UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO INTERDISCIPLINAR DE NOVAS TECNOLOGIAS NA EDUCAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA NA EDUCAÇÃO (PPGIE)
NÍVEL DOUTORADO

ALAUSE DA SILVA PIRES

A FRAMEWORK FOR MONITORING SCIENTIFIC PRODUCTION BEHAVIOR IN RESEARCH EVALUATION SYSTEMS BASED ON JOURNAL RANKING LISTS: THE BRAZILIAN CASE

PORTO ALEGRE
2020

Alause da Silva Pires

A FRAMEWORK FOR MONITORING SCIENTIFIC PRODUCTION BEHAVIOR IN RESEARCH EVALUATION SYSTEMS BASED ON JOURNAL RANKING LISTS: THE BRAZILIAN CASE

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Informática na Educação (PPGIE) da Universidade Federal do Rio Grande do Sul, como parte das exigências para obtenção do título de Doutora em Informática na Educação.

Orientador: Prof. Dr. Eliseo Berni Reategui
Coorientador: Prof. Dr. Sérgio Franco

Porto Alegre
2020

CIP - Catalogação na Publicação

Pires, Alause
A FRAMEWORK FOR MONITORING SCIENTIFIC PRODUCTION BEHAVIOR IN RESEARCH EVALUATION SYSTEMS BASED ON JOURNAL RANKING LISTS: THE BRAZILIAN CASE / Alause Pires. -- 2020.
159 f.
Orientador: Eliseo Reategui.

Coorientador: Sérgio Franco.

Tese (Doutorado) -- Universidade Federal do Rio Grande do Sul, Centro de Estudos Interdisciplinares em Novas Tecnologias na Educação, Programa de Pós-Graduação em Informática na Educação, Porto Alegre, BR-RS, 2020.

1. research evaluation systems. 2. journal rankings. 3. framework. 4. evaluation potential effects. 5. Brazilian journal ranking system. I. Reategui, Eliseo, orient. II. Franco, Sérgio, coorient. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO INTERDISCIPLINAR DE NOVAS TECNOLOGIAS NA EDUCAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA NA EDUCAÇÃO

ATA SOBRE A DEFESA DE TESE DE DOUTORADO
ALAUSE DA SILVA PIRES

Às quatorze horas do dia vinte e oito de abril de dois mil e vinte, no endereço eletrônico https://mconf.ufrgs.br/webconf/00126356, conforme a portaria 2291 de 17/03/2020 que suspende todas as atividades presenciais possíveis, nesta Universidade, reuni-se a Comissão de Avaliação, composta pelos Professores Doutores: Leandro Krug Wives , Eduardo Winter e Robert Evan Verhine para a análise da Defesa de Tese de Doutorado intitulada "**A Framework for Monitoring Scientific Production Behavior in Research Evaluation Systems Base don Journal Ranking Lists: The Brazilian Case",** da doutoranda de Pós – Graduação em Informática na Educação Alause da Silva Pires sob a orientação do Prof. Eliseo Berni Reategui e coorientação do Prof. Sérgio Kieling Franco.

A Banca, reunida, após a apresentação e argüição, emite o parecer abaixo assinalado.

[x] Considera a Tese Aprovada
( ) sem alterações;
(x) sem alterações, com voto de louvor;
( ) e recomenda que sejam efetuadas as reformulações e atendidas as sugestões contidas nos pareceres individuais dos membros da Banca;

[ ] Considera a Tese Reprovada.

Considerações adicionais (a critério da Banca):

____
Prof. Dr. Eliseo Berni Reategui
Orientador

____
Prof. Dr. Sérgio Kieling Franco
Coorientador

_______(videoconferência)__________
Prof. Dr. Leandro Krug Wives
PPGIE/ UFRGS

______(videoconferência)____
Prof. Dr. Eduardo Winter
INPI

_____(videoconferência)______
Prof. Dr. Robert Evan Verhine
UFBA

I dedicate this thesis to the Spirits of Light, who guided, protected, and strengthened me during this challenging journey.

This is also to my mother, Anelita Maria Pires. I am grateful to this great woman for believing that the greatest inheritance she could leave for her children was education.

To my father, Alberto Mateus Pires, my sister, Aniane Pires, my sister-in-law, Luisa Helena Gomes, my brother, Alerson Pires, and niece, Andressa Pires, who gave me all their love and support in this quest.

ACKNOWLEDGMENTS

I would like to express my most heartfelt gratitude and sincere thanks to my advisor Dr. Eliseo Reategui and my co-advisor, Dr. Sérgio Franco, for their continuous support of my Ph.D. research. Dr. Eliseo Reategui and Dr. Sérgio Franco have been much more than advisors; they have been professional role models and amazing friends.

Dr. Eliseo Reategui always advised me with patience, motivation, passion, and enthusiasm. He also supported and encouraged my dreams: having a better education and instilling in me that I can obtain that goal. Thanks to his support, I had the most fantastic and enriching academic experience, to include studying at Stanford University. While I was at Stanford, Dr. Eliseo Reategui gave me incredible emotional support, while also serving the roles as an amazing advisor and a great friend.

I am equally grateful to Dr. Sérgio Franco, who always encouraged me to think outside the box by showing me different perspectives and possibilities. He gave me valuable guidance, insightful comments, and timely advice to accomplish this task. His kind and wise words were extremely helpful in difficult moments.

In addition to my advisor and co-advisor, I am also grateful to Dr. Eduardo Winter, Dr. Marcelo Albano Moret, and Dr. Adelaide Faljoni-Alario. They gave me great support in getting into the Ph.D., as well as with the reference letters for my Stanford application. I am thankful for their kind words which afforded me that incredible opportunity.

Much gratitude goes to everyone at the Brazilian Coordination for the Improvement of Higher Education Personnel (CAPES) for providing the funding for my research and supporting me throughout my Ph.D. studies. I would also like to thank the Lemann Center at Stanford University for accepting me as a Visiting Student Researcher for one year and funding my year there.

I thank my fellow Ph.D. colleagues for the stimulating discussions and for all the fun we have had over these past few years through amazing Gaucho barbecues and beers. Among these fellow colleagues, a special thanks to Manuel Zunguze, who helped me with my application to Stanford.

Additionally, I thank all my friends for their unconditional support and for understanding my absences while I was focused on the Ph.D. I would like to especially thank Ariana Neustein, who helped me articulate, develop, and submit my Stanford application in English; to Andrea Midori Takai, Carina Messias, Flávia Melissa Moraes and Thaís Pereira, who encouraged me to undertake this Ph.D., and assisted me with CAPES documentation; to Eurico Miranda and Edmilson Chaves Júnior, who gently helped me with CAPES data access; to Adriana Paz, Juan Pablo Bascur and Nuria Bautista, who recently became my friends but with whom I shared many anxieties pertaining to writing a thesis.

Last but not least, I would like to thank my family. Without them, the writing of this thesis would not have been possible. They have provided me with tremendous emotional support.

Thanks for all your encouragement!

"No one is born fully-formed: it is through self-experience in the world that we become what we are."
— *Paulo Freire*

ABSTRACT

The extent and quality of research output have become key factors for university performance evaluation. Several countries introduced research evaluation systems that link funding to performance indicators as a way to enhance accountability. In general, journal rankings are an integral part of these systems. This thesis approaches the development and evaluation of a framework for monitoring scientific production behavior in settings where journal-ranking lists are at the center of research assessment. The main goal of the framework is to enable the identification of desirable and adverse patterns in academic production. Considering that Brazil has been using a specific journal ranking system (QUALIS) for more than two decades, the framework was applied in the ten-year analysis of Brazilian scientific production in eight distinct subject fields and taking the Scopus database as a reference. Results showed a decline in the proportion of Scopus-indexed articles in the areas of Social Sciences and Humanities (SSH). A few journals that remained in the system during the whole evaluation period concentrated a larger number of published articles. Overall, these journals had their QUALIS classification unchanged or improved in the ranking over the periodic evaluations. However, in general, there was a significant decrease in their citation impact. Moreover, lower-impact journals moved to the highest QUALIS categories over the years, what happened simultaneously with an increase in the number of articles in low-impact journals in all fields. These results have shown that the use of journal ranking lists may lead faculty and students to submit their papers to highly ranked journals, even though may have a low citation impact. When low-impact journals reach a high rank, they may also concentrate a high amount of published articles. In a certain way, these patterns are similar to other results found in literature, in which a significant increase in publication productivity has been followed by an impact decline. The potential effect of these evaluation models is that they may incite people to select publication venues that make them score higher according to the established criteria, regardless of their publications' visibility. Besides, this effect can be intensified once the evaluation results are linked to funds.
Keywords: research evaluation systems; journal rankings; framework, evaluation potential effects; Brazilian journal ranking system.

RESUMO

A extensão e a qualidade dos resultados de pesquisa tornaram-se fatores-chave para a avaliação do desempenho da universidade. Vários países introduziram sistemas de avaliação de pesquisas que vinculam financiamento a indicadores de desempenho como forma de aumentar a prestação de contas. Em geral, classificações de periódicos são parte integrante desses sistemas. Esta tese aborda o desenvolvimento e a avaliação de um *framework* para monitorar o comportamento da produção científica em contextos em que as listas de classificação de periódicos estão no centro da avaliação. O principal objetivo do *framework* é permitir a identificação de padrões desejáveis e adversos na produção acadêmica. Considerando que o Brasil utiliza um sistema de classificação de periódicos (QUALIS) há mais de duas décadas, o framework foi aplicado numa análise de dez anos da produção científica brasileira em oito áreas distintas, tomando o banco de dados da Scopus como referência. Os resultados mostraram um declínio na proporção de artigos indexados na Scopus nas áreas de Ciências Sociais e Humanas (SSH). Um número restrito de periódicos, que permaneceu no sistema durante o período estudado, concentrou um número maior de artigos. No geral, esses periódicos tiveram sua classificação QUALIS inalterada ou melhorada ao longo das avaliações periódicas. Entretanto, na maioria, houve uma diminuição significativa no impacto de citação desses. Além disso, aqueles de menor impacto passaram para as categorias QUALIS mais altas ao longo dos anos, o que ocorreu simultaneamente com um aumento no número de artigos em periódicos de baixo impacto em todos as áreas. Esses resultados mostraram que o uso de listas de classificação de periódicos pode levar professores e alunos a publicarem em periódicos de alta classificação, apesar de terem um baixo impacto de citação. Quando os periódicos de baixo impacto alcançam uma classificação alta, eles também podem concentrar uma grande quantidade de artigos publicados. De certa forma, esses padrões são semelhantes a outros resultados encontrados na literatura, nos quais um aumento significativo de publicações foi seguido por um declínio no impacto. O efeito potencial desses modelos de avaliação é que eles podem incitar as pessoas a selecionar meios de publicação com uma pontuação mais alta de acordo com os critérios estabelecidos, independentemente de sua visibilidade. Além disso, esses efeitos podem intensificar-se quando os resultados da avaliação são vinculados a financiamento.

Palavras-chave: sistemas de avaliação de pesquisas; classificações de periódicos; *framework*; efeitos potenciais da avaliação; sistema brasileiro de classificação de periódicos.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ABS	Association of Business Schools
A&HCI	Arts & Humanities Citation Index
AIS	Article Influence Score
ANVUR	Italian National Agency for the Evaluation of the University and Research Systems
ARC	Australian Research Council
CAPES	Brazilian Coordination for the Improvement of Higher Education Personnel
CIVR	Italian Committee for the Evaluation of Research
CERIF	Common European Research Information Format
CNEAI	Spanish National Commission for Research Evaluation
CNPq	Brazilian Science and Technology Development Council
CONECTI	Brazilian National Consortium of Education, Science, Technology and Innovation
CRIS	Current Research Information System
CRISs	Current Research Information Systems
CTC-ES	Technical and Scientific Council for Higher Education
CV Lattes	Lattes Curriculum
DCP	Database Citation Potential
ERA	Excellence in Research for Australia
euroCRIS	International Organization for Research Information
GS	Google Scholar
HES	Higher Education System
IF	Impact Factor
ISI	Institute of Scientific Information
JCR	Journal Citation Reports
LP	Lattes Platform
PNPG	National Plan for Graduate Studies
PRFS	Performance-based Research Funding System
PRFSs	Performance-based Research Funding Systems

RAE	The UK Research Assessment Exercise
REF	The UK Research Excellence Framework
SciELO	Scientific Electronic Library Online
SNIP	Source Normalized Impact per Paper
SNPG	Brazilian National Graduate System
SSH	Social Sciences and Humanities
VTR	Valutazione Triennale Della Ricerca
WoS	Web of Science

TABLE OF CONTENTS

1 INTRODUCTION

Higher Education evaluation has become pivotal in the Brazilian educational policy, following international trends (GOERGEN, 2010). It is carried out in undergraduate courses by the Brazilian System of Higher Education Evaluation (SINAES) and in graduate programs by the Brazilian National Graduate System (SNPG). The latter, in particular, follows evaluation processes defined by the Brazilian Coordination for the Improvement of Higher Education Personnel (CAPES), which is responsible for fostering graduate education and research in the country. In this research CAPES' data was used to carry out an analysis of Brazilian research output from 2007 to 2016. This analysis was based on a *framework, also developed as part of this thesis,* for monitoring scientific production behavior.

In the 1980s and 1990s, governments changed the way of interacting with colleges and universities, which became more pressured for more accountability, efficiency, and productivity, mainly in the use of publicly generated resources. This shift reflected the increasing societal requirement for making colleges and universities more responsive to national economic needs and new governmental demands for increased performance. At that moment, many countries understood that investing in the development of human capital and research through higher education was necessary for strengthening their competitive and economic positioning. Thus, the governments' interest in performance funding and budgeting for higher education increased (ALEXANDER, 2000). This increasing emphasis on governance and accountability has also led to a growing demand internationally for research evaluation systems, based not only on objective evidence, but also on transparent methods (GEUNA; MARTIN, 2003). Thus, in the recent decades, many countries adopted research evaluation systems as instruments to allocate public funds and safeguard research quality (BESLEY; PETERS, 2009; GEUNA; MARTIN, 2003; HICKS, 2012; MARTIN; WHITLEY, 2010; OANCEA; PRING, 2008). These systems were called Performance-based Research Funding Systems – PRFSs (HICKS, 2012).

CAPES started evaluating the Brazilian graduate programs in 1976. In 1980, it implemented a national evaluation system, which stands out until nowadays as an instrument of great value for the SNPG (VERHINE; DANTAS, 2009). Over the years, it has been evaluating the performance of graduate programs, which is then linked to funding. This evaluation has taken place every four years since 2013. Hence, graduate programs inform annually all the data considered essential by

CAPES to the periodic evaluation process. These data are collected and stored through an information system named ColetaCapes, which is an important system for supporting SNPG. Among all these data, scientific production is one of the main aspects evaluated by CAPES. According to Barata (2019), the evaluation process was firmly shaped around the scientific production of graduate programs, which represents about 70% of the score (30% to 40% of the score allocated to the intellectual production of academic staff and 30% to 40% to student production). Additionally, despite the variety of scientific production formats informed to CAPES, such as journal articles, books, edited and co-edited book volumes, journal articles are still the main format of publication in Brazil (MIRANDA; MUGNAINI, 2014; MUGNAINI, 2015). In summary, CAPES has been evaluating the performance of Brazilian graduate programs based mainly on their research, but more specifically research results published in journals. Because of that, this study focuses specifically on journals as the standard research output format for evaluation purposes. Evaluating the impact and prestige of these journals is a form of estimating the potential quality of the research published in them.

In 1998, CAPES developed a Brazilian journal ranking system named QUALIS, a critical instrument for supporting the periodic evaluations. In this information system, the journals in which faculty and students publish their works were listed and classified into strata indicative of quality – A1, the highest; A2; B1; B2; B3; B4; B5; C – with zero weight. As a result, one list of journals per subject field used to be provided and published every year until 2016 (BARATA, 2016). Furthermore, this ranking was carried out by the advisory committees of each of the 48 subject fields considered, following criteria previously defined by them and approved by the Technical and Scientific Council for Higher Education (CTC-ES). It sought to reflect the relative importance of different journals for a given subject field. In brief, the QUALIS system used to be characterized by formalized sets of rules and procedures organized around existing disciplines and scientific boundaries. Besides, CAPES employed this QUALIS format as periodic evaluations' support only until the last one that took place in 2017, referring to the years 2013 to 2016. A new QUALIS was approved in 2019 to be applied in the 2021 periodic evaluation.

