



HAL
open science

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)

Bastien Soulé, Raphaële Chatal

► To cite this version:

Bastien Soulé, Raphaële Chatal. 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). 2019. hal-01981966

HAL Id: hal-01981966

<https://hal.science/hal-01981966v1>

Preprint submitted on 15 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SOULÉ Bastien

Université Claude Bernard Lyon 1

Laboratoire sur les Vulnérabilités et l'Innovation dans le Sport (L-ViS – EA 7428)

UFR STAPS

27-29 Boulevard du 11 Novembre 1918

69022 Villeurbanne Cedex

CHATAL Raphaële

Université de Nantes

Centre Nantais de Sociologie (CENS – UMR CNRS 6025)

23 rue du Recteur Schmitt

BP 81 227

44312 Nantes Cedex 3

Evaluer 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)

Mots-clés : bibliométrie ; évaluation ; interdisciplinarité ; recherche ; STAPS

Résumé : cet article s'intéresse à la question sensible de l'évaluation de la recherche au sein d'une section (les STAPS) dont le caractère fondamentalement interdisciplinaire fait cohabiter des pratiques et cultures scientifiques plurielles. En focalisant notre propos sur la question de la valeur accordée aux publications d'articles dans des revues scientifiques, nous entendons mettre en évidence les effets induits par une double évolution récente : la disparition de la liste AERES de revues pour les STAPS, combinée à la priorité désormais accordée à la bibliométrie d'impact. Un examen factuel des effets générés pour les revues qui figuraient, jusqu'en 2017, dans la liste en question conduit à un constat alarmant. La bibliométrie d'impact maltraite en effet les revues de sciences sociales, et notamment les titres francophones. Etabli de longue date et dans divers contextes nationaux, ce constat est à l'origine de recommandations très claires, dans de nombreux pays (France, Angleterre, Pays-Bas, Norvège, Canada, Suisse, Allemagne, etc.), destinées à garantir une évaluation adaptée aux SHS et respectueuse de leurs spécificités. A cet égard, le virage bibliométrique opéré en STAPS s'inscrit *a contrario* d'une tendance lourde observée aux niveaux national et international, tout en mettant en péril le caractère interdisciplinaire de la section.

Research evaluation in an interdisciplinary section: the effects of a bibliometric conversion within Sports Sciences in France

Key words: bibliometrics; evaluation; interdisciplinarity; research; sports sciences

Abstract: this article addresses the sensitive issue of research evaluation within a section (Sports Sciences) whose interdisciplinary nature brings together plural scientific practices and cultures. By focusing on the question of the value given to the publication of articles in scientific journals, we intend to highlight the effects induced by a recent double evolution: the disappearance of the AERES list of journals for STAPS, combined with the priority now given to impact bibliometrics. A factual examination of the effects generated for the journals that were on the aforementioned list until 2017 leads to an alarming finding. Impact bibliometrics, whatever the preferred database, mistreats social science journals, especially non anglophone ones. This same observation, which has long been established with regard to SSH in various national contexts, has given rise to very clear recommendations in many countries (France, England, the Netherlands, Norway, Canada, Switzerland, Germany, etc.) designed to guarantee an assessment adapted to SHS and respectful of their specific characteristics. In this respect, the bibliometric shift made in french Sports Sciences goes against a heavy trend and endangers, in so doing, the interdisciplinary character of the section.

Introduction

Dans la recherche publique, l'évaluation conditionne les accréditations de certaines entités de recherche et l'obtention de financements ; à l'échelon individuel, les évolutions de carrière et les rémunérations en dépendent fortement. Cet article s'intéresse aux pratiques d'évaluation de la recherche au sein de la 74^{ème} section du CNU, les Sciences et Techniques des Activités Physiques et Sportives (STAPS), qui présente la particularité d'être fondamentalement interdisciplinaire : les membres de cette section mobilisent dans leurs enseignements et travaux de recherche diverses disciplines¹ des Sciences de la Vie et de la Santé (SVS) et des Sciences Humaines et Sociales (SHS) comme la biomécanique, l'ergonomie, l'histoire, les neurosciences, la physiologie, la psychologie, les sciences de l'éducation, les sciences de gestion ou encore la sociologie (pour s'en tenir aux principales). Les cultures comme les pratiques scientifiques y sont par définition plurielles.

Des visions et positionnements distincts coexistent à propos des « produits de la recherche » à prendre en considération lors des évaluations. La question du poids à accorder aux monographies et chapitres d'ouvrages est par exemple vive. C'est toutefois aux articles publiés dans des revues scientifiques que notre analyse est consacrée, car en STAPS comme dans de nombreuses autres sections universitaires, ce mode de valorisation constitue le principal élément d'appréciation de la productivité scientifique des entités de recherche et de leurs personnels.

Les débats relatifs à l'évaluation de la recherche prennent un relief particulier dans le périmètre restreint mais protéiforme des STAPS, faisant de cette section un observatoire

¹ Selon Collinet et Payré (2003), en référence à Berthelot, trois dimensions sont constitutives d'une discipline scientifique. 1) Une dimension *sociale et pragmatique* : à travers une institution disposant de sa propre hiérarchie et dotée d'une organisation interne particulière, chaque discipline permet à une communauté de chercheurs d'exister et de se définir. 2) Une dimension *technique* : chaque discipline renvoie à des habitudes en termes de méthode, ainsi qu'à des manières de procéder spécifiques. 3) Une dimension *cognitive* : au sein d'un espace logique d'argumentation, chaque discipline s'attaque à des problèmes, élabore des concepts et théories, partage un corpus de connaissances et de références, le tout induisant une rationalité propre. La question de savoir si les STAPS forment une discipline, au sens de Berthelot, ne se pose donc pas : elles possèdent certes la première des trois caractéristiques, mais se distinguent par une certaine dispersion des référentiels techniques et cognitifs.

privilegié de l'évaluation interdisciplinaire en acte, avec ses enjeux et difficultés (Wagner, 2016). D'autant que l'adossement aux indicateurs bibliométriques internationaux (*Journal Citation Report* ou *JCR* du *Web of Science*, *Scimago Journal Ranking* ou *SJR* de *Scopus*), de plus en plus prégnant depuis 2015, est en voie de systématisation au sein de la 74^{ème} section du CNU. Cette mobilisation d'outils dont l'intérêt et les limites sont désormais bien renseignés par les spécialistes de bibliométrie² pose plusieurs types de problèmes qui seront documentés dans la première partie de cet article. Pour ce faire, nous nous appuyerons essentiellement sur des écrits abordant les particularités de l'évaluation en SHS. Suite à cette mobilisation de la littérature, nous nous livrerons à une analyse factuelle des effets produits par le recours de plus en plus privilégié au *SJR* lors des évaluations de la recherche en STAPS. L'ambition de cet article est d'informer la communauté des effets du virage bibliométrique qui s'opère, afin de contribuer, sur des bases renseignées et clairement détaillées, au nécessaire débat sur l'évaluation de la recherche au sein de la section des STAPS.

1. Eléments de contexte : une quête d'indicateurs pour évaluer et piloter la recherche

En prônant le passage d'une gestion par les moyens à une gestion par les résultats, le *New Public Management* a bouleversé le paysage de l'enseignement supérieur et de la recherche. Un mouvement qui s'est concrétisé, en France, par la création de l'AERES (Agence d'Evaluation de la Recherche et de l'Enseignement Supérieur) en 2007 (devenue Haut Conseil de l'Evaluation de la Recherche et de l'Enseignement Supérieur ou HCERES en 2013), dans le sillage d'institutions du même type nées dans d'autres pays. Les indicateurs de performance de la recherche publique se sont dès lors multipliés afin que puissent y être adossées des inflexions et priorités politiques (Bessire & Fabre, 2014).

² La scientométrie est la mesure quantitative de l'ensemble des activités scientifiques (Gingras, 2014). La bibliométrie d'impact, pour sa part, renvoie au volet évaluatif de la scientométrie ; c'est un sous-ensemble de la scientométrie qui se focalise sur l'analyse des publications et de leurs propriétés.

Pour mener à bien ces évaluations, plusieurs métriques sont utilisées à l'échelon international, avec en tête l'*Impact Factor (IF)* des revues scientifiques fourni par le *JCR*, adossé au *Web of Science (WoS)* du groupe *Thomson-Reuters*. De tels indicateurs sont attractifs, voire providentiels en termes de prise de décision (Cronin & Sugimoto, 2014 ; Woelert, 2013) puisqu'ils rendent *a priori* commensurable, même sans expertise des domaines scientifiques considérés, le prestige relatif des revues dans lesquelles les chercheurs publient leurs travaux. Les décideurs soucieux d'optimiser les dépenses publiques ne sont pas les seuls à être friands de telles métriques : les managers d'institutions académiques (concurrentes en matière de moyens financiers, de scientifiques de renom et d'étudiants de qualité) se sont aussi mis à en faire un usage plus ou moins prononcé. Au sein même des unités de recherche, les chercheurs ont recours, de manière parfois zélée, à de tels indicateurs permettant d'« objectiver » leurs performances, quand il s'agit par exemple de dresser un bilan d'activité, de solliciter une promotion ou de candidater à un appel à projets (HEFCE, 2014). A tel point que dans le paysage académique remodelé de la fin des années 2000, cette importance grandissante accordée aux métriques internationales (principalement adossées au *WoS* et dans une moindre mesure à *Scopus*) fut décrite comme inéluctable : sur fond de quête d'excellence et de mondialisation de la recherche, d'aucuns finirent en effet par considérer ces indicateurs comme un gage de qualité et d'objectivité plus solide que les standards propres à chaque discipline (Sivertsen, 2016).

