

This item is the archived peer-reviewed author-version of:

Ambiguity in identification of peer-reviewed publications in the Finnish and Flemish performance-based research funding systems

Reference:

Pölonen Janne, Engels Tim, Guns Raf.- Ambiguity in identification of peer-reviewed publications in the Finnish and Flemish performance-based research funding systems

Science and public policy / International Science Policy Foundation - ISSN 0302-3427 - 47:1(2020), p. 1-15

Full text (Publisher's DOI): <https://doi.org/10.1093/SCIPOL/SCZ041>

To cite this reference: <https://hdl.handle.net/10067/1619960151162165141>

Ambiguity in identification of peer-reviewed publications in the Finnish and Flemish performance-based research funding systems

Janne Pölönen¹, Tim C.E. Engels², and Raf Guns²

¹Federation of Finnish Learned Societies, Snellmaninkatu 13, 00170 Helsinki, Finland

²Faculty of Social Sciences, Centre for R&D Monitoring (ECOOM), University of Antwerp, Middelheimlaan 1, 2020 Antwerp, Belgium

*Corresponding author: janne.polonen@tsv.fi

Abstract

In performance-based research funding systems (PRFSs) evidence of peer review is often considered a requirement for publications to be included. Originating from the sciences, pre-publication peer review is very common in the publishing process, also in the social sciences and humanities (SSH). Sometimes, however, it is ambiguous whether a publication is peer-reviewed or not.

In this contribution we analyse the ambiguity in identifying a journal's or publication's peer review status by comparing the classification of journals in Finland and Flanders, and by taking stock of Finnish authors' reporting of peer review status of publications. We find that ambiguity in terms of peer review status is rather common, especially in the humanities. Indeed, we find differences in peer review status in about 10 percent of all cases, both when comparing Finland and Flanders, and when comparing author-reported and centralized identification of peer review status.

1. Introduction

In 2014 researchers from two Finnish universities, the University of Tampere and the Tampere University of Technology, published a co-authored article in *Flux*, an international quarterly on networks and territories published by the University of Paris-Est Marne la Vallée. One university registered the publication as a peer-reviewed article, while the other registered it as a non-peer-reviewed article in a scholarly journal. After mutual consultation the universities' data collection personnel decided that this publication should be stored as a peer-reviewed article in the national publication information service. The journal itself, however, is not included in the Finnish national authority list of peer-reviewed publication channels. In Flanders (Belgium), unlike Finland, the expert panel responsible for classification of journals and publishers has classified *Flux* as a peer-reviewed scholarly journal.

The above case exemplifies three kinds of ambiguities related to the identification of peer-reviewed outputs in the context of performance-based research funding systems (PRFS), which we investigate in this paper based on journal lists and publication data specific to the social sciences and humanities (SSH):

- 1) To what extent do the Finnish and Flemish authority lists, both of which rely on expert evaluation to identify peer review status, disagree on the scholarly status of journals?
- 2) To what extent does the Finnish authority list of journals disagree with the researchers' self-reported status of peer review?
- 3) To what extent do co-authors affiliated to different Finnish universities disagree on the peer review status of their publications?

We are neither promoting nor contesting the validity of peer review as a criterion of research outputs but rather investigate how clear-cut the distinction between peer-reviewed and not-peer-reviewed publications is. The analysis of publication data at our disposal contributes to recognition and measuring the zones of ambiguity in identification of peer-reviewed publications but it does not provide evidence on the effects of PRFS. We use the term "grey zone" to refer to the range of outlets and outputs, the peer review status of which is ambiguous. Our results show that the PRFS criteria for peer-reviewed outputs and publication channels are not always fully consonant with the researchers' understanding of the scholarly status of their publications. Hence, PRFS criteria can be reviewed and updated to reduce the grey zone, and communication and documentation of criteria can be improved. Our findings also help to improve responsible use of metrics derived from PRFS data, if grey zones are recognized and taken into consideration when the data is used for research evaluation purposes, especially at individual level.

In the next section, we first discuss the earlier literature concerning the problem of identifying peer-reviewed publications. In section 3 and 4, we describe the background, data and methods of our study. Section 5 presents the results followed by discussion and conclusions in section 6.

2. Literature review

It is almost impossible to imagine a research evaluation or funding procedure that would not take into consideration publications, in which researchers seek to demonstrate new findings and applications of their research to other experts in the field. The practice of assessing and comparing the scientific achievements of researchers based on original research contributions was established at the French Academy of Science in the early 19th century (Crosland 1983). Since 1867, the Royal Society of London started publishing the *Catalogue of Scientific Papers* to list all scientific articles and their authors published globally since 1800. As Csiszar (2017) explains, this undertaking involved a potentially ambiguous identification of original contributions to knowledge, as well as of scholarly journals in which they are published. During the past 50 years, pre-publication peer

review has increasingly become a requirement of contributions to scientific knowledge (Baldwin 2018). Nowadays, the distinction between peer-reviewed scholarly publications and those intended for disseminating knowledge beyond academia plays a role in most expert and metrics-based evaluation and funding systems.

In recent decades, many European countries – including Finland and Flanders – have introduced PRFSs for allocating state funding to higher education institutions (Hicks, 2012; De Boer et al., 2015). In 2014, 18 EU member states had implemented some form of PRFS, in which the institutions' share of available public funding is determined, at least partly, on the basis of measurement and/or expert evaluation of scholarly publications (Jonkers & Zacharewicz, 2016; Sivertsen, 2017; Zacharewicz et al., 2018). In natural sciences and medicine, scholarly publications are typically peer-reviewed English language articles in international journals that are indexed in commercial citation databases, notably Web of Science and Scopus. Originating from the sciences, pre-publication peer review has become a generally accepted standard in the SSH as well (Sivertsen & Larsen, 2012; Mulligan et al., 2013; Tenopir et al., 2016; Rowley & Scaffi, 2017). However, a considerable share of SSH research is published in books and, because it often concerns topics of local relevance, in languages other than English (Nederhof, 1989; van Leeuwen, 2013; Sivertsen, 2018). In addition, non-scholarly or “enlightenment” literature, intended for disseminating research beyond academia, plays an important role in SSH publishing, and cannot always be unequivocally distinguished from scholarly publications (Burnhill & Tubby-Hille, 2003; Hicks, 2004).

The specificity of SSH publication patterns presents PRFSs with considerable difficulty in defining which publications should be taken into account as peer-reviewed research outputs, and how information about relevant scholarly publications can be collected (Debackere et al., 2018). During the past two decades, numerous studies have explored the representation of SSH literature in different databases, all pointing at the persisting shortcomings of Web of Science and, more recently, Scopus coverage (Nederhof et al., 1989; Hicks, 1999; Nederhof, 2006; Archambault et al., 2006; Larivière & Macaluso, 2011; Hicks & Wang, 2011; Sivertsen & Larsen, 2012; Ossenblok et al., 2012; Sivertsen, 2016a; Giménez-Toledo et al., 2017; Kulczycki et al., 2018; Aksnes & Sivertsen, 2019). In a report to the European Science Foundation, Martin et al. (2010) suggested that a comprehensive European database, “used for assessing research performance in SSH”, could be built upon institutional information sources. While this European database has not yet materialized (Puuska et al., 2018), several European countries have developed national information systems that support PRFSs with comprehensive publication data integrated from institutions (Sile et al., 2017; Sile et al., 2018).

It is not always clear which quality criteria for research output are applied in local information systems. Therefore, in several European countries PRFSs rely on authority lists of publication channels to double-check the peer review status of research outputs at the national level (Giménez-Toledo, 2017; Giménez-Toledo, 2019). Already in 2005, Norway established a PRFS with differentiated publication counts based on the type and channel of publications (Schneider, 2009; Sivertsen, 2010; 2016). Denmark, Finland and Flanders have adopted the Norwegian model with some modifications (Sivertsen, 2017; 2018b; Aagaard, 2018; Pölönen, 2018; Engels & Guns, 2018). When PRFSs take into account publications in channels that are not indexed in major international citation databases, decisions on whether these channels are peer-reviewed are typically taken at the national or regional level. This is for example the case in Denmark, Finland, Flanders and Norway, where such journals and book publishers are categorized by panels of experts in the field (Auranen & Pölönen, 2012; Sivertsen, 2016b; Verleysen, Ghesquière & Engels, 2014).

While the involvement of national research communities in the classification of publication channels is an important part of the authority list's legitimacy (Ahlgren et al., 2012), and allows an assessment of outlets without blind reliance on impact factors (Kulczycki & Rozkosz, 2017;

Pölönen & Sivertsen, 2017), expert evaluation also raises concerns about personal bias and validity (Bornmann, 2011; Haddaway et al., 2016). Analysis of the correlation between the expert-based and citation-based evaluations of journals is a well-established research track (Serenko & Dohan, 2011; Ahlgren & Waltman, 2014; Haddaway, 2016; Saarela et al., 2016; Walters, 2017), but such studies are mostly focused on the rankings and ratings of the leading journals rather than the basic delineation of peer-reviewed scholarly journals.

Some studies, however, suggest that even experts in the field may disagree whether a journal is peer-reviewed or not. In a study by Nederhof (1991), SSH departments differed in their registering of articles published in the same journals as scholarly or non-scholarly publications in the Annual reports of Dutch universities from 1980-1985. Surveying of foreign and Dutch experts concerning the scholarly status of those journals confirmed that a considerable share of them – between 11 % (experimental psychology) and 40 % (Dutch literature) – were deemed non-scholarly by the majority of respondents. Another study used EBSCO and Ulrich's serial directories as external information sources to check the self-reported peer review status of journals by Economic and Social Research Council (ESRC) award holders in UK 1984-1988 (Burnhill & Tubby-Hille, 2003). This study reported a large range of ambiguity: “only 42% of the 772 titles reported as having undergone ‘scrutiny from referees’ were also classified by the directories as ‘known to be refereed or peer reviewed’”.

The difficulties related to the identification of scholarly journals have also been observed in the context of creating national authority lists. According to Bruun-Jensen (2012), Danish researchers were opposed to the idea that the publication indicator developed for the PRFS would be based on the Norwegian authority list, because “it had not been validated with enough care: non-peer reviewed journals had snuck in, it was said, and one could find what an interviewee described as ‘the research equivalent of Donald Duck’”. There are also increasing concerns about the standards of quality control in journals that claim to be peer-reviewed, so the identification of scholarly journals also involves screening of the national authority lists for so-called predatory journals (Eykens et al. 2018).