In addition, even though CAPES states that QUALIS was conceived only to compare graduate programs within each subject field, and it should not be used out of this scope, it is well known that QUALIS has been used as a national quality indicator. Throughout the years, it has been employed to evaluate researchers, faculties, students (BARATA, 2016), as well as to allocate funds by other research funding institutions in Brazil (OLIVEIRA; AMARAL, 2017). On that subject, QUALIS seems to be working as a Performance-based Research Funding System (PRFS) in the national scenario, even if it is not its intention. Many countries, especially in Europe, implemented the PRFSs, which became the basis for greater accountability mainly to publicly funded research (WHITLEY; GLÄSER, 2007; GEUNA; MARTIN 2003; MINELLI; REBORA; TURRI, 2008). For the last three decades, PRFSs have been used as a science policy tool around the world. The objectives of these systems include the allocation of research funds to the most productive institutions, stimulation of excellence in research, improving accountability of public research, and promotion of greater research alignment to societal and economic needs (BLOCH; SCHNEIDER, 2016). Although research outputs include publications, projects, organized conferences and others, publications are in general the most significant component in the evaluated outcomes (KULCZYCKI; KORZEŃ; KORYTKOWSKI, 2017).

A key feature of PRFSs is the percentage of funding that depends on the research evaluation system (HICKS, 2012). More than 90% of the Brazilian research is developed in the universities, mainly in public ones, which do not have their own budget for science and technology activities. In general, the funding of these universities comes from the government, including research funding. While CAPES and the Brazilian Science and Technology Development Council (CNPq) are at the federal level, the State Research Support Foundations (FAPs) support research initiatives at the states level. From 2003 to 2007, funds of the CNPq were superior to those of CAPES. However, this scenario changed from 2008 to 2015, when CAPES funding gradually became expressively higher than that of CNPq (ANDES-SN, 2018). Hence, CAPES is nowadays the primary federal research funding institution in Brazil. It links funds to the performance of graduate programs according to their evaluation process. Considering QUALIS weight in the evaluation process and higher budget concentration in CAPES, the Foundation, through its research evaluation mechanism, has a significant role in directing the Brazilian research paths.

The first country to implement a PRFS was the United Kingdom in 1986 and since then, many countries have introduced and embedded them in their national research systems (BLOCH; SCHNEIDER, 2016; HICKS, 2012). In this regard, many efforts have been made to implement information systems to support research evaluation activities in different countries (WILSDON *et al.*, 2015). Following these international trends, CAPES and other institutions associated with research in Brazil have recently established a National Consortium in Education, Science, Technology and Innovation (CONECTI). For that, CAPES is planning some improvements to enable Sucupira Platform to become a Current Research Information System (CRIS), intended to help with the information management of research activity. CRIS is a database (or other information system) aimed to store, manage, and exchange contextual metadata for funded research activity conducted at a research-performing organization[1]. Some of their purposes include support for research assessment, compliance management, and assistance in the promotion and access to the research outcomes. These systems also provide a deep insight into the workflows that underpin the institutional research activity. Therefore, the development of such a system in Brazil should provide a better research information management in the country. The framework for research evaluation proposed in this thesis, once embedded in a CRIS, would provide a more realistic view of the possible effects or distortions of research evaluation systems implemented in a national level.

In general, peer review and bibliometric indicators are the two main approaches used in research evaluation systems, or even a combination of both methods. Peer review is almost ubiquitous in all science evaluation systems, especially when it comes to funding allocation. However, there are many warnings related to the reliability and validity of these reviews (GANS; SHEPHERD, 1994; BEDEIAN, 2003; REALE; BARBARA; COSTANTINI, 2007). In some countries, such as France, Germany, Switzerland, Scandinavian countries and the United States, the introduction of PRFSs has been a key driver of bibliometric activity, leading to renewed interest in bibliometrics as “an instrument of science management” (BALL; TUNGER, 2006, p. 564). Pendlebury (2008) lists more than 20 countries worldwide that regularly use bibliometric reports or “science indicator studies” to evaluate research performance and inform resource allocation. Many governments and evaluation agencies have experimented this approach, since peer review is costly and time-consuming.

Nowadays, bibliometric indicators, especially citation impact indicators, play an essential role in research evaluation. The three most important databases available for performing citation analyses are: Web of Science (WoS), Scopus, and Google Scholar (GS). The impact factor (IF) obtained from WoS has been used for many years as the main method to determine the academic value of a given article. However, given the increasing need for more reliable, fair, and inclusive instruments to evaluate research performance, other bibliometric indicators have proliferated in

---

[1]euroCRIS (2013). “Why does one need a CRIS?”. Available at: https://www.eurocris.org/why-does-one-need-cris. Accessed on: Sept. 15, 2019.

recent years. Thus, other well know citations indicators have been created, such as: Eigenfactor – based on WoS; SNIP and SJR – both based on Scopus; and h-index – nowadays computed by WoS, Scopus, and GS. Still, no bibliometric indicator by itself can capture all research dimensions because of the complexity of research communication systems. The adequacy of a journal impact measure is related to how it is used, and the type of research question addressed. An indicator may be appropriate in one context, whereas less appropriate in others (MOED, 2010).

The employment of these metrics to judge research, especially the IF, has suffered a lot of criticism for various reasons. Firstly, it is questionable that an article with no citations could be considered to have a high impact only because the journal where it was published has a positive citation record. Secondly, journals in languages other than English are often not included in the Science Citation Index of WoS (KURMIS, 2003). It has also been highlighted that the IF was created to help librarians manage journal collections, but not to measure the scientific merit of research (ARCHAMBAULT; LARIVIÈRE, 2009). As a result, there have been many demands for better mechanisms to improve the evaluation of scientific research outputs and cease the promotion of the IF, including the San Francisco Declaration on Research Assessment (DORA, 2012), The Leiden Manifesto (HICKS *et al*., 2015), and the Metrics Tide (WILSDON *et al*., 2015). All these documents claim for changes, but there has been no solution yet, according to Curry (2018), who states that beyond complaining, it is necessary to find robust, efficient and bias-free assessment methods to discover and disseminate examples of good practices, boosting the profile of assessment reform.

Furthermore, bibliometric-supported evaluations in Social Sciences and Humanities (SSH) is one of the major challenges in this scope, considering that there have been no comprehensive bibliographic data suitable for bibliometrics analysis in SSH. This limitation is mainly due to the relatively strong national and regional orientation of SSH research as well as their output diversity, which usually is not covered by the leading international databases. Hence, the coverage of WoS and Scopus is quite limited for research evaluation purposes in SSH. Both the Leiden Manifesto (HICKS *et al*., 2015) and the San Francisco Declaration on Research Assessment (DORA, 2012) have highlighted the need to take into account the diversity of research outputs across different knowledge fields.

Regarding QUALIS, it combined peer review and citation impact indicators in many subject fields until 2016. However, especially in SSH, peer review evaluation used to be the main method considered. Journal ranking systems have generated much interest and criticism in literature. In the same way, other countries such as Australia, France, Italy and Poland have also developed national journal rankings in their research evaluation systems (FERRARA; BONACCORSI, 2016; HADDOW; GENONI, 2010; VANCLAY, 2011). The legitimization and reification of journal rankings through research evaluation processes have the potential to create major behavioral changes in the academic community, as researchers realize that their careers depend on publishing in journals attributed with high rank (YOUNG *et al*., 2011). Cooper and Poletti (2011) argued that journal rankings produce a set of perverse and dysfunctional reactions that threaten to undermine long-term research quality. Moreover, academics are pushed to 'play the game,' thus changing attitude towards research evaluation (COOPER; POLETTI, 2011; ADLER; HARZING, 2009).

Along the same lines, the spread of PRFSs around the world has also generated interest in how these models impact research, but this understanding is still very limited (WHITLEY; GLASER, 2007; BUTLER *et al.*, 2010). In this regard, some potential problems of these systems are pointed out in the literature. For instance, these models can promote risk-averse behavior; induce mono-disciplinary research at the expenses of interdisciplinarity, besides discouraging certain types of research (BLOCH; SCHNEIDER, 2016). There is also a concern that they would promote a very narrow perception of usefulness for public research that neglects the role of research as knowledge resources, the importance of teaching, and research's wider democratic influence, in which only quantifiable goals of research are considered legitimate (BLOCH; SCHNEIDER, 2016).

Thus and considering that QUALIS is not only a Performance-based Research Funding System (PRFS) but also a national journal ranking list, it may be inducing some of the potential effects found in other research evaluation systems. Hence, monitoring the QUALIS system is essential in a moment that CAPES is at the same time rethinking all its evaluation process, as well as developing a national environment of research information systems, integrating federal and state systems. From this perspective, the main goal of this thesis has been to develop and evaluate a framework for monitoring scientific production behavior in settings where journal ranking lists are an important foundation for research assessment. The specific goals are as follows:

- to design the framework;
- to use the framework for monitoring Brazilian scientific production behavior;
- to contrast QUALIS with impact indicators from Scopus;

- to understand the shifting of QUALIS categories over the years.

2 HIGHER EDUCATION AND RESEARCH EVALUATION SYSTEMS

The Higher Education System (HES) has been through some reforms, including the strengthening of institutional leadership, the establishment of governing boards, the enhancement of quality, the improvement of accounting and accountability, and the implementation of performance management systems (SPORN, 2003). These reforms were stimulated by the rise of a global student market for education and research, 'massification' of higher education, rising costs of expanded HES, and pressure for management efficiency in the face of widened access and reduced resources (CURRIE, 1998).

The reforms and policies on education and research aimed to increase the competitiveness of national knowledge and research innovation. More emphasis was, therefore, given on accounting and accountability, thus changing the universities' culture of an academic, or elite, self-governance to public evaluations, which were considered more transparent, numerical and democratic (KOGAN; HANNEY, 2000, p. 10). While the HES encompasses both teaching and research, the focus of this study has been on research evaluation.

In the early 19th century, Germany was the first country to add research to the responsibilities of the universities. After 1862, the United States and, several decades later, Japan adopted the German model, but they focused more on modernization and development. This meant a combination of emphasis on research and science with the central role of the state in supporting higher education. Thus, the American model of public universities arose, which ushered the idea of public service and applied technology, besides democratization of science with a more participative departmental structure inside the universities. Therefore, variations in the concepts of research universities from Germany, the United States and Japan characterize the current research of universities around the world (ALTBACH, 2016, p. 175). In Britain, for example, research was perceived as having a symbiotic relationship with teaching. In this regard, academics should have been committed to expanding knowledge, applying the same logic to their departments and universities (TAPPER; SALTER, 2003).

In general, research universities are more expensive to operate, require more funds than other academic institutions and, with a few exceptions, are government-funded. Over time, the resources allocated to research have become more limited in relation to demand; therefore, more accountability was required from these universities, which led to the implementation of elaborated research evaluation systems in many countries (GEUNA; MARTIN, 2003). Some key objectives of these national exercises are to guide the public funding allocation based on merit, to stimulate research productivity, to support formulation of research policy and management strategies at a governmental and institutional level, to provide information on effectiveness of research management and delivery of public benefits (SCHOTTEN; EL AISATI, 2014; ABRAMO; D'ANGELO, 2015).

Whitley and Gläser (2007, p. 6) defined such systems as "organized sets of procedures for assessing the merits of research undertaken in publicly funded organizations that are implemented on a regular basis, usually by state or state-delegated agencies." Four typical outputs have been measured: volume, quality, impact, and utility. In general, these evaluations display distinct targets, such as individual researchers, groups of researchers, whole institutions, research projects, groups of projects "wrapped" in a program, research support policies, or research system as a whole (GEUNA; MARTIN, 2003).

Research evaluation systems present wide variations among countries. As to structure and governance, they differ in frequency, formalization, standardization, and transparency. Their frequency ranges from one to six or seven years, albeit not necessarily conducted on a regular schedule. Informal systems are characterized by evaluations carried out at university or department levels, while the formal ones are organized by central agencies that apply systematic rules and procedures (WHITLEY; GLÄSER, 2007; HICKS, 2012). Considering the standardization of procedures and practices, it varies significantly between fields and review panels within systems. These variations are expected when peer review judgements of quality are being made in very contrasting enquiry fields (LANGFELDT, 2001; WHITLEY; GLÄSER, 2007). Peer review and bibliometric measures are the methods mainly applied in these systems, although peer review is still the most common. When peer review is supplemented with publication and citation data or other information, the method is called 'informed peer review' (GEUNA; MARTIN, 2003). Thus, the evaluation tends to rely strongly on academic peer-review in some countries such as the United Kingdom, Italy and Portugal, while in others such as Flanders (Belgium), Denmark, Finland, Norway, and Sweden, it relies more heavily on metrics, including bibliometric approaches (MARQUES *et al*., 2017).

In less transparent systems, the evaluations are done informally by small groups of colleagues, who decide their own working procedures, but the results are not publicly available. In highly transparent systems, however, there are formal procedures to select panel members according to their reputation and expertise, as well as to make results publicly available (WHITLEY; GLÄSER, 2007; MARQUES *et al*., 2017). Another important variation among these systems concerns to their link or not to research funding allocation. Thus, some evaluation systems have not so far been directly linked to funding decisions, while in others they have had significant direct effects on resource allocation and on the proportion of employers' incomes (WHITLEY; GLÄSER, 2007, p. 8). Research evaluation systems that associate funding mechanisms with performance were developed in order to make scientific production more accountable and performance-oriented and to promote greater alignment of research with societal and economic needs. Roberts (2006) and Hicks (2012) called these systems PRFS, which seem to increase gradually. They also differ from other models of research evaluation systems, in which research funding is mostly non-competitive and of competitive project funding, for which funding allocations rely on ex ante assessment.

2.1 CHARACTERISTICS OF PERFORMANCE-BASED RESEARCH FUNDING SYSTEMS (PRFSs)

According to the definition developed by Hicks (2012), PRFSs have the following common criteria: research evaluation has to be ex post; research output has to be evaluated; government distribution of research funding has to be linked or will eventually be linked to the evaluation results; and the system has to be nationwide. However, not all the research evaluation systems are considered PRFS. In the Netherlands, for example, evaluations are carried out by the Association of the Netherlands Universities (VSNU), which uses evaluation as a management tool and not as a method of allocating funds, besides generating a relative reputation competition (GEUNA; MARTIN, 2003; GEUNA; PIOLATTO, 2016).

Distinct reasons can be related to the implementation of PRFS, such as resource concentration, international publication promotion, and general pursuit of excellence (HICKS, 2012). A key characteristic of this system is the percentage of funding associated with this evaluation. Regarding the methods, it depends on the target. Peer review and judgement based on the indicators are used for individual and department evaluations, while quantitative formulas are used for university-level evaluations. Departments or universities are usually the main targets of

PRFS (OECD, 2010). On the other hand, Hicks (2012) states that research groups are the evaluation unit with the best theoretical support, because research is conducted by groups, not by individuals or departments. Although the Netherlands' research evaluation system is not considered a PRFS, it represents an example in which evaluation is done in the level of research group or team. Moreover, their groups analyze not only the past but also the future. For that, three points call our attention: 'viability', research quality, and societal relevance. Each group is asked to provide a narrative around their plans[2].

Funding formulae generally are used to allocate funds at university-level. Overall, these formulae consider bibliometric information of output, in which the papers can be taken into account alone or based on citation information. Other variables may also be included in these formulae, such as education, socioeconomic impact, diversity-related assessments, employment of graduates; external research funding; faculty characteristics and qualifications; faculty size; graduated students; research implementation/application; international memberships; and student load (OECD, 2010; HICKS, 2012).

Although peer review presents high esteem in these kinds of systems, it is expensive and time-consuming, hence it is not commonly used. However, the academic community does not approve bibliometric methods and, therefore, they are very criticized. In order to improve these bibliometric methods, some PRFSs established weighted categories of journals, assigning more points to the journal in the top 10–20%. In addition, bibliometric indicators are in general calculated based on databases such as Scopus and WoS, which are considered inadequate for the SSH fields. Hence, pressures especially on those fields rise for more fair evaluations across heterogeneous academic disciplines (HICKS, 2012).

