recherche alimente les discussions depuis les années 1960 (Marjanovic, Hanney et Wooding, 2009). Cependant, la décision du Royaume-Uni de remplacer l'exercice d'évaluation de la recherche (RAE) (qui portait sur l'originalité, la rigueur et l'importance de la recherche) par le cadre d'excellence en recherche (REF) (qui tient compte des retombées) a relancé le débat sur le sujet. L'étude des retombées en tant que critère d'évaluation de la qualité en recherche se résume à trois grandes questions : En quoi consistent les retombées ? Comment doit-on les mesurer ? Représentent-elles une dimension de l'évaluation des recherches ?

En quoi consistent les retombées ? Dans une analyse documentaire approfondie menée en 2003 et axée sur les formes de retombées en recherche, Sandra Nutley, Janie Percy-Smith et William Solesbury ont relevé peu de théories ou de discussions sur la nature des retombées. Ils ont par contre recensé les formes suivantes de retombées en recherche (2003, p. 11) :

- Changements dans l'accès à la recherche
- Changements touchant la mesure dans laquelle les travaux de recherche sont étudiés, cités en référence ou consultés
- Citations dans les documents

- Changements dans les connaissances et la compréhension d'une question
- Changements dans les attitudes et les croyances
- Changements dans les comportements

En s'appuyant sur les travaux d'Huberman, les auteurs établissent une distinction entre l'utilisation *conceptuelle* de travaux de recherche qui, soulignent-ils, « entraîne des changements dans le degré de compréhension d'une question, les connaissances et les attitudes » [traduction libre] et l'utilisation *déterminante* des travaux de recherche qui « amène des changements dans la pratique et l'élaboration des politiques » [traduction libre] (Nutley, Percy-Smith et Solesbury, 2003, p. 11). Depuis la publication de leur étude, les discussions se sont intensifiées, ce qui a mené à l'apparition de nouvelles définitions et façons d'entrevoir les retombées en recherche.

Grâce à des fonds du Higher Education Funding Council for England (HEFCE), la London School of Economics (LSE) a lancé un projet pluriannuel qui « vise à démontrer comment la recherche universitaire en sciences sociales génère des retombées sur les politiques publiques, contribue à la prospérité économique et informe le public des enjeux stratégiques et des changements économiques et sociaux » [traduction libre] (2011). Dans un manuel ayant pour but d'aider les chercheurs à maximiser l'incidence de leurs travaux, le groupe décrit les retombées en recherche comme « une occasion d'exercer une influence, qui diffère toutefois d'un changement dans les résultats ou dans les activités en tant que conséquence de cette influence, mais qui constitue tout de même un changement dans les retombées sociales » [traduction libre] (Groupe des politiques publiques de la LSE, 2011, p. 21). Le document classe les retombées en recherche comme suit :

- les retombées dans le monde universitaire, qui correspondent à l'influence exercée par les travaux de recherche sur les parties intéressées des établissements ou universités. Ils sont mesurées par un relevé des citations dans les travaux des autres chercheurs; et
- les retombées externes, qui consistent en l'influence des travaux de recherche à l'extérieur du milieu de l'enseignement supérieur, par exemple les gens d'affaires, les gouvernements et la société civile. Ils sont mesurés par le recensement des références dans « la presse professionnelle ou les documents gouvernementaux ou par la couverture dans les médias de masse » [traduction libre] (Groupe des politiques publiques de la LSE, 2011, p. 5).

Eveliina Saari et Katri Kallio présentent également une définition des retombées en recherche. Se fondant sur la définition de Donovan (2008) selon laquelle la recherche « enrichit le capital social, économique, naturel et culturel d'un pays » [traduction libre], elles avancent que la recherche de grande qualité doit générer des retombées sur trois plans : (1) collecte et combinaison des connaissances en recherche; (2) résolution des problèmes du client; (3) réponse à une question sociétale (Saari et Kallio, 2011, p. 232).

Encadré 3. Les méthodes d'évaluation des retombées en recherche

Dans un examen des pratiques à l'échelle internationale, Grant et coll. se sont penchés sur quatre méthodes d'évaluation des retombées en recherche actuelles sur lesquelles pourrait s'appuyer le REF du Royaume-Uni. Les chercheurs ont analysé le cadre d'accessibilité et de qualité en recherche (RQF) de l'Australie qui se fonde sur l'étude de cas, le système de cotation des retombées établi par RAND pour l'Arthritis Research Campaign au Royaume-Uni qui repose sur des indicateurs, l'outil de cotation pour l'évaluation des programmes (PART) des États-Unis qui s'appuie sur l'autoévaluation et l'approche d'évaluation de la recherche en contexte (ERiC) des Pays-Bas qui combine certains éléments des autres méthodes.

Dans leur rapport, ils concluent que tous ces systèmes sont lourds et entraînent des éléments indésirables. Par contre, la méthode fondée sur l'étude de cas du RQF de l'Australie offre les solutions les plus avantageuses en vue du REF (Grant et coll., 2010, p. 69).

Comment doit-on mesurer les retombées ? Les questions concernant la mesure des retombées en recherche dépassent le cadre des difficultés associées à l'absence d'une définition claire. Yates explique qu'en Australie, « on comprend actuellement bien les retombées des travaux de recherche en éducation; on conçoit donc les travaux de recherche afin qu'ils servent aux responsables des politiques et répondent aux enjeux du système » [traduction libre] (Yates, 2005, p. 395). Dans un tel contexte, où l'on tient compte des retombées au-delà du milieu universitaire, la discussion porte sur la façon de mesurer les retombées et la question de savoir qui doit le faire. Yates souligne que les valeurs influent sur ce que l'on considère être des retombées positives et négatives. Par exemple, il est communément admis que Hitler a eu une influence fortement négative sur la société (2005). Toutefois, il n'est pas toujours possible de juger aussi facilement de la situation et certains points de vue s'opposent parfois sur le caractère positif ou négatif des retombées d'une recherche. Cette réalité met en lumière les défis que posent les paradigmes sous-tendant l'évaluation, ce que Yates appelle le problème de la diversité des valeurs (2005, p. 398, 401).

Le débat sur la façon de mesurer les retombées oppose d'un côté ceux qui estiment que l'analyse bibliométrique fournit de bons indicateurs et, de l'autre, ceux qui cherchent d'autres méthodes de mesure. À titre d'exemple, du côté des partisans de l'analyse bibliométrique, R. Tatavarti, N. Sridevi et D. P. Kothari décrivent le facteur des retombées comme « un décompte des citations dans les revues des domaines scientifique, humanitaire et technique [...] souvent utilisé en tant qu'outil pour évaluer l'importance relative d'une revue de recherche dans le domaine visé » [traduction libre] (2010, p. 1015). Ils soutiennent que le facteur des retombées figure parmi les mesures quantitatives les plus rigoureuses permettant d'évaluer la qualité en recherche. Comme mentionné ci-dessus, l'évaluation des travaux de recherche au FMI réalisée en 2011 reposait sur une analyse bibliométrique, dans laquelle le

relevé des citations fournissait une indication des retombées à court terme et les publications, des retombées à long terme (Aizenman, Edison, Leony et Sun, 2011). De l'autre côté du spectre se trouvent les études, comme celle décrite dans l'encadré 3, qui traitent d'autres méthodes pour relever les retombées en recherche.