Only few studies have systematically compared the national authority lists produced by experts in different countries. Verleysen, Ghesquière & Engels (2014) compared the Flemish, Danish, Finnish and Norwegian lists of book publishers that are taken into account for the national PRFS and discovered “both considerable agreement as well as divergence regarding the inclusion of publishers”. A comparison of the expert-based book publisher ratings from Finland with survey-based ratings from Spain showed strong discrepancies in the peer review status of nationally-oriented publishers (Mañana-Rodríguez & Pölönen, 2018). Verleysen and Engels (2015) compared the Flemish journal list to the ERIH Plus journal list, and found that 3.5% of ERIH Plus journals had been classified as non-peer-reviewed in Flanders. Pölönen et al. (2011), Saarela et al. (2016) and Pölönen & Sivertsen (2017) have compared the ratings of leading journals in Finland, Denmark and Norway but there are no studies comparing the national authority lists of journals regarding the identification of peer review status.

The notion of peer review remains challenging because of the variety of practices across fields, and because of the differences in journal, conference and book publishing (British Academy, 2007; Derricourt, 2012; Verleysen & Engels, 2013). Typically, both PRFSs and researchers define peer-reviewed publications as those that have undergone pre-publication peer review by experts in the field. Differences may concern the number of referees (one or more), their degree of anonymity vis-à-vis the authors (double-blind, single-blind or open identity), and their relation to the publication channel (editors, editorial board, reading committee, or external). Ambiguity may also result from the fact that many peer-reviewed journals and edited volumes also include items that are not peer-reviewed – editorials, opinions, comments, discussions, book reviews, and abstracts are typical

examples. Moreover, book publishers of peer-reviewed monographs and edited volumes often also publish textbooks, *libri amicorum*, and other types of books that typically do not undergo a formal peer review process.

In their study on classification of publications at Dutch law faculties, Kaltenbrunner & de Rijcke (2016) have demonstrated the importance of disciplinary differences, as in the specialties with stronger national and practical orientation (such as private or administrative law), the distinction between scholarly and professional publications appeared more controversial than in more internationally oriented fields (such as law & economics or comparative law). This is consonant with Nederhof's (1991) findings that there seems to be a stronger consensus about the scholarly status of journals in more internationally oriented SSH fields. Mañana-Rodríguez and Giménez-Toledo (2013) showed that Spanish SSH journals can be positioned on a continuum of being more social sciences-like to more humanities-like, meaning for example less frequent use of double-blind review and a larger share of outputs consisting of non-original research articles. In their study of book publication patterns, Verleysen and Engels (2014) showed that in fields with stronger orientation toward the local context, especially in the humanities, the same book publishers were used for publications aimed at academic and wider audiences.

Problems with the application of the peer review criterion are documented in various PRFS contexts. In a study of the Swedish universities' internal funding systems, Hammarfelt et al. (2016) observe that "the classification of publications as 'scholarly' [in the local database] is not always straightforward, and the definition of peer review differs across departments and authors". Dahler-Larsen (2017b, 2014) has highlighted this ambiguity with an example from his own experience concerning the correct classification, in the category of research or dissemination, of a Danish language book chapter in the university's information system that feeds the Danish PRFS. Kaltenbrunner & de Rijcke (2016) have studied the ambiguity related to the scholarly/professional distinction in the context of the Dutch national research evaluation system (SEP). As one informant, who was in charge of validating the list of scholarly publications, puts it: "there is a great mess, no one knows what is peer-reviewed and what is not... For me it's peer-reviewed when there is really a review... We have no list of peer-reviewed journals... If my researcher thinks it's a lot of work and it's scientific, I believe it".

Ambiguity can be understood, following Dahler-Larsen (2017), as "the coexistence of multiple interpretations of a phenomenon among reasonable people while there is not necessarily an easy way to choose between the interpretations or eliminate some of them". The peer review status of publications can be understood both formally – "there is really a review" – and substantially – "it's a lot of work and scientific" (Kaltenbrunner & de Rijcke, 2016). PRFSs typically employ a formal definition based on the peer review process, the existence of which can be verified externally, thereby focusing on a subset of scholarly literature that experts in the field have deemed of sufficiently high quality and importance to be published (e.g. Martin, 2010, Sivertsen & Larsen, 2012; Debackere et al., 2018). This approximation of scholarly publications does not include all outputs that researchers may consider as valid contributions to researched knowledge. In our view, it is important to recognize this ambiguity in the classification of publications in the information systems, so that they can be sufficiently inclusive, flexible and structured to support both evaluative and communicative purposes. This is particularly important in case of information systems that have been originally designed according to PRFS requirements but are later extended and developed to serve broader communicative purposes.

There is a large and growing body of literature on the possible effects of PRFSs on research performance (e.g. Auranen & Nieminen, 2010; Butler 2013; Aagaard et al., 2015; Schneider et al., 2015; Aagaard & Schneider, 2017; Himanen et al., 2009; Sandström & van den Besselaar, 2018), as well as on social and intellectual organization of research and epistemic cultures (e.g. Gläser &

Lauder, 2007; Dahler-Larsen, 2012, 2014, 2019, de Rijcke et al., 2016). It is possible to anticipate, as Kaltenbrunner & de Rijcke (2016) do, that “practices of quantifying research output may not just reproduce existing understandings of a ‘proper scholarly publication’ in a given context, but may also modify the meaning of that very notion over time”. Our study does not provide evidence on the effects of PRFSs but we recognise from this literature that producers of classifications and publication information can also engage in strategic behaviour toward the PRFS. This may involve gaming the numbers to maximise expected tangible or intangible awards (Espeland & Sauder, 2007). Sivula et al. (2015) have suggested, based on a survey to Finnish SSH researchers, “publication laundering” as one possible reaction to PRFS, meaning the manipulation of publication lists to meet the peer-review standard of scholarly publications.

In all, earlier literature provides a comprehensive framework for understanding the sources of ambiguity and challenges related to identification of peer-reviewed scholarly publications, which arguably is the most important delineating factor that PRFSs use in measuring and assessing research performance. Several studies have indicated instances where there is considerable uncertainty concerning the peer-review status of journals and book publishers, or individual articles and books. Our study contributes to this literature by measuring the range of ambiguities in identification of peer-reviewed outlets and outputs across SSH fields in the PRFS context, taking into account publication languages and types. The measures help understand how large is the share of journals and outputs, the peer-review status of which is or is not ambiguous. The investigation and recognition of these grey zones of ambiguity is important because several PRFSs, both at national and institutional level, rely on authority lists of publication channels to identify peer-reviewed publications.

3. Background on VIRT A and VABB-SHW

Our analysis is based on authority lists from Flanders and Finland, as well as publication data from Finland. Other countries that have adapted their PRFS publication indicator according to the Norwegian model – notably Denmark and Norway – have also developed authority lists. In Norway, however, the personnel in charge of developing and maintaining the authority list validate most peer-reviewed journals (level 1) instead of the expert panels, who are mainly involved in the identification of leading outlets (level 2). Expert panels validate peer-reviewed journals in Denmark but the authority list does not contain information on disapproved journals (level 0). Therefore, only the Finnish and Flemish authority lists permit us to investigate the discrepancies in expert assessment of peer review status of SSH journals. As will be explained below, the Finnish publication data supporting the PRFS include also self-reported peer-review status, which allows us – unlike the publication data in Denmark, Flanders or Norway – to investigate the peer review grey zones between authority list and self-reports, as well as between self-reports by co-authors. Another reason for restricting the analysis to Finland and Flanders is that the authors are well acquainted with the Finnish and Flemish systems and data.

The Finnish publication data in the VIRT A Publication Information Service and the Publication Forum rating of publication channels are described in Giménez-Toledo et al. (2016), Kulczycki et al. (2018), Puuska et al. (2018), Sile et al. (2018) and Pölönen (2018). Finnish universities report their publications to VIRT A, and this data is used to allocate part (13 %) of the block-grant annually to universities. The allocation is based on a funding formula, in which publications are weighted according to publication type and Publication Forum level (see below) of the publication channel. The Flemish system is the VABB-SHW, which is a parameter in the regional funding distribution mechanism, as described in Verleysen et al. (2014), Verleysen and Rousseau (2017) and Engels and Guns (2018). In both systems, field-specific panels of experts are responsible for classifying publication channels as peer-reviewed or not peer-reviewed, which we refer to as top-down identification of peer review. In the Finnish system, however, scholars also indicate at the time of

registration of a publication in their institutional CRIS whether they consider it to have been peer-reviewed prior to publication or not. We refer to this as bottom-up identification of peer review. We use the term ‘the grey zone of peer review’ to refer to publications of which the peer review status is ambiguous.

In Flanders, the authority lists of publication channels that are not included in Web of Science are used to control whether the articles and books submitted by universities to VABB-SHW satisfy the peer review criteria established for publications to be counted in the PRFS. Specifically, the channels approved by the authoritative panel (GP) need to:

- be publicly accessible;
- have an ISBN or ISSN code;
- contribute to the development of new knowledge or the application thereof; and
- apply a demonstrable peer review process by scholars who are experts in the (sub)field to which the publication belongs. It is further specified that peer review should be done by an editorial board, a permanent reading committee, external referees or by a combination of these.

In practice, the authoritative panel may also use discretion to exclude, for example, predatory journals and the publications therein (cf. Eykens, Guns & Engels 2018). The Flemish authority list distinguishes only between peer-reviewed and non-peer-reviewed channels; there is no further differentiation of peer-reviewed outlets according to prestige or impact.

In Finland, the authority list of publication channels (Publication Forum) is used to determine the weight of publications, which universities have reported to VIRTAs as being peer-reviewed, in the PRFS. The Publication Forum rating differentiates approved peer-reviewed channels to three levels according to prestige and impact: 1 = Basic, 2 = Leading and 3 = Top. Channels not meeting the level 1 criteria are assigned to level 0, so this category corresponds to the non-peer-reviewed channels in the Flemish authority lists. Range of information about journals is provided to support expert-evaluation: impact factors (JIF, SNIP and SJR) as well as ratings in Norway and Denmark.