The design of these systems considers data input from universities as well as consulting processes, in which expert panels are chosen among university representatives or field-based associations. Furthermore, many of these systems evolved from studies regarding their effects that counted with extensive formal consultations, usually researchers and universities. Transparency is another key element of PRFS. Thus, information about methods and data, such as instructions to universities concerning their submissions, formulas used to convert measures into final rankings,

[2] Available at: https://thebibliomagician.wordpress.com/2018/05/31/research-evaluation-things-we-can-learn-from-the-dutch/. Accessed on: July 20, 2019.

grades or weights as well as the final grades, are usually publicly available on the government website.

In addition, these evaluations are permanent, routinized, extended across time and space, and further defined as systematic. This kind of evaluation has stronger constitutive effects, influencing on the practices and meaning of the activity under evaluation (HAMMARFELT; DE RIJCKE, 2015). In this regard, the more complicated a system becomes, the more its indicators and metrics can be gamed (RIJCKE, *et al.*, 2016). Hence, establishing a model of the research evaluation system is a hard task, leading some authors such as those of “Leiden Manifesto” (HICKS *et al.,* 2015) to suggest that the indicators should be regularly scrutinized and updated.

Although the use of such research evaluation system has been spreading around the world, national research policy frameworks differ widely. In some countries, the distribution of block grants has been replaced by funding allocation based on these systems with the application of quantitative formulae involving outputs, research students (studentships), external funding (ROBERTS, 2006), and more recently the attempted assessment of research ‘impact’ (WATERMEYER, 2014).

2.1.1 Models of Performance-based Research Funding Systems in distinct countries

Distinct national models for evaluating the educational research in higher education have evolved in different countries. The goals of these evaluations tend to be defined by the evaluating agency according to their priorities. The first and most highly institutionalized and developed PRFS worldwide is the Research Excellence Framework (REF) from the United Kingdom. Established since 1986, it remains as an influential model to other countries, formalizing their own research evaluation exercises and thereby the allocation of research funding to higher education institutions (WATERMEYER, 2014). Its initial goal was improving selectivity in funding allocation, once there were limited resources and increased costs of research (HICKS, 2012). The UK funding structure is characterized by a “dual-support” composed of (1) funding in the form of a ‘block grant’ by the Funding Councils, allowing universities to fund infrastructural investments and support long-term, open-ended research strategies and (2) funding for clearly defined, time-bounded specific research initiatives (projects, centers, among others) by the Research Councils. The 'block grant' funding is distributed according to a formula approach that allocates money to universities, in which the department is the main evaluand. The financial support for projects uses as instruments individual doctoral grants, funding for specific research projects, programs bringing together several related projects, and multimillion-pound, multiyear research centers. The units that are being evaluated are the last activities, which differ in size and scope (MOLAS-GALLART, 2012).

In regard to the REF process, the assessment is based on the disciplinary area. It is therefore a unified framework for all subjects, which evaluates the outputs of individuals using expert peer review as well as nonacademic experts. Bibliometric data are used as support where this is deemed appropriate (MARTIN-SARDESAI *et al*., 2017). The periodicity varied from three to seven years (WHITLEY; GLÄSER, 2007, p. 6). The last REF was in 2014 covering the period from 2008 to 2013, in which 36 expert sub-panels, called Units of Assessment (UoA), reviewed the submissions, which were overseen by four main panels comprised of academic members and research users. The research of 154 UK universities was assessed, and impact scores were used for the first time. According to the quality of research, the submissions were classified as four stars (world leading), three stars (internationally excellent), two stars (recognized internationally), one star (recognized nationally) and unclassified. Furthermore, the impact of research, or even the economic and societal impact(s) of research, was introduced (WILSDON *et al*., 2015). In brief, bibliometrics plays a smaller role in the UK system and it counts with a strong and long tradition of PRFS allocation based on an elaborate peer review system.

In 1989, the research evaluation system in Spain was institutionalized through the creation of the Spanish Commission for Research Evaluation (CNEAI). The purpose was assessing the scientific production of university professors and researchers from the Higher Council for Scientific Research. The Spanish system has been described as one in which individual evaluations remain prevalent, and it is more important than organizational evaluations (CRUZ-CASTRO; SANZ-MENÉNDEZ, 2007). Funding is channeled through the salaries paid to tenured academics working in public universities and several public research establishments, such as Spanish Council for Scientific Research (*Consejo Superior de Investigaciones Científicas* [CSIC]). This process analyses the individual research output over a six-year term ("*sexenio*"), in which each applicant highlights five research contributions. It is a peer review system organized in 11 commissions that take Thomson-Reuters IF into account in their deliberations. The "*sexenio*" has been considered a basic assessment of quality and a requirement for individual promotion and participation in selection committees. Moreover, the projects, which consume sizable resources, are important evaluands in the Spanish research evaluation system (MOLAS-GALLART, 2012).

In the mid-1990s, the CNEAI explicitly recommended publishing in journals included in the Journal Citation Reports (JCR), especially with great positions in the rankings by IF, which were considered of recognized international prestige. In addition, each field has defined the requirements to obtain a positive evaluation. The CNEAI publishes results by category, discipline, and institution. Although it is not published by individual level, the institutions are aware of the individual performance, considering that the salary bonus will be based on that (OSUNA; CRUZ-CASTRO; SANZ-MENÉNDEZ, 2011).

In Australia, the government also supports university research through a dual funding system of competitive and research block grants. Competitive grant funding is awarded to universities to undertake specific research projects. Research funding is given to successful applicants following a merit-based expert peer review process. A ‘Relative Funding Model’ based on quantitative formula was implemented in 1990 to allocate block grants to universities. This model was succeeded by the Research Quantum (RQ) exercise, which allocated 5% of total operating grant funding based on performance indicators (GEUNA; MARTIN, 2003; MEEK; SUWANWELA, 2007, p. 51). The formula was initially based on external earnings, and student and publication components were later added. Publication counts, higher degree loads, and completions were also included in 1995. Each component of this formula received a weight. In 2003, the amount of funds allocated based on this formula accounted for more than half of the funding specifically targeted to research and research training (BUTLER, 2003a, 2003b).

Various evaluation schemes replaced the RQ, and they were all also based on a quantitative formula. The last one is named Excellence in Research for Australia (ERA), which is administered by the Australian Research Council (ARC), although it was developed by ARC in conjunction with the Department of Industry, Innovation and Science. The first full round of ERA occurred in 2010 and subsequent rounds of ERA were in 2012, 2015 and 2018. In this evaluation scheme, individual research outputs by academic staff and university affiliates are assessed, including published papers, authored and edited books, book chapters, conference papers, and creative works. Data are collected from individual researchers on their research activity aligned to eight discipline clusters or ‘Fields of Research’ with subsidiary ‘Units of Evaluation’. The performance is evaluated within each discipline at each university by Research Evaluation Committees (REC), comprising

Australian and international researchers[3]. Each field has the option of using peer review or metrics, such as citation information. The result is a five-level rating scale in each discipline that is reported publicly by institution and academic discipline. Initially, ERA incorporated a system of ranking journals in each discipline that was hotly contested and debated. The journal ranking was removed in 2012, however, ERA published an "acceptable journal" and lists of "publishers" (HASLAM; KOVAL, 2010; MARTIN-SARDESAI *et al*., 2017).

In 2004, Norway joined this international trend of linking performance to basic funds by introducing a Norwegian Publication Indicator, which is a system for documenting Norwegian academic publishing with the aim of measuring publication activity and allocating research funding according to publishing performance (SCHNEIDER, 2009). Only 2% of total funding for universities and university colleges come from this funding system, thus playing a marginal economic role. Some aspects were observed in the development of this indicator to avoid adverse effects already observed and described in the Australian model by Buttler (2003a, 2003b, 2004), including the shift in publication towards outlets with high acceptance rates and lower impact. Therefore, the Norwegian Publication Indicator was designed to increase research publication without a decline in impact. A slightly more sophisticated model was then developed, in which publication channels were classified in two levels. Prior to classification, some scholarly eligibility criteria were established, such as a standard external peer review process. As a rule, Level 1 comprises all channels that attend to these eligibility criteria, which can be described as 'scientific' or 'academic'. Level 2 comprehends an exclusive number of publication channels, in which each subject area is considered the most important, and preferably with an international audience. Level 2 channels constitute at most 20% of the scientific publications of a subject area total (SCHNEIDER, 2009; SIVERTSEN, 2010; AAGAARD; BLOCH; SCHNEIDER, 2015; BLOCH; SCHNEIDER, 2016).

The weighting procedure considers both level and publication forms. A Level 1 journal article yields one point, and a Level 2 article yields three points, while Levels 1 and 2 books yield 5 and 8 points, respectively. This piece of evidence is an asymmetry in the relation between these two levels for books compared to journal articles. Publication points for individual authors are based on fractional counts (*e.g.* for a Level 1 journal article with two authors, each author

---

[3] Available at: https://www.arc.gov.au/excellence-research-australia. Accessed on: July 15, 2019.

contribution counts 0.5 points). There is no limit for the fractioning. Moreover, institutions and individual researchers can nominate channels to compose each list level, but the proposals are discussed in the appropriate committee for the subject area and approved (or rejected) by a publishing committee at the Norwegian Association of Higher Education Institutions (UHR) (SCHNEIDER, 2009; SIVERTSEN, 2010; AAGAARD; BLOCH; SCHNEIDER, 2015; BLOCH; SCHNEIDER, 2016).

The Italian Department of Education and Research created specific agencies to manage evaluation processes in 2000. The first one was the Committee for the Evaluation of Research (CIVR) that performed the first evaluation exercise, the VTR 2001-2003 (*Valutazione triennale della ricerca*, Triennial Research Assessment) in 2004. VTR evaluation was originated in the European experience. It used a pure peer-review approach of a limited portion of the publications produced by researchers affiliated to universities and other research institutions (ABRAMO; D'ANGELO; DI COSTA, 2011). Thus, elected panels in each disciplinary area were in charge of evaluating research outputs in Italian universities and state-funded research agencies. For the next evaluations, the agency was encouraged to revise assessment criteria for research and to pay greater attention to the internationalization of publications, particularly in the field of social studies (MINELLI; REBORA; TURRI, 2008b; REBORA; TURRI, 2013).

The next agency is the National Agency for the Evaluation of the University and Research Systems (ANVUR), which succeeded CIVR and it is yet in operation. It carried out the second national exercise, the VQR 2004-2010 (Research Quality Assessment) in 2011, about seven years after the first one, comprising the period from 2004 to 2010. The VQR 2004-2010 was a hybrid type of evaluation exercise, based primarily on bibliometric analysis for the so-called bibliometric areas (*i.e.* hard sciences) and on peer review for the so-called non-bibliometric ones (*i.e*. SSH). In addition, unlike the previous exercise, the results determined allocation of an important financing share for individual institutions (ANVUR, 2011; ANCAIANI *et al*., 2015).

ANVUR also conducted the third national exercise, the VQR 2011-2014, which once again determined funding allocation, and the architecture was rather similar to the previous one. The most noticeable difference between them was the new criterion for determining the merit class of the examined papers, which was detailed on a Scientometrics special issue (ANFOSSI *et al.*, 2016). This agency is responsible for updating the evaluation criteria. Moreover, it collects and analyses data from participating institutions, assigned scores and published results. Considering the evaluation procedures, each research affiliated to the institutions submitted three papers (chosen among journal articles, books, book chapters, conference proceedings, etc.) that were published during the evaluated period. The process was conducted by 14 Groups of Evaluation Experts (GEV), one for each research area, coordinated by ANVUR. The research papers were evaluated using a combination of bibliometric analysis and peer review, in proportions that varied across research areas following the legal constraint that, overall, at least half of the papers were to be assigned to peer review. Based on this evaluation results, public funds for research were distributed to publicly funded institutions. This allocated funding based on Italy's PRFS rose from 7% in 2009 to 13.5% in 2013, with a further increase in the following years. The introduction of this system produced a great debate about its consequences. In 2015, a "stability law" was established on budget allocation, in which 18% of annual funding would be allocated to better performing institutions. The criteria were 70% based on the VQR results, 20% based on the scientific production of professors promoted or recruited in the period under assessment, and 10% based on international teaching activities (ANVUR, 2011; ANCAIANI *et al*., 2015; GALIMBERTI; MORNATI, 2017).

In regard to some Latin America countries, since the 1990s, Argentina has complex evaluation policies and practices due to negotiation strategies between the State and the universities. Thus, the evaluation exercises take place in the universities, as a condition for obtaining government funds, and they are done by peer evaluators (ARAUJO, 2014). In Mexico, the evaluation is made by peers taking the regulatory system, the researchers' academic and institutional backgrounds, and their scientific and technological outputs into account (ALPERIN; FISCHMAN, 2015).

2.1.2 Potential effects of Performance-based Research Funding Systems

As described in the previous paragraphs, many countries implemented research evaluation systems making use of quantitative indicators. Nevertheless, Rijcke *et al.* (2016) in their literature review state that studies about the possible effects of these systems on knowledge production are still very limited. As Butler *et al.* (2010) notes: "Assessing the impact of PRFS is a fraught exercise, which perhaps explains the paucity of broad authoritative texts on the subject." In addition, literature about this subject is spread in several relatively inaccessible outlets, sometimes in their national language and in different fields of social sciences, which make this understanding more challenging (RIJCKE *et al*., 2016).

Some studies in the literature indicate that PRFS lead to changes in publication practices, as responses of scientists to the evaluation criteria. The most recognized ones are Linda Butler's works on the potential effects of the Australian PRFS (BUTLER, 2002, 2003a, 2003b, 2004). According to her, in the mid-1990s, Australia introduced quantitative formulas for distributing research funds to universities, which comprised three elements: research income, graduate students, and publications. The last element was incorporated into the funding formulas in 1995. Thenceforth, the universities began to distribute these funds internally using the same formulas, although giving more weight to publications. The research publications collection of the Research Quantum (RQ) exercise used to be externally audited, thus universities had to prove, among other things, that the journals with the articles they were claiming were peer reviewed. A journal that was indexed by ISI was accepted as peer reviewed without question. Hence, publishing in ISI-indexed journals was the easiest course of action to take what increased the importance of ISI-indexed journal publications. In addition, the publication indicator employed in those formulas did not differentiate between the quality, visibility or impact of the publications, therefore giving little incentive to the effort of publishing in a prestigious journal (BUTLER, 2002, 2003a).

After analyzing the share of Australian publications in the WoS from 1981 to 1999, Butler observed a rapid increase in the number of publications indexed in that database but highlighted that after 1993 the largest increase was in lower-impact journals. In sum, the Australian model led to a considerable shift in publication toward outlets with high acceptance rates and lower impact (BUTLER, 2003a, 2003b). Over the years, this system promoted a general decrease in overall citation impact for Australia publications. The formulas, and in particular the publications component, were conceived by the government as a means of distributing research funds on the basis of the quality of research in Australian universities. On the other hand, publication counts are not measures of quality. Therefore, this model effect was the opposite of the intended one, which was to reward 'quality' and not quantity (BUTLER, 2004). The term "quality" is put in brackets as the studies in general only consider measures of citation impact.

In Spain, the research incentive system used to reward researchers with salary bonuses for publishing in prestigious journals, mainly on the top one third position in ISI's Journal Citation Report lists by subject category. A study carried out in Spain about scientific production for 25 years observed a remarkable increase in productivity as measured by the number of items recorded in international databases. It argued that several different causes have successively influenced this research productivity increase over time. The first cause was a change in publication behavior, thus Spanish scientists began to have more contact with their peers in the international scientific community. The second cause was an improvement in resource types, besides an increase in the mobility of Spanish researchers within the continent facilitated by the entrance of Spain in the European Economic Community. The third cause was the creation of the CNEAI, which was developed for evaluating individual research activity. This commission stimulated the publications in international journals indexed at WoS. Since then, there was a growth in production rates. This last fact showed that the policies used by the CNEAI achieved the results they were designed to bring about, *i.e.* increase productivity and internationalization of Spanish research measured almost exclusively by the number of articles at WoS. (JIMÉNEZ-CONTRERAS; MOYA ANEGÓN; LÓPEZ-CÓZAR, 2003).