En dépit des arguments en faveur de l'analyse bibliométrique, le principal problème dans la mesure des retombées repose dans la difficulté d'attribuer les retombées à la recherche. Hammersley affirme de manière plutôt catégorique que « les résultats d'une recherche ne peuvent être calculés » [traduction libre] (2008). Citant le concept de l'infiltration des connaissances de Carol Weiss, Furlong et Oancea indiquent également qu'il peut falloir plus de 20 ans pour combler l'écart entre les connaissances et leur application dans les politiques et que la mesure des retombées à court terme pose donc problème (2005). (Selon le concept de l'« infiltration des connaissances » avancé par Carol Weiss, les travaux de recherche n'ont souvent pas des retombées directes sur les politiques et la pratique. Les idées naissant de la recherche s'« infiltrent » plutôt tranquillement dans l'environnement où évoluent les politiques et la pratique, modifiant ainsi les principes admis et soulevant des questions sur une longue période.) Frances Seymour, directrice générale du Center for International Forestry Research (CIFOR), traite de ce sujet dans le billet de son blogue intitulé « Est-ce que la quête effrénée de retombées compromet la recherche ? » [traduction libre.] Voici ce qu'elle y écrit : « [...] ce qui complique cette situation, c'est la difficulté d'attribuer les retombées à la recherche, surtout dans le cas de la recherche sur les politiques et de celle touchant la gestion des ressources naturelles, et moins (par exemple) dans le cas de la recherche visant à accroître le rendement d'une culture céréalière donnée » [traduction libre] (Seymour, 2011).

Bien que leurs travaux ne le mentionnent pas explicitement, ces auteurs font allusion à la difficulté de déterminer l'origine de certains phénomènes dans des environnements complexes, où existent différents systèmes et où de multiples causes engendrent des retombées, par de nombreuses voies.

Doit-on tenir compte des retombées pour évaluer la qualité des recherches ? Les échanges sur la définition et la mesure des retombées s'appuient sur le principe selon lequel il faut tenir compte des retombées dans l'évaluation de l'excellence en recherche. Or, ce principe ne fait pas l'unanimité. Par exemple, Hammersley (2008) dénonce l'émergence d'un « modèle influencé par l'investissement » en recherche. Il s'agit d'un modèle dans lequel « on demande que les projets [de recherche] produisent un rendement et que des preuves des résultats et de l'efficacité de la génération de connaissances dans le secteur soient fournies » [traduction libre] (p. 753). Ce modèle est contraire à ce qu'il appelle un « modèle de mécénat de l'État », selon lequel « la production de connaissances était traitée comme un avantage en soit » [traduction libre] (p. 753). Le rendement incalculable sur l'investissement, les conséquences négatives sur la liberté des chercheurs et l'effet tardif des travaux de recherche sur les politiques comptent parmi les rai-

sons pour lesquelles il s'oppose au modèle. Il soutient toutefois que « le modèle influencé par l'investissement est peut-être également contraire aux engagements faisant de la recherche un moyen d'orienter les débats publics sur des enjeux stratégiques, puisqu'il intègre au cadre de la recherche des idées au sujet des résultats prévisibles » [traduction libre] (Hammersley, 2008, p. 753).

Produisant des écrits dans le domaine de l'éducation en Australie, Yates avance des arguments semblables et ajoute que, compte tenu des liens entre les processus d'évaluation et le financement, il n'est peut-être pas pertinent d'accorder de l'importance aux retombées en recherche, car cela peut amener les établissements de recherche à se concentrer sur des sujets qui produiront rapidement des résultats, ce qui entrave la liberté du monde universitaire.

Le débat sur l'examen par les pairs

« *L'examen par un comité d'experts (y compris l'examen par les pairs) : Un système dans lequel des experts portent un jugement professionnel sur le rendement de personnes ou de groupes, pendant un cycle donné, ou sur leur rendement probable à l'avenir. Il peut s'agir de groupes de chercheurs, de départements ou de consortiums. L'évaluation peut être menée entièrement par des pairs ou mettre à contribution d'autres experts tels des représentants de groupes d'utilisateurs, des profanes et des experts en finance* » [traduction libre] (Wooding et Grant, 2003, p. 20).

L'examen par les pairs est la méthode la plus souvent mentionnée dans l'évaluation de l'excellence en recherche (Jongbloed et coll., 2000; CRE, 2000 tel que cité dans Rons, De Bruyn et Cornelis, 2008). Toutefois, son efficacité et son efficacité sont l'objet de débats houleux. Les sections qui suivent décrivent certains des principaux aspects négatifs de l'examen par les pairs :

Compétition malsaine : Les auteurs des documents analysés soutiennent que l'examen par les pairs fait naître une compétition malsaine et de la rivalité entre les pairs et incite parfois même des chercheurs à retarder la publication des travaux d'autres chercheurs pour faire paraître leurs propres travaux en premier (Rowland, 2002; Roebber et Schultz, 2011). Petit-Zeman ajoute que les chercheurs utilisent cet « examen de la concurrence » pour « éliminer leurs rivaux » [traduction libre] (2003, 8^e par.). D'autres auteurs soulignent aussi que le biais et le plagiat sont problématiques.

Subjectivité : Tijssen indique que l'examen par les pairs « repose sur les opinions qualitatives de quelques éminents experts, ce qui restreint la portée de la démarche et entraîne un risque de biais attribuable à sa subjectivité » [traduction libre] (2003, p. 101). En outre, certains dénoncent l'incohérence des comités d'examen dans le monde universitaire. Selon Tijssen, la subjectivité sous-tend tous les processus d'évaluation de la recherche, même l'analyse bibliométrique. Par exemple, avant qu'on puisse effectuer le dénombrement des citations d'un document, celui-ci est soumis à un examen des pairs, lesquels déterminent s'il mérite d'être publié. Ce n'est qu'après l'examen, une démarche subjective, que le nombre de citations peut être recensé et que l'analyse des autres facteurs bibliométriques peut avoir lieu.