Publication Forum level 1 can be awarded to publication channels meeting the following criteria:

- specialised in the publication of scientific or scholarly research outcomes;
- editorial board constituted by experts;
- entire manuscripts of scholarly articles or books subject to peer review;
- registered ISSN or ISBN number.

Publication channels meeting these criteria should not be included in Level 1 if:

- over half (1/2) of the referees and authors represent a single research organisation (such as publication series or doctoral dissertation series of universities and research institutes);
- the relevance or quality of research raises questions (e.g. predatory journals).

This means that the level 1/level 0 distinction in the Finnish authority list of publication channels is not based on peer review status only. The effect of the level 0 rating of the channel is that peer-reviewed outputs published in it are counted in the PRFS with smaller weight (0.1) than those published in level 1 (1), 2 (3) and 3 (4) channels. The weight of peer-reviewed outputs in level 0 outlets is the same as for the outputs the universities have reported as not being peer-reviewed.

Finnish universities rely principally on self-reporting by the researchers for the identification of peer-reviewed articles and books. Personnel dedicated to data collection may also participate in or control the identification. Self-reporting of publications and identification of peer-reviewed outputs is in many universities also facilitated by importing metadata from international databases, such as

Web of Science or Scopus. The publication data collection guide produced by the Ministry of Education and Culture sets the criteria that articles and books have to meet to be reported as peer-reviewed publications (Ministry of Education and Culture 2018):

- the referees were independent in terms of the manuscript to be evaluated;
- the process assesses the completeness of the material and the management of the theoretical framework, the reliability of how the research has been carried out and its accuracy, the originality of the results and their novelty in relation to previous research, as characteristic of the field of science;
- the evaluation has covered the entire manuscript offered for publication rather than just an abstract or extract;
- the author has received a written referee statement of the peer review (original statement or a summary by the editorial staff/editor in chief).

Publications are also required to have an ISSN and/or ISBN. Some of these qualifications have not been in place for all peer-reviewed publication types since 2011 but have been introduced and extended gradually.

4. Data and methods

We analyse the occurrence of ambiguity in the identification of peer review in the SSH in three sets of data.

1. For the comparison of journals identified as peer-reviewed by the expert panels in Flanders and Finland we merged the 2016 versions of journal lists from both countries on the basis of ISSN. 7638 matching records were identified, for which the OECD FOS fields (OECD, 2007) were assigned on the basis of available field classifications from CRISTIN (Norway), ERIH Plus, Web of Science, and Scopus. 4505 journals assigned to any one of the SSH fields were included in this analysis. We do not include book publishers as publication channels, as earlier comparisons of the Finnish and Flemish book publisher selection have been conducted and published (Verleysen, Ghesquière & Engels, 2014).

2. For the comparison of bottom-up and top-down identification of peer-reviewed publications, we analysed all SSH publications of fourteen Finnish universities in VIRTAs from 2011 to 2015. The data include articles in journals and books, as well as monographs, while edited volumes and articles in conference proceedings were omitted. Duplicates reported by several universities to VIRTAs were identified and merged. OECD FOS fields for the SSH publications were identified on the basis of the existing field classifications in VIRTAs. This is a modified OECD FOS classification based on the content of the publication, assigned by the researchers at the time of registering the publication. In case two or more universities have assigned a publication to different fields, the publication is counted once for each different field.

3. For the comparison of identification of peer-reviewed co-publications by different universities we used Finnish 2011-2015 publications, both peer-reviewed and not peer-reviewed, counting co-publications with authors from more than one Finnish university. Each university reports such co-publications to VIRTAs, and determines their peer review status independently. In VIRTAs each publication can have only one publication type, so any discrepancies are sorted out in dialogue with the universities and are not visible in the regular VIRTAs data. In this study, however, we have had access to the original publication types reported by universities, so it has been possible to identify among university co-publications those that involved discrepancy concerning the peer review status. OECD FOS fields for the publications were identified on the basis of field classification in VIRTAs.

5. Results

5.1. Classification of journals/series as applying peer review or not applying peer review in Finland and Flanders.

A total of 4505 SSH journals/series have been classified in both Finland and Flanders. Of these journals 90.5 % have been categorized to be peer-reviewed or non-peer-reviewed channels in both countries, while 9.5 % of the journals have been evaluated differently (Table 1). The share of journals belonging to this grey zone of peer review (identified top-down) is larger in the humanities (12.6 %) than in the social sciences (7.6 %). In the social sciences, the grey zone is above the average (9.5 %) in Law, Political science, and Other social sciences. In the humanities, the share of publication channels classified differently in Finland and Flanders is below the average only in Philosophy, ethics and religion.

[Table 1 around here (tables and figures are after references)]

A majority (n=286) of the 427 journals for which we observe a discrepancy between Finland and Flanders are classified as peer-reviewed at level 1 in Finland, but have not been recognized as peer-reviewed in Flanders. Of these journals, practically all (282) are not Web of Science included journals. Furthermore, only 15 are published in Belgium and three in Finland, and only three of these journals publish exclusively in the local languages (Dutch or Finnish). Of the 141 journals approved in Flanders but assigned level 0 in Finland, 33 % are Web of Science journals (WoS journals are automatically treated as peer-reviewed in Flanders). Only 4 are published in Belgium or use the Dutch language, and none in Finland. In all, the vast majority of discrepancies results from ambiguity concerning journals published outside Flanders and Finland and in other than their national languages.

When we look specifically at publication languages, we observe that classifications in Flanders and Finland concur more often in case of English than other language journals (Table 2). The English language journals amount to 73 % of all journals but their share of journals with discrepancies is only 41 %. The opposite is true for journals published in other or multiple languages: they are over-represented among journals with discrepancies compared to their share of all journals. There are five journals publishing in German or multiple languages that have been rated as level 2 in Finland but are not considered peer-reviewed in Flanders. This may point at differences between the Flemish and Finnish expert panels in application of the peer review criteria.

[Table 2 around here]

5.2. Contrasting bottom-up reporting of peer review by authors versus top-down identification of peer review by panels responsible for producing authority lists of peer-reviewed publication channels (Finland only)

A total of 32,427 SSH publications published in 2011-2015 have been reported by Finnish universities to VIRTAs as peer-reviewed (Table 3). Of these publications 84 % have been published in channels considered to be peer-reviewed by the expert panels responsible for the Publication Forum authority list of journals and book publishers, and 16 % were published in channels considered as not peer-reviewed. The share of publications belonging to this grey zone of peer review is slightly smaller in the social sciences (15 %) than in the humanities (16 %). The smallest grey zone is attested in the fields of Psychology, followed by Media and communications, Economics and business, and Philosophy, ethics and religion (Figure 2). The largest grey zone is attested in Arts, followed by Other humanities, Educational sciences, Law and Political Science.

[Table 3 around here]

We also studied whether there is a difference in the peer review grey zone between book and journal publications, and between national language and other language publications. The 2011-

2015 publication data contains 14,012 book publications (chapters and monographs) and 9273 Finnish language publications (Table 4). It is observed that the size of the grey zone, where the top-down authority list of publication channels disagrees with bottom-up identification of peer-reviewed publications by authors, is substantially larger than the average (16 %) for both book publications (25 %) and for Finnish language publications (22 %). This holds true for most SSH fields (Figure 1).

[Table 4 around here]

[Figure 1 around here]

5.3. Co-publications that one university has reported as peer-reviewed and other university as a non-peer-reviewed article or book (Finland only)

A total of 3596 SSH co-publications (peer-reviewed and not peer-reviewed) involving authors from different Finnish universities were published in 2011-2015, and so have been reported as duplicates to VIRT.A. Of these publications, 92 % have been classified the same way by all participating universities, while 8 % of the duplicate records relate to publications with conflicting assessment regarding peer review status by at least one of the participating universities (Table 5). The share of university co-publications belonging to this grey zone of bottom-up peer review is larger in the humanities (10 %) than in the social sciences (8 %). The grey zone is largest in Philosophy, ethics and religion, Law, Political Science, Social and economic geography and Arts. It is the smallest in Psychology, followed by Economics and business and Educational sciences.

[Table 5 around here]

We also studied if there is a difference in the peer review grey zone between book and journal publications, and between national language and other language publications. The 2011-2015 publication data contains 1178 book publications (chapters and monographs) and 1379 Finnish language publications with authors from more than one Finnish university (Table 6). It is observed that the size of the grey zone, where two or more universities have reported the peer review status of the same publication differently, is substantially larger for both book publications (16 %) and for Finnish language publications (16 %) than the average for all SSH publications (8 %). This holds true for most SSH fields (Figure 2).

[Figure 2 around here]

5.4. Comparison of the three analyses

The three different analyses of ambiguity in the peer review status of publications do not show a uniform pattern across SSH disciplines (Figure 3). However, in all analyses Psychology and Economics show a small grey zone, while Arts, Other humanities, and Law are frequently located at the other end of the spectrum.

[Figure 3 around here]

The overlap of the publication channels occurring in two of the analyses to identify grey zones in peer review is considerable. 72 publication channels occur in both the Finland-Flanders comparison and the top-down versus bottom-up comparison. With the exception of two, all have been classified as not applying peer review in Finland; none are published in Finland while three are published in Belgium. The overlap in terms of channels that occur both in the Finland-Flanders comparison and the Finnish co-authored publications is limited to two journals. The overlap among the sets of channels identified through self-reporting in Finland amounts to 42 journals, the large majority of which (35) are Finnish journals.

6. Discussion and conclusions

Earlier studies have indicated instances where there is considerable uncertainty in the SSH fields concerning the peer review status of journals and publications (Nederhof, 1991; Burnhill & Tubby-Hille, 2003; Dahler-Larsen, 2014; Kaltenbrunner & de Rijcke 2016). While measuring and assessing research performance typically is focused on scholarly publications, ambiguity related to the identification of peer-reviewed publications and journals has not been investigated and measured in the PRFS context at the national level and across all SSH fields. Our study tries to fill this gap.