Parallèlement, le poids accordé à ces critères d'évaluation est jugé excessif par une part grandissante de la communauté scientifique ; conçus et exploités de manière insuffisamment réfléchie, « *ils dominent les esprits, altèrent les comportements et déterminent les carrières* » (HEFCE, 2014). Toutes disciplines confondues et quel que soit le pays, le nombre de chercheurs exprimant leur désapprobation quant à ce mode d'appréciation de leur activité, quantitatif et unidimensionnel, va croissant (RNAAS, 2013 ; HEFCE, 2014 ; CDHSS, 2010 ;

Bornmann & Werner, 2016). Les imprécisions et simplifications induites ont de telles conséquences négatives qu'elles ont conduit à formuler, en 2013, la *Declaration on Research Assessment* (DORA³) (Pulverer, 2015), puis, en 2015, le manifeste de Leiden⁴, ensemble de principes et de recommandations pour un usage raisonné et responsable de la bibliométrie à des fins d'évaluation.

Un instrument supplémentaire est apparu dans plusieurs pays au cours de la décennie 2000 : les listes de revues relevant des sciences sociales et des Humanités, parfois agrémentées de classements (en trois rangs le plus souvent). L'objectif était alors double : disposer d'un outil alternatif pour évaluer les productions scientifiques, ajusté au domaine des SHS confronté aux biais disciplinaires et linguistiques du *WoS* et de *Scopus* (Hicks, 1999 ; Katz, 1999 ; Pons-Novell & Tirado-Fabregat, 2010)⁵ ; mais aussi, dans une certaine mesure, inciter les chercheurs à publier dans des revues internationales (les mieux classées au sein de ces listes) afin d'accroître la reconnaissance et la compétitivité de leurs entités de recherche (Pontille & Torny, 2013). En France, à partir de 2008, l'AERES a ainsi impulsé la création de listes de revues relevant des SHS, en la justifiant par le caractère inapproprié des bases de données (BDD) internationales pour évaluer l'impact et la productivité des entités de recherche de ce domaine. Très présente à l'origine, la hiérarchisation des revues composant ces listes AERES a été abandonnée dans près de 90% des disciplines, au motif notamment qu'il était suffisant de délimiter un périmètre de supports éditoriaux présentant des garanties en termes de scientificité et de sélectivité (Williams & Galleron, 2011).

2. Revue de littérature

³ <https://sfedora.org/read/>

⁴ <https://vimeo.com/133683418>

⁵ Ces listes n'existent pas seulement dans les pays non-anglophones (Espagne, France, etc.), mais aussi dans certains pays anglophones soucieux de ne pas voir déconsidérées les revues scientifiques nationales (comme l'Australie).

2.1. De quelques invariants internationaux en matière de diffusion scientifique en SHS

Figures reconnues de la recherche en sciences sociales sur les activités physiques et sportives, Terret et Vigarello (2012) dressent le constat suivant : « *dans le domaine des sciences sociales, les critères de qualité (« excellence ») d'une production scientifique ne sont pas forcément, à ce jour, la publication en langue anglaise. La reconnaissance de cette spécificité (qui ne doit pas nous exonérer de l'effort de faire connaître nos travaux en publiant occasionnellement en anglais) est nécessaire* ». Pour mieux comprendre l'exception dont il est question et décentrer l'analyse de la seule question linguistique, il convient d'appréhender d'autres spécificités des sciences sociales en matière de diffusion des travaux de recherche. Sivertsen (2016) souligne à ce propos l'existence de particularités transnationales. Quel que soit le pays, les objectifs de recherche en sciences sociales sont souvent ancrés dans le local et/ou tournés vers l'impact sociétal. Par conséquent, les chercheurs souhaitant remplir correctement leurs missions, présenter clairement leurs résultats et toucher efficacement les publics adéquats sont amenés à privilégier une diffusion de leurs travaux *via* des canaux hétérogènes : articles dans des revues internationales, mais aussi nationales, professionnelles ou grand public, ouvrages, rapports, etc. Des convergences internationales se dessinent ainsi : en sociologie et en histoire, par exemple, 45% des publications en Belgique et en Norvège sont des articles dans des revues scientifiques ; un peu moins d'un tiers sont des chapitres d'ouvrages, mais la moitié des chercheurs utilisent ce type de support pour diffuser leurs travaux. Les deux tiers des sociologues qui écrivent dans des revues scientifiques utilisent en parallèle d'autres supports de publication.

Par ailleurs, dans tous les pays, publier dans sa langue natale est bien plus fréquent en SHS que dans les sciences expérimentales ou médicales (Sivertsen, 2016). En Norvège comme en Belgique, la majorité des publications en SHS sont en anglais et dans des revues scientifiques ; mais la majorité des chercheurs de ce champ publient aussi des ouvrages et/ou

articles dans leur langue natale. Il s'agit en fait d'une norme observable dans la plupart des pays. Se tourner vers l'international sert à échanger avec un petit nombre d'experts, dans des domaines spécialisés, afin de contribuer à des développements théoriques et méthodologiques. Alors que publier dans des revues nationales (un tiers à 50% des productions d'articles) permet d'échanger autour de thématiques spécifiques à des contextes ou intérêts nationaux (renvoyant à une des raisons d'être des sciences sociales). Pour résumer, en SHS et plus particulièrement en sciences sociales, les différents supports de valorisation se complètent, davantage qu'ils n'entrent en concurrence (Sivertsen, 2016). Ces normes et pratiques, également constatées dans de nombreux autres contextes nationaux⁶ (Van Leeuwen, 2006 ; Nederhof, 2006 ; Ossenblok et al., 2012 ; HEFCE, 2014 ; Ochsner et al., 2016a ; Pölönen et al., 2017), contribuent à soustraire à la mesure bibliométrique une partie considérable de la production en SHS (notamment celle qui contribue à la pertinence sociétale de ce domaine⁷). La diversité des cultures scientifiques, consubstantielle des STAPS, devrait inciter à tenir compte de telles spécificités. Pour autant, la description faite par Terret et Vigarello (2012) semble indiquer que l'on est loin du compte : *« les modes d'évaluation de l'excellence scientifique sont fondés en priorité sur des modèles bibliométriques qui, malgré leurs limites reconnues en sciences sociales, sont imposés à tous comme une norme »*. Pour dépasser une explication en termes de particularisme des SHS, la teneur des *modèles bibliométriques* dont il est question doit être abordée : dans quel but et comment sont-ils construits ? Que permettent-ils d'appréhender ? De quelles limites sont-ils porteurs ?

⁶ À ce sujet, dans le document *Catégorisation des revues en Economie et en Gestion* de la section 37 du Comité national du CNRS, on peut lire p. 4 : *« Des travaux spécifiques sur la France peuvent être de grande qualité sans nécessairement être acceptés dans des revues de langue anglaise. Dans certains cas, la langue française ou la publication pour un public francophone peuvent permettre une expression bien plus riche. Les revues françaises jouent donc un rôle important dans la diffusion des connaissances, en particulier auprès de la communauté des économistes et gestionnaires au-delà de la sphère académique. Cette diffusion est même parfois plus large que celle de certaines revues anglo-saxonnes à audience infra-nationale ou très spécialisée »*.

⁷ À cet égard, publier dans des revues peu prestigieuses (au regard de leur *IF*) n'est pas nécessairement synonyme de qualité scientifique discutable : c'est même parfois la condition *sine qua non* pour diffuser des résultats de recherche pertinents sur le plan social et économique, mais pas ou peu accueillis dans les revues phare (Chavarro et al., 2017).

2.2. La mécanique des métriques internationales

Eugene Garfield a créé l'*IF*⁸ dans les années 1950 pour permettre aux bibliothécaires de cibler leurs abonnements vers les revues scientifiques publiant les articles les plus cités (Pulverer, 2015 ; Bornmann & Werner, 2016). Bien qu'il ait très tôt envisagé l'utilisation de cet indicateur à des fins d'évaluation des chercheurs (Pontille & Torny, 2017), son intention initiale n'était pas de contribuer à la mesure de la qualité de la recherche ou de la valeur des contenus d'une revue. La banalisation de cet indicateur assez rudimentaire à des fins d'évaluation relève donc d'un usage détourné, ce qui explique bien des inadéquations et approximations largement documentées (Pendlebury, 2009 ; Hicks et al., 2015 ; HEFCE, 2014 ; Gingras, 2014 ; Pansu et al., 2013 ; Pontille & Torny, 2013).