In the UK, changes in publication patterns were also found regarding their research evaluation system - The UK Research Assessment Exercise (RAE) - before three different periodic evaluations – 1992, 1996 and 2001. A study demonstrated that the researchers reacted to the evaluation criteria, thus changing the publication patterns. In 1992, the approach was counting total publications, and the result was that UK scientists substantially increased their article production. In 1996, RAE stimulated more "quality". Therefore, there was a gradual increase in the number of papers in journals with a relatively high citation impact. From 1997 to 2000, the institutions promoted more collaborations among their active research staff, thus producing more intensively, albeit not more publications, which induced once more an increase in quantity but not necessarily in "quality" (MOED, 2008).

All these studies regarding changes of publication patterns experienced by different research evaluation systems show how researchers responds to funding stimuli. Based on that premise, other countries seek to develop their PRFS criteria taking past experiences into account. Norway established its PRFS in 2005, considering a publication point-based performance indicator. A study about the Norway model demonstrated that its publication activity and international visibility have been growing over a long period. Additionally, this increase was considerably higher after the model implementation, but the citation impact has remained quite stable (AAGAARD; BLOCH; SCHNEIDER, 2015; SCHNEIDER; AAGAARD; BLOCH, 2015). However, another study at the individual researcher level found that the average of publication points per researcher has decreased from 2004 to 2012. At the same time, average publication counts and number of coauthors per paper have increased substantially. The Publication Indicator activated a larger share of researchers either to begin publishing on a regular basis or to shift publication activity toward the types of scientific channels covered by the Indicator (BLOCH; SCHNEIDER, 2016).

Based on the Norway optimistic result, many countries in Europe, such as Denmark and Finland, have implemented a similar or equal model. Therefore, Ingwersen and Larsen (2014) investigated the patterns of Danish publications before and after the introduction of the Norwegian publication point-based performance indicator in 2008. The study demonstrated an increase in the research article productivity after the PRFS implementation. However, they found linear progress of the citation impact, which happened independent of the PRFS introduction.

Nevertheless, other researches have recently questioned these casual effects assigned to performance-based evaluation systems. As such, Osuna, Cruz-Casto and Sanz-Menéndez (2011) investigated previous research about the growth of international scientific publications from Spain credited to the establishment of the new evaluation system. They concluded that "the growth in Spanish publications cannot be attributed indisputably to the establishment effect of Research Evaluation Systems, but rather to the increase of expenditure and number of researchers in the Spanish R&D system along with some maturation effects." According to them, the research incentive system role is minor and indirect, and it is not the primary explanation for the overall growth in Spanish publication output. Furthermore, they call attention to the applied methodologies in this kind of study as well as the need for being more cautious in cause and effect studies at the national level.

Along the same lines, Van Den Besselaar, Heyman and Sandstrom (2017) redid and extended the Butler (2003a, 2003b) analyses. They have stated that her conclusions were incorrect. As reported by them, there was an improvement in Australia's research output, besides a significant increase in its "quality" after the PRFS implementation. Considering that Butler's studies have influenced both policy discussions and designs of PRFS around the world, Van Den Besselaar, Heyman and Sandstrom (2017) work called the attention of international experts, who discussed this subject deeply from different angles in a specific volume of Journal of Informetrics. Sivertsen and Aagaard (2017) summarized the main aspects in that long and deep discussion. As reported by them, a key issue in such discussion is to what extent changes in research behavior can be attributed to a specific policy mechanism that is extremely complex, involving different potential factors, in which disciplinary cultures interact with both local, national and international incentive structures. In addition, they have stated that there is not a specific methodological design to address all the challenges found in this kind of study that can solve all fragilities. In conclusion, Butler's, on one hand, stands out as the strongest design and the most convincing in-depth contextual knowledge. On the other hand, the claims of Sandström and Van Den Besselaar are based on studies with methodological or conceptual limitations. Thus, their discussions show limited support among international experts (SIVERTSEN; AAGAARD, 2017).

In addition to modifying publication patterns, these exercises have been reported as fostering a more strategic evaluation of academics' careers and for creating pressure for higher productivity (BUTLER *et al.*, 2010; HICKS, 2012), significantly limiting researchers' autonomy (ELTON, 2000; MARTIN; WHITLEY, 2010; TAPPER; SALTER, 2003). Many academics view these research assessment exercises as a major source of anxiety and uncertainty (MCCARTHY; SONG; JAYASURIYA, 2017; MARTIN; WHITLEY, 2010; YOKOYAMA, 2006), as they are put under pressure not only 'to lift their publication output', but also 'to tailor them to fit the most valued types of publications' (PARKER, 2008, p. 383).

Other factors than money play a role in the effect degree of PRFS, such as Butler's, highlighting the researcher's reputation (BUTLER *et al.*, 2010). The scientific fields and academic career stages of scientists may also mediate the institutions and researchers' responses degree to the implementation of these systems (LAUDEL, 2006; WHITLEY, 2003). These systems are also known to increase the pressure on scientists to publish and reinforce the "publish or perish" culture among the scientific community, which leads to "inflation of publications", but without necessarily improving their quality, decreasing returns in the long run and also resulting in "salami-slicing" effects (GEUNA; MARTIN, 2003; LIEFNER, 2003; QIU, 2010).

In short, PRFS are dynamic systems and they can be gamed. Thus, distinct weaknesses or distortions of performance measures are usually exploited to learn how to game it. A study states that an equilibrium solution would be a periodic revision of performance measures or reassignment of agents. However, if both principals and agents are learning over time, the dynamic is likely to become more complex (HEINRICH; MARSCHKE, 2010, p. 203). Additionally, many countries have adopted this kind of system with little addressing the political objective. In Brazil's case, a new method, based on the last one, was implemented in 2018 without empirical knowledge or evidence about the aggregated potential or individual effects of the previous model. Studying this system will allow us to identify its fragilities and thus to contribute to the design of a new one.

2.2 PEER REVIEW AND BIBLIOMETRIC INDICATORS AS EVALUATION TOOLS

Peer review has been the primary method for research evaluation exercises (ABRAMO; D'ANGELO; DI COSTA, 2011), as well as for quality control in sciences, social sciences, arts and humanities (BORNMANN, 2011a; LEE *et al*., 2013). According to Lee *et al.* (2013, p. 2), the basic principle of this method is: "experts in a given domain appraise the professional performance, creativity, or quality of scientific work produced by others in their field or area of competence*.*" As reported by Geisler and Abdallah (2000, p. 219), "this process represents the ultimate power exercised by experts who police themselves and who evaluate each other."

Peer review was designed to encourage peer impartiality by involving the use of a "third party", *i.e*. someone neither affiliated directly with the reviewing entity (university, research council, academic journal, etc.) nor too closely associated with the person, unit, or institution being reviewed (LEE *et al*., 2013). In general, critiques of this method arise from violations of that impartiality, promoting bias in the allocation of resources (DAY, 2015; GALLO; SULLIVAN; GLISSON, 2016).

Regarding national research evaluation exercises, research products submitted by institutions are evaluated by appointed panels of experts. This peer review approach shows some common limitations described in literature, as follows:

- it provides the most significant weight to research output with more quality (ABRAMO; D'ANGELO, 2011);
- it is time-consuming and costly (ABRAMO; D'ANGELO, 2011);
- it is subject to many biases and distortions (HORROBIN, 1990; DAY, 2015; GALLO; SULLIVAN; GLISSON, 2016);
- evaluation parameters may be predefined, but each panel member uses their criteria to mark them (BORNMANN, 2011b);
- it is criticized for taking a conservative view and not being receptive to new ideas (BORNMANN, 2011b);

- disagreement among peers is known to be common and wide (JAYASINGHE; MARSH; BOND, 2001; MARSH; JAYASINGHE; BOND, 2008).

All these limitations led many governments to seek for relatively quick, easy and inexpensive alternatives. In this scenario, bibliometric indicators appeared to be an interesting alternative, being applied on their own or in combination with peer review. Bibliometric techniques are used to quantify and measure the performance of books, articles, and other media of communication. Thus, bibliometric measures, such as the number of publications and citations, are widely used as performance indicators in research policy and within the research system (LANE; LARGENT; ROSEN, 2014, chapter 21). While publication counts serve as an indicator of the amount of new scientific knowledge produced by researchers, the citations received by scientific publications are used as indicators of scientific impact (RUSSELL; ROUSSEAU, 2010). Thereby, these indicators are often used in funding decisions, appointments and promotions of researchers, with substantial implications in the science-policy arena.

The advantages of using bibliometric indicators include providing more objective information about the scientific performance, allowing the assessment of many documents, and most of the indicators are easily interpretable for specialists. Considering their disadvantages or other limitations, literature shows their lower acceptance in arts and humanities and social sciences due to inadequate journal coverage in these areas in the databases. A far distance between bibliometric methods and research practices in the disciplines has also been seen. There has also been a limitation on the usual citation normalization process, especially for interdisciplinary research, since it can be challenging to define the field (MOED, 2005; ELSEVIER, 2013).

In sum, the two methods offer different points of view on a common problem. Therefore, some studies emphasize that they should be considered complementary and, wherever possible, used concurrently, especially in small scale evaluations. Thus, more comprehensive and reliable bibliometric data should be used to guide and support peer decisions as to budget allocations and to the definition of research agendas and strategic goals (RUSSELL; ROUSSEAU, 2010). The metrics should also be combined in order to provide a multidimensional view of the research (CRONIN; SUGIMOTO, 2014, p. 386).

However, literature shows a proliferation of metrics, which leads to stronger and more frequent reactions regarding their use, especially in the evaluation of a researcher's performance and output. As a result, there have been many calls for something better to improve the evaluation of research outputs and cease the strengthening of metrics, such as the IF. These calls resulted in some documents such as the San Francisco Declaration on Research Assessment (DORA, 2012), the Leiden Manifesto (HICKS *et al*., 2015), and the Metrics Tide (WILSDON *et al*., 2015). In summary, these documents provide higher weight to peer reviews and draw attention to the indiscriminate use of these indicators due to their limitations. Although these documents claim for changes, Curry (2018) states no solution has been found yet.

2.2.1 Citation impact indicators

The citation idea works as a fundamental indicator of impact, albeit not of quality, which is an issue of considerable debate (MINGERS; LEYDESDORFF, 2015b). Bibliometricians usually comprehend the citation rate as a proxy measure of scientific impact or of impact on the relevant scientific communities. In most competitive areas, and somewhat more slowly in SSH, citation-based indicators have been incorporated in the daily routines of virtually all research groups. They provide information on an individual level or research units such as researchers, research groups, research institutions, countries as well as journals, which are the most frequently studied unit. The leading international bibliographic databases used to calculate these indicators are WoS, Scopus, and Google Scholar (GS). Thus, the indicators have different values depending on the research unit, chosen database, and time period within which publications or citations are counted (WALTMAN, 2016).

Wouters *et al.* (2015, p.9-10) divide the basic citation impact indicators into size-dependent and size-independent. The size-dependent indicators are the total number of citations, number of highly cited publications and h-index, while the size-independent include the average number of citations per paper and the proportion of highly cited publications. Each of these indicators are discussed by the same authors as follows:

> - Total number of citations
> - Average number of citations per publication. The best-known indicator based on the idea of counting the average number of citations per publication is the journal IF, which counts the average number of citations received by the publications in a journal. Indicators based on average citation counts are frequently used, but they are also criticized in the literature. Citation distributions

> tend to be highly skewed (e.g. ALBARRÁN *et al*., 2011; SEGLEN, 1992), and therefore the average number of citations in a set of publications may be strongly influenced by one or a few highly cited publications. This is observed by Aksnes and Sivertsen (2004) at the level of countries and by WALTMAN *et al.* (2012a) at the level of universities. Due to the skewness of citation distributions, suggestions are often made to replace or complement indicators based on average citation counts by alternative indicators (e.g. AKSNES; SIVERTSEN, 2004; BORNMANN; MUTZ, 2011; LEYDESDORFF; OPTHOF, 2011; WALTMAN *et al*., 2012a). Indicators based on the idea of counting highly cited publications are a frequently suggested alternative.
>
> - Number of highly cited publications, in which a certain threshold needs to be chosen to determine whether a publication is counted as highly cited or not. The idea of counting highly cited publications has been suggested by Martin and Irvine (1983), Plomp (1990, 1994), and Tijssen et al. (2002). The i10-index reported by GS is based on the idea of counting highly cited publications.
> - Proportion of highly cited publications.
> - h-index (or Hirsch index) is defined as: a research unit has an h-index if each of its publications have at least h citations and the other publications each have no more than h citations. The h-index was introduced in 2005 (HIRSCH, 2005) and has quickly become very popular. A large number of variants and extensions of the h-index have been proposed in the literature, of which the g-index (EGGHE, 2006) is probably the one that is best known. Some counterintuitive properties of the h-index are highlighted by Waltman and Van Eck (2012a).

Furthermore, Wouters *et al.* (2015) also emphasize that new indicators have been proposed in the literature, but all of them is understood as variants or extensions of these basic indicators. In addition, another study classifies the citation indicators into the first generation, which is composed of the basic indicators and second and third generations (CRONIN; SUGIMOTO, 2014, p. 386). The second generation is characterized by relative or normalized indicators, with a correction for biases (*e.g.* differences in citation practices between subject fields). The third generation of indicators is based on advanced network analysis using parameters such as network centrality.

As aforementioned, journals are the most frequent research units in studies of citation impact indicators. The focus on quality of research has led to an interest on the quality of the publishing journal itself. In large-scale evaluations, such as REF from the UK, in which a vast number of papers need to be graded, the practice of judging an article by the journal in which it has been published became endemic. This fact is also true in Brazil. Thus, the next section focuses specifically on some citation impact indicators for journals.

2.2.1.1 Citation impact indicators for journals

The IF was the first basic journal metric, besides the best-known indicator of journals' citation impact (GARFIELD, 1972). Its initial purpose was to help researchers search in literature and not to evaluate research. In addition, its use rapidly expanded from journals to the evaluation of individuals, groups, units, universities, institutions, fields, countries, geopolitical, regions, etc. The IF is published every year for all journals included in Thompson Reuters' WoS. It is based on the number of citations in a year to papers published in the previous two years.

In addition, the IF indicator causes a lot of debate in the literature. In general, it is not about the indicator itself but more on the use of this metric for assessing individual publications based on the journal in which they have appeared (WALTMAN, 2016; MINGERS; YANG, 2017). Many studies have therefore highlighted its limitations (GLÄNZEL; MOED, 2002; HARZING; VAN DER WAL, 2009):

- IF depends heavily on the research field, because there are significant differences among fields in citation density.
- It is calculated based on a two-year window, which is a short period for many disciplines. There is also a lead time between submitting a paper and having it published, which may be two years. The five-year IF addresses this criticism and it is superior regarding it. An empirical comparison between the two-year and the five-year IF has been presented by Campanario (2011).
- There is a lack of transparency on how it is calculated, which casts doubt on the results.
- It can be deliberately distorted by "gaming of the metric." Some examples are the practice of publishing many review articles that are more highly cited; publishing short reports or book reviews that are cited but are not included in the count of papers; publishing yearly overviews of the research published in the journal or pressuring authors to gratuitously reference excessive papers from the journal (LOWRY *et al*., 2013; MOED, 2000; WILHITE; FONG, 2012).

Considering these limitations and others related to the use of such metric in individual performance evaluation, the use of IF in research evaluation contexts has been heavily criticized (SEGLEN, 1989; AMIN; MABE, 2003; DORA, 2012; HICKS *et al.*, 2015).

Likewise, another well-known basic citation impact indicator is the h-index. H-index and IF are widely seen as the two most popular bibliometric indicators. The h-index can be used for journals, individual researchers, or departments. The novelty about h-index is that it summarizes both impacts, in terms of citations, and productivity in terms of number of papers. Moreover, WoS, Scopus or GS routinely calculate it. Some limitations of h-index were identified, and a range of modifications have been suggested in literature (MINGERS; YANG, 2017).

These basic citation indicators for the journals discussed before do not correct the differences in citation density among fields. Thus, a fundamental principle of citation analysis is that citation counts of publications from different areas should not directly be compared since they have different citation patterns. In general, sciences have a much greater citation density than social sciences or humanities. For example, a study showed that molecular biology presented citation rates ten times greater than computer sciences. These differences were also observed within a multidisciplinary field, such as business and management. Studies have proposed different approaches to normalize citation impact indicators for field differences, and some of them were based on average citations and others on highly cited publications (IGLESIAS; PECHARROMÁN, 2007; MINGERS; LEYDESDORFF, 2015a; WOUTERS *et al*., 2015).