In this study we measure the range of ambiguity related to identifying peer-reviewed journals and outputs in a PRFS context on the basis of the national authority lists of publication channels from Flanders and Finland (2016 versions), as well as national publication data of Finnish universities for the period 2011-2015. We use the term “grey zone” to refer to the range of outlets and outputs, the peer review status of which is ambiguous. According to our study, 9.5 % of the SSH journals have been evaluated differently in Flanders and Finland (analysis 1), 16 % of the Finnish universities’ peer-reviewed SSH output were published in not approved channels (analysis 2), and the peer review status of 8 % of the Finnish university co-publications in the SSH was reported differently by the universities involved (analysis 3). As we expected, the grey zone is larger in the humanities than the social sciences. Moreover, in all SSH fields the ambiguity with regard to the peer review status of publication channels is more common among book publications than journal articles and in national language publications compared to other languages. In summary, the distinction between peer-reviewed and non-peer-reviewed publications in the context of a PRFS is not entirely clear cut. These ambiguities have various reasons, some of which are also important limitations for our analysis.

The criteria for the approved publication channels, including pre-publication peer review by experts in the field, are quite similar in Flanders and Finland (analysis 1). It is important to remember, however, that some of the differences between journal lists in Flanders and Finland result from different handling of Web of Science journals, which are automatically approved in Flanders but not in Finland. The Flemish and Finnish panels are also allowed to assign journals that are branded by their editors/publishers as applying peer review into a category of non-approved journals for a variety of more or less discretionary reasons. These include, for example, serious doubts about the trustworthiness of such a declaration and/or the quality of journals (e.g. predatory journals), journals that are considered local (used only by researchers of one institution), or journals that are considered marginal from the perspective of the country/region. It is also possible that expert panels have applied the peer review criteria with different degrees of formality and strictness, or that peer review standards are adapted to disciplinary context. The latter practice is akin to cognitive contextualization in expert evaluation (Lamont 2010), and could explain why some journals have been considered among the leading outlets in Finland but are not considered to be peer-reviewed in Flanders. National authority lists may also involve strategic behaviour where panels take into consideration the potential funding effects the approval of outlets may have for their field or institution in the PRFS.

One pertinent reason for the differences in the national authority lists may also be the absence of evidence of the peer review process of the publication channel. In some cases information about the peer review may simply have been temporarily unavailable on the website at the time of the assessment. As such, channels that are classified as not applying peer review should be considered merely channels for which the identification of peer review has not been successful thus far. The authority lists are dynamic in a sense that in both Flanders and Finland it is possible to request re-evaluation of an outlet for approval. The comparison of the Flemish and Finnish authority lists suggests that in case of SSH journals, ambiguity is present especially in case of journals publishing

in languages other than English or the expert panels' national languages (e.g. journals publishing in French, German, Spanish, and Italian). Thus, our analysis also points at the difficulty of identifying peer-reviewed journals across linguistic and cultural boundaries. The same has been observed also in the case of book publishers (Mañana-Rodríguez & Pölönen, 2018). It may also be that, especially in the social sciences, journals publishing in other languages than English are regarded as less relevant outlets for research publications.

PRFSs typically rely on researchers' and universities' self-reporting of peer-reviewed publications, and/or authority lists of peer-reviewed publication channels. The Finnish PRFS uses both methods, so it is possible also to measure the difference between peer-reviewed output as identified by researchers/universities (bottom-up) and by the authority list of publication channels (top-down). Our results show that identification of peer review status based on an authority list does not always accord with self-reports (analysis 2). Our study also shows that even co-authors of the same article or monograph may differ in their assessment of whether it is peer-reviewed or not (analysis 3). The data enabled us to detect and measure author level ambiguity in the case of Finnish university co-publications (analysis 3), which represent a relatively small share of SSH output.

The peer review criteria used in Finland for the approved publication channels (Publication Forum level 1-3) and for the reporting of peer-reviewed articles and monographs to the data collection are basically in agreement: both require pre-publication peer review by independent experts in the field. Some of the discrepancies between top-down and bottom-up definition of peer-reviewed publications in Finland is explained by the fact that level 0 also includes some peer-reviewed channels if considered local, not relevant, or questionable. In this sense, the authority list actually functions in the Finnish PRFS not to exclude outputs but to differentiate between more and less valued outputs that the universities have reported and submitted to the funding model. In principle, universities are responsible for complying with the formal criteria of data collection but the extent to which the researchers' self-reporting is checked and controlled by the data collection personnel varies between universities. Universities may also assist data collection with metadata imports, however self-reporting plays an important role especially in case of national language journal outputs and book publications not indexed in Web of Science or Scopus. It is possible that in some universities the Publication Forum level of the journal or book publisher is considered already in the identification of peer-reviewed articles or monographs (clearly a potentially limiting factor to our analyses 2 and 3).

It is important to remember that the data collection definition of peer-reviewed publication is formal, emphasizing the existence of a recognizable pre-publication process. Some researchers may not know the formal criteria, and consider any output in a peer-reviewed channel as peer-reviewed. Researchers may also think that "it's a lot of work and scientific", so their paper or book merits to count as peer-reviewed (Kaltenbrunner & de Rijcke, 2016), whether or not the procedure behind the publication technically counts as peer-reviewed. Many SSH scholars, for example, discuss their manuscripts during local or external seminars which tend to improve their manuscripts considerably before they are submitted. Such feedback by peers typically does not comply with definitions of pre-publication peer review processes, yet may well convince some authors that peer review has taken place. In the case of books and local journals, punctual editorial feedback might also be considered as peer review by some authors. Researchers may also have considerable interest in the classification of their publications as peer-reviewed in the local CRIS, because institutions may use this data in internal research assessments and monitoring, as well as evaluations for funding allocation, recruitment and bonuses (Hammarfelt et al. 2016; Wahlfors & Pölönen, 2018). There is certainly room for human error in applying the peer review criterium, but also gaming the system(s) by presenting not-refereed publications as peer-reviewed is possible (Sivula et al. 2015).

Overall, ambiguity of peer review status is more common in the humanities than in the social

sciences. This is an expected finding in light of earlier studies (Nederhof, 1991; Mañana-Rodríguez and Giménez-Toledo 2013; Verleysen and Engels 2014; Kaltenbrunner & de Rijcke, 2016), as in the humanities and law the same channels are often used for communicating research within academia as well as to professional and general audiences. In most social sciences disciplines the differentiation of academic and other publication channels is more pronounced, hence leading to less ambiguity in peer review status of publications. Where the ambiguity is present, it mostly relates to book publications (chapters and monographs) and publications in the local language, pointing to the difficulty of a balanced multilingualism in science (Sivertsen 2018a). The relatively high incidence of ambiguity among national language book publications suggests that researchers used to peer review practiced in journals may have difficulties in recognizing peer review in books, or that some researchers regard monographs and chapters as valued research contributions even though they do not meet the formal criteria for peer-reviewed publications. We speculate that discrepancies might occur especially in publications with co-authors from different fields and publication cultures.

It is interesting to note, given the relatively large grey zone attested in national language book publications, that both Flanders and Finland have introduced labels for peer-reviewed publications to promote peer review standards and transparency among national publishers (Kulczycki et al., 2019). The Guaranteed Peer Reviewed Content or GPRC label (Verleysen & Engels, 2013) created by the Flemish Publishers Association is a label for individual books that are published by Flemish publishers and have undergone a peer review process prior to publication. The GPRC label is required for chapters, monographs and edited volumes by Flemish publishers to be taken into account in the PRFS. In Finland, the Federation of Finnish Learned Societies (TSV) grants a right to use a label for peer-reviewed scholarly publications (the development of which was inspired by GPRC) to Finnish publishers of academic/scholarly books, book series and journals that adhere to a series of requirements, both concerning the peer review process itself and documentation related to the review process (Kulczycki et al. 2019). The label can be applied to monographs, as well as individual chapters and articles in books or journals. The label is not a prerequisite for inclusion of publications in the PRFS but it supports the data collection process by helping researchers and other involved personnel to identify peer-reviewed monographs and articles reported to the local CRISes. One interesting topic for further research is the effect of peer review labels on peer review practices in outlets as well as the peer review definitions of outputs.

Our study underscores the fact, pointed out in the earlier literature, that the classification as peer-reviewed or non-peer-reviewed of outlets and outputs is potentially messy and contested in the research community (Dahler-Larsen, 2014; Kaltenbrunner & de Rijcke 2016), and even among co-authors of research publications. This ambiguity results in peer review grey zones, also in the PRFS context. PRFSs relying on quantitative monitoring and comparison of the universities' publication activity on the basis of national publication data need to make the distinction between peer-reviewed and non-peer-reviewed outputs in some relatively formal and transparent manner. Such PRFSs typically rely on authority lists of journals and book publishers, because the peer-reviewed scholarly literature can be identified, even in SSH fields, on the basis of channels with competent peer review procedures (Sivertsen & Larsen 2012). This is in our view a reasonable approach; however, it is important to acknowledge that identification of peer-reviewed channels is not always clear-cut. Authority lists also involve a degree of inaccuracy: some peer-reviewed or otherwise valued outputs end up being undervalued in the PRFS because the channel is not approved in the national authority list. Likewise, non-peer-reviewed outputs in approved channels are overvalued in the PRFS, unless a mechanism is in place for identifying these publications. In Flanders and Finland, the PRFSs rely on the 4-page rule and on researchers' self-reporting, respectively. Overall, PRFSs using national publication data intend to cover comprehensively the peer-reviewed research outputs typical of SSH fields, notably book and national language publications, which are not

covered in the international databases (Web of Science or Scopus). Yet, our study demonstrates that most ambiguity about what counts as valued peer-reviewed research output in the PRFS also relates to these kinds of publications.

Our study concerns the PRFSs in Finland and Flanders, however similar ambiguities regarding the peer review status of publications may also be present in other European PRFSs that measure and assess research performance on the basis of scholarly publications. Problems related to identification of peer-reviewed publications are also present in other evaluation contexts, such as project funding, recruitment and career advancement, as well as institutional funding-schemes and bonus systems (Hammarfelt et al. 2016). Important issues related to merits and incentives, such as the relative value of peer-reviewed publications versus publications aimed at professional and general audiences, or the value of book and national language publications versus English language journal outputs, are present and relevant in all evaluation contexts. Nevertheless, in most evaluation contexts, and especially those relying on expert-evaluation, the relative assessment of a specific output remains opaque and any disagreements that might exist between reasonable people remain undisclosed. The PRFS, on the contrary, decides on the basis of formalistic and transparent rules and definitions the value of each output in the national level funding scheme (Dahler-Larsen, 2014; Marini 2018). Therefore, it also brings underlying debates and assumptions about quality and impact of research and different kinds of outputs to the public sphere. Public discussion may not resolve all disagreements but it can help to develop the PRFS and influence research evaluation practices more generally. In this sense, publication indicators used in the PRFS, but also other research evaluation systems, can have important constitutive effects by establishing and shaping of meanings and practices (Dahler-Larsen 2014).