Tableau 1 : limites et effets de l'utilisation de l'<i>IF</i> à des fins d'évaluation de la recherche	
--	--

Valeur moyenne associée	L' <i>IF</i> ne dit rien de l'impact effectif d'un article publié dans une revue, ni de sa valeur intrinsèque (des articles publiés dans une revue à fort
-------------------------	---

⁸ Le mode de calcul de l'*Impact Factor* est détaillé à l'adresse suivante : https://guides.biblio.polymtl.ca/mise_en_valeur_de_la_recherche/impact_revues

à une revue et non à son contenu	<i>IF</i> reçoivent souvent un faible nombre de citations). L' <i>IF</i> d'une revue dépend d'un nombre restreint d'articles très cités (répartition de type Pareto)
Recensement de citations sans connaissance de leur origine	Un propos provocant peut être cité à de nombreuses reprises, même si sa contribution scientifique s'avère faible Plus largement, citations de teneur très différente mises sur le même plan
Influence de la démographie des communautés scientifiques	<i>IF</i> d'une revue très fortement corrélé au nombre combiné de chercheurs et de revues au sein d'une communauté scientifique (en mathématiques, chercheurs assez peu nombreux, <i>IF</i> des revues spécialisées faible ; en sciences médicales, nombreux chercheurs et nombre restreint de revues spécialisées, <i>IF</i> élevés)
Calcul peu reproductible	Ecart reposant sur ce qui est effectivement pris en considération pour établir le numérateur (nombre de citations d'articles d'une revue sur une période donnée) et le dénominateur (nombre d'articles publiés par cette même revue sur la période considérée)
Illusion de précision	Au regard des approximations soulevées, niveau de précision (trois décimales) usurpé et servant surtout à faciliter la hiérarchisation des revues ainsi que leur classement par quartiles (lui-même artificiel et sujet à caution)
Incitation aux manipulations	Auteurs soumettant un article à une revue parfois invités à citer plus que nécessaire des articles récemment publiés dans ladite revue Plus largement, développement d'une « culture de la citation » anticipant certains effets bibliométriques générés (de la part d'éditeurs, auteurs, experts)

Autant d'insuffisances ayant récemment amené le CNRS (2015) à mettre en avant sa volonté d'en finir avec les nombreux usages impropres de la bibliométrie, l'utilisation abusive et biaisée des indicateurs de citation et de mesure d'impact.

Plusieurs spécificités des sciences sociales se répercutent négativement sur les scores d'impact des revues de ce champ, particulièrement bas quelle que soit la BDD prise en considération (Gingras, 2014) : nombre élevé de revues scientifiques en SHS ; rythme de publication plus faible que dans les sciences expérimentales et biomédicales (Pansu et al., 2013) ; fréquentes références à des ouvrages et chapitres d'ouvrage, aux côtés des mentions faites à des articles de revues (Gingras, 2014)⁹ ; calcul de l'*IF* sur une période de 2 ans peu cohérent avec la temporalité des avancées théoriques en sciences sociales (plus lente qu'en

⁹ 50% des références dans les articles relevant des sciences sociales renvoient à des livres et non à des revues. « Les contributions à des ouvrages collectifs, qui sont fréquentes et importantes en sciences sociales, sont ainsi dévalorisées, simplement parce qu'on ne peut pas leur attribuer de facteur d'impact » (Gingras, 2014, p.40).

sciences de la nature)¹⁰. L'*IF* est donc « aveugle » à la majorité des citations faites à des travaux relevant des disciplines des sciences sociales, *a fortiori* quand l'auteur écrit en français, privilégie des méthodes qualitatives (Vogel et al., 2017) et des approches interprétatives (Andrews, 2008). Il ne fonctionne en fait que pour les disciplines dans lesquelles les BDD contiennent non seulement les travaux citant, mais aussi la majorité des travaux cités. Une double condition présente dans près de 100% des cas dans les disciplines relevant des SVS, mais seulement dans 40 à 60% des cas en mathématiques ou en économie, et dans une proportion encore moindre dans les sciences sociales et les Humanités (Krull & Tepperwien, 2016). Gingras (2014) en conclut que les BDD comme le *WoS* mais aussi *Scopus* ne constituent pas des standards pouvant s'appliquer à tous les domaines disciplinaires.

Les partisans de l'usage du *SJR* à des fins d'évaluation de la recherche en SHS voient dans cette métrique une alternative à l'*IF*, au motif que *Scopus* serait davantage intégrateur de revues relevant des sciences sociales et/ou publiées dans d'autres langues que l'anglais. Tous périmètres disciplinaires confondus, *Scopus* couvre effectivement davantage de revues scientifiques que le *WoS* (près de 23000 titres en 2016, contre environ 13000 pour son concurrent). La couverture des sciences sociales par le *SJR* demeure néanmoins limitée (Sivertsen, 2016) et peu satisfaisante (Mongeon & Paul-Hus, 2016), du fait de l'hétérogénéité des supports de diffusion des connaissances utilisés en SHS (Hicks, 2004 ; Archambault et al., 2006 ; Engels et al., 2012 ; Sivertsen, 2014), mais aussi de faiblesses structurelles propres à *Scopus* (comme la couverture incomplète des articles parus dans certaines revues) (CNRS, 2015). En résultent une faible précision de *Scopus* pour les SHS¹¹ (Hicks & Wang, 2011;

¹⁰ Un exemple est fourni par Gingras (2015, p. 73) : « le taux de citations moyen sur deux ans des articles de la revue médicale *The Lancet* en 1980 était de 2.4, et celui de *l'American Sociological Review* de 1.8 ; mais, calculés sur une période de dix ans, les résultats sont très différents. L'impact de la revue de sociologie (20.9) dépasse alors grandement celui de la revue médicale (14) ».

¹¹ A titre d'exemple, en Norvège, au sein de la catégorie des revues scientifiques internationales (telles que les définissent les chercheurs en SHS), le taux de couverture dans le *SJR* est inférieur à 50% (Sivertsen, 2016).

Nederhof, 2006) et de fortes divergences entre disciplines au sein de ces dernières (Sivertsen, 2016). En 2009, Schöpfel et Prost (2009) observaient que la seule différence significative entre le *SJR* et le *JCR*, à propos de l'intégration des titres français, était la meilleure couverture du premier pour la médecine et la pharmacologie. Au final, les biais générés par *Scopus* sont donc proches de ceux, bien connus, du *WoS*, et les classements de revues diffèrent du reste très peu entre ces deux référentiels (Manana-Rodriguez, 2015). Sivertsen (2016) décrit même une tendance au déclin du taux de couverture des revues de SHS par *Scopus*, du fait du rythme élevé de création de nouveaux titres. Mobiliser le *WoS* ou *Scopus* à des fins d'évaluation dans une section interdisciplinaire introduit donc des biais qui défavorisent les SHS (Mongeon & Paul-Hus, 2016).

L'ouverture du *SJR* aux revues non-anglophones reste également limitée. Harzing et Alakangas (2016) soulignent même l'invisibilité des chercheurs non-anglophones sur le *WoS* et sur *Scopus*.

L'un des avantages concurrentiels mis en avant à propos du *SJR* est sa capacité à classer par quartiles les revues au sein de nombreux périmètres disciplinaires et thématiques, pour une meilleure prise en compte des spécificités de chaque champ. Un argument de prime abord recevable pour niveler les effets démographiques évoqués *supra*, qui font que les revues de certaines disciplines (physiologie par exemple) ont mécaniquement un score d'impact supérieur à d'autres (psychologie par exemple). Toutefois, la manière dont *Scopus* ventile les revues entre périmètres est floue, quand elle ne génère pas de flagrantes erreurs¹². Par ailleurs,

¹² La revue *Mobilization* est par exemple uniquement classée au sein du périmètre des transports, alors que c'est une revue de sciences politiques et de sociologie (Schreiber, 2013). Dans le domaine qui nous intéresse, de nombreuses revues sont associées à des champs sans lien avec leur ligne éditoriale : la revue *Loisir & Société* à « Métaux et Alliages » ; la revue *Sociologie & Sociétés* à « Sciences Biologiques et de l'Agriculture » ; le *Sociology of Sport Journal*, la revue *STAPS* et le *Journal of Sport Management* à « Médecine du Sport et Orthopédique » et « Physiothérapie, Thérapie Sportive et Réadaptation » ; la revue *Réseaux* à « Ingénierie Electrique et Electronique » ; les revues *European Studies in Sports History* et *Sport in History* à « Physiothérapie, Thérapie Sportive et Réadaptation » ; la revue *Gender & History* à « Géographie, Développement et Planification » ; etc. Par ailleurs, une revue spécifique au champ des STAPS en France (*Science & Motricité*) est enregistrée sous deux noms, avec des classements différents. Monnaie courante, ces

comme déjà évoqué à propos du *JCR* (cf. tableau 1), le classement par quartiles proposé par le *SJR* ne permet pas une comparaison interdisciplinaire fiable (Schreiber, 2013 ; Manana-Rodriguez, 2015). L'instabilité dans le temps du *SJR* est également problématique en termes d'évaluation : les scores et classements par quartiles qui en découlent évoluent sans offrir une traçabilité stable des *rankings* antérieurs, générant des modifications rétrospectives de classement (Manana-Rodriguez, 2015).

Enfin, même si elle fait l'objet de critiques (Rossner et al., 2007), la méthode de calcul de l'*IF* est simple et connue. A l'inverse, la méthodologie et l'algorithme qui structurent un indicateur comme le *SJR* ne sont pas seulement plus (et trop) compliqués (Falagas et al., 2008) ; ils sont aussi flous. Cette métrique est censée intégrer deux variables pondératrices qui la distinguent de l'*IF* : le prestige des revues citant les articles d'autres revues (exprimé par un bon classement dans le *SJR*) ; la proximité thématique (évaluée via *Scopus*) entre les revues citant et les revues citées (HEFCE, 2014). Sans juger du bien-fondé de la proposition, la « mécanique » mise en œuvre pour ce faire demeure à bien des égards mystérieuse, à propos de ce qui est réellement mesuré comme de la manière dont s'effectue la mesure (Manana-Rodriguez, 2015). Le caractère récursif de l'algorithme à l'origine du premier indicateur interroge les spécialistes de bibliométrie, qui déplorent, d'une part, un manque de définition de la dimension « prestige », et, d'autre part, une absence d'éléments méthodologiques clairs quant à cette mesure. On peut néanmoins inférer que ce qui confère du prestige à une revue scientifique, et par conséquent aux revues que cette dernière cite, c'est le nombre de citations qu'elle reçoit de la part d'autres journaux évalués selon le même processus (Manana-Rodriguez, 2015). Dans les faits, en dépit de la publication de l'algorithme¹³, nul n'est en mesure, à l'extérieur du groupe *Elsevier* (propriétaire de la BDD *Scopus*), de valider ou répliquer les résultats mis en avant *via* cette métrique. Or, quelles que soient les métriques

imprécisions jettent un certain discrédit sur le *SJR*, fragilisant notamment les classements par quartile proposés au sein de chaque périmètre disciplinaire ou thématique.