Even considering the same field, one should also be careful to contrast papers from different years, because an article published firstly had more years to attract citations. Another desirable approach is to consider differences in publication type. For instance, journals with a larger number of review papers will be more cited than a journal with regular research articles. Similarly, editorials and book reviews generate citations, but they might not be counted as papers. Regarding differences in citation counts, literature presents some normalizations approaches, such as field, percentile, and source normalizations (WALTMAN, 2016; MINGERS; YANG, 2017). The concept of scientific field is also a key issue to calculate normalized citation impact indicators. Researches have different opinions on how this concept should be undertaken, for instance, through predefined database fields, disciplinary classification systems, or sophisticated computer algorithms to define fields or with citing-side normalization approaches that do not establish fields explicitly (WOUTERS *et al*., 2015).

Taking into account concerns about normalizations as well as the relative prestige of citing journals, several new and more complex indicators have been developed and form what Cronin and Sugimoto (2014) called as second and third generations of citation indicators. Some of these indicators are specific to data sources. The most well know ones are the Eigenfactor, which is calculated based on WoS; and SNIP and SJR, both based on Scopus. Eigenfactor and SJR are indicators that measure the prestige of citations, attributing more weight to citations from high-impact sources, while SNIP normalizes differences in citation characteristics among the fields (WOUTERS *et al*., 2015).

The basic idea of Eigenfactor is that a single citation from a high-quality journal may be more valuable than multiple citations from peripheral publications. This indicator measures the importance of a citation by the influence of the citing journal, divided by the total number of citations appearing in that journal. It is considered a measure of centrality, like that used by Google to return search results. When ranking web pages, Google's PageRank algorithm considers not only how many hyperlinks a web page receives, but also from where those hyperlinks come. Eigenfactor algorithm does something similar. Instead of ranking websites, it ranks the journals, and instead of using hyperlinks, it employs citations in the academic literature as tallied by Journal Citation Reports (JCR). This indicator also corrects differences across disciplines and journals in the propensity to cite other papers (BERGSTROM, 2007).

JCR also includes a related metric called the Article Influence Score (AIS), which is the Eigenfactor divided by the proportion of papers in the database belonging to a journal over five years. AIS is similar to five-year JIF; however, unlike this one, the article influence indicator provides more weight to citations from high-impact journals than to those from low-impact ones. Self-citations at the level of journals are also not counted both for AIS and Eigenfactor (MINGERS; YANG, 2017).

The SJR indicators are fairly identical to the article influence indicator, but their value is normalized by the total number of citations in the citing journal for the year in question. The first version of SJR was introduced in 2010 (GONZÁLEZ-PEREIRA; GUERRERO-BOTE; MOYA-ANEGÓN, 2010) and a revised one in 2012 (GUERRERO-BOTE; MOYA-ANEGÓN, 2012), which is currently included at Scopus. This last version feature is that the weight of a citation depends not only on the citation impact of the citing journal, but also on a measure of the thematic closeness of the citing and cited journals. This calculation provides more weight to a citation from a citing journal that is thematically close to the cited journals than a citation from a more distant citing journal (WOUTERS *et al*., 2015; MINGERS; YANG, 2017). Furthermore, some limitations of the SJR and Eigenfactor are pointed out in literature, such as the values for "prestige" as challenges for interpretation, as they are not a mean citation value, but only make sense in comparison with others. They are not normalized for the fields (LANCHO-BARRANTES; GUERRERO-BOTE; MOYA-ANEGÓN, 2010).

Another indicator is the Source Normalized Impact per Paper (SNIP), which uses a source normalization approach to correct for differences in citation practices between scientific fields. It was introduced by Henk Moed in 2010, although in 2012, Waltman *et al.* (2013) provided a modified version, which is the one nowadays reported by Scopus. The SNIP values are calculated by the Centre for Science and Technology Studies (CWTS) of Leiden University. This indicator strength is that it does not require a field classification system, in which the boundaries of fields are explicitly defined. Regarding this indicator calculation, Mingers and Yang (2017, p. 327) reports:

> It firstly calculates a three-year IPP (impact per paper, effectively a three-year JIF). It then calculates the "database citation potential" (DCP) for the particular journal, by finding all the papers in year n that cite papers from the journal in the preceding ten years and calculating the mean of the number of references in those papers to papers within the database, i.e. Scopus. Then, the DCP for all journals in the database is calculated and the median of these values is noted. The DCP for the journal is then divided by the median to relativize it to journals, creating a relative DCP (RDCP). If this value is above 1, then the field has greater citation potential; if it is less than 1, the field has lower citation potential. Finally, SNIP = IPP/RDCP. If the field has high density, RDCP will then be above 1 and the IPP will be reduced or vice versa, if the field has low density. The currently implemented version of SNIP has two changes (WALTMAN *et al*., 2013): the DCP has been calculated using the harmonic mean rather than the arithmetic mean, and the relativization of the DCP is now decreased.

All these citation impact indicators for journals have been often used as a substitute for publication-level citation statistics in evaluation exercises, a practice that has been rejected by many bibliometricians. This repudiation occurs since the distribution of citations over the publications in a journal is highly skewed, which means that any journal-level indicators are not representative of the citation impact of a typical publication in a journal. Moreover, bibliometric indicators may lead the user to valuable insights, but only when there is knowledge of their methodologies and respect for their limitations (WOUTERS *et al*., 2015).

2.2.2 Peer review versus bibliometrics

Correlating bibliometrics with peer review has been a central subject of Bibliometrics since the very beginning. Initially, the studies used to empirically investigate to what extent the number of citations correlates with peer judgment of either the quality or the influence of scientific work. Nowadays, researches are carried out to investigate the validity of peer review rather than bibliometrics. In general, the literature shows the correlation between these two methodologies with varied strengths. This variation also happens between fields or even varies within fields. Additionally, this correlation is weaker in most fields in humanities, applied fields, technical sciences, and social sciences (WOUTERS *et al*., 2015).

In Italy, Abramo and D'Angelo (2011) compared informed peer review with bibliometrics in the national research evaluation system (VTR). They concluded that bibliometrics was superior for the natural and formal sciences across a range of criteria – accuracy, robustness, validity, functionality, time, and cost. However, they recognized that bibliometric indicators are not yet sufficiently robust to inform peer-review in SSH. Abramo, D'Angelo and Di Costa (2011) suggest that bibliometric indicators should integrate or completely substitute the classic peer review in the Italian national research evaluation for hard sciences. Another study also about the VTR, relative to the period of 2004-2010, observed a fair to good agreement between informed peer review and bibliometric analysis and absence of statistical bias between both in Economics, Management, and Statistics (BERTOCCHI *et al*., 2015).

Another study contrasted a range of citation metrics and peer judgment of a researcher's influence on six fields of public health in Australia. A moderate positive correlation was found for four fields, but no relationships or negative relationships were observed for the other two fields. Regarding the latter cases, the authors conclude a peer understanding of research influence within these fields differed from visibility in the mainstream and peer-reviewed scientific literature. Finally, they advised combining both peer review and metrics in research evaluation processes (DERRICK *et al*., 2011).

In Norway, a case study of research groups at the University of Bergen verified the correlation between bibliometric indicators with evaluation ratings provided by peer reviews. All the analyses showed a positive relationship but relatively weak. They have stated that peer ratings cannot be considered standards to which bibliometric indicators should be expected to correspond. Hence, the correlation can be positive if the aspects assessed by peers correspond to those reflected through bibliometric indicators. The explanations for a weak relationship were the shortcomings of both methods and lack of comparability (AKSNES; TAXT, 2004).

Regarding the UK national research evaluation, there have been many studies comparing citation metrics and peer review in the past decades (THOMAS; WATKINS, 1998; NORRIS; OPPENHEIM, 2003, 2010; TAYLOR, 2011; CLERIDES; PASHARDES; POLYCARPOU, 2011; BUTLER; MCALLISTER, 2009). According to Traag *et al.* (2019), most of the studies comparing citation metrics and peer review in the UK found correlations of about 0.7 or higher, on the order of 0.9, although the results vary from field to field. The authors, however, called attention to The Metric Tide report (WILSDON *et al*., 2015), in which the correlations were usually on the order of 0.4, thus showing a lack of agreement between metrics and peer review based on some statistical evidence. This report analyzed the possible role of citation metrics in the latest REF in the UK. Based on that result, they concluded that metrics should support, not supplant, expert judgment. Considering the significant ambiguity in this discussion on the agreement between metrics and peer review, Traag *et al.* (2019) investigated the statistical analysis presented in The Metric Tide report. They stated that the correlations tests in most of these studies are generally basic, leading to problematic interpretations. Therefore, they provide clarity in such discussion.

Following the authors, four critical points should be considered on this kind of analyzes: firstly, the level of aggregation in which individual level constitutes the lowest and research institutions level, the highest. Secondly, the use of either a size-dependent or a size-independent perspective, taking the size of an institution into account or not. The authors considered especially relevant this point when reporting correlations. Thirdly, correlations may not be the most informative measure, thus one should consider the suitability of other measures. Fourth, peer review is subject to uncertainty, which should be taken into consideration in the interpretations. Considering all these points, Traag *et al.* (2019) conclude that particularly in Physics, Clinical Medicine and Public Health, metrics agree quite well with peer review and may offer an alternative to peer review.

In brief, comparisons between peer review and metrics depend on which specific dimensions and forms of peer review are being related to exactly what bibliometric indicators. In the literature there is not a common methodology in this kind of analyzes, resulting in different outcomes with respect to the correlation strength as well as different interpretations.

2.3 SOCIAL SCIENCES AND HUMANITIES IN RESEARCH EVALUATION ENVIRONMENTS

Disciplinary fields have different modes of impacting academic communities, the economy, and society. This statement is especially true for SSH due to their organizational and epistemic characteristics and their diversity on types of outcomes (WHITLEY *et al.*, 2000; BASTOW; DUNLEAVY; TINKLER, 2014). SSH researchers study societal and cultural issues that may have a direct impact on policymakers, managers, people in the legal system, and general public. Hence, this requires specific types of communication that are generally related to national or regional topics and in the national language. Nederhof stated: “Societies differ, and therefore results from humanities or social science studies obtained in one country may not always be very useful to researchers in other countries” (NEDERHOF, 2006, p. 83).

In this regard, some features of SSH make the research evaluation of fields much more challenge. Firstly, SSH traditionally assign large weight to books, book chapters and monographs, which get cited more often than journal articles (HICKS, 2004; NEDERHOF, 2006). Second, some SSH disciplines are characterized by a more pronounced national and regional orientation (NEDERHOF, 2006). Third, SSH targets a more heterogeneous audience, including not only scholarly readers but also a non-scholarly public (HICKS, 2004). The bibliometric approach is consequently limited in these fields, thus some studies showed that these methods cannot readily be used for SSH (HICKS, 2004; LARIVIÈRE; GINGRAS; ARCHAMBAULT, 2006; NEDERHOF, 2006). Although databases such as the WoS and Scopus have made considerable advances in increasing the coverage of archival journals and articles in SSH, it is still very limited (HICKS; WANG, 2011), specially for outputs of non-English-speaking countries (LARIVIÈRE; MACALUSO, 2011). Similar difficulties take place in Computer Sciences and Engineering, considering that many of their publications are in conference proceedings, which are also less covered by those bibliographic databases. Another problem in these fields is that the same work may be published multiple times, both in conference proceedings and on a journal (WOUTERS *et al.*, 2015).

These differences in SSH publication patterns from those observed in scientific, technical, and biomedical fields are well known and discussed in literature. However, bibliometric studies in SSH recently have driven more attention to the topic of internal diversity. Studies at disciplinary levels show that there is an interdisciplinary variety in terms of publication patterns across the spectrum of SSH. One example of patterns is the differences between most disciplines belonging to social sciences and those classified as humanities. Use of international journals, English as a publication language and more frequent co-authorship became more predominant in social sciences, while publishing in books and chapters and use of national or regional languages retained a central position in humanities (ENGELS; OSSENBLOK; SPRUYT, 2012; OSSENBLOK, 2016; PUUSKA, 2014; SIVERTSEN, 2009). Other works have noted that these publication patterns differ between the disciplines of the SSH, while they are similar across countries within the disciplines (VAN LEEUWEN, 2006; SIVERTSEN, 2016a).

Furthermore, SSH disciplines present different citation behavior and culture. Therefore, a citation window above three years is more appropriate for their research assessment. A study suggests using almost a 10-year citation window, which is inappropriate for evaluation purposes since it leads to an obsolete publication set (GLÄNZEL, 1996). In general, SSH journals are more transdisciplinary, which leads to methodological problems such as field normalization (HICKS, 2004).

Although all these limitations regarding research evaluation in SSH have been pointed out, the growing pressure of accountability, the prevailing government practices based on New Public Management and the availability of quantitative data had led many governments to implement bibliometric methods also in the SSH during the last decades (KEKÄLE, 2002; HAMMARFELT; DE RIJCKE, 2015; HAMANN, 2016). Nevertheless, peer review is still the most important evaluation methodology, especially in SSH. Many efforts have been made to make it more sophisticated, methodologically controlled, as well as free from unwanted biases, distortions, and unexpected side effects. These initiatives are around the notion of originality, unorthodox science, or interdisciplinarity (GUETZKOW *et al*., 2004; HAMMARFELT, 2011). Unlike many systems, there are efforts to classify and evaluate non-indexed journals (mainly in national languages), besides the classification of books and publishers (BONACCORSI; CÍCERO; FERRARA, 2015).

In addition, considering that SSH outputs are poorly represented in international databases, several countries have made many efforts to improve coverage, creating national databases and repositories, mainly those with a performance-based funding model, such as Spain, Norway, Denmark, Belgium (Flanders), and Finland (SĪLE *et al*., 2017; GIMÉNEZ-TOLEDO *et al*., 2019). Norway was the first country to establish a national information system in 2005 with complete quality-assured bibliographic data covering all peer-reviewed scholarly publishing in the total higher education sector. The driver behind this system has been the Norway research evaluation system called the "Norwegian model", which requires bibliographic data for publication indicators that serve a performance-based funding formula (SCHNEIDER, 2009; SIVERTSEN, 2010; AHLGREN; COLLIANDER; PERSSON, 2012; AAGAARD; BLOCH; SCHNEIDER, 2015). Those databases are used not only for evaluative purposes, but also to improve information retrieval for scholars and broad access to publications in multiple languages. They also allow to centralize and systematize SSH scholarly and non-scholarly outputs, with acquisition of knowledge about productivity and publication behavior in diverse SSH disciplines.

2.3.1 Internationalization in Social Sciences and Humanities

Internationalization has been addressed by higher education policy as a central characteristic of research activities, affecting all the scientific disciplines with different rate and pace. The drivers of research internationalization include globalization of economies, increased competition for excellent researchers and research funds, and need to improve reputation and visibility at the knowledge frontier (VEUGELERS *et al*., 2005; REALE *et al*., 2012). Concerning internationalization of SSH, it is noteworthy that the SSH disciplines are "international" in their origin. The SSH "nationalization" arose close to the idea of democratization in education, culture, and social life (SIVERTSEN, 2016a).

In general, research evaluations apply two main criteria for "internationalization" and "research quality." One is the publication coverage by commercial indexing databases, such as Scopus and WoS. The other is to publish in journals and English, specifying preferences in terms of language and output type. Thus, some research evaluation or performance-based funding models treat separately the publications indexed in Scopus or WoS as the main criteria in SSH, and others select a limited number of international journals for indexing, thus stimulating more publications on them. As aforementioned, although some improvements have been made in terms of coverage and output diversity in these commercial databases to attend the specificities of SSH, it is far from complete. They have shown no capability of keeping up with the rapid development of new international and specialized journals in these fields (CURRY; LILLIS, 2010, p. 6; SIVERTSEN, 2016a).