In countries that have established a PRFS, universities rely for some part of their research funding on their performance as measured by the funding model indicators, including publications. This creates a strong incentive for the universities to employ the PRFS publication indicator(s) in their internal funding, evaluation and management systems (Sivertsen & Schneider, 2012; Aagaard et al., 2014; Aagaard, 2015; Wahlfors & Pölönen, 2018; Woelert & McKenzie, 2018; Krog Lind, 2019). The problem is that a PRFS indicator, which is used for funding-allocation at macro level, can tolerate some degree of ambiguity in the identification of peer-reviewed outputs, as long as the indicator is not excessively biased against particular universities. At the individual level, however, it is of great consequence if a peer-reviewed or otherwise valued research output is not recognized because the channel is assigned to level 0 or is not approved. It is important that the evaluators and researchers being evaluated are aware of the limitations, including ambiguities, related to PRFS indicators. Therefore, we very strongly advise against the mechanistic use of the PRFS publication indicator, or authority lists, for individual level assessments. This is fully consonant with the DORA declaration (2012) and the Leiden Manifesto (Hicks et al., 2015), which recommend that indicators should inform but not supplant expert opinion based on publication contents.

It has been pointed out in the literature that information systems play a key role in mediating between generalized standards, such as PRFS definition of peer-reviewed output, and local practices (Kaltenbrunner & de Rijcke, 2016). In order to serve both communicative and evaluative purposes, information systems need to recognize the range of ambiguity in definition of scholarly publications and be sufficiently inclusive and flexible in its coverage of outputs. Large disagreement between the formal criteria and the researchers' understanding of the scholarly status of their publications may reduce the usefulness of the information system as communication infrastructure. It may also raise doubts about the validity of the data and undermine legitimacy of the measurements and assessments based on the data (Krog Lind, 2019). Therefore, it is important that the formal PRFS criteria for peer-reviewed research output are at regular intervals reviewed and adjusted, if necessary, to reduce discrepancies. Legitimacy can also be compromised if it is believed that non-scholarly publications are frequently classified as peer-reviewed outputs for the purpose of gaming

the numbers (Sivula et al., 2015). In most European countries that have introduced a PRFS based on institutional data, a public web-based interface is provided for browsing the national level publication information. This provides the universities, as well as individual researchers, with the option of controlling and verifying the data. Transparency is indeed seen as an important part of the legitimacy of the PRFS.

The ambiguity regarding what counts or does not count as a peer-reviewed research publication concerns individual researchers as well. The most important document by which researchers present their research output to various evaluations is the list of publications, in which they may be required to indicate all or some number of most important peer-reviewed publications. Our results imply that the task of identifying peer-reviewed publications can be more difficult than is often expected, and could be a research integrity concern. In Finland, the Guidelines for Responsible conduct of research established by the Finnish Advisory Board of Research Integrity indicate “exaggerating one’s own scientific and scholarly achievements, for example, in a CV or its translation, in a list of publications, or on one’s homepage” as a potential violation of responsible conduct of research (TENK 2012). “Misrepresenting research achievements” is also one of the unacceptable practices indicated in The European Code of Conduct for Research Integrity (ALLEA 2017).

Dahler-Larsen (2017b) has drawn attention to what he calls “a massive wave of conflicting rules and guidelines” that create “a disturbing ambiguity in academia”. We suggest that ambiguity is likely to increase also due to diversification of qualities that are being sought in publications. In addition to traditional quality standards (solidity, originality, and scientific relevance), societal impact and open availability are increasingly important considerations. In the near future, the peer review status of a publication concerning research funded by certain European funders will possibly determine if it is subject to Plan S open access requirements and non-compliance sanctions (Else, 2018). It will be interesting to see if the new open access policies influence how researchers define the peer review status their outputs. The Plan S implementation plan possibly relies on the Directory of Open Access Journals (DOAJ) as a trusted source of peer-reviewed open access journals. Thus, a new grey zone may be emerging, as not all DOAJ indexed journals have been approved as peer-reviewed outlets in the national authority lists.

To summarize, PRFSs typically treat peer review as a baseline criterion for scholarly publications and rely on authority lists and/or self-reporting to determine the peer review status of publications. Our analysis shows that grey zones of ambiguity exist in the identification of peer-reviewed publication channels as well as in self-reporting of peer-reviewed outputs. PRFSs relying on national publication data take into account large numbers of publications – the entire output of concerned institutions – so the distribution of funding between institutions is not much affected by the fact that a certain degree of ambiguity concerns the peer-review status of a small part of the outputs of each university. In case of vast majority of outputs this ambiguity is not present. Yet, approximation of scholarly publications with peer review status may exclude, especially in the SSH, some outputs that researchers consider as valid contributions to researched knowledge. This limitation of the PRFS data should be considered if the data is used in research evaluation, especially at the individual level. If information systems supporting the PRFS also serve communicative purposes, they should be designed to include all outputs that researchers consider relevant contributions to research and dissemination, even if they may not be taken into account in the PRFS. Also the PRFS criteria for peer-reviewed outputs, and authority list criteria for peer-reviewed publication channels, should be regularly reviewed to ensure that outputs are taken into account adequately. In case of authority lists, it is important that expert evaluation is supported with the most comprehensive and up-to-date information concerning publication channels. Development of international information sources and expert collaboration is needed for improving reliable identification of peer-reviewed journals and book publishers across linguistic and cultural boundaries.

References

- Aagaard, K. (2015) 'How incentives trickle down: Local use of a national bibliometric indicator system', *Science and Public Policy*, 42/5: 1–13.
- Aagaard, K. (2018) 'Performance-based Research Funding in Denmark: The Adoption and Translation of the Norwegian Model', *Journal of Data and Information Science*, 3/4: 20–30.
- Aagaard, K., Bloch C. and Schneider, J. W. (2015) 'Impacts of performance-based research funding systems: The case of the Norwegian Publication Indicator', *Research Evaluation*, 24/2: 106–117.
- Aagaard, K., Bloch, C, Schneider, J, Henriksen, D., Kjeldager Ryan, T. & Lauridsen, P. (2014). *Evaluering af den norske publiceringsindikator*. Aarhus Universitet: Aarhus.
- Aagaard, K. & Schneider, J. (2017) 'Some considerations about causes and effects in studies of performance-based research funding systems', *Journal of Informetrics*, 11/3: 923–926.
- Ahlgren, P., Colliander, C., & Persson, O. (2012) 'Field normalized rates, field normalized journal impact and Norwegian weights for allocation of university research funds', *Scientometrics*, 92/3: 767–780.
- Ahlgren, P. & Waltman, L. (2014) 'The correlation between citation-based and expert-based assessments of publication channels: SNIP and SJR vs. Norwegian quality assessments', *Journal of Informetrics*, 8/4: 985–996.
- Aksnes, D. W. & Sivertsen, G. (2019) 'A Criteria-based Assessment of the Coverage of Scopus and Web of Science', *Journal of Data and Information Science*, 4/1: 1–21.
- ALLEA (2017) *The European Code of Conduct for Research Integrity*. All European Academies: Berlin.
- Archambault, É., Vignola-Gagné, É., Côté, G., Larivière, V. & Gingras, Y. (2006) 'Benchmarking scientific output in the social sciences and humanities: The limits of existing data-bases', *Scientometrics*, 68/3: 329–342.
- Auranen, O. & Nieminen, M. (2010) 'University research funding and publication performance—An international comparison', *Research Policy*, 39/6: 822–834.
- Auranen, O. & Pölönen, J. (2012) *Classification of scientific publication channels: Final report of the Publication Forum project (2010–2012)* Federation of Finnish Learned Societies: Helsinki.
- Baldwin, M. (2018) 'Scientific Autonomy, Public Accountability, and the Rise of “Peer Review” in the Cold War United States', *Isis*, 109/3: 538–558.
- Bornmann, L. (2011) 'Scientific peer review', *Annual Review of Information Science and Technology*, 45/1: 197–245.
- British Academy (2007) *Peer Review: The Challenges for the Humanities and Social Sciences*.
- Bruun-Jensen, C. (2004) 'Making Lists, Enlisting Scientists: the Bibliometric Indicator, Uncertainty and Emergent Agency', *Science Studies*, 24/2: 64–84.
- Burnhill, P.M. & Tubby-Hille, M. E. (2003) 'On Measuring the Relation between Social Science Research Activity and Research Publication. *Research Evaluation*, 4/3: 130–152.
- Butler, L. (2003) 'Explaining Australia's increased share of ISI publications – the effects of a funding formula based on publication counts', *Research Policy*, 32/1: 143–155.