¹³ <https://www.scimagojr.com/SCImagoJournalRank.pdf>

développées, la moindre des exigences est que les données sur lesquelles elles s'appuient puissent être auditées par la communauté évaluée par ce biais (Pulverer, 2015).

2.3. L'indépendance des BDD et des producteurs de métriques

Thomson-Reuters, propriétaire de l'*ISI* et du *WoS*, n'est pas producteur de contenus scientifiques. A l'inverse, *Elsevier* produit diverses métriques (dont le *SJR*) et est par ailleurs le leader mondial de l'édition scientifique (avec un Chiffre d'Affaires de plus de 2,5 milliards d'euros en 2014 et une profitabilité très élevée). La suspicion de conflit d'intérêt, soulevée par des experts en bibliométrie comme Bergstrom et West (2018), fait peser un doute sur l'objectivité des métriques proposées par *Scopus*, qui appartient au groupe *Elsevier*. A titre d'exemple, la dernière métrique lancée par *Scopus Metrics* (*CiteScore*, fin 2016) a pour effet perturbant de « dégrader » la valeur des revues du groupe *Nature* (propriété de *Springer*, concurrent d'*Elsevier*) de 40% par rapport à ce que suggérait jusque-là l'*IF* de ces revues. Inversement, *CiteScore* fait bénéficier les titres du groupe *Elsevier* d'un gain de 25%, selon le même critère. Bergstrom et West (2018) ont établi ce constat en comparant les deux métriques sur la base d'un échantillon de près de 10 000 revues scientifiques. En soi, le double positionnement d'*Elsevier* est problématique étant donné que cette entreprise contrôle des métriques qui évaluent ses propres revues et celles d'éditeurs concurrents. En France, le CNRS a officiellement sensibilisé la communauté scientifique quant à l'ambiguïté de cette situation : « *les usages bibliométriques demeurent aujourd'hui encore largement dépendants des stratégies éditoriales et commerciales de l'édition scientifique internationale* » (CNRS, 2015, p.14).

3. Problématique et méthode

Au regard de cette revue de littérature, l'*IF* comme le *SJR* présentent d'importantes fragilités structurelles qui rendent problématique voire imprudente leur utilisation à des fins d'évaluation de la recherche. Il convient donc d'élaborer des référentiels d'évaluation en phase avec les modes de production et de diffusion des savoirs en vigueur au sein des différentes disciplines. C'est précisément la raison pour laquelle l'AERES a encouragé, à la fin des années 2000, la constitution de listes de revues.

Or, la liste AERES de revues pour le périmètre des STAPS, créée fin 2009 et révisée en 2012 au terme de consultations élargies au sein de la communauté des enseignants-chercheurs de la section, a été supprimée en 2017, sur décision d'un groupe de travail mandaté par le HCERES pour réviser le référentiel d'évaluation des productions scientifiques des unités de recherche en STAPS. L'un des motifs invoqués étant que le *SJR*, plus inclusif que le *JCR*, convenait à l'ensemble des disciplines constitutives de la section et fournissait un outil d'évaluation équitable. Au regard de notre revue de littérature, les sciences sociales sont vulnérabilisées par cette double évolution (recours systématique au *JCR* et au *SJR* à des fins d'évaluation ; disparition de la liste de revues du périmètre STAPS), quand bien même la possibilité est laissée de faire subsidiairement référence aux listes de revues constituées par le HCERES pour d'autres disciplines des SHS¹⁴. Une telle évolution a de quoi surprendre, tant elle s'inscrit *a contrario* de la recommandation phare des spécialistes de la bibliométrie d'impact consistant à ne pas placer les BDD internationales au cœur du système d'évaluation de la recherche.

Afin d'aller au-delà du constat établi sur la base de la littérature scientifique, nous chercherons à savoir si la convocation accrue de la bibliométrie d'impact, telle qu'elle est constatée dans la section des STAPS, constitue une tendance plus généralement observable dans les pratiques d'évaluation de la recherche : est-elle recommandée, pour les disciplines

¹⁴ Au moment de l'écriture de cet article, une seule liste de ce type est consultable sur le site du HCERES (pour l'économie et la gestion).

relevant des SHS, dans les rapports institutionnels, les référentiels et autres « guides de bonne pratique » en matière d'évaluation produits dans différents pays ? Il est en effet envisageable qu'en dépit des mises en garde émanant de chercheurs, les organisations en charge de la gestion et du développement de la recherche gardent en ligne de mire, sur fond de quête d'excellence et d'optimisation du rang des entités de recherche, les métriques internationales, exerçant du coup des incitations à s'appuyer sur ces indicateurs, y compris dans des domaines s'y prêtant peu.

De manière plus concrète, nous entendons ensuite mettre en lumière les effets induits par l'invitation de la bibliométrie en STAPS, conjuguée à la disparition de la liste AERES de revues pour cette section. Pour ce faire, nous procéderons à un examen factuel du devenir des revues qui figuraient, jusqu'en 2017, dans la liste en question, en prenant comme nouvel étalon le quartile du *SJR*, supposé plus favorable que le *JCR* et son *IF* aux revues de SHS et/ou non anglophones.

Tableau 2 : modes spécifiques de mise à l'épreuve du terrain par objectif		
Objectifs	Indicateurs	Sources d'information
Disposer d'éléments de comparaison Cerner les tendances lourdes en matière d'évaluation en SHS	Recommandations et « bonnes pratiques » préconisées en matière d'évaluation de la recherche par les instances académiques et agences d'évaluation	En France : documents AERES/HCERES ; rapports du CDHSS et de l'Académie des Sciences ; rapports ministériels (2010-15) A l'international : rapports produits aux Pays Bas, en Grande Bretagne et au Canada
Observer les effets du changement de référentiel d'évaluation au sein de la section STAPS pour les travaux en SHS	Comparaison du classement <i>SJR</i> par quartile avec les classements de la liste AERES STAPS	Informations fournies par le <i>Scimago Journal Ranking</i> (période 2011-2016) à propos des revues de sciences sociales figurant sur la dernière liste AERES STAPS (2012)

Le recueil de données a aussi pris la forme d'une « attention flottante » (Wacheux, 1999), depuis 2014, en tant qu'expert HCERES pour l'évaluation des unités de recherche en STAPS et directeur d'un laboratoire interdisciplinaire en SHS évalué par l'AERES en 2014. Une démarche inspirée de celle de Bessire et Fabre (2014) dans leur étude sur l'évaluation de la

recherche en sciences de gestion. Ce corpus permet une analyse multi-scalaire par bien des aspects incomplète, mais néanmoins renseignée, de l'évolution en cours au sein de la section STAPS.

4. Résultats

4.1 Analyse des préconisations institutionnelles pour l'évaluation de la recherche en SHS

Les conclusions émanant de la quinzaine de documents analysés ci-après sont fortement redondantes. Adossés à des expériences concrètes d'évaluation, ces rapports tirent aussi les leçons des résultats de recherches menées sur la bibliométrie d'impact. Les métriques dominantes sont considérées comme inadaptées à une évaluation transparente, claire, précise, juste, constructive et responsable dans les SHS, fragilisant les prétentions de comparaisons interdisciplinaires élargies. La recommandation la plus fréquemment formulée est de recourir à des critères et indicateurs pluriels, adaptés à chaque champ disciplinaire.

Le tableau 3 synthétise le contenu des documents produits en France, alors que le tableau 4 s'intéresse à quatre rapports d'autres pays.