Moreover, the literature highlights that SSH show a core-periphery structure in terms of research capacity and publication outlets, and even more so regarding scholarly recognition and prestige (as measured by citations and prizes). This is dominated by North America (United States and Canada) and Europe, which together produce more than 80% of the articles registered in the world's leading Citation Indexes. Overall, the globalization of research has mostly favored the core countries. In contrast, periphery countries showed diminished autonomy and increased dependence on North America and Europe (MOSBAH-NATANSON; GINGRAS 2014). Sivertsen reinforces in his articles that indexing a publication in those commercial databases should not be used as a criterion for research quality or an internationalization indicator in SSH (SIVERTSEN, 2016a; SIVERTSEN, 2019). Additionally, the author states that publication patterns are more deeply rooted in scholarly norms, methods and practices, hence it is not just a question of new trends versus old traditions.

Given the commercial nature of these databases as well as their limitations, some studies seek to understand different internationalization patterns in SSH. A study in Norway of a complete representation of the entire production in SSH, including papers not covered by Scopus or WoS, identified two patterns of internationalization in the SSH. First, there was an increasing multi-authorship and international collaboration in research over the latest years, in contrast to the single-authored articles that traditionally predominated in the SSH. Second, there was a movement of articles from a core of close-at-hand disciplinary journals to fully international journals that represent specialties within disciplines or a cross-disciplinary thematic scope, which was called de-concentration and specialization. At the same time, the author noticed an increase in the concentration of articles in more general disciplinary journals at the national level. As a conclusion, the study states that national and international journals in the SSH have different roles, thus not representing competing alternatives in the publication pattern, but rather supplementary (SIVERTSEN, 2016a).

Regarding the language, Latin was the first of several international languages used in SSH during several centuries. It was the predominant language over the pre-modern period in Europe. Over the years and in every part of the world, transnational connections and collaborations have gained more weight, and English has become the *lingua franca* of international communication. Fluency in English gives a clear advantage to the researchers that are native speakers over non-native ones. The negative consequences of this linguistic inequality are more significant in SSH compared to natural sciences, considering the greater importance of written expression and interpretative analysis as well as less firmly standardized (and universally canonized) conceptual formulations. Hence, in countries that have switched to English, such as the Netherlands, younger scholars who have never written any academic work in their native language seem to have lost all sense of nuance, depth, vividness. The quality of their writing is poor, and the problems arose from this fact are hardly ever acknowledged.

Likewise, a study investigated patterns in the language and type of SSH publications in non-English speaking European countries. They found that publication patterns differ both between and within fields. In addition, they observed an increase in the share of articles and publications in English for all countries. In conclusion, the study highlighted that internationalization policies in non-English speaking countries should consider various starting points and cultural heritages (KULCZYCKI *et al*., 2018).

Thus, an essential characteristic of the knowledge circulation in SSH is linguistic diversity. Therefore, requiring the mediation of translation in SSH is more often than in the natural sciences. At the same time, publication in national languages is a condition for reaching a non-academic audience, which also checks the quality and relevance of the SSH research. These publications are also essential for keeping contact between academics and society, and for returning the public investment in academia by providing easy access to research results. The language choice depends on the research relevance in the international scenario versus its significance for the local or national community. Still, the same project may also contribute to both dimensions (SAPIRO; SEILER-JUILLERET, 2016; SIVERTSEN, 2016a).

3 THE EVALUATION SYSTEM OF BRAZILIAN GRADUATE PROGRAMS

For most of the developing countries, higher education institutions were not integrally linked to indigenous cultures and, in many cases, were imposed by colonial rulers (ALTBACH, 2016, p. 175). In this regard, Brazil, which was a colony of Portugal, created its first higher education college only in 1808 after the arrival of the king of Portugal, John VI. Regarding graduate studies, although few experiences took place in the country before the 1960s, it was only after this period, during the military regime, that the universities implemented Master degree programs in different areas of knowledge. These programs were a result of the creation, in the 1950s, of two federal agencies for financial support: the CNPq, focused on research and linked to the Brazilian Ministry of Sciences, Technology and Innovation (MCTI) that nowadays is the Brazilian Ministry of Sciences, Technology, Innovation and Communication (MCTIC); and the other was the CAPES, focused on graduate programs and associated with the Brazilian Ministry of Education (MEC). Considering these agencies, CAPES is the one that both evaluates all the graduate programs of the country, attributing a grade to them, and links scholarships and funds (SÁ BARRETO; DOMINGUES; BORGES, 2014).

The university sector is the site with the most research activities in Brazil. Once CAPES detains most part of the resources to research and it is responsible for evaluating all graduate programs, linking funding to performance, it has a pivotal role in research improvement in Brazil. CAPES created the national evaluation system of Brazilian graduate programs in 1980, which is a fundamental instrument of the Brazilian National Graduate System (SNPG). A central issue in this evaluation is to differentiate the graduate programs among them in the same subject field, which depends on establishing specific criteria for each subject field. This evaluation process came up due to the great demand of CAPES for establishing parameters to distribute funds among the increasing number of graduate programs (CASTRO; SOARES, 1986). Before the development of this evaluation process, there were not comparable and available systematic data that could be used to define the number of scholarships to be distributed among the different programs. Thus, from that year on, CAPES began to collect and evaluate data of graduate programs (*Banco de Metadados* – CAPES).

In order to evaluate and monitor all graduate programs, CAPES organized committees from different subject fields, counting on renowned researchers, and establishing the Technical and

Scientific Council for Higher Education (CTC-ES). Over the years, the committees and CTC-ES formed important discussion forums, establishing quality standards for research and academic career for legitimizing study subjects, theories and methodologies, as well as valuing certain publication patterns and interaction with the international academic community (COUTINHO, 1996). The goals of the evaluation by these committees are to stimulate and reward excellence in the graduate programs.

The first evaluation system criteria in 1980 were not necessarily quantifiable, being expressed in levels. Level A implied a very good program; B, good; C, fair; D, poor; and E, insufficient. Over time, these criteria lost the power to distinguish the programs, and the oldest ones used to be at the highest level. CAPES did not use to disclose the results of these periodic evaluations publicly (FERREIRA; MOREIRA, 2002; CAPES, 2011). Throughout the years, this system experienced some improvements at different times, but the more representative one took place in 1998. Thus, the prior levels gave place to grades ranging from 1 to 7. CAPES included more quantitative indicators, such as the number of papers published by the programs (FERREIRA; MOREIRA, 2002; CAPES, 2011). Another significant change was the design of a research evaluation system, more specifically a journal rating system named QUALIS. Additionally, CAPES established an evaluation form unified for all subject fields. In this new form, the programs were evaluated according to seven different aspects, such as the Programme proposal; Academic staff; Research activity; Training activity; students' body; thesis and dissertations; and scientific outputs. Hence, all these improvements made it the first quantitative and qualitative evaluation system in Brazil (SCHWARTZMAN, 2013). In 2007, that evaluation form was updated, thus being composed of five items, which were Programme proposal; Academic staff; Student, thesis and dissertations; Scientific outputs; and Social engagement and impact. This last form was maintained until 2016.

Among all evaluated elements so far, the scientific production always had a high weight, which mostly determined the graduate programs' grade (SOUZA; PAULA, 2002). Therefore, if a researcher intended to contribute to the distinction of their graduate program, they should observe the publication formats with more relevance to their program subject field like journals, books, annuals conferences, among others. Despite the different publication formats and efforts of some subject fields to evaluate books, the journal is yet the dominant format of publication in Brazil (MIRANDA; MUGNAINI, 2014; MUGNAINI, 2015).

Considering that peer review is costly and time-consuming, many governments and evaluation agencies have tried an approach based on the classification of journals by rating. This categorization can be in merit classes following an ordinal scale approach or ranking the journals by assigning to each journal a number representing its position in a full ordering, following a cardinal method (ROUSSEAU, 2002). As aforementioned, Brazil, through CAPES, also created its journal ranking system (QUALIS). Once the current evaluation process is firmly shaped around the scientific production of graduate programs, more specifically journals, studying QUALIS is of extreme importance to understand possible patterns of the Brazilian scientific production.

3.1 THE BRAZILIAN JOURNAL RANKING SYSTEM

QUALIS is a tool established by CAPES in 1998 to rank the journals used for publication by faculty and students. It is based on the journal's perceived quality, which works as an indirect quality indicator of the articles published on it. Initially, the purpose of developing this system was qualifying the journals and not only counting the number of published articles. Hence, there was a need to introduce indicators that could express the levels of competition and competitiveness among graduate programs, moving them up toward a higher international insertion not only of knowledge production but also patterns of human resources qualification (HOSTINS, 2006). Although QUALIS is just a piece of all evaluation processes performed by CAPES, it is the main one and stands out as an essential quality indicator of the Brazilian scientific production (OLIVEIRA; AMARAL,2017).

Concerning the significance of QUALIS, Barata (2016) highlights that even though this system was conceived only to support the evaluation of graduate programs, allowing to compare them, it has been used by the scientific community in different ways. For instance, scientific editors have been using QUALIS to evaluate scientific production and look for funding; agencies use it to approve these funds. Research funding institutions have also been using QUALIS to evaluate researchers, to grant them or not scholarships or other types of funds. Universities and research institutes have been using QUALIS to evaluate their faculty and researchers. In the light of these considerations, CAPES through the QUALIS system exerts a strong influence on scientific production in the country. Therefore, it is a powerful tool capable of changing both organizational structures and cultures and individual scientific activities.

The first QUALIS model rated the journals in three groups, according to the circulation of journals, and each group into three strata, based on their impact or relevance. The groups were international, national, or local, with levels, A, B, and C. In addition, each subject field had the autonomy to decide its QUALIS criteria. Thus, some subject fields mostly applied bibliometric indicators from the primary databases, especially WoS. Some of them adopted only qualitative metrics, while others mixed both methodologies, although all fields preserved the authority of peer committees. Over time, the subject fields were using only a few strata among the nine ones to rank the journals. Thereby, the system became inefficient to rank the journals adequately. As a result, CAPES made a new modification on QUALIS in 2008. Since then, the new ranking comes to classify the journals into eight different categories, seeking to reflect the relative importance of different journals for a subject field. A1 was the highest category, followed by A2; B1; B2; B3; B4; B5; C – with zero weight (BARATA, 2016). As aforementioned, CAPES maintained this last classification system until the periodic evaluation referring to the years 2013 to 2016.

In 2019, CAPES conceived a new QUALIS system[4] as a part of its initiative to restructure all of the evaluation processes. Now, this system is a unique classification list for all Brazilian subject fields, which means that each journal receives only one qualification. The new criteria comprise mainly bibliometric indicators that consider the number of journals' citations. The new classification strata range from A1, the highest; A2; A3; A4, B1; B2; B3 and B4, the lowest. As a result, only one QUALIS list will be developed based on the Brazilan publications from 2017 to 2020 of all subject fields. This list will be used for the first time to support the next periodic evaluation, in 2021. It's composition and use is therefore beyond of the scope of this thesis.

The QUALIS system considered here was composed by a list of journals in which graduate programs published their researches, one list by each subject field. Hence, a journal was listed and classified only if faculties or students published on it in the year or period of classification. Additionally, these lists used to go through an annual update process and the same journal, when classified into two or more subject fields, might be classified into different categories of QUALIS (BARATA, 2016). At the end of each periodic evaluation, CAPES used to update all lists and to generate productivity indicators for each graduate program. These indicators were both in absolute

---

4 Aprimoramento do processo de avaliação da pós-graduação (2019). Available at: https://www.capes.gov.br/images/novo_portal/documentos/DAV/avaliacao/18072019_Esclarecimentos_Qualis2.pdf. Accessed on May, 20, 2020.

(publication quantity) and relative numbers to the points that each subject field attributes to different categories of QUALIS.

The highest category A1 of QUALIS, in general, yielded 100 points; A2 ranged from 80 to 70 points; B1 from 70 to 60; B2 from 60 to 50; B3 from 50 to 20; B4 from 30 to 10; and B5 from 10 to 5. In 2012, the CTC-ES established that the number of journals ranked in the A1 and A2 categories needs to be lower than 25%, in which A1 is lower than A2; A1 and A2 categories plus B1 category less than 50%. All subject fields should apply these rules. These hallmarks demonstrated how the Brazilian journal ranking system is much based on transparent quantitative metrics, signaling the mathematical path to the graduate programs that seek to maintain or improve their grade and consequently to guarantee their funds.

Nowadays, CAPES convenes 49 subject field-based peer committees. The work of each committee is headed by a "subject field coordinator", who counts with the support of some ad hoc consultants. The organization into these subject fields in Brazil arose as a way of organizing and managing information regarding science and technology (DE SOUZA, 2006). Hierarchically, nine major fields cover the 49 subject fields, which, in turn, group basic subject fields and specialties[5]:

1st level – Major fields: agglomeration of different subject fields, due to the affinity of their objects, cognitive methods and instrumental resources reflecting specific socio-political contexts;

2nd level - Subject Fields: set of interrelated knowledge, collectively constructed, gathered according to the nature of the research object for teaching, research and practical applications;

3rd level – Basic Subject Fields: segmentation of the subject field established according to the object of study and methodological procedures recognized and widely used;

4th level - Specialty: thematic characterization of research and teaching activity. The same specialty can be classified in different major fields, subject fields and basic subject fields.

Among other responsibilities, the peer committees have the autonomy to develop appropriate criteria of journal ranking for their subject fields. These criteria are set in normative documents by subject field and are revised every three to four years, and previously approved by

5 Tabela de Áreas do Conhecimento. Available at: https://www.capes.gov.br/images/documentos/documentos_diversos_2017/TabelaAreasConhecimento_072012_atualizada_2017_v2.pdf. Accessed on May, 20, 2020.

the CTC-ES. In sum, the peer review mechanism predominates in the Brazilian journal ranking model. The use or not of bibliometric indicators is up to each subject field. A study analyzed 49 normative documents and verified that 81% of the subject fields use the Impact Factor (IF) based on the WoS as an indicator, of which 74% consider it as the leading indicator. Furthermore, 56% of the subject fields use the SJR indicator based on Scopus, of which 89% consider it only as a support to define the rank. Other indicators used are h-index and Cites per Doc, both of Scopus, as well as the h-index of GS. Seven fields have not reported using any bibliometric indicator in their evaluation. The authors also detailed some of the checked qualitative criteria. One is if the journal has been indexed on other databases deemed relevant to the field and if the journal has a responsible editor and editorial board, as well as an International Standard Serial Number (ISSN). The diversity of authors' institutional affiliation is also considered. The language of publication as well as the publication format and periodicity are also verified (OLIVEIRA; AMARAL, 2017).

As mentioned, although CAPES conceived QUALIS as a tool to compare the graduate programs, the academic community in Brazil and other funding institutions have been applying these lists as a national quality indicator of the research in the country. Moreover, considering that most of the research in Brazil occurs at universities, in the end, they depend on this research evaluation system to receive their resources. Thus, it seems this system has been working as a PRFS in the national scenario.

3.1.1 QUALIS as a research evaluation system

In general, research evaluation systems differ on how they are organized and governed, as well as on their implications for resource allocation decisions, which is the main difference (WHITLEY; GLÄSER, 2007). In this section, the features of QUALIS are pointed out, taking the characteristics of research evaluation systems into account, which were described by Whitley and Gläser (2007). According to the author, regarding their governance and structure, these systems differ in frequency, formalization, standardization, and transparency. QUALIS used to have a periodicity of three years, but since 2013, it became four years. It is a standardized system, because it shows normalized evaluation methods and practices. In regard to formalization, QUALIS can be considered a highly formal system, which, according to the author, involves central agencies with systematic rules and procedures. In addition, it relies mainly on peer review, and there are general evaluation procedures and practices across the subject fields. At the same time, some of its evaluation practices vary considerably, which is usual in judgments of research quality in very different subject fields. In terms of transparency, QUALIS may be deemed a transparent system since all rules are in normative documents, which are publicly available. Each coordinator of subject field also suggests a panel member list with renowned researchers, which CAPES needs to approve before the evaluation process, and that is made public later. Furthermore, the foundation updates and publishes all 49 journals lists every year.

Considering the consequences of research evaluation systems for research funding in public sciences, Whitley and Gläser (2007) states that they can affect the resource allocation directly or even in the proportion of employers' incomes. To better understand these effects, the author distinguishes and contrasts two types of evaluation systems, weak and strong. QUALIS can be framed as a strong research evaluation system. In the author's view, this kind of system stands out by highly formalized sets of rules and procedures, and structure around existing disciplines and scientific boundaries. In addition, the evaluations are usually ranked on a standard scale and publicly available. The results directly affect funding decisions, presenting a considerable impact on the management of universities and similar organizations. Weak systems are characterized by informality, little standardization of procedures or criteria, and limited publishing of results or official ratings. Moreover, they have far less impact on funding allocations, and changes tend to be incremental rather than radical.

