

Les impacts d'une évaluation bibliométrique standardisée : le cas des publications de sciences sociales au sein d'une section universitaire multidisciplinaire

The impacts of a standardized bibliometric evaluation: The case of social science publications in a multidisciplinary university section

Los impactos de una evaluación bibliométrica estandarizada: el caso de las publicaciones de ciencias sociales en una sección universitaria multidisciplinar

Bastien Soulé

Volume 22, Number 3, 2020

URI: <https://id.erudit.org/iderudit/1081290ar>

DOI: <https://doi.org/10.7202/1081290ar>

[See table of contents](#)

Publisher(s)

Faculté d'éducation, Université de Sherbrooke

ISSN

1911-8805 (digital)

[Explore this journal](#)

Cite this article

Soulé, B. (2020). Les impacts d'une évaluation bibliométrique standardisée : le cas des publications de sciences sociales au sein d'une section universitaire multidisciplinaire. *Nouveaux cahiers de la recherche en éducation*, 22(3), 110–131. <https://doi.org/10.7202/1081290ar>

Article abstract

In France, Hcéres has been placing impact bibliometrics at the heart of the evaluation process of sports sciences (STAPS) research units since 2018. This article sets out to analyze the effects of this shift in the valuation of social science productions, during laboratory evaluations as well as on the scholarly practices of the players involved. The evaluation reports of six laboratories as well as twelve interviews with teacher-researchers show that the new guide to the products of research profoundly alters their activity. It is applied with a zeal that reinforces the centrality of *SCImago Journal Ranking*, which has become a normative tool of appraisal. The teacher-researchers concerned are developing dual strategies in order to cope with these new requirements while also giving themselves alternative spaces for expression and recognition.

Les impacts d'une évaluation bibliométrique standardisée: le cas des publications de sciences sociales au sein d'une section universitaire multidisciplinaire

Bastien Soulé

Université Claude Bernard Lyon 1

Résumé

En France, le Hcéres place depuis 2018 la bibliométrie d'impact au cœur du processus d'évaluation des unités de recherche en STAPS. L'objectif de cet article est d'analyser les effets de ce virage en matière de valorisation des productions en sciences sociales, lors des évaluations de laboratoire comme sur les pratiques scientifiques des principaux intéressés. Les rapports d'évaluation de six laboratoires et douze entretiens réalisés auprès d'enseignants-chercheurs montrent que le nouveau guide des produits de la recherche altère en profondeur leur activité. Il est appliqué avec un zèle qui renforce la centralité du *Scimago Journal Ranking*, devenu outil normatif d'appréciation. Les enseignants-chercheurs concernés élaborent des stratégies duales pour composer avec ces nouvelles exigences tout en se ménageant des espaces d'expression et de reconnaissance alternatifs.

Mots-clés: évaluation, France, recherche, sciences sociales, STAPS

The impacts of a standardized bibliometric evaluation: The case of social science publications in a multidisciplinary university section

Abstract

In France, Hcéres has been placing impact bibliometrics at the heart of the evaluation process of sports sciences (STAPS) research units since 2018. This article sets out to analyze the effects of this shift in the valuation of social science productions, during laboratory evaluations as well as on the scholarly practices of the players involved. The evaluation reports of six laboratories as well as twelve interviews with teacher-researchers show that the new guide to the products of research profoundly alters their activity. It is applied with a zeal that reinforces the centrality of *SCImago Journal Ranking*, which has become a normative tool of appraisal. The teacher-researchers concerned are developing dual strategies in order to cope with these new requirements while also giving themselves alternative spaces for expression and recognition.

Keywords: evaluation, France, research, social sciences, STAPS

Los impactos de una evaluación bibliométrica estandarizada: el caso de las publicaciones de ciencias sociales en una sección universitaria multidisciplinar

Resumen

En Francia, desde 2018 Hcéres sitúa la bibliometría de impacto al centro del proceso de evaluación de las unidades de investigación de STAPS. El objetivo de este artículo es analizar los efectos de este cambio en materias de la valoración de las producciones en ciencias sociales, tanto en las evaluaciones de laboratorio como en las prácticas científicas de los actores principales. Los informes de evaluación de seis laboratorios y doce entrevistas realizadas a docentes-investigadores muestran que la nueva guía de productos de investigación afecta profundamente su actividad. Se aplica con un celo que refuerza la centralidad del *Scimago Journal Ranking*, el cual se ha convertido en una herramienta de evaluación normativa. Los docentes-investigadores afectados desarrollan estrategias duales para hacer frente a estas nuevas exigencias, estableciendo al mismo tiempo, espacios alternativos de expresión y de reconocimiento.

Palabras clave: evaluación, Francia, investigación, ciencias sociales, STAPS

1. Introduction

Évaluer consistant à accorder de la valeur, cette opération suppose que l'on dispose d'étalons de mesure (Herreros et Milly, 2004). Dans la recherche publique, des référentiels tendent à cadrer de plus en plus fortement les processus d'évaluation d'unités de recherche (UR). Leur existence n'épuise pas les désaccords, à propos de leur élaboration, de leur interprétation ou de leur application. Au sein même des disciplines, la divergence des pratiques de production scientifique implique des dissonances, concernant par exemple le degré de normalisation des pratiques académiques ou l'incitation à l'internationalisation des publications (Gozlan, 2016). Sous couvert d'objectivation et de transparence, l'évaluation de la recherche s'avère être une fabrique sociale faite d'appropriations plurielles des «guides», matérialisant des rapports de force évolutifs (Gozlan, 2016). En particulier, la maîtrise du processus de définition de la qualité de la recherche (critères de sélection des supports de publication, définition des indicateurs statistiques permettant le classement des revues) assure une position dominante en définissant la norme évaluative, instrument d'exercice d'un contrôle au sein des disciplines scientifiques (Tiberghien et Beauvois, 2008). Façonner ou écrire les règles du jeu, au sens bourdieusien du terme, procure l'avantage de définir ce qui va compter pour légitimer les agents du champ scientifique (des conditions d'entrée à la «consécration»), assez strictement codifié (Desrochers, Paul-Hus, Haustein, Costas, Mongeon, Quan-Haase ... Larivière, 2018).

La diversité des points de vue quant à ce qui vaut (plus ou moins) et ce qui ne vaut pas est accrue lorsque les cultures scientifiques sont multiples (Zarka, 2009). En France, au sein de la 74^e section du Conseil national des universités (les sciences et techniques des activités physiques et sportives – STAPS) se côtoient les champs des sciences de la vie et de la santé (SVS) et des sciences humaines et sociales (SHS). Cette particularité en fait un observatoire privilégié de l'évaluation interdisciplinaire de la recherche, avec ses enjeux et difficultés (Andrews, 2008; Wagner, 2016). Sans surprise, les désaccords y sont structurels quant aux «produits de la recherche» (selon la nomenclature du Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur – Hcéres¹) à prendre en compte lors de l'évaluation des UR. La question du poids à accorder aux monographies et chapitres d'ouvrages est par exemple vive. Les articles publiés dans des revues scientifiques attisent également les débats quant à la manière de les considérer. Or, le Hcéres a pris

1 Autorité administrative indépendante française chargée d'évaluer les établissements d'enseignement supérieur et de recherche, les formations et les laboratoires, le Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur a succédé en 2014 à l'Agence d'évaluation de la recherche et de l'enseignement supérieur (AERES). Créée en 2006, cette dernière a été fortement critiquée au sein de la communauté scientifique. Les griefs ont atteint leur paroxysme au début des années 2010: manque de transparence et de légitimité scientifique, introduction d'un système de notation aux conséquences jugées délétères, etc. La création du Hcéres visait à remédier à ce hiatus en affichant «des principes d'indépendance, de simplicité de fonctionnement et de procédures ainsi que de légitimité scientifique et de transparence» (déclaration de M^{me} Fioraso, ministre de l'Enseignement supérieur et de la Recherche, dans le quotidien *Le Monde* du 3 janvier 2013).

un virage radical en 2018, officialisant la centralité des indicateurs bibliométriques² internationaux (*Journal Citation Report* ou *JCR* du *Web of Science*, et surtout *Scimago Journal Ranking* ou *SJR* de *Scopus*) lors de l'évaluation des UR relevant des STAPS. Le Haut Conseil a concomitamment décrété la disparition de la liste de revues scientifiques relevant du champ des STAPS, qu'il avait créée en 2009. Charge aux chercheurs en sciences sociales de faire valoir d'éventuels référencements de revues sur d'autres listes disciplinaires produites par le Hcéres. Cet avènement illustre la standardisation de l'évaluation de la recherche au sein de cette jeune section universitaire.