Tableau 3 : synthèse de l'analyse documentaire (échelle nationale)				
	Rapport CDHSS (2010)	Rapport de l'Académie des Sciences (2011)	Référentiels d'évaluation AERES/HCERES	Rapport Pumain & Dardel

			(2014-2015)	(2014)
Respect des spécificités disciplinaires	<p><i>Problème majeur : insuffisante prise en compte d'une spécificité des SHS (place des livres et ouvrages édités) dans les indicateurs bibliométriques</i></p> <p><i>Bonnes pratiques de l'évaluation devant être définies par discipline</i></p>	<p><i>Disciplines ayant toutes leurs spécificités en termes de publication et d'usage de la bibliométrie</i></p> <p><i>La bibliométrie ne permet pas de comparer des chercheurs de différentes disciplines</i></p>	<p><i>SHS à considérer à travers leurs particularités disciplinaires (place centrale des ouvrages par exemple)</i></p> <p><i>Etre respectueux de la diversité des cultures et usages scientifiques</i></p> <p><i>D'autres langues que l'anglais peuvent constituer le véhicule reconnu de la recherche¹⁵</i></p>	<p><i>Ne pas imposer un modèle unique, admettre une variabilité de critères en fonction des disciplines</i></p> <p><i>S'adapter aux logiques et cultures des communautés disciplinaires</i></p>
Pertinence des BDD internationales pour l'évaluation en SHS	<p><i>Différences importantes dans la manière dont les BDD reflètent la recherche en SHS</i></p> <p><i>Différences encore plus importantes dans la manière dont elles représentent chaque discipline</i></p>	<p><i>Toutes les disciplines (notamment les SHS) ne sont pas couvertes dans les BDD, inappropriées (absence de bonnes BDD en SHS)</i></p> <p><i>Perspectives effectives d'utilisation de la bibliométrie en SHS limitées à de très rares domaines</i></p> <p><i>Utilisation d'indicateurs bibliométriques absurde pour évaluer en SHS ; peut mener à de graves erreurs d'appréciation</i></p>	<p><i>Plusieurs secteurs disciplinaires confrontés à la faible présence des revues de référence dans les BDD bibliométriques</i></p> <p><i>Utilisation d'indices bibliométriques pour mesurer la performance inadaptée dans un grand nombre de domaines</i></p> <p><i>Les listes de revues sont classantes dans les disciplines pour lesquelles les BDD internationales sont majoritairement acceptées</i></p>	<p><i>Rôle de la bibliométrie à moduler très fortement selon les disciplines</i></p> <p><i>BDD encore largement inadaptées dans la plupart des disciplines des SHS</i></p>
Place de l'évaluation qualitative par les pairs	<p><i>Postulat de la disjonction entre les qualités respectives des articles et des revues qui les publient</i></p> <p><i>Bonne évaluation devant reposer sur la lecture des productions</i></p>	<p><i>Seule l'évaluation qualitative peut apprécier l'originalité et l'esprit d'innovation d'un chercheur</i></p>	<p><i>Qualité d'une production irréductible à des éléments quantitatifs : valeur non universelle d'éléments calculés automatiquement</i></p> <p><i>Evaluation qualitative et collégiale par les experts devant rester la perspective principale</i></p>	

L'injonction à ne pas se fier outre-mesure aux indicateurs bibliométriques fournis par le WoS

¹⁵ Dès le début des années 2000, la section 37 du CNRS constatait sa propre difficulté, voire son incapacité à accorder la place qui leur revient aux revues de langue française : « *l'ignorance de la langue française conduit les classifications internationales à sous-évaluer dans une certaine mesure les revues publiant des articles en français. Le classement actuel tente de corriger, au moins en partie, ce biais. Nous sommes toutefois parfaitement conscients que l'effort aurait pu être encore poussé plus loin, notamment dans certaines sous-disciplines* » (section 37 du CNRS, *Classement des revues à comité de lecture en économie et gestion*, 29 avril 2003).

ou *Scopus* est claire, en raison de leur caractère non exhaustif, de l'existence d'autres BDD plus adaptées ou de formes de publications échappant à leur hégémonie (*Peer Community In*, revues nationales ou régionales, supports de diffusion auprès du monde socio-économique, revues professionnelles ou techniques, etc.). Le caractère parfois très biaisé des différents facteurs d'impact fournis par ces BDD est clairement évoqué, notamment à propos des revues en *open access*. Il est donc conseillé d'utiliser ces indicateurs non étalonnés avec discernement et de leur préférer l'appréciation qualitative, notamment pour déceler l'originalité et le caractère novateur des recherches¹⁶.

Tableau 4 : synthèse de l'analyse documentaire (échelle internationale)

	<i>Royal Netherlands Academy of Arts &</i>	<i>University of Waterloo Working Group on Bibliometrics</i> (2016)	<i>Higher Education Funding Council for England</i> (2014)
--	--	---	--

¹⁶ D'autant que les articles scientifiques innovants, permettant, à terme, des avancées significatives dans leurs domaines respectifs, sont fréquemment publiés dans des revues à *IF* modéré et tardent à obtenir la reconnaissance académique qui leur est due (Wang et al., 2017).

	<i>Sciences</i> (2013)	<i>Federation for the Humanities and Social Sciences</i> (2017)	
Respect des spécificités disciplinaires	<p><i>Procédures d'évaluation devant tenir compte des différences entre et au sein des domaines, notamment en SHS (grande variété et hétérogénéité)</i></p> <p><i>Mission des SHS devant se refléter dans ses indicateurs de qualité. Diffusion nécessaire des résultats via des supports adaptés aux professionnels, politiques et décideurs qui ne lisent pas les revues scientifiques. Cibler ces groupes externes via des publications adaptées</i></p>	<p><i>Différences de cultures telles que des comparaisons robustes et sensées entre disciplines ne sont pas possibles : par ex., les articles en SHS citent en moyenne moins de références ; les articles à auteurs multiples (générant un nombre élevé de citations) sont moins communs en SHS...</i></p> <p><i>Utilisation de la bibliométrie limitée par les différences dans les pratiques de publication et de citation entre disciplines : chercheurs en SHS publiant moins dans des revues internationales et/ou anglophones, plus dans des revues régionales et ouvrages, moins saisis dans les systèmes de suivi des citations</i></p> <p><i>Impacts scientifiques pluriels et complexes en SHS, ne se limitant pas à ce que mesure la bibliométrie</i></p>	<p><i>Un seul modèle ne peut convenir à tous : un système de recherche mûr doit être à géométrie variable (jugements d'experts, indicateurs quantitatifs et qualitatifs)</i></p> <p><i>Evaluer en tenant dûment compte du contexte et de la diversité disciplinaire. Qualité académique hautement contextuelle. Penser en termes de qualité de la recherche, pas en visant une définition/mesure unique de qualité</i></p> <p><i>Concevoir et utiliser des mesures tenant mieux compte de la diversité de la recherche ; prendre des mesures sensibles et significatives pour refléter cette dernière</i></p>
Pertinence des BDD internationales pour l'évaluation en SHS	<p><i>Indicateurs de qualité devant à la fois couvrir la qualité scientifique et la pertinence sociétale</i></p>	<p><i>Une taille unique ne convient pas à tous ; applicabilité de la bibliométrie différente selon les disciplines</i></p> <p><i>Bibliométrie dévalorisant les productions en langue non anglaise, les travaux interdisciplinaires ou d'échelle régionale</i></p> <p><i>Indicateurs bibliométriques disant peu des impacts élargis (économiques, sociétaux...), indirects et à long terme</i></p>	<p><i>SHS créant des défis particuliers pour les analyses bibliométriques</i></p> <p><i>Orientation de la recherche en SHS poussant à publier dans des revues nationales/régionales non couvertes par les BDD</i></p> <p><i>Rôle important des livres contribuant également à la couverture limitée des publications SHS par les BDD</i></p>
Place de l'évaluation qualitative par les pairs		<p><i>Nécessité d'un ensemble d'indicateurs diversifiés et souples (qualitatifs et quantitatifs)</i></p> <p><i>Pour un usage responsable, bibliométrie devant intervenir comme outil support à d'autres modes d'évaluation (qualitatifs, par des pairs experts des spécificités disciplinaires)</i></p>	<p><i>Impossibilité d'évaluer la qualité scientifique en utilisant uniquement des indicateurs quantitatifs</i></p> <p><i>Faire preuve de prudence lorsqu'on examine à l'aide des BDD : problèmes de couverture, biais... Aucun ensemble de chiffres ne permet de saisir les jugements multiples et nuancés sur la qualité de la recherche</i></p> <p><i>Les métriques doivent appuyer, et non supplanter le jugement d'experts et l'examen par les pairs, principale base</i></p>

De nombreux autres rapports et comités avancent des analyses et recommandations semblables, en Suisse¹⁷, en Allemagne¹⁸, au niveau de l'Union Européenne¹⁹ ou encore de l'UNESCO²⁰. Les risques d'une définition étroite de la valeur de la recherche et du recours systématisé à un jeu limité d'indicateurs sont clairement posés : cela peut conduire à accorder la priorité à ce qui est facilement mesurable, au détriment de ce qu'il est important et/ou souhaitable de faire en recherche²¹.

4.2 Analyse des effets du virage bibliométrique opéré en STAPS pour les sciences sociales

Nous avons testé les effets de la double évolution soulignée sur un échantillon précis : les revues de sciences sociales présentes sur la liste AERES STAPS 2012. Quatre domaines ont été distingués : sociologie, anthropologie et sciences politiques (n=60 revues dont 38% de titres anglophones) ; management et gestion (n=24 à 54% anglophones) ; histoire (n=26 à 69% anglophones) ; sciences de l'éducation (n=50 à 82% anglophones)²². Une recherche systématique de chaque revue a été faite dans le *SJR* : présence ou absence, quartile (en retenant le meilleur en cas de référencement dans plusieurs domaines), écarts avec le classement 2012 (en considérant les 3^{ème} et 4^{ème} quartiles du *SJR* comme le niveau 3 AERES). Quatre états ont été attribués à chaque revue : stabilité, disparition, déclassement et

¹⁷ Projet *Mesurer les Performances de la Recherche* de la Conférence des Recteurs des Universités Suisses (CRUS).

¹⁸ *Etudes Wissenschaftsrat du German Council of Science and Humanities : Steering group report on the pilot study research rating in Chemistry and Sociology* (2008) ; *Recommendations for rankings in the system of higher education and research. Part 1: Research* (2004) ; etc.

¹⁹ *Rapport Assessing Europe's University Based-Research* (2010) produit par la Direction Générale de la recherche de la Commission Européenne.