Expertise des évaluateurs : Les auteurs des documents analysés soulèvent des questions au sujet des évaluateurs, de leurs qualifications et capacités ainsi que des personnes décidant qui possède les compétences voulues pour faire partie d'un comité d'examen ou être un évaluateur. Tout comme la diversité des valeurs constitue un problème dans l'évaluation des retombées en recherche, les biais philosophiques ou méthodologiques sont également une source de préoccupation (Yates, 2005, p. 397).

Longueur et coût élevé du processus : Lawrence O'Gorman (2008, p. 3) émet des critiques sur la longueur du processus d'examen par les pairs et le temps considérable exigé et avance que ces facteurs dissuadent les chercheurs d'accepter d'examiner des documents. L'examen par les pairs représente également une démarche coûteuse. Comme mentionné dans l'encadré 1 (voir la page 3, les critiques à l'endroit du RAE du Royaume-Uni dénonçaient notamment son coût élevé, ce qui a mené à la proposition d'un système fondé sur des mesures. Mark Ware traite de ce problème et présente des solutions de rechange créatrices adaptées aux besoins du marché pour faciliter l'examen par les pairs, par exemple la possibilité de créer un système de devise virtuelle dans lequel les chercheurs accumuleraient des crédits en fonction des documents qu'ils examinent (2011, p. 48).

Conservatisme : Comme l'indiquait une section antérieure, les chercheurs des domaines nouveaux et novateurs arrivent difficilement à trouver des personnes ayant l'expertise voulue pour évaluer leurs travaux. O'Gorman avance l'idée selon laquelle une seule façon de penser domine souvent les petits domaines de recherche, ce qui complique encore plus le processus d'examen pour les chercheurs qui travaillent à ouvrir de nouveaux horizons (2008, p. 5).

Objectif imprécis : Les méthodes d'examen par les pairs sont très répandues, mais leur objectif est souvent imprécis. Ware souligne ce qui suit : « Pour commencer, il est difficile de décider comment améliorer le système [d'examen par les pairs] sans s'entendre sur ses objectifs (2006). Cette démarche vise-t-elle à sélectionner les meilleurs documents à publier dans une revue scientifique ? À minimiser (à défaut d'éliminer) la fraude et les autres comportements répréhensibles ? À améliorer la qualité des documents publiés ? À améliorer la qualité des travaux de recherche ? À agir comme un filtre en écartant les travaux de mauvaise qualité ou en décidant non pas si un document sera publié, mais plutôt dans quelle publication il le sera ? » [traduction libre] (p. 33.)

Il s'agit d'une question importante, car la validité de l'examen dépend de son objectif. Si l'examen n'a pas de visée précise, les résultats peuvent être mal interprétés.

D'autres critiques sur l'examen par les pairs relevées dans l'analyse documentaire concernent l'incapacité de détecter la fraude, le manque de transparence dans la sélection des membres des comités, la non-garantie de l'application de normes méthodologiques élevées, la lenteur et l'incompétence rattachées au processus ainsi que l'absence d'analyse de la pertinence des recherches sur le plan de l'élaboration des politiques (Boaz et Ashby, 2003; Yates, 2005).

Malgré les lacunes relevées, de nombreux chercheurs estiment que l'examen par les pairs est l'élément fondamental de l'évaluation de la recherche, car aucune autre méthode ne permet d'évaluer la qualité d'une recherche si ce n'est par la lecture du document de recherche et la formulation d'un avis éclairé (Becker, Bryman et Sempik, 2006, p. 19). À titre d'exemple, un rapport commandé par un groupe d'organismes de financement dans le cadre d'un examen de l'évaluation en recherche au Royaume-Uni décrit les conclusions de neuf ateliers au cours desquels 142 chercheurs et directeurs de la recherche provenant d'établissements d'enseignement supérieur se sont réunis pour discuter de la qualité en recherche et de leurs vues à l'égard de différentes méthodes d'évaluation. Les participants devaient classer quatre types de systèmes d'évaluation en recherche, soit des systèmes fondés sur des mesures, sur des cotes de rendement passé, sur des autoévaluations et sur des examens par un comité d'experts. Les conclusions démontraient que l'examen par un comité présentait « les points les plus positifs et les points les plus négatifs » [traduction libre] par rapport aux autres systèmes (Wooding et Grant, 2003, p. 25). Cependant, lorsque les participants étaient répartis en groupes et devaient concevoir le système d'évaluation idéal en recherche, 22 des 29 groupes ont fondé leur système sur un processus d'examen par les pairs ou par un comité d'experts.

Reconnaissant les lacunes de l'examen par les pairs, de nombreux établissements et chercheurs ont décidé de faire l'essai de différents types de processus d'examen afin de réduire le biais et d'accroître l'efficacité de la démarche. Ware énumère certains de ces processus : attribuer une cote aux examens menés et soumettre des commentaires aux examinateurs, fournir des listes de vérification ou des gabarits aux examinateurs, former les examinateurs et leur offrir des récompenses (2011, p. 34). Quelques nouveaux modèles d'examen par les pairs sont également mis à l'essai, par exemple la rétribution par l'auteur (Rowland, 2002), l'examen par les pairs après la publication, l'examen ouvert et l'examen en cascade (Ware, 2011). (Pour la rétribution de l'auteur, Rowland lance une mise en garde en indiquant qu'une telle formule désavantagerait certains chercheurs, dont de nombreux établis dans les pays du Sud. Dans l'examen en cascade, on demande à l'auteur dont le document a été rejeté s'il « souhaite joindre à son document les rapports des examinateurs afin de le soumettre ailleurs, ce qui réduit le travail d'examen requis par la revue spécialisée où le document sera ensuite soumis » [traduction libre] (Ware, 2011, p. 35).) La technologie a aussi contribué à améliorer le processus, car Internet a facilité les échanges et le suivi des articles, des examens et des examinateurs. Néanmoins, de nombreux auteurs affirment que la technologie offre beaucoup plus de possibilités de parfaire le système.

Faisant référence au débat global sur l'examen par les pairs, Ware indique que « plutôt que de voir la fin d'une méthode [...], nous assistons à de vastes débats, à des essais et à l'adoption de mesures novatrices dans un nombre croissant de revues scientifiques. On peut soutenir que, loin d'être dans une situation de crise, l'examen par les pairs s'ouvre à de nouvelles formules et n'a jamais provoqué une telle effervescence » [traduction libre] (2011, p. 49).

Le débat sur l'utilisation des mesures

L'utilisation des mesures dans l'évaluation de l'excellence en recherche est fortement contestée. Selon Andras, les mesures comprennent ce qui suit : « [...] le nombre de publications dans une liste d'endroits définie (revues scientifiques et conférences) par des personnes, des départements ou des universités, le nombre de fois où ces publications sont citées pendant une période donnée, les indices calculés à l'aide de ces données (indice h), les paramètres découlant des citations et des tableaux d'attribution de paternité de travaux et les mesures de parts du marché » [traduction libre] (Moed, 2005; Moed et coll., 2004; HEFCE, 2008a rapporté dans Andras, 2011, p. 90).