Soulé et Chatal (2018) ont très tôt montré un premier impact de cette conversion bibliométrique: la valeur attribuée aux revues publiant des articles de sciences sociales (au sens large, intégrant les sciences de gestion, de l'intervention, etc.) s'est vue bouleversée. Le *SJR* «maltraite» les revues de sciences sociales (notamment francophones), nombreuses à disparaître des radars ou à être déclassées. Il survalorise à l'inverse de nombreux supports anglophones qui n'étaient pas, jusqu'alors, considérés comme des revues de premier plan par la communauté scientifique. L'analyse de ces auteurs conforte les écrits soulignant que l'ouverture du *SJR* aux revues de sciences sociales et/ou non anglophones, mise en avant pour justifier le recours désormais prioritaire à cette mesure, est très partielle (Schöpfel et Prost, 2009; CNRS, 2015; Sivertsen, 2016; Harzing et Alakangas, 2016). Les biais générés par *Scopus* et le *SJR* sont en fait proches de ceux du *WoS* et de l'*IF*, renseignés de longue date (Pansu, Dubois et Beauvois, 2013; Pontille et Torny, 2013; Higher Education Funding Council for England, 2014; Gingras, 2014; Hicks, Wouters, Waltman, de Rijcke et Rafols, 2015); les classements de revues diffèrent du reste très peu entre ces deux référentiels (Manana-Rodriguez, 2015). Mobiliser l'un (*IF*) ou l'autre (*SJR*) à des fins d'évaluation dans une section interdisciplinaire, en espérant objectiver et fiabiliser les comparaisons, introduit donc des biais défavorables aux SHS (Schreiber, 2013; Mongeon et Paul-Hus, 2016). Pourtant, la «boîte noire» que constitue le *SJR*, dont le fonctionnement est flou (Falagas, Pitsouni, Malietzis et Pappas, 2008), l'algorithme et la pertinence probabiliste contestés (Larivière et Sugimoto, 2018; Foulley, Celeux et Josse, 2017), la traçabilité et la reproductibilité douteuses (Manana-Rodriguez, 2015; Delgado-López-Cózar et Cabezas-Clavijo, 2013) est bel et bien devenue centrale dans l'évaluation des UR en STAPS.

2 Calculés sur la base des citations d'articles scientifiques, les indicateurs bibliométriques sont un moyen d'analyse de la diffusion et de l'impact des publications savantes. L'appréciation permise portait initialement sur les revues, afin de cerner les plus centrales dans chaque domaine de recherche (pour s'y abonner, les lire ou y publier); elle s'est étendue aux chercheurs, ainsi qu'à leurs structures (laboratoires, universités) voire aux pays d'appartenance, dans une perspective de hiérarchisation rendue possible par la quantification.

Entre cette analyse et 2020, trois vagues d'évaluation³ ont été pilotées par le Hcéres. L'occasion d'aborder la manière dont la mutation décrite altère, concrètement, les attentes des experts et l'évaluation des laboratoires est le premier objectif assigné à cet article: décrire les effets du changement de référentiel sur la valorisation des productions.

L'occasion aussi d'un pas de côté pour s'intéresser à la façon dont les évalués composent avec ce référentiel standardisé, vivent l'évaluation et se sont adaptés à cette nouvelle donne. Le second objectif consiste donc à rendre compte des pratiques des enseignants-chercheurs (EC) concernés au regard de leurs productions scientifiques.

Ainsi, deux objectifs interreliés sont assignés à cette recherche descriptive, réalisée dans une perspective exploratoire. Pour analyser les effets pluriels de ce changement de référentiel, envisagé comme un processus systémique, la valorisation des productions en sciences sociales sera d'abord appréhendée lors des évaluations de laboratoire (rédaction du rapport d'activité, échanges avec les comités d'experts, appréciations et recommandations émanant de ces derniers). Par la suite, les répercussions en matière de stratégies de publication, et, plus largement, de pratiques scientifiques des principaux intéressés seront abordées. Notre attention se focalise sur les équipes de sciences sociales au sein de la section des STAPS, particulièrement affectées par la normalisation de l'évaluation de leurs productions scientifiques.

2. Contextualisation: un virage bibliométrique à contre-courant

L'évolution décrite en introduction a de quoi surprendre tant elle s'inscrit *a contrario* des analyses et recommandations émanant des spécialistes de la bibliométrie d'impact (Gingras, 2014). La littérature grise n'est pas en reste. Pour s'en convaincre, il suffit de consulter les rapports internationaux émanant de diverses institutions s'étant penchées sur l'évaluation de la recherche en sciences sociales (*cf.* Soulé et Chatal, 2018), à l'image de la Fédération des sciences humaines (2017) au Canada. Ils soulèvent explicitement les effets pervers générés par le recours à la bibliométrie pour évaluer la recherche:

- faible pertinence des bases de données (BDD) internationales pour l'évaluation en SHS;
- non-respect des spécificités disciplinaires, en matière notamment de diffusion des travaux de recherche (il existe des formes pertinentes de publication échappant à l'hégémonie des BDD internationales: revues nationales ou régionales, professionnelles ou techniques, etc.);

3 À l'échelon national, cinq vagues d'évaluation (A, B, C, D, E) correspondant à cinq zones géographiques permettent d'évaluer, selon un rythme quinquennal, l'ensemble des établissements supérieurs, formations et équipes de recherche de chaque zone.

- injonction à ne pas se fier outre-mesure aux indicateurs fournis par le *WoS* ou *Scopus*, dont l'utilisation doit rester prudente et assortie d'appréciations qualitatives destinées à déceler l'originalité et le caractère novateur des recherches;
- risques liés au recours systématisé à un jeu limité d'indicateurs: définition étroite de la valeur de la recherche, encouragement aux comportements opportunistes (sélection des actions génératrices d'utilité personnelle comme l'émiettement des publications⁴ ou la cosignature d'articles sans contribution significative au travail);
- glissement consistant à accorder la priorité à ce qui est facilement mesurable, au détriment de ce qu'il est important et/ou souhaitable de faire en recherche.

Les acquis des travaux relatifs à l'évaluation de la recherche et à la bibliométrie d'impact ne sont guère mis à profit dans les réflexions menées au sein de la section des STAPS. Au contraire, la difficulté inhérente à l'évaluation multidisciplinaire est contournée par le recours providentiel à des indicateurs bibliométriques (*SJR*, *JCR*)⁵. Il est à ce titre révélateur que le guide des produits de la recherche pour les STAPS publié en avril 2018 par le Hcéres soit le seul à ne pas s'encombrer de mises en garde relatives aux éventuels travers des outils placés au cœur du dispositif d'évaluation. Les onze autres guides mis en ligne sur le site du Haut Conseil en 2018 (en sciences de la terre, physique, chimie, mathématiques, informatique, physiologie, neurosciences, biologie, agronomie, sciences économiques et de gestion, histoire, etc.) évoquent tous, sans exception, des points de vigilance vis-à-vis de l'usage des BDD internationales à des fins d'évaluation: absence de grandes revues généralistes; caractère biaisé des différents facteurs d'impact fournis; sur-cotation ou sous-cotation de certains journaux; existence de BDD alternatives; etc.

3. Les effets délétères de l'évaluation standardisée sur les chercheurs

Les développements qui précèdent sont annonciateurs: en terre multidisciplinaire, l'évaluation de travaux de sciences sociales au prisme du *SJR* et/ou du *JCR* est mécaniquement porteuse de sérieuses distorsions. Elle entraîne leur déconsidération,

4 On parle aussi de «saucissonnage» éditorial (*salami slicing* en anglais), pratique consistant à rédiger plusieurs articles à partir d'un seul jeu de données ou d'une expérience unique, plutôt que de privilégier une publication plus étoffée, *a priori* de meilleure qualité (Larivière et Sugimoto, 2018).