Therefore, QUALIS has become progressively stronger over the years after the characteristics of weak and strong systems have been considered, as well as the changes of QUALIS. The impact of strong systems is likely to be especially noticeable, according to Whitley and Glässer (2007). Hence, the author pointed out some significant consequences of institutionalizing such systems, as follows:

- Researchers will seek to contribute to the collective goals of their fields as understood by current elites. Consequently, as this type of evaluation system becomes more influential, the researchers' work should become more integrated around these goals.
- Peer review develops and applies standard criteria of research quality for the field, which becomes institutionalized as dominant over time. Thereby, there is a judgment centralization across individual researchers, universities, and other research organizations.

Thus, the degree of uncertainty about the intellectual importance of research strategies and outputs for collective goals (WHITLEY *et al.* 2000, p. 123-124) should decline.

- A possible decline in the diversity of intellectual goals and approaches within sciences over time, especially where they challenge current orthodoxies. This effect arises as a result of the centralization and standardization of research goals and evaluation criteria throughout scientific fields.
- Disciplinary standard and goals may inhibit the development of new fields and objectives. Thus, intellectual innovations may tend to be focused on current sciences and their concerns. Hence, strong research evaluation systems are less likely to intellectual and organizational innovations, since they reinforce conservative tendencies in determining intellectual quality and significance.
- The standardization, formalization, and publication of quality rankings intensify the stratification of individual researchers and research teams.

Regarding QUALIS, this is a comparative evaluation in which graduate programs in the same subject field are contrasted. Therefore, if a researcher desires to have their program well evaluated, they have to follow the goals established by the subject field of their graduate program. These goals are public for the entire academic community in specific documents prepared by a restrict academic elite. These facts can lead researchers to become more aware of the need to compete with others, a reduction in scientific diversity, besides other effects such as those aforementioned by Whitley and Glässer (2007).

From these perspectives, rethinking a new research evaluation system demands to understand the real role of this system in the national scene, but paying close attention to these international points of view about these same issues. Thus, establishing tools to monitor this system is an approach to identify these possible systemic effects over the years.

3.2 INSERTION OF QUALIS IN THE NATIONAL PLAN FOR GRADUATE STUDIES

The National Plan for Graduate Studies (PNPG) is a set of public policies for graduate studies created in Brazil from 1974 to broaden and improve the graduation programs in the country. CAPES is responsible for elaborating and executing these plans, and currently, the VI PNPG is being performed by the institution, covering the period from 2011 to 2020. In such way, between 1974 and 1989, the PNPG I, II and III were carried out, but between 1990 and 2004 the institution developed the IV PNPG that was not officially published, and only in 2005 the V PNPG was elaborated, which remained until 2010 (FERREIRA; MOREIRA, 2002).

In this scenario, CAPES developed QUALIS in 1998, precisely within the period in which no national plan had been officially published to guide the sector development, but that it is considered a period of significant changes in the economic, political, and educational contexts. At that moment, the evaluation system in Brazil was going through a reformulation process to increase its insertion in the international environment of the scientific knowledge production, then shifting the focus of policies from teaching to research and scientific output (KUENZER; MORAES, 2005). According to Hostins (2006), there is a need of introducing indicators that could express the levels of competition and competitiveness among graduate programs. As pointed out by Coutinho *et al.* (2012), the scientific research in Brazil usually takes place mainly inside graduate programs at universities.

Another arrangement of QUALIS in eight different categories arose in 2008, when the criteria adopted ten years before became inefficient to segregate the journals well, and thus perform an adequate evaluation. This QUALIS implementation happened during the V PNPG term, which the main goal was to promote growth in the national graduate system (BRAZIL, 2005, p. 9). Again, in this PNPG, there is a higher focus on results and products. Hence, concerning all evaluation processes, two requirements were defined as central: students and research outputs. They include the journals ranked by QUALIS, as well as books, patents, and technologies, which are also considered in the evaluation procedure by CAPES.

The recent VI PNPG (2011-2020) reinforced the need for peer review in the evaluation process to guarantee quality and introduced new parameters and procedures. The purpose was to improve the current model and to correct distortions. One of these distortions was the development of a strong group traditionalism coupled with the accommodation of programs. Another one was productivism result and quantity dominance (BRAZIL, 2010, p. 22). In regard to the evaluation criteria, VI PNPG suggested considering both the impact and relevance of the scientific outputs in the frontier of knowledge (BRAZIL, 2010, p. 37). Additionally, it reported the historical advance of Brazilian scientific productivity and its increasing inclusion in the WoS and Scopus databases until 2009, thus presenting data by knowledge fields, such as the number of articles, count of citations, and median of IF (BRAZIL, 2010, p. 224-231).

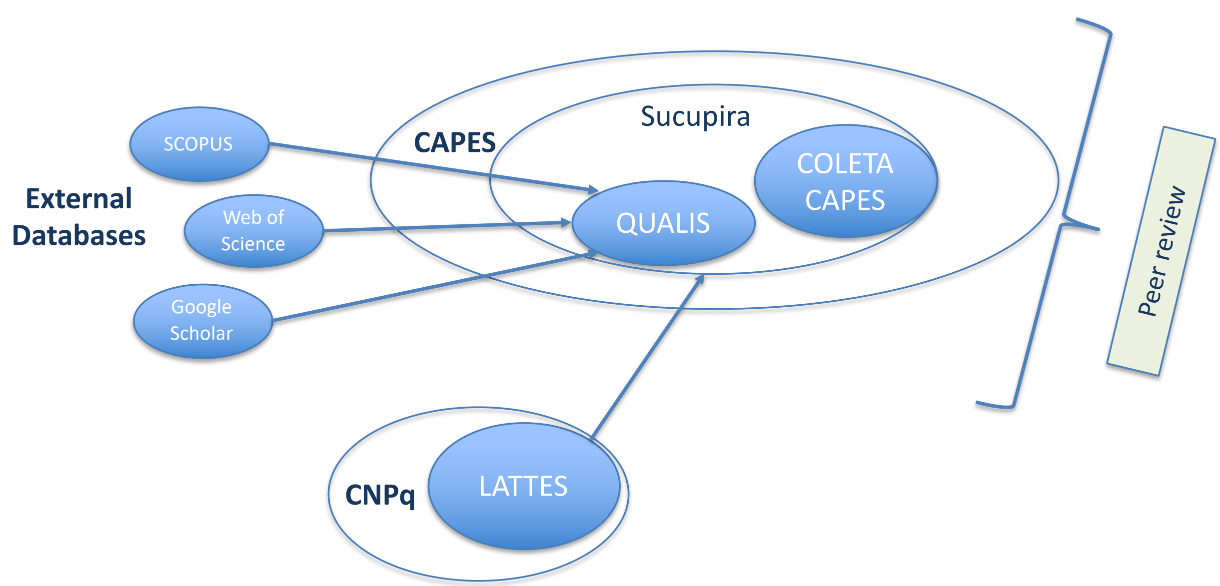
It is clear the essential role given to scientific production linked to graduate programs in the national scenario and the effort to improve the evaluation system as a whole and mainly QUALIS. The last national plan called the attention to other types of research outputs and the need of evaluating them. CAPES also conceived tools to assess books and technological products. It is important to emphasize that all these tools follow the same idea of QUALIS for ranking journals, stratifying the outputs in different categories. Once there is a considerable increase in the graduate system over the years, and, consequently, in the number of outputs, these evaluations become increasingly complex on a large scale. Thus, they demand more people and funding to guarantee a minimum of quality.

4 INFORMATION TECHNOLOGY AND RESEARCH EVALUATION IN THE BRAZILIAN CONTEXT

As the evaluation criteria and methods have been modified and refined along years, new information systems were developed in Brazil, improving data management. Some of these systems supported the evaluation of graduate programs, as well as research management in the country. In order to follow international trends in research evaluation exercises, CAPES has designed the implementation of a Current Research Information System (CRIS), which purpose is to improve the quality of received data and accessibility to end-users.

Figure 1 shows the current Brazilian research information systems and their interactions. These systems are the Lattes Platform (LP), which embodies the Curriculum Lattes, managed by the Brazilian Science and Technology Development Council (CNPq); Sucupira Platform, which comprises the ColetaCapes and the Journal Ranking System (QUALIS), operated by CAPES. In addition, external databases that are sources of bibliometric indicators are used as a support tool in the evaluation process performed by peer review in CAPES. These components in the context of a CRIS are detailed in the next sections.

Figure 1 – Information systems that support the evaluation of graduate programs as well as research management in Brazil



Source: The author

4.1 LATTES PLATFORM (LP)

LP is a virtual environment developed by the CNPq. It was established to manage CNPq funds, therefore integrating curriculum, research group and institution databases from across the country (BALANCIERI *et al*., 2005). It includes information systems, databases, and web portals. Its main component is Lattes Curriculum (CV Lattes) that stores, manages and, searches curricula of students, professors, and researchers, standardizing and centralizing personal, professional and academic information of the Brazilian scientific community.

4.1.1 Lattes Curriculum (CV Lattes)

Regarding the implementation of CV Lattes, paper forms and DOS-based systems began being used for curriculum systems in the 1990s. From 1999 onwards, a more modern system replaced the old one after debates and surveys with consultants on the real need to develop this database, as well as on the establishment of a curriculum model that would meet the requirements for research funding agencies (CNPQ, 2016). Furthermore, academic results such as academic degrees, professional resumes, publications, received funds, academic positions, awards, etc. are among the information gathered in CV Lattes.

Undoubtedly, this platform is the major source of information on Brazilian researchers and it has a high potential for gathering information (ALVES; YANASSE; SOMA, 2011). Lane (2010) recognized the Lattes Platform (LP) as a complete and highly qualified database. Nowadays, CV Lattes has been used as a source of information by funding agencies for evaluating researcher's performance, projects, graduate programs, etc. Hence, it is a rich and powerful database that presents innumerous potential applications (scientific, technological, economical, etc.). This database is also systematically updated, which allows the scientific community itself to monitor the quality and fidelity of the system information. This control takes place since the professional curriculums are analyzed to award funding. Although the CV Lattes system is public and accessible online (http://iattes.cnpg.br/), access is restricted. Currently, an alpha-numeric password is required in order to search for a curriculum to avoid automatic searches (ALVES; YANASSE; SOMA, 2011). Furthermore, CV Lattes is available in HTML format in the CNPq website, displaying information only in a personal way, *i.e.* the registered information is individually associated with each person.

4.2 SUCUPIRA PLATFORM

In 2013, CAPES launched Sucupira Platform, a new and important tool for collecting information, conducting analyses and evaluations online. Nowadays, it is considered the SNPG base reference. Moreover, it makes the information, processes and procedures that CAPES performs in SNPG available at real time and with much more transparency for the entire academic community. It also provides the managerial-operational participation of all processes, allowing greater engagement of the deans and coordinators from the graduate programs. The platform name is a tribute to Professor Newton Sucupira – author of the Brazilian document that conceptualized, structured and institutionalized Brazilian graduate studies in its present format (https://www.capes.gov.br/avaliacao/plataforma-sucupira).

Through the Sucupira Platform, any person can search for information of the graduate programs as they are being filed. These actions provide more guarantee of transparency and, consequently, higher data reliability, considering that faculty and students can check if their graduate program filed the data correctly. In addition, this continuous availability of information throughout the year enables peer-review committees to prepare or endorse the indicators to be used in the evaluation process in advance. Furthermore, all data used by peer review in the evaluation process are accessible to any public person, which permit in some subject fields doing simulations or comparisons that are essential to transparency and isonomy (OLIVEIRA; AMARAL, 2017).

4.2.1 ColetaCapes and QUALIS

In the Sucupira Platform, CAPES incorporated other systems, in which ColetaCapes and QUALIS stand out. ColetaCapes was firstly known as DataCapes and was the first database implemented by CAPES with the purpose of receiving annual information on the performance of all graduate programs. As a result of internet development, CAPES updated ColetaCapes in 1999 by releasing the permission to register students' names as well as linking them to their scholarships, dissertations and thesis. In 2004, CAPES improved that system once again and changed its software architecture to Java, which allowed an integration database with LP. Every year all graduate programs include the most important data of students and faculty members for their graduate program's evaluation, especially about their scientific production (RIBEIRO, 2008; SIQUEIRA, 2019).

QUALIS was established by CAPES in 1998, but it was completely restructured in 2008. This system is applied for ranking the journals used by faculty and students to publish their work. As a result, a list of these journals used to be generated and published by each subject field of CAPES (BARATA, 2016). QUALIS was designed to meet the specific needs of the evaluation system and is based on information provided through ColetaCapes. Currently, the Foundation has just launched a new ranking system, which gives more emphasis to bibliometric indicators from the main international bibliometric databases, mainly Scopus, to evaluate the journals.

4.3 BIBLIOMETRIC DATABASES

The term "bibliometrics" was firstly introduced in 1969 as a substitute for "statistical bibliography" (HOOD; WILSON, 2001). It was defined as "the application of mathematical and statistical methods to books and other media of communication" (PRITCHARD, 1969, p. 349). Databases developed by commercial establishments or public or private institutions have provided both the source of raw data, as well as the analytical tools for many bibliometric studies. In 1988, Burton (1988, p. 43) summarized the progress made in electronic databases. This includes extensive increases in database coverage, rapid development of new databases, and release of a wide variety of user-friendly tools to improve and facilitate access to existing services. Hood and Wilson (2003) reported that the number of databases was over 12,900* (with the number of records in these databases at over 16,800 million) in 2001. Hence, these databases showed a large-scale development and use. For every established academic discipline, one may find one or more databases nowadays (HOOD; WILSON, 2003). Although there are more available databases, they usually cover only a limited number of scientific fields, and some of them do not contain data on the references of publications, and therefore they cannot be used to calculate citation impact indicators. Considering databases that provide citation counts, the three most popular multidisciplinary databases are WoS, Scopus and Google Scholar (WALTMAN, 2016). Moreover, regarding a nationwide database, there is the Scientific Electronic Library Online (SciELO) in Brazil (MUGNAINI, 2013).

Databases with bibliometric information on published scientific literature play an important role in the field of quantitative studies of science and in the development and application of Science and Technology indicators. In Brazil, these databases and their indicators have been used by some subject fields for performing the evaluation of their journals over the last years. Recently, after the implementation of a new research evaluation system, these databases have gained a more critical role and are used more broadly in all subject fields.

Regarding the WoS database, it was considered for a long time the major source of citations, covering papers in around 12,000 journals in all domains of science. It was originally produced and upheld by the Institute of Scientific Information (ISI), afterwards by Thomson Reuters and today by Clarivate Analytics. It consists of a subscription-based database that comprises a few citation indices. The best-known include Science Citation Index (SCI), Social Sciences Citation Index (SSCI), and Arts & Humanities Citation Index (A&HCI), covering journals and book series. Nowadays, proceedings and books are also covered by another citation indices – Conference Proceedings Citation Index and a Book Citation Index (MINGERS; LEYDESDORFF, 2015b; WALTMAN, 2016). Moreover, the WoS provides the possibility to search for publications, citations and h-indexes.

Scopus was created in 2004 as a very similar rival database to WoS, available from Elsevier, which rapidly became a good alternative (VIEIRA; GOMES, 2009). This database stands out as the largest searchable citation and abstract source of searching literature, which is continually expanded and updated (REW, 2009). It contains over 50 million article references from 20,000 peer-reviewed journals, covering also conferences and books (MINGERS; LEYDESDORFF, 2015b). Scopus features have been studied less extensively when compared to WoS, but some studies have pointed out relevant issues. Among them, the literature assigned to Scopus, the presence of incorrect DOIs to publications, issues in the accuracy of author's identifiers, and duplicate publications (WALTMAN, 2016). Scopus provides the possibility to search for by publications, citations, h-indexes, besides metrics such as CiteScore, SJR, and SNIP.