²⁰ *World Social Science Report. Knowledge Divides* (2010) produit par l'*International Social Science Council* de l'UNESCO. Source : www.unesco.org/shs/wssr

²¹ D'où la recommandation de la Direction Générale de la recherche de la Commission Européenne : « *Mesurer ce qui compte plutôt que compter ce qui peut être facilement mesuré* ». Source : https://ec.europa.eu/research/science-society/document_library/pdf_06/assessing-europe-university-based-research_en.pdf

²² Cinq revues figurent dans deux groupes : *Genèses* (sociologie, anthropologie et sciences politiques et histoire), *Formation Emploi* (sciences de l'éducation et sociologie, anthropologie et sciences politiques), *Politiques & Management Public* (sociologie, anthropologie et sciences politiques et gestion), *Histoire de l'éducation* (histoire et sciences de l'éducation) et *History of Education* (histoire et sciences de l'éducation). Le corpus est donc composé de 155 titres.

promotion. Pour les deux derniers états, l'écart a été quantifié (nombre de rangs perdus ou gagnés).

Une première analyse de ce que produit le passage au *SJR* a été réalisée en avril-mai 2016, offrant une visibilité sur les classements par quartile des années 2011 à 2014. Une seconde analyse, effectuée en février-mars 2018, a permis de solidifier cette comparaison tout en étendant la période couverte aux classements 2015 et 2016.

Tableau 5 : effets sur le référencement et le classement des revues de sociologie, anthropologie et sciences politiques

Etat de la revue	Nombre	Pourcentage
Disparue	17	28%
Déclassée	15	25%
Inchangée	24	40%
Promue	4	7%
Total	60	100%

7% des revues de ce groupe disciplinaire bénéficient d'un classement amélioré *via* le *SJR* ; cependant, 28% ne sont plus référencées et un quart se trouvent rétrogradées. A ce propos, il est à noter que les 15 revues déclassées (parmi lesquelles 13 francophones) ne perdent pas seulement un rang, mais, en moyenne, 1.4 échelon (moyenne de l'écart entre le classement AERES STAPS 1, 2 ou 3 et le classement *SJR* Q1, Q2 ou Q3/Q4).

Pour les revues anglophones, le passage au *SJR* n'a pas d'incidence majeure, mais il assoit une très nette domination : 81% d'entre elles, lorsqu'elles sont couvertes par le *SJR*, sont classées Q1 ; aucune n'est classée Q4 ; et ce sont les seules à progresser en termes de classement (n=4).

A l'inverse, 76% des titres francophones (y compris les plus prestigieux) sont concernés par une disparition ou un déclassement. Sur les 22 titres francophones figurant dans le *SJR*, un seul est classé Q1, 18% Q2, 73% Q3 et 9% Q4.

Ces résultats attestent d'une différence de traitement, en termes de présence comme de classement dans le *SJR*, entre revues francophones (très largement confinées au Q3) et anglophones (trustant le Q1).

Etat de la revue	Nombre	Pourcentage
Disparue	11	46%
Inchangée	8	33%
Promue	5	21%
Total	24	100%

Près de la moitié des revues de gestion et de management auparavant référencées dans la liste AERES STAPS ne sont pas couvertes par le *SJR*. Ces disparitions concernent une revue anglophone sur 13 et 10 revues francophones sur 11 (seule la *Revue Française de Gestion* reste référencée). Par ailleurs, seules des revues anglophones sont promues. Enfin, un effet très manifeste est constaté en termes de classement des revues répertoriées dans le *SJR* : 77% des revues anglophones sont classées Q1 et Q2 (fréquemment au terme de promotions d'un ou deux rangs) contre 0% des revues francophones.

Etat de la revue	Nombre	Pourcentage
Disparue	6	23%
Déclassée	9	35%
Inchangée	7	27%
Promue	4	15%
Total	26	100%

Pour les revues d'histoire, il est moins question d'un traitement différencié des revues anglophones et francophones que d'un déclassé généralisé à ces deux sous-ensembles.

35% des revues repérées sont déclassées et 23% (presque exclusivement anglophones) disparaissent. 15% seulement (toutes anglophones) sont promues.

Tableau 8 : effets sur le référencement et le classement des revues de sciences de l'éducation

Etat de la revue	Nombre	Pourcentage
Disparue	4	8%
Déclassée	6	12%
Inchangée	22	44%
Promue	18	36%
Total	50	100%

Particularité notable, plus d'un tiers des revues de sciences de l'éducation sont promues. Une fois encore, cela concerne exclusivement des revues anglophones (n=17), à une exception près. La moitié des revues promues gagne 2 rangs, passant d'un classement 3 AERES au Q1 du *SJR*. A deux exceptions près, toutes les revues francophones relevant de cette discipline sont *ipso facto* écartées en étant classées dans le Q4 ou en disparaissant.

D'un point de vue général, l'édition annuelle d'un score d'impact assorti d'un classement par quartile amplifie l'ambiguïté de l'outil. En effet, les variations interannuelles de quartile sont fréquentes, en particulier pour les revues francophones et/ou classées Q3 et Q4. A titre d'exemple, au sein du sous-échantillon des revues de sociologie, anthropologie et sciences politiques, 42 variations (majoritairement à la baisse) sont constatées pour les revues francophones, dont 31 entre les Q3 et Q4 ; alors que seules 10 variations (majoritairement à la hausse) sont observables pour les revues anglophones. Le constat interpelle d'autant plus qu'à l'évidence, la qualité d'une revue scientifique n'oscille pas radicalement d'une année sur l'autre. Ne pas disposer d'un outil stabilisé introduit un aléa problématique du point de vue, notamment, des stratégies de publication (ciblage des revues). Cela rend aussi inconfortable la

rédaction de bilans d'activité : faut-il mentionner le classement de la revue au moment de la soumission d'un article, de son acceptation ou de sa publication ?

Notre examen en deux temps (2016 et 2018) dévoile par ailleurs des changements rétrospectifs de classement, pour une seule et même année de référence : par exemple, en 2016, une revue classée Q2 pour l'année 2013 peut figurer au sein du Q3 pour la même année en 2018. Sur les 76 revues de sciences sociales répertoriées par le *SJR*²³, 15 modifications rétrospectives de ce type ont été observées, touchant 20% des titres. Dans le détail, 13 changements à la hausse ont été constatés (revues anglophones exclusivement) et 2 à la baisse (un titre francophone et un germanophone). De tels changements renforcent le caractère instable et évolutif d'un outil censé objectiver la mesure.

En prenant comme point de départ une liste de revues élaborée au terme de consultations élargies au sein de la communauté des STAPS, le constat dressé est en tous points opposé à celui qui souligne l'ouverture du *SJR* aux revues de sciences sociales et/ou non anglophones. Les effets produits au sein des quatre groupes de revues explorés devraient inciter à la prudence : il semble inopportun, à ce jour, de faire du *SJR* un outil d'évaluation systématique pour la section des STAPS. Celui-ci s'apparente en effet à une « boîte noire » qui maltraite les revues de sciences sociales (notamment francophones) qui, jusqu'à 2017, faisaient référence dans cette communauté ; il survalorise à l'inverse de nombreuses revues anglophones qui n'étaient pas, jusqu'alors, considérées comme des supports de premier plan. Le déclassement produit est d'autant plus problématique qu'il se combine à la propension de plus en plus forte, lors des évaluations réalisées au sein de la 74^{ème} section du CNU et par le HCERES, à considérer les classements Q1 et Q2 des revues comme des indicateurs de qualité particulièrement élevée des articles publiés. Entreprendre ce virage et l'imposer à des disciplines hétérogènes, dont les représentants ne partagent pas les mêmes points de vue sur

²³ Exclusion faite des revues de sciences de l'éducation, qui n'ont été intégrées au corpus qu'en 2018.

l'excellence et les modes de diffusion adéquats des résultats de recherche, éloigne automatiquement d'une évaluation fine par les pairs.

Une double évolution, plusieurs interprétations

Au terme de cette analyse, comment interpréter la double évolution constatée (suppression de la liste de revues produite par l'AERES pour les STAPS et centralité simultanément accordée au *JCR* et au *SJR* à des fins d'évaluation) ? Si l'ignorance des effets générés par ces deux décisions liées est en cause, cet article pourra modestement contribuer à prendre la mesure des conséquences pour les entités de recherche relevant des SHS, lorsqu'elles seront soumises à évaluation à travers le nouveau « guide des produits de la recherche » pour les STAPS.

Une interprétation possible consiste à considérer que la liste de revues publiée par l'AERES en 2012 n'avait pas de valeur. Ou encore qu'elle servait de refuge pour les spécialistes des SHS de la section STAPS, leur permettant de ne pas affronter la compétition scientifique internationale et les éloignant de l'excellence. Quoique caricaturale, cette façon de voir semble relativement prégnante au sein de la communauté STAPS, sur fond de faible maturité épistémologique et de méconnaissance des travaux de recherche sur la bibliométrie d'impact. Le « guide des produits de la recherche » pour les STAPS publié en juin 2018 par le HCERES est en effet 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 11 autres guides mis en ligne sur le site du Haut Comité²⁴ au moment de l'écriture de cet article évoquent tous, sans exception, des points de vigilance vis-à-vis de l'usage des BDD internationales à des fins d'évaluation : sous-représentation des revues de certaines disciplines ; absence de plusieurs grandes revues généralistes ; sur-cotation ou sous-cotation de certains journaux ; etc.