5 Selon un schéma faisant écho au «solutionnisme technologique», qui renvoie aux solutions séduisantes mais simplistes s'imposant d'elles même car dans l'air du temps (Morozov, 2013). Ce recours s'appuie généralement sur une appréhension et une définition restreintes des problèmes à résoudre, qui nécessiteraient pourtant un diagnostic approfondi du fait de leur complexité. Le solutionnisme technologique conduit à considérer la nouveauté (surtout si elle est adossée à des *data* et un algorithme) comme intrinsèquement vertueuse, source systématique de médiation, tout en sous-estimant ou refusant de considérer les conséquences inattendues voire les effets pervers générés (Morozov, 2013). Depuis le milieu des années 2010, le *SJR* bénéficie ainsi d'un engouement singulier au sein de la section des STAPS: l'attribution d'un score d'impact aux revues, leur classement par quartiles (dans de nombreuses catégories thématiques ou disciplinaires), assorti de codes couleur allant du vert au rouge, n'y sont sans doute pas étrangers (en dépit de la méconnaissance quant à la genèse des scores d'impact, *cf. supra*).

qui s'accompagne d'effets délétères sur les évalués: altération des priorités, défiance, sentiment d'injustice, etc. Plusieurs auteurs ont analysé les conséquences plurielles du recours à la standardisation en matière d'évaluation, dans divers secteurs professionnels.

3.1 Une orientation prescriptive sans explicitation des objectifs

La genèse de la normalisation décrite fait écho à l'évaluation prescriptive (De Gaulejac, 2012): construite à partir d'indicateurs censés mesurer les résultats de l'activité en toute neutralité, elle est fondée sur des présupposés non explicités et s'appuie sur des outils importés de l'extérieur dont la mécanique est largement méconnue. Aux sources de ce type d'évaluation, on note la préférence de la mesure à l'analyse, de la comparaison à la qualité intrinsèque d'une recherche, ou encore de la normalisation à la prise en compte de l'originalité. Sous couvert d'objectivation, de gain de temps ou de pacification des évaluations (Berry, 2009), Tiberghien et Beauvois (2008, p. 141) évoquent la «tyrannie métrologique du facteur d'impact», avec son lot d'effets systémiques contre-productifs.

Cette impossibilité de discuter les critères est susceptible de générer une perte de sens, voire un sentiment de violence symbolique. Pour être légitimes, les référentiels doivent être en phase avec les réalités professionnelles vécues, d'où la souhaitable co-élaboration de ces étalons de valeur (Herreros et Milly, 2004). Dans le cas des STAPS, le Hcéres a mis les chercheurs en sciences sociales devant le fait accompli; les indicateurs n'ont pas été mis en discussion puis validés avec des spécialistes du domaine. Cette manière d'exercer le pouvoir en imposant sa représentation du monde par le biais de dispositifs d'évaluation non discutables, alors même qu'ils se révèlent partiels et bouleversent les équilibres antérieurs, ne peut que susciter méfiance et rejet (De Gaulejac, 2012).

3.2 Mesurer la performance ou apprécier le travail réalisé?

Lorsque l'on pense évaluer le travail, on évalue souvent (de manière plus ou moins précise) le résultat du travail (Dejours, 2011). En se focalisant sur la partie mesurable de la performance, l'évaluation est dès lors susceptible d'engendrer une appréciation très éloignée de la réalité, propice à l'émergence d'un sentiment d'injustice. Miser essentiellement sur l'aspect quantitatif a aussi tendance à altérer le travail effectué et priorisé.

La performance est de plus en plus systématiquement mesurée au moyen d'outils quantitatifs; le travail, moins facile à appréhender dans sa complexité, résiste à cette approche. S'approcher d'une juste mesure du travail nécessite la combinaison du jugement d'utilité (technique, sociale, économique) et du jugement de beauté (respect des règles de l'art d'une activité). Selon Dejours (2003), la conjonction de ces deux formes de jugement permet d'évaluer le travail réalisé de la manière la moins approximative, et de rétribuer le travail accompli en accordant une reconnaissance indispensable à l'équilibre mental.

Seuls les pairs, qui connaissent l'art du métier, sont à même de se prononcer sur ce jugement de beauté.

3.3 Des stratégies d'adaptation, jusqu'à la mise en parenthèses du «travail bien fait»

Les évalués ne subissent pas passivement l'évaluation mais en adaptant leurs comportements aux paramètres portés à leur connaissance. De moyen destiné à appréhender si l'on fait bien ou mal son travail, l'indicateur peut devenir finalité à atteindre, quoi qu'il en coûte (De Gaulejac, 2012). Des dérives et incitations à la manipulation sont intrinsèquement liées à ce dessein. Que l'on bascule ou non dans ces comportements répréhensibles (fraude, plagiat, mais aussi dissimulation ou caviardisation des données, etc.), les stratégies en question peuvent devenir désespérantes à long terme. Les libertés prises par rapport à l'éthique constituent l'aboutissement logique d'une intense mise en tension de soi (entre ce qu'il faut faire pour être bien évalué et ce que l'on considère être un travail bien fait) (De Gaulejac, 2012). S'appliquer des critères auxquels on n'adhère pas explique sans doute la montée d'un dépit au sein des professions scientifiques. Plus largement, si l'évaluation semble incontournable, elle peut finir par imposer sa propre logique, générant alors des effets peu questionnés et mal anticipés sur le cheminement et l'orientation des professionnels (Nizet, 2013).

3.4 Les effets de la quête d'excellence

Sur fond de «lutte des places», l'évaluation est cependant acceptée, voire attendue par ceux qui, se considérant performants, ont intérêt à ce que la mesure objective leur supposée supériorité. La désignation de quelques vainqueurs s'accompagne, symétriquement, de celle de nombreux perdants, générant une déception chronique (De Gaulejac, 2012). À plus long terme, la logique de l'excellence se traduit par l'attribution d'une «petitesse», négatif de la «grandeur» de quelques-uns. Stimulant pour les uns, ce processus est stigmatisant pour la grande majorité (dont les nombreux candidats à la récompense) (Herreros et Milly, 2004).

4. Éléments méthodologiques

Afin de cerner les répercussions des évolutions décrites, nous décrivons la manière dont se sont déroulées les évaluations des six UR en STAPS comprenant des EC en sciences sociales⁶, évaluées entre 2018 et 2020. Une certaine variété caractérise cet

⁶ Pour la vague D (2017-2018): Technique et enjeux du corps (TEC) (Université Paris Descartes). Pour la vague E (2018-2019): Unité de recherche pluridisciplinaire sport santé société (URePSSS) (Université de Lille, Artois, Littoral côte d'Opale) et Complexité, innovation, activités motrices et sportives (CIAMS) (Université Paris sud – Orléans). Pour la vague A (2019-2020): Centre de recherches en sciences sociales, sport et corps (CRESCO) (Université Toulouse 3), Sport et environnement social (SENS) (Université Grenoble Alpes) et Laboratoire sur les vulnérabilités et l'innovation dans le sport (L-Vis) (Université Lyon 1) (vague A – 2019-2020).

échantillon: taille et effectifs (d'une dizaine à une cinquantaine de titulaires), degré d'interdisciplinarité (intra-SHS ou avec les SVS), portage monotutelle ou unité multisite. Le recueil d'informations s'est focalisé sur la prise en considération, dans le processus d'évaluation, des produits de la recherche (par les évaluateurs et les évalués): ceux qui sont placés au cœur, constituant des pivots pour l'évaluation; ceux qui, à l'inverse, sont considérés comme secondaires ou disqualifiés; la manière également dont le jugement s'adosse aux différentes catégories de productions scientifiques, et dont se construisent les appréciations. Les recommandations formulées par les évaluateurs, ainsi que les arguments auxquels elles sont adossées, sont aussi disséqués.