- Crosland, M. (1983) 'Scientific credentials: record of publications in the assessment of qualifications for election to the French Académie des Sciences', *Minerva*, 19: 605–631.
- Csiszar, A. (2017) 'How lives became lists and scientific papers became data: cataloguing authorship during the nineteenth century', *British Journal of History of Science*, 50/1: 23–60.
- Dahler-Larsen (2012) *The Evaluation Society*. Stanford University Press: Stanford.
- Dahler-Larsen, P. (2014) 'Constitutive Effects of Performance Indicators: Getting beyond unintended consequences', *Public Management Review*, 16/7: 969–986.
- Dahler-Larsen, P. (2017a) 'Theory-Based Evaluation Meets Ambiguity: The Role of Janus Variables', *American Journal of Evaluation*, 39/1: 1–18.
- Dahler-Larsen, P. (2017b) 'The new configuration of metrics, rules, and guidelines creates a disturbing ambiguity in academia'. *LSE Impact Blog*.
<http://blogs.lse.ac.uk/impactofsocialsciences/2017/07/13/the-new-configuration-of-metrics-rules-and-guidelines-creates-a-disturbing-ambiguity-in-academia/>
- Dahler-Larsen, P. (2019) *Quality: From Plato to Performance*, Palgrave Macmillan.
- De Boer, HF., Jongbloed, B., Benneworth, P., Cremonini, L., Kolster, R., Kottmann, A., Lemmens-Krug, K. & Vossensteyn, H. (2015) *Performance-based funding and performance agreements in fourteen higher education systems - Report for the Ministry of Education, Culture and Science*, CHEPS, University of Twente: Enschede.
<https://ris.utwente.nl/ws/portalfiles/portal/5139542/jongbloed+ea+performance-based-funding-and-performance-agreements-in-fourteen-higher-education-systems.pdf>.
- de Rijcke, S., Wouters, P.W., Rushforth, A.D., Franssen, T.P. & Hammarfelt, B. (2016) 'Evaluation practices and effects of indicator use—a literature review', *Research Evaluation*, 25/2: 161–169.
- Debackere, K., Arnold, E., Sivertsen, G., Spaapen, J. & Sturn, D. (2018) *Mutual Learning Exercise: Performance-Based Funding of University Research*. European Union: Brussels.
- Derricourt, R. (2012) 'Peer review: Fetishes, fallacies and perceptions', *Journal of Scholarly Publishing*, 43/2: 137–147.
- DORA Declaration (2012) *The San Francisco Declaration on Research Assessment*. Retrieved July 3, 2018 from <https://sfdora.org/read/>.
- Else, H. (2018) 'Radical open-access plan could spell end to journal subscriptions', *Nature*, 561: 17–18.
- Engels, T. C. E., & Guns, R. (2018) 'The Flemish performance-based research funding system: A unique variant of the Norwegian model', *Journal of Data and Information Science*, 3/4: 45–60.
- Eykens, J., Guns, R. & Engels, T. C. E. (2018). *Comparing VABB-SHW (version VIII) with Cabells Journal Blacklist and Directory of Open Access Journals: Report to the Authoritative Panel*. ECOOM: Antwerp.
- Giménez-Toledo, E., Mañana-Rodríguez, J., Engels, T., Guns, R., Kulczycki, E., Ochsner, M. Pölönen, J., Sivertsen, G. & Zuccala, A. (2019) 'Taking scholarly books into account, part II: a comparison of 19 European countries in evaluation and funding', *Scientometrics*, 118/1: 233–251.
- Giménez-Toledo, E., Mañana-Rodríguez, J., Engels, T. C. E., Ingwersen, P., Pölönen, J., Sivertsen, G., Verleysen, F. T. & Zuccala, A. A. (2016) 'Taking scholarly books into account: current developments in five European countries', *Scientometrics*, 107/2: 685–699.

- Giménez-Toledo, E., Mañana-Rodríguez, J. & Sivertsen, G. (2017) 'Scholarly book publishing: Its information sources for evaluation in the social sciences and humanities', *Research Evaluation*, 26/2: 91–101.
- Gläser, J. & Laudel, G. (2007) 'The Social Construction of Bibliometric Evaluations'. In R. Whitley & J. Gläser (eds.), *The Changing Governance of the Sciences*, pp. 101–123. Springer: Dordrecht.
- Gläser, J. (2012) 'How Does Governance Change Research Content? On the Possibility of a Sociological Middle-Range Theory Linking Science Policy Studies to the Sociology of Scientific Knowledge', *The Technical University Technology Studies Working Papers*.
- Haddawy, P., Hassan, S.-U., Asghar, A., & Amin, S. (2016) 'A comprehensive examination of the relation of three citation-based journal metrics to expert judgment of journal quality', *Journal of Informetrics*, 10/1: 162–173.
- Hammarfelt, B., Nelhans, G., Eklund, P. & Åström, F. (2016) 'The heterogeneous landscape of bibliometric indicators. Evaluating models for allocating resources at Swedish universities', *Research Evaluation*, 25/3: 292–305.
- Hellqvist, B. (2010) 'Referencing in the humanities and its implications for citation analysis', *Journal of the American Society for Information Science and Technology*, 61/2: 310–318.
- 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. (2012) 'Performance-based university research funding systems', *Research Policy*, 41/2: 251–261.
- Hicks, D. & Wang, J. (2011) 'Coverage and overlap of the new social science and humanities journal lists', *Journal of the American Society for Information Science and Technology*, 62/2: 284–294.
- Hicks, D., Wouters, P. F., Waltman, L., de Rijcke, S., and Rafols, I. (2015) 'The Leiden Manifesto for research metrics: use these 10 principles to guide research evaluation', *Nature*, 520/7548, 429–431.
- Himanen, L., Auranen, O. & Puuska, H.-M. (2009) 'Influence of research funding and science policy on university research performance: A comparison of five countries', *Science and Public Policy*, 36/6: 419–430.
- Jonkers, K. & Zacharewicz, T. (2016) *Research Performance Based Funding Systems: a Comparative Assessment*. European Union: Brussels.
- Kaltenbrunner, W. & de Rijcke, S. (2016) 'Quantifying 'Output' for Evaluation: Administrative Knowledge Politics and Changing Epistemic Cultures in Dutch Law Faculties', *Science and Public Policy*, 44/2: 1–10.
- Krog Lind, J. (2019) 'The missing link: How university managers mediate the impact of a performance-based research funding system', *Research Evaluation*, 28/1: 84–93.
- Kulczycki, E., Engels, T.C.E., Pölonen, J., Bruun, K., Dušková, M., Guns, R., Nowotniak, R., Petr, M., Sivertsen, G., Istenič Starčič, A. & Zuccala, A. (2018) 'Publication patterns in the social sciences and humanities: The evidence from eight European countries', *Scientometrics*, 116/1: 463–486.
- Kulczycki, E. & Rozkosz, E. (2017) 'Does an expert-based evaluation allow us to go beyond the

Impact Factor? Experiences from building a ranking of national journals in Poland', *Scientometrics*, 111/1: 417–442.

Kulczycki, E., Rozkosz, E. A., Engels, T., Guns, R., Hołowiecki, M., & Pölönen, J. (2019) 'How to identify peer-reviewed publications: Open-identity labels in scholarly book publishing', *PloS one*, 14/3: e0214423.

Lamont, M. (2010) *How Professors Think: Inside Curious World of Academic Judgment*. Harvard University Press: Cambridge Massachusetts.

Larivière, V. & Macaluso, B. (2011) 'Improving the coverage of social science and humanities researchers' output: The case of the érudit journal platform', *Journal of the American Society for Information Science & Technology*, 62/12: 2437–2442.

Mañana-Rodríguez, J. & Giménez-Toledo, E. (2013) 'Scholarly publishing in social sciences and humanities, associated probabilities of belonging and its spectrum: a quantitative approach for the Spanish case', *Scientometrics*, 94/3: 893–910.

Mañana-Rodríguez, J. & Pölönen, J. (2018) 'Scholarly book publishers' ratings and lists in Finland and Spain: Comparison and assessment of the evaluative potential of merged lists', *Aslib Journal of Information Management*, 70/6: 643–659.

Marini, G. (2018) 'Tools of individual evaluation and prestige recognition in Spain: how sexenio 'mints the golden coin of authority'', *European Journal of Higher Education*, 8/2: 201–214.

Ministry of Education and Culture (2018) *Publication Data Collection Instructions for Research*. Retrieved July 3, 2018 from <https://wiki.eduuni.fi/download/attachments/39984924/Publication%20data%20collection%20instructions%20for%20researchers%202018.pdf?version=1&modificationDate=1528478853333&api=v2>

Mulligan, A., Hall, L. & Raphael, E. (2013) 'Peer review in a changing world: An international study measuring the attitudes of researchers', *Journal of the American Society for Information Science and Technology*, 64/1: 132–161.

Nederhof, A. J. (1989) 'Books and chapters are not to be neglected in measuring research productivity', *American Psychologist*, 44/4: 734–735.

Nederhof, T. (2006) 'Bibliometric monitoring of research performance in the Social Sciences and the Humanities: A Review', *Scientometrics*, 66/1: 81–100.

Nederhof, T. & Zwaan, R. A. (1991) 'Quality Judgments of Journals as Indicators of Research Performance in the Humanities and the Social and Behavioral Sciences', *Journal of the American society for information science*, 42(5): 332–340.

Nederhof, A. J., Zwaan, R. A., De Bruin, R. E., & Dekker, P. J. (1989) 'Assessing the usefulness of bibliometric indicators for the humanities and the social sciences: A comparative study', *Scientometrics*, 15/5-6, 423–435.

Organisation for Economic Co-operation and Development (2007) *Revised field of science and technology (FOS) classification in the Frascati manual*. Working Party of National Experts on Science and Technology Indicators, Organisation for Economic Co-operation and Development: Paris.

Ossenblok, T. L. B., Engels, T. C. E. & 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 (2005–9)', *Research Evaluation*, 21/4: 280–290.

Puuska, H.-M., Guns, R., Pölönen, J., Sivertsen, G., Mañana-Rodríguez, J. & Engels, T. (2018). *Proof of concept of a European database for social sciences and humanities publications: Description of the VIRT-A-ENRESSH pilot*, CSC & ENRESSH.

Pölönen, J. (2018) Applications of, and Experiences with, the Norwegian Model in Finland, *Journal of Data and Information Science*, 3/4: 31–44.

Pölönen, J., Leino, O. & Auranen, O. (2011) 'Coverage and Ranking of Journals: Comparison of six data sources'. *European Network of Indicator Designers (ENID) Conference in Rome, 7th-9th September 2011*.

Pölönen, J. & Sivertsen, G. (2017) 'Experiences with the rating of publication channels for the Norwegian Model: With a response to a proposal for automated ratings from Saarela et al. (2016)', *22nd Nordic Workshop on Bibliometrics And Research Policy, 9.-10.11.2017 Helsinki*.

Rowley, J. & Sbaiffi, L. (2017) 'Academics' attitudes towards peer review in scholarly journals and the effect of role and discipline', *Journal of Information Science*, 2017, pp. 1–14.