Les mesures sont considérées comme étant des solutions de rechange moins coûteuses et lourdes et plus objectives et transparentes que l'examen par les pairs (Wooding et Grant, 2003, p. 22-23). Coryn explique que le nombre de citations recensées est perçu comme un indice de rendement dans l'évaluation en recherche, selon le principe voulant que « les travaux fréquemment cités puissent être considérés méritoires ou importants, puisque la mesure dans laquelle ils sont utilisés (c'est-à-dire cités) représente une indication de leur apport à l'enrichissement des connaissances » [traduction libre] (2006, p. 115). Cependant, même les partisans de l'utilisation des mesures conviennent que celles-ci présentent de nombreuses limites et que beaucoup d'améliorations sont nécessaires pour éliminer leurs faiblesses (Andras, 2011, p. 103). Les paragraphes qui suivent décrivent certains des principaux points faibles des mesures.

Validité : L'argument probablement le plus important à l'encontre des mesures est celui voulant qu'elles ne constituent pas de véritables indicateurs de la qualité, mais qu'elles en soient plutôt des indices. Bridges établit une distinction entre les indicateurs de type A et de type B. Selon lui, les indicateurs de type A sont des mesures de la qualité intrinsèque et ceux de type B, des indices de qualité. Dans le cas d'un diamant, il soutient qu'un indicateur de type A pourrait porter sur l'éclat du diamant tandis que le prix constituerait un indicateur de type B. L'auteur affirme que l'évaluation fondée sur les indicateurs de type B peut être faussée; dans son exemple du diamant, le prix fluctue pour des raisons qui ne reposent pas seulement sur la qualité du diamant. De même, il estime que les mesures représentent des indices de la qualité en recherche. Bridges soutient ce qui suit : « Dans tout système d'évaluation de la qualité, il faut voir à ce que cette évaluation soit fondée sur des indicateurs de type A, sur les caractéristiques intrinsèques à la qualité du travail plutôt que sur des indicateurs qui ne sont pas inhérents à la qualité, dont l'utilisation faussera la pratique sans pour autant améliorer la qualité » [traduction libre] (2009, p. 511).

Boaz et Ashby soulèvent la même question en indiquant qu'on émet un « postulat erroné lorsqu'on affirme que tous les travaux de recherche publiés dans des revues scientifiques ou cités par d'autres sont justes, fiables, valides, exempts de biais, non frauduleux ou de qualité suffi-

sante » [traduction libre] (2003 tel que cité dans National Center for the Dissemination of Disability Research, 2005, p. 2). Après tout, le nombre de citations recensées et les autres facteurs d'analyse bibliométrique dépendent de l'examen par les pairs et d'autres décisions subjectives. En effet, un document est d'abord soumis à un tel examen avant d'être accepté par une revue scientifique, présenté à une conférence ou utilisé sous d'autres formes qui influent sur l'analyse bibliométrique. Les décisions des revues scientifiques de publier le document s'appuient également sur des facteurs n'ayant pas trait à sa qualité même, par exemple sur la réputation de l'auteur ou la popularité du sujet. Ware indique également que « presque tout compte rendu de travaux originaux, même peu rigoureux, finira par être publié dans une revue scientifique si l'auteur est assez persévérant » [traduction libre] (2011, p. 29). Par conséquent, l'analyse bibliométrique peut devenir non pas une indication de la qualité de la recherche, mais plutôt une mesure de la « quantité », par exemple un facteur démontrant la volonté du chercheur de faire publier ses travaux ou le taux élevé d'acceptation d'une revue scientifique (COSEPUP, 1999 tel que cité dans National Center for the Dissemination of Disability Research, 2005, p. 2).

Les essais cliniques de médicaments fournissent un exemple de l'indication erronée de la qualité pouvant découler de l'analyse bibliométrique. Ernest House (2008) affirme que les sociétés pharmaceutiques ont une telle mainmise sur l'industrie qu'elles peuvent se servir des publications comme elles l'entendent pour satisfaire leurs besoins. Il indique que ces sociétés sont trois fois plus susceptibles de présenter des résultats d'essais positifs, bien que parfois incomplets, aux fins de publication et rédigent même parfois les rapports pour les chercheurs afin que les études paraissent favorables (p. 418). Lorsque le nombre de publications fournit une indication de l'excellence en recherche, il est possible qu'un document de recherche soit considéré de grande qualité alors que, dans les faits, le nombre de publications et l'analyse bibliométrique rendent compte uniquement de la capacité d'une société pharmaceutique de déjouer le système. House signale que les éditeurs de revues scientifiques, comme Jeffrey Drazen du *New England Journal of Medicine*, reconnaissent maintenant les problèmes liés à la publication des résultats d'essais cliniques et proposent de suivre de nouvelles lignes directrices pour la publication, qui comprennent les mesures suivantes : communication de la totalité des données, transparence concernant les intérêts financiers, soumission des plans originaux des essais (afin de relever les changements et les modifications) et déclaration signée des auteurs indiquant qu'ils ont bel et bien rédigé le rapport (House, p. 423).

Par ailleurs, traitant de la recherche en éducation, Rons, De Bruyn et Cornelis dénoncent l'étroitesse et les lacunes de l'analyse bibliométrique ainsi que les risques qui se posent lorsque le financement est lié à l'évaluation. Les auteurs soutiennent que l'analyse bibliométrique ne prend pas en compte certains aspects de la qualité tels : « [...] les aspects de la formation et les liens de vos chercheurs à l'échelle internationale, la culture de recherche dans laquelle sont baignés les étudiants, les retombées de la recherche sur la société, l'économie, la culture, les poli-

tiques gouvernementales, l'aide au développement, la culture scientifique des citoyens et les organismes à but non lucratif » [traduction libre] (2008, p. 46).

Citations négatives : Les travaux de Bridges (2009) de même qu'un document du Higher Education Funding Council for England (2011) soulèvent un problème concernant les citations négatives, soit les documents cités en raison de leur faible qualité. Certains diront que ces documents contribuent à l'enrichissement des connaissances. Toutefois, l'évaluation visant à trouver des recherches de grande qualité peut être faussée si elle intègre comme mesure le nombre de documents ayant été cités, car ceux-ci peuvent faire l'objet de nombreux commentaires défavorables en raison de leur faible qualité. De même, au sujet du nombre de téléchargements d'un document de recherche comme mesure de la qualité, Bridges (2009) rappelle qu'un document peut être jugé de mauvaise qualité après avoir été téléchargé. Les citations négatives, comme les autres éléments négatifs mentionnés ci-dessus, remettent en cause la validité des mesures comme indicateurs de la qualité.