Pour cerner la manière dont est construite l'évaluation, l'idéal aurait été de recueillir le point de vue des experts mandatés par le Hcéres et des membres des UR concernées. Nous avons néanmoins pris le parti de ne pas directement aborder la question auprès des évaluateurs. En effet, lorsqu'ils acceptent une expertise de laboratoire, ces derniers sont tenus à un devoir de discrétion, de confidentialité et de secret quant aux procédures auxquelles ils prennent part pour le compte du Haut Conseil. Nous avons donc dû nous contenter des traces de leur activité de jugement que constituent les rapports d'évaluation⁷. Bien qu'imparfaite, cette analyse documentaire est révélatrice de la manière dont les experts, après avoir interagi avec les enseignants-chercheurs (EC) des UR évaluées, posent collégialement un diagnostic quant à leur production scientifique. Se dévoilent par ce biais des interprétations, des hiérarchisations et des conseils qui seront analysés au prisme du guide des produits de la recherche pour les STAPS. Par ailleurs, l'évaluation étant appréhendée comme un processus, d'autres documents ont été intégrés au corpus: dossiers d'autoévaluation rédigés par chaque UR, échanges écrits en amont de la visite du comité Hcéres (questions préparatoires appelant une réponse écrite en amont et/ou orale lors de l'audition), réponses écrites de chaque UR en réaction à la prise de connaissance de la version *bêta* du rapport d'évaluation. Le total représente en moyenne 130 pages de documents par UR collectées lors de ce premier temps de recueil.

Douze entretiens semi-directifs d'une durée moyenne d'une heure ont ensuite été menés auprès d'EC relevant des SHS au sein des UR évaluées: principalement sociologues et historiens (n = 7), et, dans une moindre mesure, philosophes, spécialistes des sciences de gestion et psychologues (n = 5). Cet échantillon est constitué à parité d'hommes et de femmes, de maîtres de conférences et de professeurs des universités. Les interviewés avaient de 30 à 61 ans. Dans une logique exploratoire, l'objectif était de bénéficier de leur regard sur le processus d'évaluation de leur UR, mais aussi d'aborder les impacts pluriels de la dynamique d'évaluation au sein de la section (sur les stratégies de publication,

⁷ Ils se déclinent en deux versions systématiquement analysées: l'une, synthétique (6-7 pages), est mise en ligne sur le site du Hcéres; l'autre, plus développée (20-30 pages), n'est pas rendue publique.

la reconnaissance et la satisfaction au travail) et ses éventuels effets pervers. Le guide d'entretien utilisé figure dans le tableau 1.

Les matériaux provenant de l'analyse documentaire et des entretiens ont été ventilés selon un codage thématique (Huberman et Miles, 1991) prédéterminé par les guides utilisés au stade de la collecte, sans que le passage des données brutes aux données organisées revête un caractère strictement confirmatif, restreint aux «boîtes» préconçues (Bardin, 2013). Par exemple, l'élaboration de stratégies duales en matière de publication,

Tableau 1
Guide d'entretien

Catégories	Sous-thèmes et indicateurs
Élaboration du dossier d'autoévaluation de l'UR (focalisation sur les produits de la recherche)	<ul style="list-style-type: none"> · Préparation du bilan scientifique: organisation au sein de l'UR et des sous-équipes, formes de valorisation des productions scientifiques. · Appropriation et interprétation du guide des produits de la recherche en STAPS: choix des critères (listes Hcéres/AERES, JCR, SJR).
Déroulement de la visite du comité	<ul style="list-style-type: none"> · Composition du comité (représentation disciplinaire, profil des experts). · Choix de présentation du bilan scientifique de l'UR. · Nature des échanges avec les experts.
Indices de qualité des articles dans des revues scientifiques	<ul style="list-style-type: none"> · Référentiels mobilisés par le comité pour établir son appréciation. · Nomenclature utilisée.
Divergences dans la prise en compte des produits de la recherche	<ul style="list-style-type: none"> · Discussions avec le comité à propos des référentiels. · Recomptages et redressements éventuels opérés par le comité. · Prise en compte des revues nationales, chapitres et ouvrages.
Appréciation qualitative portée sur la production de l'UR	<ul style="list-style-type: none"> · Portée, originalité, innovation, interdisciplinarité.
Point de vue sur l'incitation à internationaliser les productions	<ul style="list-style-type: none"> · Portée et impact des recommandations à publier dans des revues du SJR et/ou du WoS. · Effets vertueux (références, diffusion, réseau). · Difficultés générées. · Compromis scientifiques pour se conformer aux normes internationales.
Influence exercée sur les stratégies de publication	<ul style="list-style-type: none"> · Objectifs actuels de publication de l'interviewé et de l'UR (quantitatifs et en matière de supports). · Orientation des choix de supports, évolutions depuis 10 ans. · Stimulation produite, effets pervers ou contre-productifs.
Supports de publication, reconnaissance académique et accomplissement professionnel	<ul style="list-style-type: none"> · Revues et éditeurs faisant autorité dans le domaine de l'interviewé-e · Supports de publication faisant «rêver» les interviewés. · Appréciation de ce support au prisme du guide des produits de la recherche. · Décalages éventuels et problèmes induits en matière de reconnaissance ou de satisfaction au travail.
Connaissance des indices bibliométriques de la part de l'interviewé	<ul style="list-style-type: none"> · Mode de calcul de l'IF et du score d'impact SJR. · Recul critique sur les indicateurs et les listes de revues utilisés.
Point de vue sur la centralité des quartiles du SJR dans l'évaluation	<ul style="list-style-type: none"> · Description de l'évolution des critères. · Intentions des décideurs, effets sur les différentes disciplines des STAPS.
Mode d'évaluation souhaitable et/ou idéal	<ul style="list-style-type: none"> · Éléments relatifs à la procédure, au profil des experts, aux critères retenus, aux indices de qualité, au type d'appréciation formulé, etc.

non anticipée lors du cadrage de la recherche, a émergé pendant l'analyse des entretiens; de même, l'ardeur avec laquelle un certain nombre d'experts s'approprient le référentiel d'évaluation, allant jusqu'à en faire une application excessive, constitue une trouvaille non escomptée. Un contrôle du matériau résiduel a été entrepris pour s'assurer que les interprétations produites n'étaient pas trop fortement dépendantes du cadrage initial (Mukamurera, Lacourse et Couturier, 2006), conduisant par exemple à écarter une partie du corpus.

Pour l'analyse documentaire, les ingrédients des diagnostics posés, mis en correspondance avec le référentiel d'évaluation faisant foi au sein de la section, ont structuré l'analyse et rendu possible le repérage de l'application faite de ce dernier (incluant des décalages et réinterprétations): sélection et priorisation des éléments factuels auxquels adosser l'appréciation; hiérarchisation des différents produits de la recherche (sur une dialectique centralité/relégation); justification des recommandations formulées. À l'issue de ce codage essentiellement descriptif, l'interprétation a été réalisée à l'aune du cadre conceptuel fourni dans la partie 2.

L'analyse du matériau procédant des entretiens s'est déroulée de manière itérative, dès la phase de recueil. Fonction de l'approche de référence (basée pour l'essentiel sur la partie 3), les unités de sens retenues confirment et modifient, à la marge, ce cadrage conceptuel. À travers le guide d'entretien, «ensemble de questions à poser qui avaient été formalisées, développées, subdivisées et rédigées de manière à ce qu'elles puissent être utilisées dans une grille d'analyse ancrée dans le cadre conceptuel» (Mukamurera *et al.*, 2006, p. 117), l'organisation progressive des données obtenue par le biais d'une codification de teneur interprétative (Huberman et Miles, 1991) a permis l'atteinte d'un niveau jugé satisfaisant de plausibilité et de cohérence, assurant l'intelligibilité du discours.

5. La fabrique de l'évaluation de la recherche en sciences sociales au sein des STAPS

Cette première section consacrée aux résultats détaille le déroulement des évaluations, appréhendées comme des cheminements jalonnés de choix aboutissant à une appréciation assortie de recommandations. Au regard des apports procurés par la littérature, nous avons été particulièrement attentifs à la manière dont les invitations à la pondération et au discernement, en matière de critériologie, sont intégrées par les experts. La fabrique de l'évaluation laissant une certaine latitude lors de la mise en œuvre d'un référentiel pour le compte du Hcéres, l'interprétation et l'appropriation qui en sont faites sont d'autant plus intéressantes à observer.