²⁴ Ils concernent les sciences de la terre et de l'univers, la physique, la chimie, la mécanique, le génie civil, les mathématiques, l'informatique, la robotique, la physiologie, les neurosciences, la biologie, l'immunologie, l'agronomie, les sciences économiques et de gestion, l'histoire, l'histoire de l'art et l'archéologie.

La dérive vers le « populisme évaluatif » conduit à considérer qu'il est préférable de disposer de chiffres imparfaits plutôt que de ne pas être en mesure de quantifier (Gingras, 2014). Confrontées à la complexité et à des difficultés d'identification des réels facteurs de succès, les organisations tendent selon Di Maggio et Powell (1983) à copier celles qui sont considérées comme performantes et/ou ayant réussi, plutôt que de se mettre en quête de solutions originales. Dans le domaine académique, la jeunesse d'une discipline génère souvent un sentiment d'insuffisante reconnaissance. Celui-ci peut inciter à s'engager dans une quête de crédibilité par le biais d'indicateurs présentés comme objectifs, indiscutables, calqués sur les pratiques des sciences exactes et utilisés comme « machine à légitimer » (Bessire & Fabre, 2014). Observé en sciences de gestion (Bennis & O'Toole, 2005 ; Durand & Dameron, 2011 ; FNEGE 2010), ce phénomène d'isomorphisme mimétique semble également présent dans les STAPS. Il se double, pour les travaux réalisés en SVS, d'un attrait pour les index bibliométriques en vogue dans les « disciplines mères ».

Une interprétation politique peut également être formulée, dans la continuité du point précédent. L'élaboration et le choix des infrastructures informationnelles dédiées à la bibliométrie valorisent, dans une activité de recherche, ce qui est considéré comme essentiel et marginalisent ce qui est perçu comme accessoire (Pontille & Tornay, 2013). L'isomorphisme devient coercitif dès lors qu'en dépit de l'absence de consensus (autour de valeurs ou de cadres cognitifs)²⁵, un alignement des comportements est néanmoins visé. La quête de performance caractéristique du *New Public Management* pousse en fait à éluder la question éthique et celle de la justice des rapports sociaux. Les métriques deviennent alors des institutions pourvoyeuses de procédures gouvernant « *les interactions et les comportements des acteurs et des organisations* » (Lascoumes & Le Galès, 2005, p. 15). « *Machines à mesurer* » (Durand & Dameron, 2011, p. 561), ces classements contribuent à distribuer des

²⁵ Les chercheurs en SHS méconnaissent les indicateurs bibliométriques, leur préférant des critères comme la « bonne réputation » d'une revue ou sa présence dans une des listes de l'AERES (Pansu et al., 2013).

ressources et à véhiculer une certaine représentation des problèmes (Lascoumes & Le Galès, 2005). A cet égard, au sein de la section des STAPS, le recours aux métriques internationales constitue bel et bien une source de pouvoir, tant ces dernières sont constitutives de valeurs, d'identités et de moyens d'existence pour les acteurs de la recherche (HEFCE, 2014). De ce point de vue, la construction des référentiels d'évaluation, puis leur imposition renvoient à des stratégies de domination aux effets déjà observés dans les sciences du sport (Mierzejewski, 2005 ; Wagner, 2016).

Conclusion : recommandations pour l'évaluation de la recherche au sein des STAPS

Il convient de rester prudent quant aux conclusions de cet article, qui n'ont de valeur que rapportées à un contexte ; or ce dernier évolue rapidement car l'évaluation de la recherche est un processus particulièrement dynamique, en France, depuis la fin des années 2000 (Lallier & Topalov, 2017). Néanmoins, la convergence entre les données empiriques récoltées et les analyses disponibles dans la littérature scientifique permet d'affirmer que le virage bibliométrique observé s'inscrit à contre-courant des évolutions repérées au niveau international.

Aussi, la vocation de cet article est de susciter une prise de conscience des dérives potentielles d'un système d'évaluation avant tout focalisé sur les indicateurs bibliométriques. Il s'agit *in fine* d'éviter que le stimulant défi épistémologique que représente l'évaluation de la recherche au sein d'une section interdisciplinaire ne soit tout simplement contourné par un recours insuffisamment raisonné aux métriques internationales. Il ne saurait pour autant être question d'adopter une posture anti-évaluation et de revendiquer un isolationnisme des SHS. Une position nuancée, au sens de Pansu et al. (2013), nous semble à même de rendre l'évaluation rigoureuse, juste et efficace.

Pour les STAPS, l'horizon principal devrait être d'assumer la diversité consubstantielle de la section et la complexité inhérente des évaluations à réaliser, en se dotant d'outils d'évaluation indépendants et adaptés à l'ensemble de la section. Dans le nouveau « guide des produits de la recherche » pour les STAPS, la concession à la multidisciplinarité prend la forme d'une ouverture aux revues susceptibles d'être listées dans d'autres référentiels publiés sous l'égide du HCERES. *A priori* louable, ce principe consiste cependant à se dessaisir de responsabilités qui incombent à la section. D'une part, une dilution thématique est induite, pouvant être interprétée comme l'aveu d'une absence de réelle spécificité des STAPS : la place généralement accordée aux revues spécialisées dans le sport et l'activité physique dans les autres référentiels disciplinaires est pour le moins ténue. D'autre part, ce choix rend une grande partie de la communauté dépendante du bon vouloir de collègues relevant d'autres disciplines des sciences sociales d'établir ou non une liste de revues.

Un compromis consisterait à coupler aux métriques de type *SJR* ou *IF*, qui semblent convenir aux représentants de certaines disciplines du périmètre des STAPS, une liste restreinte de revues de SHS traitant spécifiquement du sport, de l'activité physique, du corps et de l'éducation physique, dans lesquelles des chercheurs relevant des STAPS publient effectivement et dont la qualité scientifique est reconnue. La constitution d'une telle liste nécessite une consultation élargie, voire un travail d'enquête conçu et piloté par un groupe attentif à la pluralité disciplinaire. Une perspective cohérente avec la recommandation formulée sans relâche, depuis une dizaine d'années, par les spécialistes de l'évaluation de la recherche : pour définir des critères de qualité, il convient de recourir à des procédures *bottom up* mobilisant des représentants de chaque discipline, dans leur diversité, plutôt que de chercher à uniformiser les dits critères par le biais d'un transfert d'une discipline à l'autre (Ochsner et al., 2016b).

Le récent « guide des produits de la recherche » en STAPS génère une situation d'autant moins tenable qu'en novembre 2018, un communiqué de presse²⁶ du HCERES est venu formaliser son soutien aux principes et recommandations de la *DORA* et du Manifeste de Leiden (cf. notes de bas de page 3 et 4). Un nouveau pas franchi par le Haut Comité vers la relativisation et la progressive mise à distance des indicateurs bibliométriques, au profit de la prise en compte du contenu des travaux et de la portée de leurs résultats. Des métriques comme l'*IF*, le *H index* ou le *SJR* peuvent dès lors éventuellement (et non systématiquement) venir soutenir (à condition d'intégrer leurs limites) des appréciations avant tout qualitatives, supposant une consultation attentive des produits de la recherche évalués. Cette relégation sanctionne autant les effets pervers dont sont porteurs les outils bibliométriques dominants que l'absence de transparence quant au fonctionnement de nombre d'entre eux.

Bibliographie

Académie des Sciences (2011). *Du bon usage de la bibliométrie pour l'évaluation individuelle des chercheurs* (Rapport du 21 janvier 2011 à M^{me} la Ministre de l'Enseignement supérieur et de la Recherche. Repéré à <http://www.academie-sciences.fr/pdf/rapport/avis170111.pdf>

Andrews, D. (2008). Kinesiology's inconvenient truth and the Physical Cultural Studies imperative. *Quest*, 60, 45-62.

Archambault, E., Vignola-Gagné, E., Côté, G., Larivière, V. et Gingras, Y. (2006). Benchmarking scientific output in the social sciences and humanities: the limits of existing

²⁶ <https://www.hceres.fr/ACTUALITES/Toutes-les-actualites/Hceres-Du-bon-usage-des-criteres-d-evaluation-de-la-recherche>

- databases. *Scientometrics*, 68(3), 329-342.
- Bennis, W.G. et O'Toole, J. (2005). How business schools lost their way. *Harvard Business Review*, may, 96-104.
- Bergstrom, C. and West, J. (2018). *Comparing Impact Factor and Scopus CiteScore*. Eigen Factor Project, University of Washington. Repéré à <http://eigenfactor.org/projects/posts/citescore.php>
- Bessire, D. et Fabre, P. (2014). *New public management* et évaluation de la recherche publique : l'invention du « produisant ». *Finance Contrôle Stratégie*, 17(4). Repéré à <http://journals.openedition.org/fcs/1554>
- Bornmann, L. et Werner, M. (2016). The journal Impact Factor and alternative metrics. *EMBO J*, 17(8), 1094-1097.
- Chavarro, D., Tang, P. et Rafols, I. (2017). Why researchers publish in non-mainstream journals: training, knowledge bridging and gap-filling. *Research Policy*, 46(9), 1666-1680.
- 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.
- Collinet, C. et Payré, S. (2003). Les sciences en STAPS, questions vives. Dans C. Collinet (dir.), *La recherche en STAPS*. Paris : PUF, 9-47.
- Conseil pour le Développement des Humanités et des Sciences Sociales (2010). *Pour des sciences humaines et sociales au cœur du système d'enseignement supérieur et de recherche*. Rapport remis à M^{me} Valérie Pécresse, Ministre de l'Enseignement supérieur et de la Recherche.
- Cronin, B. et Sugimoto, C.R. (dir.) (2014) *Beyond bibliometrics: harnessing multidimensional indicators of scholarly impact*. Cambridge, MA : MIT Press.
- DiMaggio, P.J. et Powell, W.W. (1983). The iron cage revisited: institutional isomorphism and