Incitation négative : Sastry et Bekhradnia (2006), Bridges (2009) et d'autres auteurs ont également soulevé la possibilité que l'analyse bibliométrique incite les chercheurs à mettre l'accent sur la publication ou à mener leurs recherches de façon à accroître la probabilité que leurs résultats soient publiés. Décrivant les pratiques en publication en Afrique du Sud, Robert Tijssen, Johann Mouton, Thed N van Leeuwen et Nelius Boshoff indiquent que « les méthodes pour évaluer la réussite dans les universités des pays développés s'appuient généralement sur les réalisations atteintes (et donc les récompenses accordées), soit sur la publication des résultats de recherche dans des revues internationales courantes à comité de lecture » [traduction libre] (2006, p. 171). Toutefois, les sujets d'intérêt en Afrique du Sud ne retiennent pas nécessairement l'attention des revues scientifiques établies dans les pays du Nord. Les chercheurs se retrouvent donc devant un choix à faire : doivent-ils profiter des « récompenses » ou plutôt risquer de les perdre et s'en tenir aux programmes de recherche qui importent dans leur secteur ?

Fausse objectivité : L'objectivité est souvent présentée comme l'un des avantages des systèmes fondés sur les mesures. Il est vrai que le calcul est une démarche objective. Toutefois, la sélection des renseignements à prendre en compte dans cette démarche ne l'est pas nécessairement. Le choix des personnes qui décident des données à intégrer aux calculs suscite également des interrogations (Wooding et Grant, 2003, p. 23).

Les auteurs des documents analysés soulignent que les revues scientifiques et les autres formes de présentation des travaux de recherche, qui sont le fondement des mesures, font la belle part aux résultats de recherche rédigés en anglais (Yule, 2010; Liang, Wu et Li, 2001; Bridges, 2009). Les documents traitant de la publication des travaux en Chine et en Afrique du Sud indiquent que la majorité des travaux publiés dans ces pays paraissent dans des revues scientifiques locales (Liang, Wu et Li, 2001; Tijssen, Mouton, van Leeuwen et Boshoff, 2006). Comme mention-

né précédemment, le cas de l’Afrique du Sud démontre que, même lorsque la langue ne pose pas problème, les sujets, les méthodes et la présentation des travaux des chercheurs de ce pays ne correspondent pas toujours aux exigences ou aux intérêts des revues scientifiques qui sont dirigées depuis les pays du Nord où y sont établies. Cette situation réduit les probabilités que les résultats des travaux soient publiés dans des revues ou à des événements internationaux considérés de grande envergure, ce qui diminue par le fait même le nombre de citations des travaux des chercheurs des pays du Sud (Tijssen, Mouton, van Leeuwen et Boshoff, 2006, p. 172). Dans ce cas, le fait de prendre en compte seulement les mesures pour évaluer la qualité des recherches peut donner faussement l’impression que les travaux effectués dans les pays du Sud sont de faible qualité alors que, dans les faits, le petit nombre de citations illustre peut-être une différence dans les programmes de recherche.

Domaines de recherche : Coryn (2006), Bridges (2009), Kenna et Berche (2011) ainsi que d’autres auteurs se questionnent sur la pertinence des mesures et leur comparaison entre les domaines et les disciplines. Les données bibliométriques varient selon les disciplines. Un document de l’OCDE souligne également le fait que l’analyse bibliométrique ne tient pas compte des documents qui se situent dans la zone « grise », qui sont d’une « importance cruciale dans le cas des travaux interdisciplinaires et des percées novatrices » [traduction libre] (1997, p. 9).

Graham (2008) souligne par ailleurs que les chercheurs dans les domaines des arts, des sciences humaines et des sciences sociales en Australie ont rejeté les facteurs des retombées de Thomson ISI en tant qu’indices de classement des revues scientifiques, les jugeant inappropriés dans le cadre de leurs travaux. Au sujet du domaine des arts, Coryn ajoute ce qui suit : « Pour publier un document sur une oeuvre, il faut peut-être l’entendre, la voir, la lire ou la découvrir autrement, par exemple sur scène ou à l’occasion d’autres événements publics. Dans le domaine de la création artistique, l’évaluation de la publication selon une vision étroite tenant compte uniquement des documents écrits entraîne des illogismes; un document rédigé par un peintre au sujet d’une de ses oeuvres est recensé parmi les publications alors que l’oeuvre en elle-même n’est pas prise en compte. De même, l’article d’un critique au sujet d’une oeuvre musicale est considéré comme une publication alors que la présentation de l’oeuvre musicale et l’oeuvre en elle-même ne comptent pas et ne peuvent être cités » (Strand, 1998 tel que mentionné dans Coryn, C. L., “The Use and Abuse of Citations as Indicators of Research Quality”, 2006, p. 117).

Innovations et nouveaux chercheurs : Comme souligné ci-dessus, les mesures posent particulièrement problème dans les nouveaux domaines. Andras, qui appuie l’utilisation des mesures, soutient ceci : « En ce qui concerne la présentation des résultats de recherche dans les revues scientifiques de prestige, les travaux de la science normale sont beaucoup plus susceptibles d’être publiés, car il est bien plus probable que la communauté scientifique accepte d’emblée les résultats de ces travaux que de ceux de la science révolutionnaire, qui peuvent semer la controverse » [traduction libre] (2011).

Au Royaume-Uni, les débats au sujet du RAE et du REF ont également fait ressortir les difficultés auxquelles sont confrontés les jeunes chercheurs, car les revues scientifiques sont parfois plus enclines à publier un article rédigé par un auteur bien connu que celui d'un nouveau chercheur, malgré la qualité de son travail. Ainsi, les mesures sont plus susceptibles d'avantager les chercheurs chevronnés par rapport aux moins expérimentés.

Globalement, l'analyse bibliométrique offre des perspectives intéressantes. Wooding et coll. indiquent que des discussions soutenues se tiennent sur la façon de concevoir des mesures qui soient davantage axées sur le domaine et tiennent compte des travaux de recherche menés dans d'autres contextes, rédigés dans d'autres langues et réalisés par de nouveaux chercheurs. Toutefois, les chercheurs sont fortement d'avis que les résultats de l'analyse bibliométrique ne doivent pas et ne peuvent pas remplacer l'avis des experts. Lorsqu'il y a lieu, ces résultats peuvent servir d'outil complétant les évaluations sur l'excellence en recherche (van Raan, 1993; OCDE, 1997; David, 2008; Tatavarti, Sridevi et Kothari, 2010). D'autres comme Becker, Bryman et Sempik soutiennent que la qualité en recherche ne peut être jugée que « par des personnes pratiquant l'art ancien de la lecture » (Becker, Bryman et Sempik, 2006, p. 19).