5.1 La mise en ordre de bataille

La rédaction du document d'autoévaluation fait émerger un enjeu spécifique pour les unités et sous-équipes de sciences sociales: comment faire valoir et reconnaître leur production, notamment la partie généralement importante non couverte par les BDD internationales?

Certaines unités de notre échantillon, dirigées par des psychologues ou des spécialistes des SVS, ne laissent guère de marge de manœuvre aux collègues concernés. Elles vont au-delà des prescriptions du Hcéres, précisant non seulement l'*IF* de chaque revue, le quartile dans lequel elle se situe dans chaque domaine du *SJR*, mais aussi le nombre de citations reçues par chaque article dans *Google Scholar*. Dans ce type d'unité zélée, il n'est pas fait allusion aux listes de revues produites par le Hcéres ou l'AERES, ce qui ampute paradoxalement les bilans de 10 à 20 articles pouvant légitimement y figurer. *«On nous fait comprendre à chaque réunion que les chapitres et les ouvrages, ça ne sert à rien, que seules les revues internationales et référencées comptent, et parmi elles surtout les Q1.»* Quand des EC en sciences sociales expriment leur souhait de faire valoir des chapitres d'ouvrage (comme les y invite le guide des produits de la recherche en STAPS), il leur est rétorqué que *«ce n'est pas une bonne stratégie de valorisation des publications»*.

Plus modérées, d'autres unités dirigées par des EC relevant des SVS s'inscrivent dans le compromis et valorisent des productions ne s'avérant pas centrales au prisme du *SJR*. Un dialogue interne pousse alors à signaler, dans le rapport d'autoévaluation, des articles publiés dans des revues anciennement référencées par l'AERES. La majorité des unités constitutives de notre échantillon se conforment au guide des produits de la recherche en STAPS, tout en s'efforçant de ne pas subir un cadre évaluatif en partie décalé par rapport à leurs activités. L'enjeu est notamment de faire valoir une production scientifique diversifiée: articles scientifiques dans des revues internationales (en précisant ou non le quartile), mais aussi nationales (mention aux listes AERES ou Hcéres); ouvrages et contributions à des ouvrages collectifs; activités de diffusion de connaissances; rapports de recherche. Au sein des sous-équipes de sciences sociales, la stratégie consiste à garder, si possible, la maîtrise de la rédaction et de la présentation du bilan correspondant à leur activité, afin de valoriser au mieux les productions. La principale zone d'incertitude réside dans la prise en considération ou non, par le comité d'experts, des revues ne figurant pas dans le *SJR*. En la matière, les EC concernés avouent espérer un soutien de la part des experts représentant leur champ disciplinaire, tout comme ils comptent sur eux pour rappeler l'importance des chapitres et ouvrages dans une production en sciences sociales; une double attente généralement déçue.

5.2 La polarisation sur les revues figurant dans les deux premiers quartiles du *SJR*

Le *SJR* constitue l'instrument central (parfois quasi exclusif) des appréciations formulées. Outre un périmètre (inclusion ou non des revues), cette mesure permet de situer les titres selon qu'ils figurent dans le premier quartile (Q1), le Q2, le Q3 ou le Q4 des revues du champ disciplinaire ou thématique considéré. La norme implicite qui se dégage de la lecture des rapports distingue deux sous-groupes de revues:

- la qualité et l'excellence des productions scientifiques sont définies à l'aune des revues des Q1 et Q2. La quantité et la proportion des articles scientifiques dans des revues du Q1, et, dans une moindre mesure, du Q2, sont systématiquement utilisées comme levier prioritaire pour armer les jugements. Elles permettent de situer les unités et sous-équipes entre deux bornes (production excellente et remarquable, ou modeste et insuffisante) par le biais d'une objectivation qui n'est ensuite altérée qu'à la marge par le recours à d'autres indicateurs;
- les revues des Q3 et Q4, si elles sont parfois quantifiées dans les rapports, ne sont jamais nommées comme gage de qualité. Quelque peu déconsidérées, elles remplissent une fonction différente: boucler le périmètre des «revues référencées».

Selon les rapports, les revues ne figurant pas au *SJR* mais présentes sur les listes Hcéres ou les anciennes listes AERES peuvent être considérées comme référencées, ou non; cette prise en compte est à la discrétion du comité d'experts. Il arrive qu'elles soient exclues des bilans, au terme de fastidieux redressements prenant des airs d'épuration. Quoi qu'il en soit, ces revues sont considérées comme de moindre qualité que celles référencées dans le *SJR*, quel que soit le quartile. En fait, leur évocation sert surtout à souligner le nécessaire effort à entreprendre pour viser, à l'avenir, des revues du Q1 et du Q2 du *SJR*.

Une recommandation générique est formulée dans chaque rapport, sans exception: augmenter le nombre d'articles publiés dans des revues scientifiques du Q1 (et dans une moindre mesure du Q2). Pour ce faire, il *convient de* «sélectionner des cibles qualitatives de revues indexées *SJR* ou *WoS*» ou de privilégier des «supports référencés à fort impact». Aucune unité ne déroge à ce conseil, y compris quand la production est jugée bonne, très satisfaisante ou de qualité. Une véritable logique de conversion se dessine: «*dans un contexte d'excellence*», des injonctions à faire des choix sont formulées dans le but de «clarifier sa stratégie de publication à l'international dans les bases référencées». Il s'agira alors de «cibler des supports de renom», voire de «se recentrer à l'avenir exclusivement sur des cibles référencées».

La description de certains échanges avec le comité de visite fait écho à ce que Robert (2014) nomme une procédure de disciplinarisation: «*On nous a dit vous avez un bon bilan de publications, mais les listes AERES n'existent plus, maintenant il faut faire du Scimago. Ça devient urgent, ça se passera pas pareil dans 5 ans. L'évangélisation a surtout été forte pendant le huis clos avec les doctorants [...]. Un membre du comité m'a dit "on les a secoués, on leur a dit qu'ils n'avaient pas compris que ça n'existait plus, les listes" [...]. Le délégué scientifique a proposé de revenir nous former au SJR. Sans aucun argument autre pour soutenir cet outil que le fait que les listes AERES, c'était fini.*»

5.3 L'injonction à l'internationalisation

Le caractère insuffisamment international des productions et la pression à accroître la diffusion à l'étranger constituent un autre invariant des rapports d'évaluation. Par définition, cet aspect est redondant avec le point précédent, puisque les revues classées dans le Q1 et le Q2 du *SJR* sont, sauf exception, éditées en langue anglaise.

Les communications lors de colloques et de congrès, presque invisibles dans les rapports d'évaluation, ne sont évoquées que par rapport à cet aspect. C'est aussi vrai, dans une moindre mesure, concernant les ouvrages et chapitres d'ouvrage, dont il est seulement fait mention s'ils sont édités par de «*grands éditeurs anglophones*» (Palgrave Macmillan, Routledge, Sage, etc.). C'est un indice de qualité censé témoigner de l'ambition des auteurs, exactement comme pour les revues scientifiques ciblées: à ce propos, ne sont nommément cités dans les rapports d'évaluation que des titres anglo-saxons, y compris quand les membres des unités ont publié dans des revues de référence comme *Sociologie du travail* ou *Ethnologie française*. Une sorte de «fuite en avant» se dessine: il faut accroître sans cesse cette production à l'international, érigée en idéal.

Ce processus s'accompagne d'une déconsidération de ce qui n'est pas diffusé et communiqué en anglais, langue à l'aune de laquelle se définit l'internationalisation; écrire ou présenter une communication en espagnol ou en portugais est plutôt commenté comme une étrangeté. L'incitation à se tourner davantage vers l'international et à en faire une priorité est omniprésente, y compris quand il est stipulé que la proportion d'articles et de chapitres publiés en langue étrangère est très satisfaisante. L'effort est attendu de la part de toutes les UR, et son accentuation demandée.

5.4 Les autres produits de la recherche, variables d'ajustement

Dans le guide des produits de la recherche en STAPS publié en 2018, il est recommandé de prendre en considération diverses formes de valorisation: les articles scientifiques, mais aussi ceux publiés dans des revues professionnelles ou techniques, les monographies et

ouvrages scientifiques, ou encore les chapitres d'ouvrage. Plus largement, un équilibre entre ces formes complémentaires de valorisation est encouragé.