Saarela, M., Kärkkäinen, T., Lahtonen, t., Rossi, T. (2016) 'Expert-based versus citation-based ranking of scholarly and scientific publication channels', *Journal of Informetrics*, 10/3: 693–718.

Sandström, U. & van den Besselaar, P. (2018) 'Funding, evaluation, and the performance of national research systems', *Journal of Informetrics*, 12/1: 365–384.

Schneider, J. W. (2009) 'An outline of the bibliometric indicator used for performance-based funding of research institutions in Norway'. *European Political Science*, 8/3: 364–378.

Schneider, J. W., Aagaard, K. & Bloch, C.W. (2015) 'What happens when national research funding is linked to differentiated publication counts? A comparison of the Australian and Norwegian publication-based funding models', *Research Evaluation*, 25/2:1–13.

Serenko, A., & Dohan, M. (2011) 'Comparing the expert survey and citation impact journal ranking methods: Example from the field of artificial intelligence', *Journal of Informetrics*, 5/4: 629–648.

Sile, L., Guns, R., Sivertsen, G., & Engels, T. C. E. (2017) *European Databases and Repositories for Social Sciences and Humanities Research Output*. ECOOM & ENRESSH: Antwerp.

Sile, L., Pölönen, J., Sivertsen, G., Guns, R., Engels, T. C. E., Arefiev, P., Dušková, M., Faurbæk, L., Holl, A., Kulczycki, E., Macan, B., Nelhans, G., Petr, M., Pisk, M., Soós, S., Stojanovski, J., Stone, A., Šušol, J. & Teitelbaum, R. (2018) 'Comprehensiveness of national bibliographic databases for social sciences and humanities: findings from a European survey', *Research Evaluation*, 4/1: 310–322.

Sivertsen, G. (2010) 'A performance indicator based on complete data for the scientific publication output at research institutions', *ISSI Newsletter*, 6/1: 22–28.

Sivertsen, G. (2016a) 'Publication-Based Funding: The Norwegian Model'. In: M. Ochsner et al. (eds.), *Research Assessment in the Humanities: Towards Criteria and Procedures*, pp. 71–90. Springer International Publishing.

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

Sivertsen, G. (2017) 'Unique, but still best practice? The Research Excellence Framework (REF) from an international perspective', *Palgrave Communications*, 3, 17078.

- Sivertsen, G. (2018a) 'Balanced multilingualism in science', *BiD: textos universitaris de biblioteconomia i documentació*, 40.
- Sivertsen, G. (2018b) 'The Norwegian Model in Norway', *Journal of Data and Information Science*, 3/4: 2–18.
- Sivertsen, G. & Larsen, B. (2012) 'Comprehensive bibliographic coverage of the social sciences and humanities in a citation index: an empirical analysis of the potential', *Scientometrics*, 91/2: 567–575.
- Sivula, A., Suominen, J. & Reunanen, M. (2015) "'A1 alkuperäisartikkeli tieteellisessä aikakauslehdessä", Uusien julkaisukäytänteiden omaksuminen ihmistieteissä 2000-luvulla', *Kasvatus & Aika*, 9/3: 149–171.
- TENK (2012) *Responsible conduct of research and procedures for handling allegations of misconduct in Finland*. The Finnish Advisory Board on Research Integrity: Helsinki.
- Tenopir, C., Levine, K., Allard, S., Christian, L., Volentine, R., Boehm, R., Nichols, F., Nicholas, D., Jamali, H. R., Herman, E. & Watkinson, A. (2016) 'Trustworthiness and authority of scholarly information in a digital age: Results of an international questionnaire', *Journal of the American Society for Information Science and Technology*, 67/10: 2344–2361.
- Van Leeuwen, T. N. (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.
- Van Leeuwen, T. (2013) 'Bibliometric Research Evaluations, Web of Science and the Social Sciences and Humanities: A Problematic Relationship?', *Bibliometrie - Praxis und Forschung*, 2.
- Verleysen, F. T. & Engels, T. C. E. (2013) 'A Label for Peer-Reviewed Books', *Journal of the American Society for Information Science and Technology*, 64/2: 428–430.
- Verleysen, F. T. & Engels, T. C. E. (2014) 'Internationalization of peer reviewed and non-peer reviewed book publications in the Social Sciences and Humanities', *Scientometrics*, 101/2: 1431–1444.
- Verleysen, F.T. & Engels, T.C.E. (2015) 'ERIH Plus in 2014: Stand van zaken en tijdschriftselectie vergeleken met het VABB-SHW'. ECOOM: Antwerpen.
- Verleysen, F. T., Ghesquière, P. & Engels, T. C. E. (2014) 'The objectives, design and selection process of the Flemish Academic Bibliographic Database for the Social Sciences and Humanities (VABB-SHW)'. In: W. Blockmans et al. (eds.) *The use and abuse of bibliometrics*, pp. 115–125. Academiae Europaea; Portland Press.
- Verleysen, F.T. & Rousseau R. (2017) 'How the existence of a regional bibliographic information system can help evaluators to conform to the principles of the Leiden Manifesto', *Journal of Educational Media & Library Sciences*, 54/1: 97–109.
- Wahlfors, L. & Pölönen, J. (2018) 'Julkaisufoorumi-luokituksen käyttö yliopistoissa', *Hallinnon Tutkimus*, 37/1: 7–21.
- Walters, W. (2017) 'Do subjective journal ratings represent whole journals or typical articles? Unweighted or weighted citation impact?', *Journal of Informetrics*, 11/3, 730–744.
- Woelert, P. & McKenzie, L. (2018) 'Follow the money? How Australian universities replicate national performance-based funding mechanisms', *Research Evaluation*, 2018, 3/1: 1–12.
- Zacharewicz, T., Lepori, B., Reale, E. & Jonkers, K. (2018) 'Performance-based research funding

in EU Member States—a comparative assessment’, *Science and Public Policy*, 1–11.

Table 1: Classification of journals as peer-reviewed (1) or not peer-reviewed (0) in Finland and Flanders.

Field of science	Journals	FLA 0 - FIN 1	FLA 1 - FIN 0	Grey zone
All SSH fields	4505	286	141	9.5 %
Social sciences	2789	100	111	7.6 %
5.1 Psychology	466	2	18	4.3 %
5.2 Economics and business	687	18	26	6.4 %
5.3 Educational sciences	276	6	13	6.9 %
5.4 Sociology	344	15	7	6.4 %
5.5 Law	279	22	13	12.5 %
5.6 Political science	220	15	13	12.7 %
5.7 Social and economic geography	197	5	2	3.6 %
5.8 Media and communications	221	11	9	9.0 %
5.9 Other social sciences	99	6	10	16.2 %
Humanities	1716	186	30	12.6 %
6.1 History and archaeology	324	43	3	14.2 %
6.2 Languages and literature	748	86	13	13.2 %
6.3 Philosophy, ethics and religion	404	25	1	6.4 %
6.4 Arts	180	25	10	19.4 %
6.5 Other humanities	60	7	3	16.7 %

Table 2: Discrepancies in classification of social sciences and humanities journals between Flanders and Finland according to publishing language

Language	Journals	Share of all humanities journals	Share of humanities journals with discrepancies	Share of all social sciences journals	Share of social sciences journals with discrepancies
English	3348	52.3 %	24.1 %	87.8 %	57.3 %
Multiple languages	559	23.1 %	35.2 %	5.8 %	19.4 %
French	184	8.2 %	14.4 %	1.5 %	5.2 %
German	117	5.0 %	10.6 %	1.1 %	2.4 %
Dutch	83	3.4 %	3.7 %	0.9 %	3.3 %
Italian	63	3.1 %	3.7 %	0.4 %	2.4 %
Spanish	57	2.0 %	1.9 %	0.8 %	4.7 %
Swedish	9	0.3 %	0.9 %	0.1 %	0.0 %
Finnish	5	0.0 %	0.0 %	0.2 %	0.9 %
Other languages	80	2.5 %	5.6 %	1.3 %	4.3 %

Table 3: Bottom-up and top-down classification of publications as peer-reviewed or not peer-reviewed in Finland

Field	All publications	Publications with discrepancy	Grey zone
All SSH	32427	5108	16 %
Social sciences	20998	3186	15 %
5.1 Psychology	1956	128	7 %
5.2 Economics and business	4896	626	13 %
5.3 Educational sciences	3822	756	20 %
5.4 Sociology	3013	476	16 %
5.5 Law	2218	400	18 %
5.6 Political science	1656	299	18 %
5.7 Social and economic geography	651	89	14 %
5.8 Media and communications	1107	128	12 %
5.9 Other social sciences	1679	284	17 %
Humanities	11948	1963	16 %
6.1 History and archaeology	3057	488	16 %
6.2 Languages and literature	3872	587	15 %
6.3 Philosophy, ethics and religion	2654	334	13 %
6.4 Arts	1301	346	27 %
6.5 Other humanities	1064	208	20 %

Table 4: Bottom-up and top-down classification of publications as peer-reviewed in Finland for book publications and Finnish language publications.

Field	Book publications	Top-down vs. Bottom-up discrepancy	Grey zone	Finnish language publications	Top-down vs. Bottom-up discrepancy	Grey zone
All SSH	14012	3449	25 %	9273	2032	22 %
Social sciences	7417	2046	28 %	5292	1253	24 %
5.1 Psychology	198	53	27 %	250	53	21 %
5.2 Economics and business	989	301	30 %	348	122	35 %
5.3 Educational sciences	1546	517	33 %	1116	320	29 %
5.4 Sociology	1276	348	27 %	996	243	24 %
5.5 Law	1172	275	23 %	1076	197	18 %
5.6 Political science	815	209	26 %	489	135	28 %
5.7 Social and economic geography	210	56	27 %	134	30	22 %
5.8 Media and communications	491	83	17 %	327	40	12 %
5.9 Other social sciences	720	204	28 %	556	113	20 %
Humanities	6767	1427	21 %	4122	792	19 %
6.1 History and archaeology	2031	380	19 %	1458	228	16 %
6.2 Languages and literature	2166	428	20 %	1058	169	16 %
6.3 Philosophy, ethics and religion	1460	233	16 %	775	121	16 %
6.4 Arts	578	240	42 %	461	172	37 %
6.5 Other humanities	532	146	27 %	370	102	28 %

Figure 1: Share of book and Finnish language publications classified differently as peer-reviewed by the universities and authority list of publication channels in Finland.