Conclusions et questions à approfondir

Comme dans le cas du secteur de l'enseignement supérieur au Royaume-Uni, des compressions budgétaires sont appliquées dans le domaine du développement, où l'on cherche également à accroître la justification de l'emploi des fonds, la gestion axée sur les résultats et l'optimisation des ressources. Cette réalité a amené tous les établissements de recherche et les organismes subventionnaires de recherche à mettre davantage l'accent sur la qualité des travaux de recherche réalisés. Par conséquent, l'évaluation de l'excellence en recherche et les débats concernant l'examen par les pairs, l'analyse bibliométrique et les retombées comme dimension de l'excellence en recherche prennent de plus en plus d'importance dans le domaine de la recherche pour le développement.

Toutefois, la conception de méthodes d'évaluation de l'excellence en recherche n'est pas une démarche simple. L'excellence est définie différemment selon les personnes (Becker, Bryman et Sempik, 2006; Martí et Villasante, 2009; Furlong et Oancea, 2005). Il faut donc instaurer des méthodes d'évaluation comportant de multiples couches et de multiples dimensions. En mettant au point les systèmes d'évaluation de l'excellence en recherche, les évaluateurs doivent veiller à ce que ceux-ci soient transparents, justes, pertinents et moins lourds, qu'ils puissent être comparés les uns aux autres et qu'ils soient reconnus par les pairs (Wooding et Grant, 2003, p. 15). Ces méthodes doivent également être conçues en fonction d'un objectif précis défini dans le cadre de l'évaluation.

La présente analyse visait à offrir un aperçu d'un sujet plutôt vaste et complexe. Bien entendu, il demeure de nombreuses lacunes et questions pouvant faire l'objet de futures recherches, par exemple les suivantes :

- Qu'entend-on par « excellence en recherche » dans le domaine de la recherche pour le développement international ? Même si cette question faisait l'objet de la présente analyse en grande partie, les documents recensés et analysés indiquent qu'une bonne part de la réflexion sur le sujet a été menée au sein des organismes de recherche des pays du Nord.
- Que signifie chacune des dimensions ou chacun des critères de la qualité traités dans le présent document ? Comment peut-on les relever ou les mesurer dans le cadre d'une évaluation de l'excellence en recherche ? (Que signifie la pertinence dans la recherche interdisciplinaire ? Comment la démontrer ou la mesurer ?)
- Quels sont les avantages, les inconvénients et les compromis nécessaires pour distinguer clairement l'« excellence » de la « qualité » en recherche ?
- Comment doit-on évaluer la qualité des recherches interdisciplinaires et des recherches novatrices ou de découverte ?
- Comment évaluer les retombées de la recherche, si ceux-ci sont jugés comme étant une composante de l'excellence en recherche ?

La présente analyse documentaire a permis de relever quelques exemples provenant des pays du Nord, qui peuvent être approfondis pour examiner comment évaluer l'excellence en recherche dans le domaine du développement. À titre d'exemple, même si divers aspects du REF du Royaume-Uni ont été contestés, ce mécanisme constitue un exemple de modèle d'évaluation présentant des caractéristiques générales utiles : les résultats et les retombées des recherches ainsi que l'environnement de recherche. De plus, il n'aborde pas les détails propres à l'examen par des comités selon la discipline et offre donc de la souplesse dans un processus qui suit autrement des critères précis. Il convient toutefois de noter que les points de vue des établissements et des chercheurs des pays du Sud font particulièrement défaut dans le présent rapport. Étant donné que la recherche pour le développement international cible presque exclusivement les pays du Sud, il est essentiel de connaître l'opinion et l'expérience de ces parties. Les futures analyses devront examiner davantage les perspectives des chercheurs et des établissements de recherche des pays du Sud au sujet de l'excellence en recherche.

Enfin, les questions soulevées dans la présente analyse s'étendent bien au-delà du monde universitaire. L'évaluation de l'excellence en recherche est souvent liée au financement et, par conséquent, à la génération de nouvelles connaissances et à l'innovation, éléments tous deux essentiels à l'évolution des sociétés et non seulement du monde universitaire. Boaz et Ashby soulignent ce qui suit : « Après tout, comparativement à d'autres sources de connaissances sur lesquelles s'appuient les responsables des décisions, la recherche a entre autres comme point fort qu'il s'agit d'un produit dont la qualité est assurée et qui répond à des normes définies. Plus généralement, la notion de la qualité en recherche doit contribuer à donner aux chercheurs et aux utilisateurs des recherches l'assurance que les données probantes servent à l'élaboration des politiques et à la pratique » [traduction libre] (2003, p. 2).

Bibliographie

- Aagaard-Hansen, J., & Svedin, U. (2009). Quality Issues in Cross-disciplinary research: towards a Two-pronged approach to evaluation. *Social Epistemology*, 165-176.
- Aizenman, J., Edison, H., Leony, L., & Sun, Y. (2011). *Evaluating the Quality of IMF Research: A Citation Study*. Washington : Independent evaluation Office - IMF.
- Andras, P. (2011). Research: metrics, quality, and management implications . *Reserach Evaluation*, 90-106.
- Arnold, D. (2008). Cultural Heritage As a Vehicle for Basic Research in Computing Science: Pasteur's Quadrant and a Use-Inspired Basic Research Agenda. *COMPUTER GRAPHICS forum, Volume 27 (number 8)*, 2188–2196.
- Arunachalam, S. (2009). *Social Science Research in South Asia An analysis of the published journal literature*.
- Becker, S., Bryman, A., & Sempik, J. (2006). *Defining 'Quality' in Social Policy Research: Views, Perceptions and a Framework for Discussion*. Suffolk: Lavenham: Social Policy Association.
- Boaz, A., & Ashby, D. (2003). *Fit for purpose? Assessing research quality for evidence based policy and practice*. Retrieved 2011, from ESRC UK Centre fro Evidence Based Policy and Practice: Working Paper 11: <http://www.kcl.ac.uk/content/1/c6/03/46/04/wp11.pdf>
- Boix Mansilla, V., Feller, I., & Gardner, H. (2006). *Quality assessment in interdisciplinary research education*. Surrey: Research Evaluation.
- Booth, D. (2011, April). *Working with the grain and swimming against the tide: Barriers to uptake os research finding on governance and public services in low-income Africa*. Retrieved July 7, 2011, from UK Department for International Development : <http://www.dfid.gov.uk/r4d/PDF/Outputs/APPP/appp-working-paper-18.pdf>
- Bridges, D. (2009). Research quality assessment in education: impossible science, possible art? . *British Educational Research Journal* , 497-517.
- Coryn, C. L. (2006). The Use and Abuse of Citations as Indicators of Research Quality. *Journal of MultiDisciplinary Evaluation* , 115-121.
- Coryn, C. L., Hattie, J. A., Scriven, M., & Hartmann, D. J. (Dec 2007). Models and Mechanisms for Evaluating Government-funded Research: An International Comparison. *American Journal of Evaluation*, 437-457.