Dans les faits, la hiérarchie est évidente et la complémentarité appelée de ses vœux par le Hcéres s'efface. L'ancrage de l'appréciation relative à la publication d'articles scientifiques est si puissant qu'il tend à éclipser l'ensemble des jugements portés sur les autres valorisations. Pour preuve, la quasi-absence de monographies, chapitres d'ouvrage ou articles de diffusion de connaissances n'entache en rien l'appréciation portée sur la performance d'une équipe si elle publie suffisamment d'articles dans des revues des Q1 et Q2 du *SJR*.

Il arrive même fréquemment qu'une activité soutenue en matière de diffusion de connaissances soit interprétée comme une source de déséquilibre et de fragilité. La vulgarisation, quand elle est décrite comme importante, n'est pas pour autant qualifiée de point fort; elle est plutôt désignée comme un refuge allant à l'encontre d'une production de qualité.

5.5 Livres et chapitres d'ouvrage: les grands oubliés de l'évaluation

La relégation des monographies et contributions à des ouvrages collectifs est attestée par le faible volume accordé à ces aspects dans les rapports d'évaluation⁸. Généralement, il est seulement fait mention du nombre de ces produits, éventuellement assorti d'une évocation de l'éditeur (s'il n'est pas français).

L'absence d'ouvrage, ou la présence d'un seul ouvrage au bilan d'une unité ou d'une sous-équipe, n'empêche en rien de qualifier sa production de remarquable ou exceptionnelle. C'est le cas d'une sous-équipe ayant publié un ouvrage, quand l'autre sous-équipe de la même unité, comptabilisant six ouvrages, est décrite comme étant en grande difficulté en matière de production scientifique. Concernant la seconde entité, il est précisé que «*ces derniers éléments contrebalancent en partie le nombre de publications scientifiques*». Un aveu sans doute involontaire, mais parlant: les ouvrages et chapitres d'ouvrages ne seraient pas des publications scientifiques. En conséquence de quoi, une autre unité présentant un bilan impressionnant de 20 ouvrages publiés ne se verra pas décerner de point fort sur cet aspect.

Les «éditions de référence» vers lesquelles il est recommandé de s'orienter sont anglo-saxonnes: aucun éditeur français n'est cité, si ce n'est pour conseiller de «*mieux cibler*» son

8 Dans les rapports détaillés d'évaluation, le nombre de mentions aux termes chapitre/ouvrage/monographie/livre a été comparé au nombre d'occurrences des termes article/revue. Ont été écartés les doublons (liés aux formules du type chapitre d'ouvrage ou article de revue) ainsi que les évocations d'articles et revues professionnels ou de vulgarisation. Les rapports obtenus sont de 1 à 3 concernant deux UR de l'échantillon, 1 à 4 pour deux autres, 1 à 5 pour une UR et 1 à 7 dans le cas le plus extrême.

éditeur (PUF, Seuil). Amusé, un interviewé rapporte cette remarque d'un expert, à propos d'un ouvrage: «Ah, Routledge, voilà, ça c'est bon!»

Les livres, et *a fortiori* les contributions à des ouvrages collectifs, font presque systématiquement l'objet d'une déconsidération fortement ressentie par les interviewés. Si leur existence est mentionnée et quantifiée par les experts, ils ne pèsent pas réellement dans l'avis exprimé: «ils s'en foutent, ça n'intéresse personne»; «ce n'est pas abordé, encore moins valorisé». C'est même avec une certaine gêne, dans certaines UR, que ces produits de la recherche sont mis en avant dans les dossiers d'autoévaluation .

6. Tiraillements professionnels et adaptations stratégiques

Inscrite dans une temporalité élargie, cette seconde section consacrée aux résultats aborde la manière dont les critères d'évaluation imposés au sein de la section des STAPS altèrent les pratiques professionnelles des enseignants-chercheurs et leur rapport au travail.

6.1 Quelle place pour la reconnaissance et la satisfaction du travail bien fait?

Les situations décrites posent un problème d'épanouissement professionnel. Pour certains sociologues et historiens, la notion de travail bien fait n'est guère compatible avec l'injonction croissante à ne publier que dans des revues figurant dans le Q1 ou le Q2 du *SJR*. En effet, la plupart des interviewés soulignent que la reconnaissance disciplinaire, dans la plupart des sciences sociales, provient de canaux et supports pas ou peu pris en considération dans l'évaluation à l'œuvre en STAPS: ouvrages, revues francophones de référence, etc. Une disjonction s'instaure, générant l'impression de ne pas faire son métier correctement, du point de vue à la fois de sa section de rattachement (les STAPS) et de sa discipline mère. Des risques psychosociaux sont en suspens quand l'antagonisme est trop marqué entre les critères d'excellence imposés de l'extérieur et la conception d'un travail de qualité, pour soi et ses pairs. «Du fait de devoir aller dans les deux sens, t'es jamais vraiment satisfait de ton travail, en tout cas moi ça me met mal, t'es bien nulle part.»

Plusieurs interviewés relatent le caractère difficilement supportable de cette perte de sens du travail de recherche, affirmant malgré tout s'essayer au grand écart pour faire carrière et/ou ne pas empêcher leurs doctorants de trouver une place à l'université. Un EC fait part d'une grande frustration liée à l'incitation indirecte à négliger certaines facettes de son métier (comme la portée sociale des recherches menées). En résulte une mise en tension entre ce qu'il faut faire pour être bien évalué et ce que l'on considère comme du travail bien fait. La sociologie du travail, tout comme les travaux de psychopathologie, soulignent qu'en cherchant à améliorer ses indicateurs de performance, on peut perdre

le sens profond du métier et la valeur accordée à son activité. À terme, de telles stratégies peuvent devenir désespérantes.

C'est du reste pour se soustraire à la dévalorisation, plus que par conviction du bien-fondé d'une valorisation internationale grâce aux revues cotées, que plusieurs EC interviewés s'efforcent de «jouer le jeu». *«Je me force à faire un Q1 de temps en temps, pour leur montrer que je sais faire, mais que je ne veux pas le faire plus.» «Dans le labo, on écoute, on ne placardise pas ceux qui publient en Q1. Les autres sont quantité négligeable.»*

6.2 L'élaboration de stratégies duales: «jouer le jeu» pour s'épanouir ailleurs

Face au durcissement des critères, les interviewés sont nombreux à avoir été dans l'obligation d'élaborer une stratégie duale de publication. *«Jouer sur deux tableaux»* est une forme d'adaptation à une injonction contradictoire. La stratégie consiste à se fixer un objectif en matière de publications dans les classements internationaux (par an, ou sur la durée du quinquennal). Il s'agit de remplir le contrat pour soi et ses collègues, de *«montrer patte blanche»* pour ensuite *«se sentir libre», «se faire plaisir»* à côté en contribuant par exemple à des ouvrages collectifs ou à des numéros spéciaux de revues échappant aux BDD. Même les collègues qui se disent résistants sur les questions d'évaluation acceptent le compromis consistant à *«faire l'attendu»*. Un interviewé explique que cet espace d'expression secondaire est ce qui lui permet de *«tenir»*.

Publier dans des revues du premier quartile du *SJR* est en fait une manière de disposer d'un *«ticket»* pour se permettre de s'éloigner des critères normatifs. Ce n'est pas forcément une source d'intense satisfaction, ni une entreprise très difficile, abstraction faite de la nécessaire écriture en anglais (aucun interviewé n'est en mesure de rédiger, en autonomie, un article dans cette langue). Les expertises faisant suite à la soumission d'articles dans des revues classées dans les Q1 ou Q2 ne poussent pas toujours les auteurs dans leurs retranchements, à condition de se plier à des structurations et formats d'expression plus restrictifs que dans la plupart des revues francophones. *«La plupart du temps, je vais te le dire franchement, ça n'apporte rien pour ton travail [...]. J'ai publié dans le Scandinavian Journal of... je ne sais plus trop quoi (rires): les retours étaient catastrophiques! C'est une revue pas habituée à publier des travaux de sciences sociales. Ça te fait vraiment pas avancer.»*

Les expertises les plus constructives ont pour la plupart été reçues à la suite de la soumission d'articles dans des revues francophones de renom. Concrètement, le débat scientifique approfondi et la stimulation intellectuelle se produisent surtout lorsque sont visées des revues *«faisant rêver»*: *Revue française de sociologie, Mouvements, Genèses, Nouvelles questions féministes, Actes de la recherche en sciences sociales, Vingtième*

siècle, Intellectica, Sociologie du travail, Histoire de l'éducation, etc. Autant de périodiques scientifiques considérés comme plus difficiles d'accès, pour un EC, que ceux figurant au Q1 ou au Q2 du *SJR*.