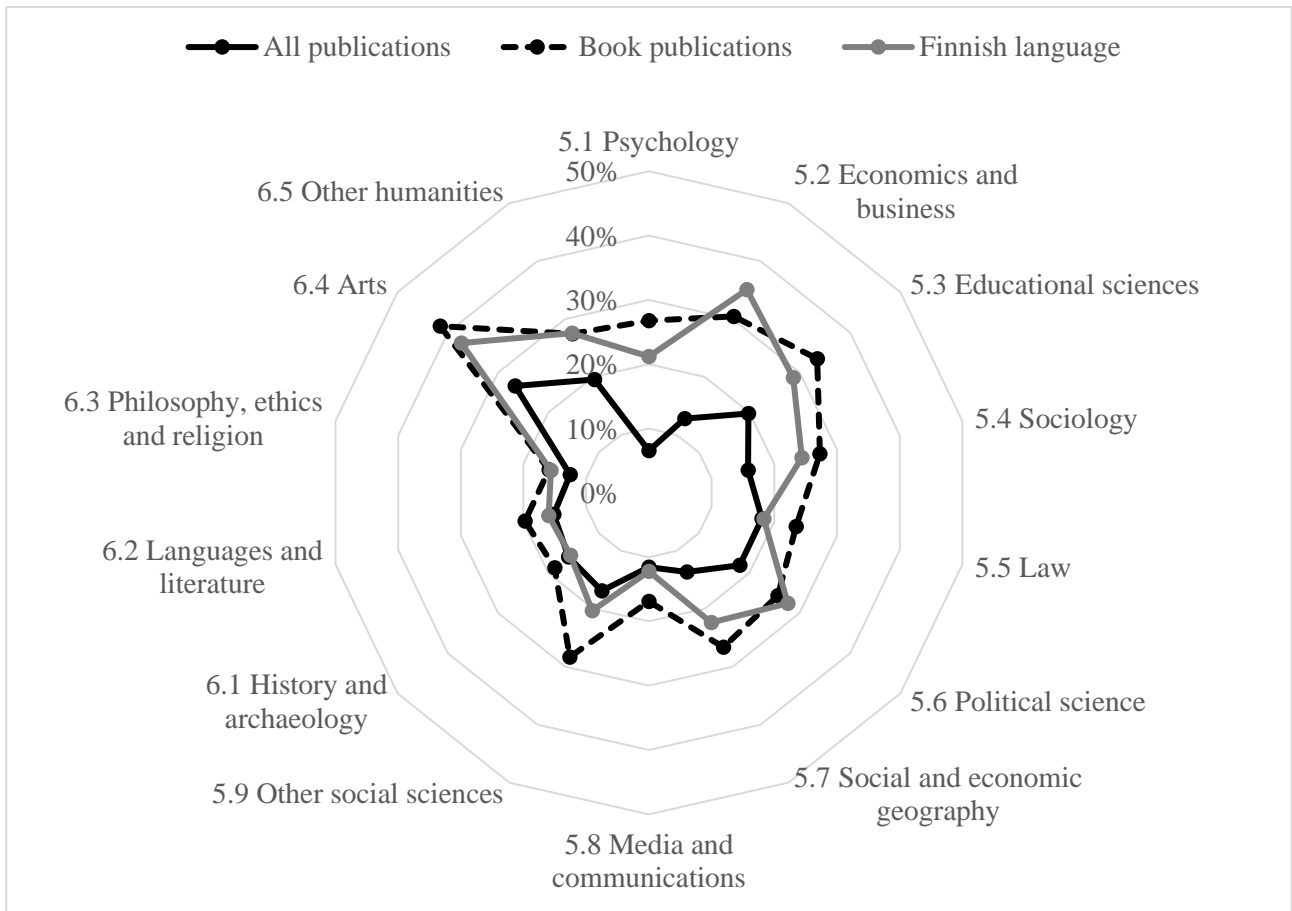


Table 5: Classification of SSH co-publications as peer-reviewed or not peer-reviewed by different universities in Finland

	Publications with authors from 2 or more universities	Discrepancy in peer review status between 2 or more universities	Grey zone
All SSH	3596	303	8 %
Social sciences	3436	284	8 %
5.1 Psychology	544	12	2 %
5.2 Economics and business	858	66	8 %
5.3 Educational sciences	676	52	8 %
5.4 Sociology	428	44	10 %
5.5 Law	157	22	14 %
5.6 Political science	211	28	13 %
5.7 Social and economic geography	89	11	12 %
5.8 Media and communications	174	18	10 %
5.9 Other social sciences	299	31	10 %
Humanities	949	103	11 %
6.1 History and archaeology	149	12	8 %
6.2 Languages and literature	335	33	10 %
6.3 Philosophy, ethics and religion	132	22	17 %
6.4 Arts	162	20	12 %
6.5 Other humanities	171	16	9 %

Table 6: Classification as peer-reviewed of SSH co-publications by different universities in Finland for book publications and Finnish language publications

Field	Book publications	Top-down vs. Bottom-up discrepancy	Grey zone	Finnish language publications	Top-down vs. Bottom-up discrepancy	Grey zone
All SSH	1178	194	16 %	1379	214	16 %
Social sciences	1059	180	17 %	1059	161	15 %
5.1 Psychology	39	8	21 %	59	6	10 %
5.2 Economics and business	160	35	22 %	116	23	20 %
5.3 Educational sciences	297	38	13 %	296	34	11 %
5.4 Sociology	170	27	16 %	190	31	16 %
5.5 Law	79	17	22 %	91	14	15 %
5.6 Political science	95	18	19 %	90	19	21 %
5.7 Social and economic geography	42	8	19 %	32	9	28 %
5.8 Media and communications	57	10	18 %	69	9	13 %
5.9 Other social sciences	120	19	16 %	116	16	14 %
Humanities	446	68	15 %	426	65	15 %
6.1 History and archaeology	77	6	8 %	73	6	8 %
6.2 Languages and literature	157	18	11 %	135	16	12 %
6.3 Philosophy, ethics and religion	85	17	20 %	71	17	24 %
6.4 Arts	46	14	30 %	72	16	22 %
6.5 Other humanities	81	13	16 %	75	10	13 %

Figure 2: Share of co-publications in books and national language classified differently as peer-reviewed or not peer-reviewed by the universities in Finland.

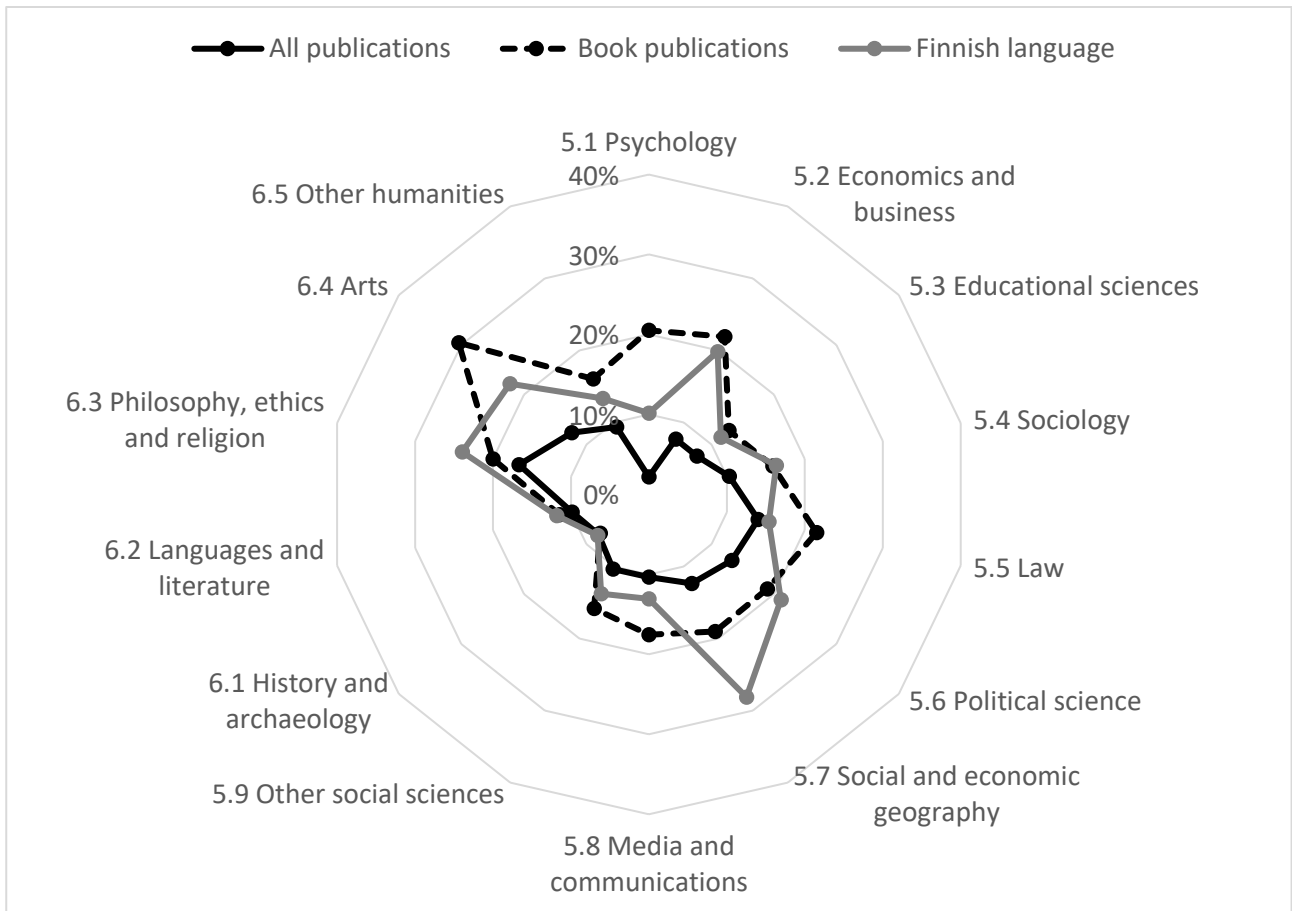


Figure 3: Comparison of peer review grey-zones across fields