- collective rationality in organizational fields. *American Sociological Review*, 48, 147-160.
- Durand, T. et Dameron, S. (2011). Where have all the business schools gone? *British Journal of Management*, 22, 559-563.
- Engels, T.C., Ossenblok, T.L. et Spruyt, E.H. (2012). Changing publication patterns in the social sciences and humanities, 2000-2009. *Scientometrics*, 93(2), 373-390.
- 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.ideas-idees.ca/sites/default/files/impact_report_en_final.pdf
- FNEGE (2010). L'évaluation des enseignants-chercheurs en Sciences de Gestion. Rapport de la FNEGE, février. Repéré à www.fnege.net
- Gingras, Y. (2014). *Les dérives de l'évaluation de la recherche. Du bon usage de la bibliométrie*. Paris : Editions Raisons d'Agir.
- Gingras, Y. (2015). Dérives et effets pervers de l'évaluation quantitative de la recherche : sur les mauvais usages de la bibliométrie. *Recherche en soins infirmiers*, 121(2), 72-78.
- 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.
- Hicks, D. et Wang, J. (2011). Coverage and overlap of the new Social Sciences and Humanities journal lists. *Journal of the American Society for Information Science and Technology*, 62(2), 284-294.
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S. et Rafols, I. (2015). Bibliometrics: the Leiden Manifesto for research metrics. *Nature*, 520, 429-431.
- Hicks, D. (1999). The difficulty of achieving full coverage of international social science

- literature and the bibliometric consequences. *Scientometrics*, 44(2), 193-215.
- Hicks, D. (2004). The four literatures of social science. Dans H.F. Moed, W. Glänzel et U. Schmoch (dir.), *Handbook of Quantitative Science and Technology Research*. Dordrecht : Kluwert Academic, 476-496.
- Higher Education Funding Council for England (2014). *The metric tide. Independent review of the role of metrics in research assessment and management*. Rapport repéré à www.hefce.ac.uk/rsrch/metrics
- Katz, J.S. (1999). *Bibliometric indicators and the Social Sciences*. Brighton: University of Sussex.
- Krull, W. et Tepperwien, A. (2016). The Four 'I's : Quality Indicators for the Humanities. Dans M. Ochsner, S.E. Hug et H.D. Daniel, *Research assessment in the humanities : towards criteria and procedures*. Springer Open, 165-179. Repéré à <https://link.springer.com/content/pdf/10.1007%2F978-3-319-29016-4.pdf>
- Lallier, J. et Topalov, C. (2017). Qui organise l'évaluation dans les sciences humaines et sociales en France ? Une approche par les profils de carrière. *Sociologie*, 8(2). Repéré à <http://journals.openedition.org/sociologie/3111>
- Lascoumes, P. et Le Galès, P. (2005). Introduction : l'action publique saisie par les instruments. Dans P. Lascoumes et P. Le Galès, *Gouverner par les instruments*. Paris : Presses de Sciences Po, 11-44.
- Manana-Rodriguez, J. (2015). A critical review of SCImago Journal & Country Rank. *Research Evaluation*, 24, 343-354.
- Maurel, M.C. (2011). L'évaluation en sciences humaines et sociales : comment mesurer ce qui compte ? Colloque *Évaluation des productions scientifiques : des innovations en SHS ?* CNRS, Paris, 9-10 juin.
- Mierzejewski, S. (2005). *Le Corps académisé. Genèse des STAPS (1968-1982)*. Thèse de

doctorat, Université Paris X Nanterre.

Mongeon, P. et Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106, 213-228.

Nederhof, A. (2006). Bibliometric monitoring of research performance in the Social Sciences and the Humanities: a review. *Scientometrics*, 66(1), 81-100.

Ochsner, M., Hug, S.E., et Daniel, H.D. (dir.) (2016a) *Research assessment in the humanities: towards criteria and procedures*. Springer Open. Repéré à <https://link.springer.com/content/pdf/10.1007%2F978-3-319-29016-4.pdf>

Ochsner, M., Hug, S.E., et Daniel H.D. (2016b) Humanities scholars' conceptions of research quality. Dans M. Ochsner, S.E. Hug et H.D. Daniel, *Research assessment in the humanities : towards criteria and procedures*. Springer Open, 43-69. Repéré à <https://link.springer.com/content/pdf/10.1007%2F978-3-319-29016-4.pdf>

Ossenblok, T.L., Engels, T.C. et Sivertsen, G. (2012). The representation of the social sciences and humanities in the Web of Science. A comparison of publication patterns and incentive structures in Flanders and Norway. *Research Evaluation*, 21(4), 280-290.

Pansu, P., Dubois, N. et Beauvois, J.L. (2012). *Dis-moi qui te cite et je saurai ce que tu vaux. Que mesure vraiment la bibliométrie ?* Grenoble : Presses Universitaires de Grenoble.

Pendlebury, D.A. (2009). The use and misuse of journal metrics and other citation indicators. *Archivum Immunologiae et Therapiae Experimentalis*, 57(1), 1-11.

Pölonen, J., Engels, T., Guns, R., Sivertsen, G. et Verleysen, F. (2017). SSH journal publishing in Flanders and Finland. *Research Evaluation in the Social Sciences and Humanities 2017 Congress*, University of Antwerp, Belgium, 6-7 july, 44-48.

Pons-Novell, J. et Tirado-Fabregat (2010). Is there life beyond the ISI Journal lists? The international impact of Spanish, Italian, French and German economics journals. *Applied Economics*, 42(6), 689-699.

- Pontille, D. et Torny, D. (2013). La manufacture de l'évaluation scientifique. Algorithmes, jeux de données et outils bibliométriques. *Réseaux*, 177, 25-61.
- Pontille, D., et Torny, D. (2017). Infrastructures de données bibliométriques et marché de l'évaluation scientifique. Dans P.M. Menger et S. Paye (dir.), *Big data et traçabilité numérique. Les sciences sociales face à la quantification massive des individus*. En ligne : <http://books.openedition.org/cdf/5002> en ligne : <http://books.openedition.org/cdf/5002>
- Pulverer, B. (2015). DORA the brave. *EMBO J*, 34(12), 1601-1602.
- Pumain, D. et Dardel, F. (2014). L'évaluation de la recherche et de l'enseignement supérieur. Rapport à M^{me} la Ministre de l'Enseignement supérieur et de la Recherche. Repéré à https://cache.media.enseignementsup-recherche.gouv.fr/file/Actus/98/8/Rapport_Pumain_Dardel_295988.pdf
- Rossner, M., Van Epps, H. et Hill, E. (2007). Show me the data. *The Journal of Cell Biology*, 179(6), 1091-1092.
- Royal Netherlands Academy of Arts and Sciences (2013). *Towards a framework for the quality assessment of social science research*. Report of the Committee « Quality Indicators in the Social Sciences ». Repéré à <https://www.knaw.nl/en/news/publications/towards-a-framework-for-the-quality-assessment-of-social-science-research>
- 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. (2014). Scholarly publication patterns in the social sciences and humanities and their coverage in Scopus and Web of Science. Dans E. Noyons (dir.), *Proceedings of*

the science and technology indicators conference 2014 Leiden. Leiden: Centre for Science and Technology Studies, 598-604.

Sivertsen, G. (2016). Patterns of internationalization and criteria for research assessment in the social sciences and humanities. *Scientometrics*, 107(2), 357-368.

Terret, T. et Vigarello, G. (2012). Sciences humaines et sciences exactes : le déséquilibre de l'évaluation. L'exemple des Sciences et techniques des activités physiques et sportives (STAPS). *Esprit*, juillet, 26-29.

University of Waterloo Working Group on Bibliometrics (2016). White Paper on bibliometrics: measuring research outputs through bibliometrics. Report. Waterloo, Ontario: University of Waterloo. Repéré à <https://uwspace.uwaterloo.ca/bitstream/handle/10012/10323/Bibliometrics%20White%20Paper%202016%20Final%20March2016.pdf>

Van Leeuwen, T. (2006). The application of bibliometric analyses in the evaluation of social science research. Who benefits from it, and why it is still feasible. *Scientometrics*, 66(1), 133-154.

Vogel, R., Hattke, F. et Pettersen, J. (2017). Journal rankings in management and business studies: what rules do we play by? *Research Policy*, 46(10), 1707-1722.

Wacheux, F. (1999). *Méthodes qualitatives et recherche en gestion*. Paris : Economica.

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.

Wang, J., Veugelers, R. et Stephan, P. (2017). Bias against novelty in science: a cautionary tale for users of bibliometric indicators. *Research Policy*, 46(8), 1416-1436.

Williams, G. et Galleron, I. (2016). Bottom Up from the bottom: a new outlook on research evaluation for the SSH in France. Dans M. Ochsner, S.E. Hug et H.D. Daniel, *Research*

assessment in the humanities : towards criteria and procedures. Springer Open, 181-198.

Repéré à <https://link.springer.com/content/pdf/10.1007%2F978-3-319-29016-4.pdf>

Woelert, P. (2013). The economy of memory: publications, citations, and the paradox of effective research governance. *Minerva*, 51(3), 341-362.