Pour plus de la moitié des EC interviewés, la publication rêvée n'est pas un article de revue; ils seraient davantage comblés par un ouvrage largement diffusé, qui susciterait un important intérêt et laisserait une trace durable, constituant une référence utile à la communauté. Un EC explique sa stratégie: produire *a minima* dans des revues correspondant à la critériologie de la section STAPS, «*pour qu'on me laisse tranquille*», et consacrer le peu de temps résiduel à la rédaction de cet ouvrage. Ce pragmatisme est aussi lié aux responsabilités des EC envers les doctorants. «*Avec les doctorants, c'est impossible, on va à l'essentiel, là où ça paie pour espérer faire carrière... Avec les plus jeunes, on standardise davantage, du fait du recentrage produit par les experts.*»

Un jeune EC reconnaît avoir totalement intériorisé les règles dictées par le Hcéres: «*[...] pendant ma thèse, pour savoir où publier, j'ai choisi l'outil qui m'évaluera [...]. Si tu joues contre, tu ne peux pas être recruté, je n'ai pas voulu me pénaliser moi-même. Et c'est toujours celui dont je me sers aujourd'hui*». Il est toutefois remarquable que les plus jeunes EC interviewés élaborent aussi des stratégies duales: maximiser ses chances de recrutement, pour pouvoir ensuite se consacrer à l'écriture de contributions à des ouvrages collectifs ayant du sens par rapport à ses objets de recherche; se projeter, à terme, sur la perspective d'un ouvrage («*quand j'estimerai avoir joué le jeu, je prendrai le temps, peu importe si c'est reconnu ou pas dans une carrière, pour écrire un livre*»).

Parfois, des dilemmes surviennent. Un EC fait état de tiraillements: fréquemment sollicité pour participer à de stimulants projets dont la valorisation prend la forme d'ouvrages collectifs, il n'ignore pas que ce format le pénalise. «*Je ne vois pas comment je ne pourrais pas participer à ça... Ce que je voudrais, c'est trouver un équilibre entre les deux, il y a des projets super intéressants scientifiquement, sans valorisation dans des revues.*»

7. Discussion

Les processus analysés revêtent la plupart des atours de l'évaluation prescriptive. Dans les mesures établies, le jugement d'utilité est confiné à un strict périmètre. Il éclipse celui de beauté, qui faute d'être établi au sein de la section tend à être recherché à l'extérieur, au contact de pairs relevant des sciences sociales. Pour autant, l'adaptation est nécessaire; allant jusqu'à certaines compromissions éthiques afin d'amasser, à travers des publications, du capital symbolique (Desrochers, Paul-Hus, Haustein, Costas, Mongeon, Quan-Haase ... Larivière, 2018), elle est à l'origine de mises en tension de soi et de son identité professionnelle. Adhérer partiellement aux critères, en d'autres termes «jouer le jeu» (ce qui nécessite des efforts conséquents) permet seulement de ne pas perdre, sans

réel espoir de faire partie des «grands»; car en dépréciant une large part des productions en sciences sociales, la critériologie en vigueur contribue à la dévalorisation de leurs auteurs. Il ressort des échanges avec les interviewés un sentiment de non-reconnaissance qui altère leur quotidien, en dehors du contexte spécifique de l'évaluation. Sans équivoque, le champ sémantique qui se dessine au fil des entretiens (stigmatisation, mépris, humiliation, douleur, injustice, dévalorisation, violence, etc.) ne peut qu'interpeler.

Un certain nombre d'effets pervers est généré en ce qui a trait à la production et la diffusion de connaissances au sein de la section des STAPS. Sous couvert d'excellence, la nouvelle donne opérée éloigne les spécialistes des sciences sociales de ce qu'ils considèrent comme une recherche et une valorisation de qualité, c'est-à-dire équilibrées, tournées vers des mondes distincts auxquels correspondent différents objectifs et types de littérature: revues internationales indexées, mais aussi livres, revues nationales et presse non scientifique (Hicks, 2005). En minorant ou ignorant les trois derniers pans, l'évaluation bibliométrique donne une image déformée du domaine des sciences sociales, déconnectée des problématiques contextuelles et des intérêts locaux qui constituent pourtant une partie de sa raison d'être (Archambault et Larivière, 2010). Du reste, en promouvant sa propre vision de l'excellence scientifique, la section des STAPS dans son ensemble s'inscrit à contre-courant non seulement de la relativisation de la toute-puissance des indicateurs bibliométriques⁹, mais aussi de l'exigence d'une visibilité publique et d'un effet sociétal accrus de la recherche publique; autant d'éléments constitutifs d'un changement d'*illusio* en cours (Desrochers *et al.*, 2018). Ces choix poussent en toute logique les acteurs à élaborer des stratégies adaptatives: il s'agit par exemple, du fait de la surévaluation des titres anglophones, de cibler les revues bien classées, dont certaines sont décrites comme n'étant pas particulièrement sélectives (*Sport in Society*, *International Journal for the History of Sport*, etc.). En conséquence, l'expertise la plus rigoureuse, émanant des titres phares de la discipline (notamment francophone, en histoire et sociologie) peut de la sorte être contournée. L'intériorisation de critères d'évaluation survalorisant les articles dans des revues indexées favorise aussi l'enfermement disciplinaire; un paradoxe au sein d'une section revendiquant son interdisciplinarité. Par ailleurs, les formats longs de valorisation deviennent de moins en moins tentants, quand bien même Charle (2009) rappelle que pour les humanités et une partie des sciences sociales, «l'innovation durable et l'impact le plus profond sur les communautés de recherche ne passent pas par cette forme article». Enfin, sur le plan épistémologique, le quadrillage bibliométrique fait tendre la recherche vers une science «hyper-normale» poussant au conformisme (Berry, 2009) plutôt qu'à la remise en cause et à l'originalité.

9 Et d'un appel à leur utilisation responsable: cf. le manifeste de Leiden, ou la *San Francisco Declaration on Research Assessment*.

8. Conclusion

La pierre angulaire de l'évaluation des UR en STAPS est constituée, pour ce qui est des SHS, du *Scimago Journal Ranking*. C'est en particulier vers ses deux premiers quartiles que converge l'appréciation de l'excellence. Le corollaire de cette focalisation est l'attention toute particulière accordée à l'internationalisation de la valorisation des recherches, qui dépasse le cadre des articles de revues pour concerner également les ouvrages et communications lors de manifestations scientifiques. Les décomptes et agrégations chiffrées, qui font l'objet d'analyses et commentaires assez brefs, occupent une très large surface dans les rapports d'évaluation, occultant l'analyse qualitative des produits de la recherche (leur portée, leur originalité), pourtant censée prévaloir. En publiant dans des revues internationales essentiellement lues par des pairs à l'étranger, les SHS se déconnectent de la culture et de la société environnantes, perdant de ce fait une de leurs raisons d'être. D'autres modèles peuvent être évoqués: au Québec, le rayonnement des travaux (en particulier grâce aux ouvrages) est pris en compte.

Considéré comme radical, peu adapté et insuffisamment mis en débat, le guide des produits de la recherche en STAPS a généré, lors de sa publication en 2018, une certaine défiance matérialisée par des demandes officielles de révision émanant de plusieurs sociétés savantes. En réaction, le Hcéres a promis des évolutions, promesse à ce jour restée lettre morte. Au sein de notre échantillon d'UR évaluées, les comités d'experts font bien peu de cas de ces réserves; au contraire, ils semblent réinterpréter le guide en question, pour en faire une application particulièrement zélée. On assiste dans bien des cas à des entreprises de purification des bilans scientifiques, convergeant vers un indicateur transformé en outil normatif, mis sur un piédestal alors même que nul n'est à même d'en expliquer la mécanique. La considération accordée aux publications échappant à la forme de l'article et/ou n'étant pas couvertes par le *SJR* est réduite à la portion congrue. En tout état de cause, le jugement de la plupart des comités est armé par un référentiel qui n'est pas scrupuleusement respecté.