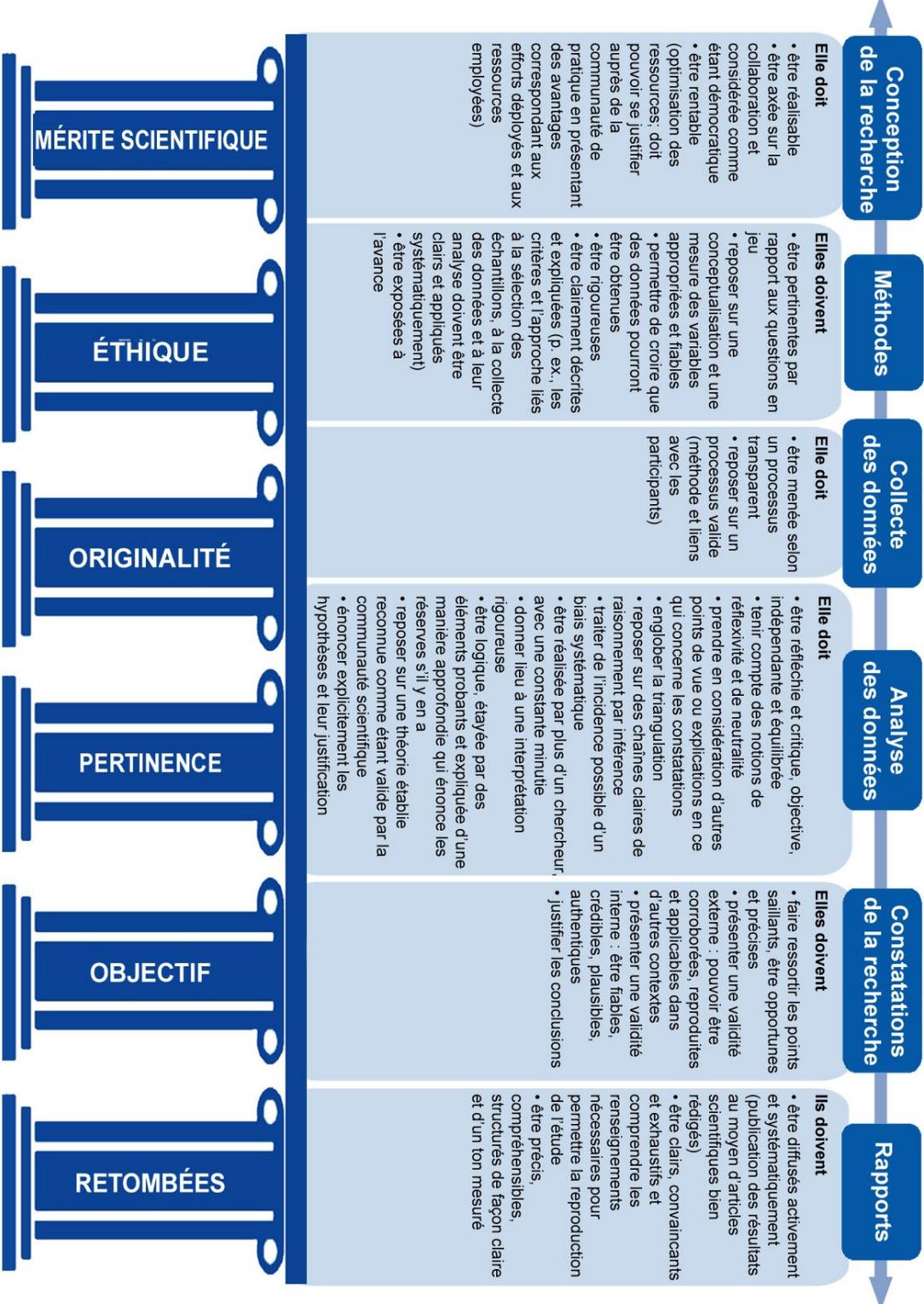
- Coryn, C. (Pre-publication chapters from Coryn, C. L. S. (2013)). What is Social Science Research and why would we want to evaluate it? In C. Coryn, *Evaluating social science research: A handbook for researchers, instructors, and students*. (pp. 1-32). New York, NY: Guilford.
- David, M. E. (2008, 7, 1,). Research Quality Assessment and the Metrication of the Social Sciences. *European Political Science*, 52-63, Palgrave Macmillan Ltd.
- Dictionary.com. ((n.d)). Retrieved November 24, 2011, from Dictionary.com Unabridged.: <http://dictionary.reference.com/browse/research>
- Donaldson, J. S. (2009). In search for blueprint for an evidence-based global society. In S. I. Donaldson, A. Christina, & M. M. Mark, *What counts as credible evidence in applied research and evaluation practice?* (pp. 2-17). Thousand Oaks, California: Sage.
- Donovan, C. (2007). The qualitative future of research evaluation. *Science and Public Policy*, 585-597.
- Fielding, N. (2010). Elephants, gold standards and applied qualitative research. *Qualitative Research*, 123-127.
- Furlong, J., & Oancea, A. (2005). *Assessing quality in Applied and Practice-based Educational Research*. Oxford: Oxford University Department of Educational Studies.
- Graham, L. (2008). Rank and File: Assessing research quality in Australia. *Educational Philosophy and Theory*, 811-815.
- Grant, J., Brutscher, P.-C., Kirk, S. E., Butler, L., & Wooding, S. (2010). *Capturing Research Impacts: A review of international practice*. Cambridge, UK: Rand Europe.
- Groundwater-Smith, S., & Mockler, N. (2007). Ethics in practitioner research: an issue of quality. *Research Papers in Education*, 199-211.
- Hammersley, M. (2008). Troubling criteria: A critical commentary on Furlong and Oancea's framework for assessing educational research. *British Educational Research Journal*, 747-762.
- Higher Education Funding Council for England. (2011). *Consultation on draft panel criteria and working methods*. England: HEFCE.
- Higher Education Funding Council of England. (n.d.). *Higher Education Funding Council of England*. Retrieved January 25, 2012, from Research Excellence Framework : <http://www.hefce.ac.uk/research/ref/>