L'autre résultat marquant est l'élaboration, de la part des enseignants-chercheurs, de stratégies duales leur permettant de composer avec ces nouvelles exigences tout en se ménageant des espaces d'expression et de reconnaissance alternatifs. L'enjeu de la reconnaissance est indissociable des processus d'évaluation; à ce titre, comme pour ce qui est du sens de l'évaluation en contexte professionnel, les perspectives exclusivement normatives montrent de sérieuses limites (Nizet, 2013). Présentée comme nécessaire, l'adaptation observée génère des tensions et n'est sans questionner les interviewés quant au sens de leur métier. En effet, devoir se plier à des critères d'évaluation auxquels on n'adhère pas, pour espérer se réaliser et obtenir une satisfaction professionnelle à côté, est à l'évidence délétère quant au bien-être au travail.

Références

- Andrews, D. (2008). Kinesiology's inconvenient truth and the Physical Cultural Studies imperative. *Quest*, 60, 45-62.
- Archambault, E. et Larivière, V. (2010). The limits of bibliometrics for the analysis of the social sciences and humanities literature. *2010 World social science report. Knowledge divides* (p. 251-254). Paris: UNESCO-International Social Science Council.
- Bardin, L. (2013). *L'analyse de contenu*. Paris: Presses Universitaires de France.
- Berry, M. (2009). Les mirages de la bibliométrie, ou comment scléroser la recherche en croyant bien faire. *Revue du MAUSS*, 33(1), 227-245.
- Charle, C. (2009). L'évaluation des enseignants-chercheurs. Critiques et propositions. *Vingtième siècle. Revue d'histoire*, 102(2), 159-170.
- CNRS (2015). *Fiabiliser l'analyse des publications scientifiques: vers l'optimisation des choix de bases bibliographiques et d'outils de métrique*. Paris: Direction de l'information scientifique et technique.
- Cronin, B. et Sugimoto, C. R. (dir.) (2014). *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact*. Cambridge, MA: MIT Press.
- De Gaulejac, V. (2012). *La recherche malade du management*. Versailles: Éditions Quæ.
- Dejours, C. (2003). *L'évaluation du travail à l'épreuve du réel. Critique des fondements de l'évaluation*. Paris: INRA Éditions.
- Dejours, C. (2011). La psychodynamique du travail face à l'évaluation: de la critique à la proposition. *Travailler*, 25(1), 15-27.
- Delgado-López-Cózar, E. et Cabezas-Clavijo, A. (2013). Ranking journals: could Google Scholar Metrics be an alternative to Journal Citation Reports and Scimago Journal Rank? *Learned Publishing*, 26(2), 101-114.
- Desrochers, N., Paul-Hus, A., Haustein, S., Costas, R., Mongeon, P., Quan-Haase, A., Bowman, T. D., Pecoskie, J., Tsou, A. et Larivière, V. (2018). Authorship, citations, acknowledgments and visibility in social media: Symbolic capital in the multifaceted reward system of science. *Social Science Information*, 57(2), 223-248.
- Deslauriers, J.-P. (1991). *Recherche qualitative. Guide pratique*. Montréal, QC: McGraw-Hill.
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A. et Pappas G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *The FASEB Journal*, 22, 338-342.
- Federation for the Humanities and Social Sciences (2017). *Approaches to assessing impacts in the Humanities and Social Sciences*. Ottawa, Canada. Repéré à https://www.federationhss.ca/sites/default/files/sites/default/uploads/policy/2017/impact_report_en_final.pdf
- Fouley, J.-L., Celeux, G. et Josse, J. (2017) *Empirical Bayes approaches to PageRank type algorithms for rating scientific journals*. Paris: INRIA Éditions.
- Gingras, Y. (2014). *Les dérives de l'évaluation de la recherche. Du bon usage de la bibliométrie*. Paris: Raisons d'agir.
- Gozlan, C. (2016). Les sciences humaines et sociales face aux standards d'évaluation de la qualité académique. Enquête sur les pratiques de jugement dans une agence française. *Sociologie*, 7(3), 261-280.
- Harzing, A. W. et Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics*, 106(2), 787-804.
- Herreros, G. et Milly, B. (2004). Les conditions de pérennisation de la qualité à l'hôpital: de l'intérêt de la latéralité. *Gestions hospitalières: la revue du management hospitalier*, 188, 809-815.

- Hicks, D. (2005). The four literatures of social science. Dans H. F. Moed, F. Henk, W. Glänzel et U. Schmoch, *Handbook of quantitative science and technology research* (p. 473-496). Dordrecht: Kluwer Academic.
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S. et Rafols, I. (2015). Bibliometrics: the Leiden Manifesto for research metrics. *Nature*, 520, 429-431.
- Higher Education Funding Council for England (2014). *The metric tide. Independent review of the role of metrics in research assessment and management*. Repéré à <http://www.hefce.ac.uk/rsrch/metrics>
- Huberman, A. et Miles, B. (1991). *Analyse des données qualitatives: recueil de nouvelles méthodes*. Bruxelles: De Boeck.
- Larivière, V. et Sugimoto C. R. (2018). *Mesurer la science*. Montréal, QC: Presses de l'Université de Montréal.
- Manana-Rodriguez, J. (2015). A critical review of SCImago Journal & Country Rank. *Research Evaluation*, 24, 343-354.
- Mongeon, P. et Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, 106, 213-228.
- Morozov, E. (2013). *To save everything, click here. The folly of technological solutionism*. New York, NY: Public Affairs.
- Mukamurera, J., Lacourse, F. et Couturier, Y. (2006). Des avancées en analyse qualitative: pour une transparence et une systématisation des pratiques. *Recherches qualitatives*, 26(1), 110-138.
- Nizet, I. (2013). Compte rendu de [Paquay, L., Van Nieuwenhoven, C. et Wouters, P. (dir.) (2010). *L'évaluation, levier du développement professionnel. Tensions, dispositifs, perspectives*. Bruxelles: De Boeck]. *Nouveaux cahiers de la recherche en éducation*, 16(2), 186-190.
- Pansu, P., Dubois, N. et Beauvois, J.-L. (2013). *Dis-moi qui te cite et je saurai ce que tu vaux. Que mesure vraiment la bibliométrie?* Grenoble: Presses Universitaires de Grenoble.
- Pontille, D. et Tornay, D. (2013). La manufacture de l'évaluation scientifique. Algorithmes, jeux de données et outils bibliométriques. *Réseaux*, 177, 25-61.
- Robert, P. (2014). Critique de la logique de la «gestionnarisation». *Communication & organisation*, 45, 209-222.
- Schöpfel, J. et Prost, H. (2009). Le JCR facteur d'impact (IF) et le SCImago Journal Rank Indicator (SJR) des revues françaises: une étude comparative. *La psychologie française*, 54(4), 287-305.
- Schreiber, M. (2013). Uncertainties and ambiguities in percentiles and how to avoid them. *Journal of the American Society for Information Science and Technology*, 64(3), 640-643.
- Sivertsen, G. (2016). Patterns of internationalization and criteria for research assessment in the social sciences and humanities. *Scientometrics*, 107(2), 357-368.
- Soulé, B. et Chatal, R. (2018). Évaluer la recherche dans une section universitaire interdisciplinaire : les effets de la conversion bibliométrique au sein des sciences et techniques des activités physiques et sportives (STAPS). *STAPS*, 122(4), 929.
- Tiberghien, G. et Beauvois, J.-L. (2008). Domination et impérialisme en psychologie. *Psychologie française*, 53, 135-155.
- Wagner, U. (2016). The publishing game. The dubious mission of evaluating research and measuring performance in a cross-disciplinary field. *Scandinavian Sport Studies Forum*, 7, 63-88.
- Zarka, Y. C. (2009). L'évaluation: un pouvoir supposé savoir. *Cités*, 37(1), 113-123.