- House, E. R. (2008). Blowback: Consequences of Evaluation for Evaluation. *American Journal of Evaluation, Volume 29*(Number 4), 416-426.
- Kenna, R., & Berche, B. (2011, June). Normalization of peer-evaluation measures of group research quality across academic disciplines. *Research Evaluation, 107-116*.
- Lardone, M., & Roggero, M. (n.d.). *Study on monitoring and evaluation of the research impact in the public policy of Policy Research Institutes (PRIs) in the region*. Retrieved July 2011, from Evidence Based Policy Development Network: <http://www.ebpdn.org/download/download.php?table=resources&id=3013>
- Laudel, G., & Glaser, J. (2006). Tensions between evaluations and communication practices. *Journal of Higher Education Policy and Management, 289-295*.
- Liang, L., Wu, Y., & Li, J. (2001). Selection of databases, indicators and models for evaluating research performance of Chinese universities. *Research Evaluation, 105-113*.
- London School of Economics. (2011). *Maximizing the impact of academic research*. Retrieved 2011, from Impact of Social Sciences: <http://blogs.lse.ac.uk/impactofsocialsciences/>
- LSE Public Policy Group. (2011). *Maximizing the impacts of your research: A Handbook for Social Scientists*. London : LSE Public Policy Group.
- Malterud, K. (2001). Qualitative research: standards, challenges, and guidelines. *The Lancet, 483-488*.
- Marjanovic, S., Hanney, S., & Wooding, S. (2009). *A historical reflection on research evaluation studies, their recurrent themes and challenges*. Cambridge: RAND Corporation.
- Martí, J., & Villasante, T. R. (2009). Quality in Action Research: Reflections for Second-Order Inquiry. *Systemic Practice and Action Research, 383-396*.
- National Center for the Dissemination of Disability Research. (2005). *What are the standards for Quality Research - Focus technical brief*. Retrieved September 22, 2011, from National Center for the Dissemination of Disability Research: <http://www.ncddr.org/kt/products/focus/focus9/>
- Nutley, S., Percy-Smith, J., & Solesbury, W. (2003). *Models of research impact: a cross-sector review of literature and practice*. London: Learning and Skills Research Centre.
- O'Gorman, L. (2008, January). *The (frustrating) state of peer review*. Retrieved July 2011, from International Association for Patter Recognition Newsletter: <http://iapr.org/docs/newsletter-2008-01.pdf>

- O'Neil, M. (2002). Commentary: We May Need a New Definition for Research Excellence. *University Affairs*.
- Organisation de coopération et de développement économiques. (1997). *The evaluation of scientific research: Selected Experiences*. Paris: OECD.
- Petit-Zeman, S. (2003, January 16). Trial by peers comes up short. *The Guardian*, p. <http://www.guardian.co.uk/science/2003/jan/16/science.research>.
- RAND Corporation. (2011, November). *RAND's Standards for High-Quality Research and Analysis*. Retrieved January 2012, from RAND Corporation: http://www.rand.org/standards/standards_high.html
- REF 2014. (July 2011). *REF 2014: Assessment Framework and guidance on submission*. UK: HEFCE.
- Rodrik, D. (2011, June 10). *Dani Rodrik's weblog: Unconventional thoughts on economic development and globalization*. Retrieved June 2011, from A rejection letter I would like to receive from a journal one day: http://rodrik.typepad.com/dani_rodriks_weblog/2011/06/a-rejection-letter-i-would-like-to-receive-from-a-journal-one-day.html
- Roebber, P. J., & Schultz, D. M. (2011, April 12). *Peer Review, Program Officers and Science Funding*. Retrieved July 25, 2011, from PLoS One : <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3075261/>
- Rons, R., De Bruyn, A., & Cornelis, J. (2008). Research evaluation per discipline: a peer-review method and its outcomes. *Research Evaluation* , 45-57.
- Rowland, F. (2002, October). *The Peer Review Process*. Retrieved July 2011, from Joint Information Systems Committee : http://www.jisc.ac.uk/uploaded_documents/rowland.pdf
- Saari, E., & Kallio, K. (2011). Developmental Impact Evaluation for Facilitating Learning in Innovation Networks. *American Journal of Evaluation* , 227-245.
- Sastry, T., & Bekhradnia, B. (2006). *Using metrics to allocate research funds: initial response to the Government's consultation proposals*. Oxford: Oxford Higher Education Policy Institute (HEPI).
- Seymour, F. (2011, June 27). *Does the pressure for impact compromise research?* Retrieved June 2011, from Forests Blog: <http://blog.cifor.org/3439/does-the-pressure-for-impact-compromise-research/>
- Spencer, L., Ritchie, J., Lewis, J., & Dillon, L. (2003). *Quality in Qualitative Evaluation: A framework for assessing research evidence*. UK Government Chief Social Researcher's Office .

- Stephen, C., & Daibes, I. (2010). *Defining features of the practice of global health research: an examination of 14 global health research teams*. Retrieved July 4, 2011, from Global Health Action : <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2903310/>
- Tatavarti, R., Sridevi, N., & Kothari, D. (2010). Assessing the quality of university research – the RT factor. *General Science*, 1015-1019.
- Tijssen, R. J., Mouton, J., van Leeuwen, T. N., & Boshoff, N. (2006, December). How relevant are local scholarly journals in global science? A case study of South Africa. *Research Evaluation*, Volume 15 (number 3), 163-174.
- Tijssen, R. (2003). Scoreboards of research excellence. *Research Evaluation*, 91-103.
- Wagner, C. S., Roessner, J., Bobb, K., Thompson Klein, J., Boyack, K. W., Keyton, J., et al. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Infometrics*, 14-26.
- Ware, M. (2011). Peer review: recent experience and future direction. *New Review of Information Networking*, 23-53.
- Webb, C. (1993). Feminist research: definitions, methodology, methods and evaluation. *Journal of Advanced Nursing*, 416-423.
- Wooding, S., & Grant, J. (2003). *Assessing Research: The Researcher's View*. Retrieved from RA Review: <http://www.ra-review.ac.uk/reports/assess/AssessResearchReport.pdf>
- Wu, H., Ismail, S., Guthrie, S., & Wooding, S. (2011). *Alternatives to Peer Review in Research Project Funding*. Cambridge, UK: RAND Corporation .
- Yates, L. (2005). Is Impact a measure of Quality? Some Reflections on the Research Quality and Impact Assessment Agendas. *European Educational Research Journal*, Volume 4(Number 4), 391-403.
- Yule, M. (2010). *Assessing Research Quality*. Ottawa: International Development Research Centre - Peace Conflict and Development.

Annexe 1. Éléments conceptuels courants de l'excellence en recherche au fil du processus de recherche



Liste de sigles

ERiC – Approche d'évaluation de la recherche en contexte des Pays-Bas

HEFCE – Higher Education Funding Council for England

CRDI – Centre de recherches pour le développement international

FMI – Fonds monétaire international

OCDE – Organisation de coopération et de développement économiques

PART – Outil de cotation pour l'évaluation des programmes

RAE – Exercice d'évaluation de la recherche

RAISS – Système de cotation des retombées pour l'Arthritis Research Campaign

REF – Cadre d'excellence en recherche

RQF – Cadre d'accessibilité et de qualité en recherche