In addition to Scopus, Google Scholar (GS) was also launched in 2004. It works diversely, searching the web for documents that refer to papers and books. GS automatically establishes relationships between cited and citing sources, playing the role of a citation index in the web retrieval. A wide number of GS's scientific citations are accessible in full text and free of charge in the web environment. The documents are ranked by their relevance to the searched queries, which is based on citation relations and not just on the visiting or linking rates on the Internet. Hence, GS has two main advantages: it is an information retrieval tool with precise and advanced algorithms, and also a mean to provide resource evaluation opportunities through citation analysis techniques (FRIEND, 2006; NOTESS, 2005).

In general, several studies compare those three databases based on coverage, selected country data, institutions, selected journals, publication types, subject areas, among others. However, coverage by countries, institutions, journals as well as across subject disciplines is the most common subject. This varies significantly between these databases. Overall, GS outperforms both WoS and Scopus in terms of scientific literature coverage. Besides scientific journals, it covers citation and bibliographic data of non-serial resources such as conferences, research projects, dissertations, pre-prints, and books (CHEN, 2010; GEHANNO *et al*., 2013; HARZING, 2013). Orduña-Malea *et al.* (2015) estimated the coverage of GS to be about 160-165 million documents. Martín-Martín *et al*. (2014) highlights the coverage in a diversity of types, languages and countries. Khabsa and Giles (2014) estimated that GS indexes about 100 million English-language documents, representing almost 87% of all English-language scholarly documents available on the web. On the contrary, earlier studies criticized its coverage in terms of lack of comprehensiveness, quality control and transparency (JACSO, 2005a, 2005b, 2005c; WOUTERS *et al*., 2015), and weakness in covering older works (MEIER; CONKLING, 2008).

A subscription is necessary to access WoS or Scopus through a web interface. Most institutions, as well as professional bibliometric centers, often have this subscription for full access to those databases, which is required for advanced citation analyses at a larger scale (WALTMAN, 2016). Unlike WoS or Scopus, GS is a freely available citation database; therefore, it has an important role in the democratization of citation analysis (HARZING; VAN DER WAL, 2008, p. 12). On the other hand, performing large-scale citation analyses using GS is more difficult, because the only way to access it is through its web interface (WALTMAN, 2016). A software tool called Publish or Perish (HARZING, 2007) is usually used for performing analyses on GS.

4.4 CURRENT RESEARCH INFORMATION SYSTEMS (CRISs)

Changes on how research is being conducted worldwide have led many countries to invest in technological improvements related to data collection, networking, storage, and management, hence creating a shift towards the paradigm of data-intensive science (HEY; TANSLEY; TOLLE, 2009). An important driver in these developments seems to be the introduction of performance-based funding systems, because they require standardized data at the institutional level (SIVERTSEN, 2016b).

First of all, clarifying the terms information system, research information and CRIS is essential. An information system is understood as a human-machine complex, in which interrelated components work together to collect, process, store, and disseminate information to support decision making, coordination, control, analysis, and visualization in an organization. Research information means those of research activities, *i.e.* the so-called metadata about projects, publications, published data sets, infrastructures and people/teams. Jeffery *et al.* (2014, p. 6-7) described the main elements of research information as follows:

> a. Research output, including various kinds of text-based scientific publications, data sets, patents, software, devices, designs, artistic works and performances and a wide array of other types.
> b. Information on the processes, workflows and methods utilized during the research process, *e.g.* observations, experiments, and simulations with their associated measurements.
> c. The people involved in research activities, including various categories of R&D personnel, *i.e.* researchers, research administrators/managers, technical and support staff participating in research projects.
> d. Organizations involved in research activities, for example, research-performing organizations and research funders of various sectors, and their internal structure (*e.g.* schools, departments, institutes).
> e. Research projects, which refer to planned research activities aiming at the accomplishment of specific tasks under resource and time constraints. Projects typically, but not necessarily, rely on some sort of funding support.
> f. The research funding environment that supports research, for example, structured funding programs with competitive allocation procedures executed by national and international public bodies or non-governmental organizations or direct state grants to research-performing organizations, covering, for instance, salaries of permanent personnel and basic operational costs.
> g. Facilities and equipment that are utilized for research purposes. Facilities include research infrastructures that can be physical (*e.g.* buildings, synchrotrons, telescopes, vessels, supercomputers) or virtual (*e.g.* software systems), single-sited or distributed.
> h. Services related to research activities and/or provided through research infrastructures or by organizations using facilities or equipment. Services can be targeted to other researchers, for example making facilities and

equipment available for experiments, or to third parties like industry (*e.g.* identification of materials through spectroscopic methods).
i. Events related to research activities, such as scientific conferences and workshops or observation or experiment periods.
j. Measurements and indicators concerning research activities, covering research outputs, outcomes and impacts and input side research funding.

Finally, CRIS, also known as Research Information Management or RIM System (RIMS), is considered a system where various types of research information are maintained and interlinked (JEFFERY *et al.*, 2014). The basis of this system was presented by the EC. Although CRIS has a long history, more than 50 years, it has not been so far used for information management, evaluation and presentation of research. During the 2000s, the use of CRIS for evaluation and benchmarking increased, as well as for management and decision support in universities (JEFFERY, 2012).

CRIS collects a wide range of metadata about all aspects of the research activity carried out at an institution, providing a deep insight into the workflows that underpin the institutional research activity. It also allows effective teamworking across institutional research support units (CASTRO, 2019). According to the International Organization for Research Information (euroCRIS), CRIS has been developed to "assist users in their recording, reporting and decision-making concerning the research process, whether they are developing programs, allocating funding, assessing projects, executing projects, generating results, assessing results or transferring technology." Therefore, this type of system has been required by many funding organizations to manage funds to programs, awarded projects and outputs associated with those funded projects, as well as to manage and interlink people and organizational units (JEFFERY; ASSERSON, 2010).

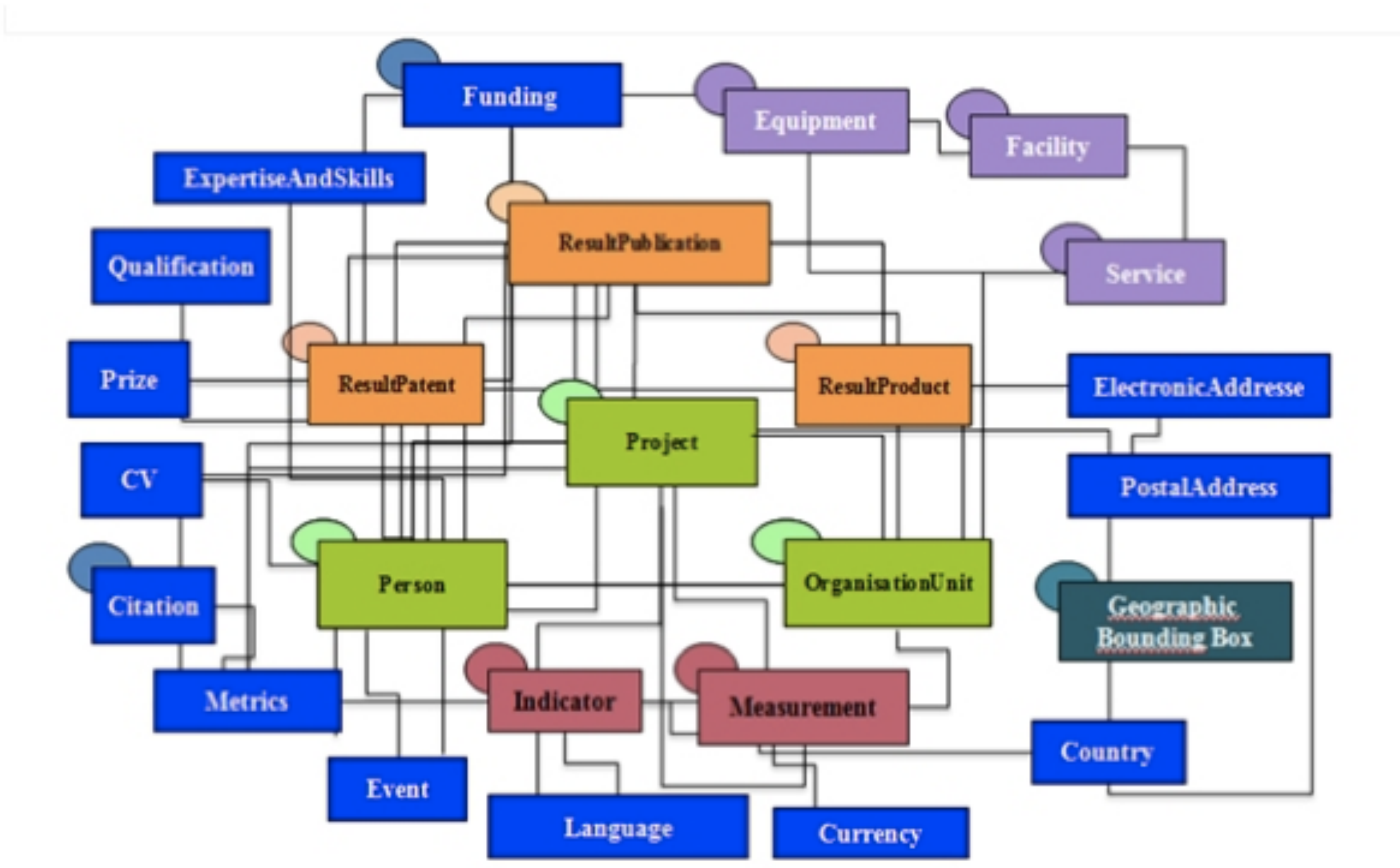
In this scenario, interoperation (exchange) of research management data and a similar report among research management information systems are a basic principle. In order to get that, in the 1980s, the EC convened a group of delegated national experts for developing a standard for interoperation, and The Common European Research Information Format (CERIF) was the established format. This data model arose from the need for a standardized CRIS (ZIMMERMAN; JEFFERY, 2004), considering that each research information user had developed its own information system. Furthermore, the number of systems for managing scholarly publications used to be limited and heterogeneous, therefore reducing the efficiency and effectiveness of research and research communication's management (JEFFERY; ASSERSON, 2010).

According to euroCRIS, CERIF is: "A concept about research entities and their relationships – Specification (Conceptual Level). A description of research entities and their relationships – Model (Logical Level). A formalization of research entities and their relationships – Database Scripts (Physical Level)."

CRISs activities and developments in Europe are tightly interrelated with CERIF, which describe the research involved entities and their manifold relationships in an efficient and scalable way (JÖRG; HÖLLRIGL, 2012). A traditional CERIF has the following entities: person, project, organization, publication, patent, product, funding, equipment, and facility (Figure 2). An organization is represented by attributes and holds relationships with others. The relevant organizations, their attributes and relationship descriptions as such compose the model of the domain for setting up an information system (JÖRG, 2010).

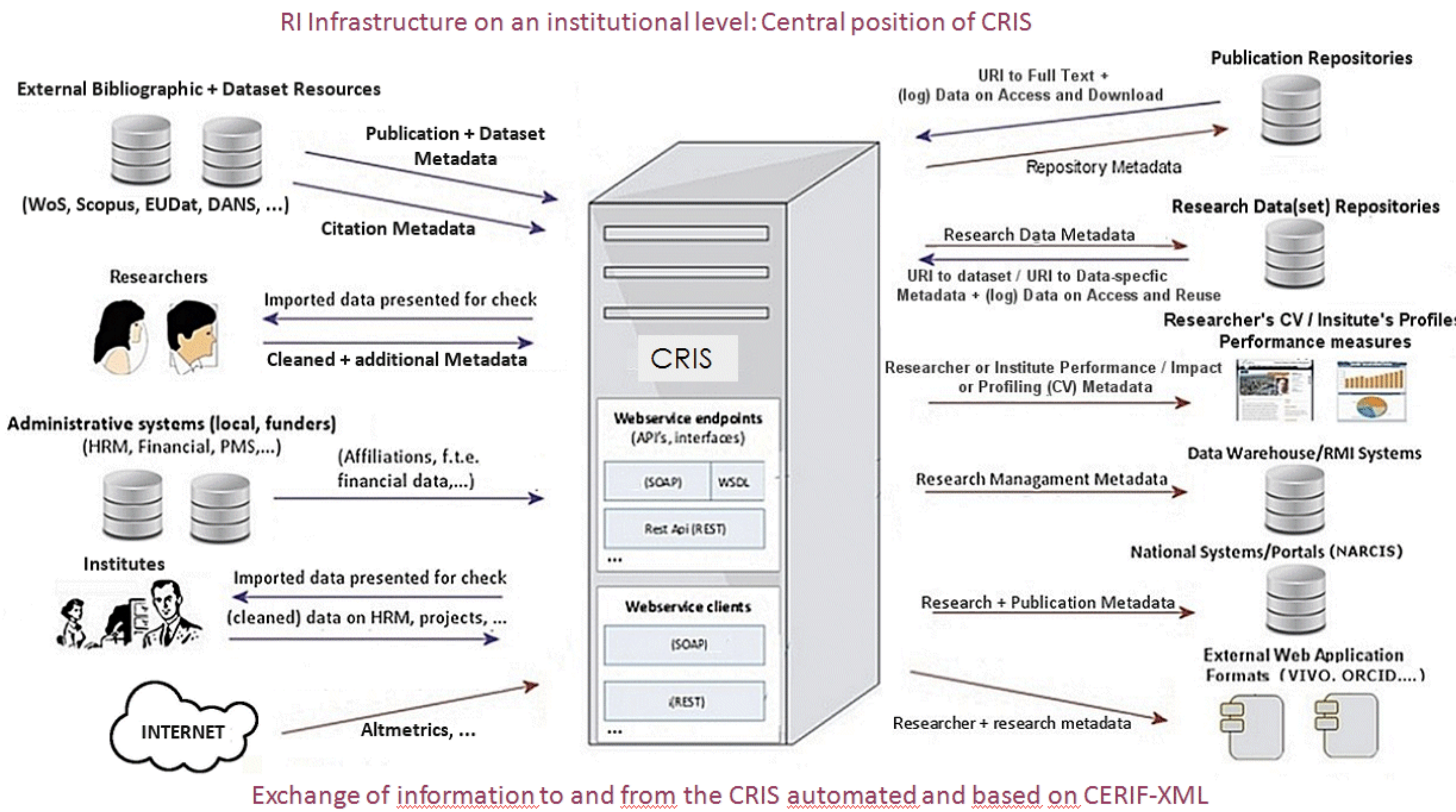
Figure 3 shows a graphical representation of CERIF-CRIS that takes a central position in the Research Information Infrastructure. It is possible to visualize that CRIS plays a critical role with information sharing and provides the required flexibility for multiple stakeholders' needs. In addition, among others, there is an automatic import of data from external online sources, such as Scopus and WoS, and repositories, as well as the storage by CRIS of research outputs, outcomes, impacts and activities from those sources, repositories or via manual data entry by researchers.

Figure 2 – The Common European Research Information Format (CERIF)



Source: https://www.eurocris.org/cerif

Figure 3 – Current Research Information System (CRIS)



Source: https://www.eurocris.org/sites/default/files/images/CRIS-central-big.gif

In Europe, many higher education institutions and other research institutions developed CRISs, besides creating national repositories for the publications and products of their scientists. The funding agencies also developed their own CRIS, which intended to provide information on projects, persons, institutions, and financial resources (HORNBOSTEL, 2006). In the USA, CRIS was founded within the National Institute of Food and Agriculture at nearly the same time as in the European Union (LEIVA-MEDEROS *et al.*, 2017). By 2012, the CERIF was accepted as the national 'standard' in nine countries with a variety of organizations using it, such as funders, universities and research laboratories (JEFFERY, 2012).

The spread of CERIF-CRIS implementation was pushed forward by the launch of a set of programs in the area of research information management in the UK alongside their existing program of repositories and subsequent research data program. These programs were formed from the agreement between universities, research funders and administrators, and researchers of JISC (the UK organization that provides ICT services to the higher education sector) (JEFFERY, 2012). CRISs on the institutional level have become widespread recently, both in locally and commercially developed solutions. In the UK, these systems have become critical elements for institutional reports to research funders and Government's Research Assessment Framework (REF). Additionally, they have the most significant number of fully operational CRISs at institutions, which often coexist with institutional repositories (IR) aimed to expose the institutional research outputs to the world. CRIS coupled with IR, considering the interoperability between them, has been deemed the most widely implemented and effective configuration (CASTRO, 2018).

Norway is one of a few countries that has a fully integrated non-commercial CRIS at the national level, named CRISTIN. It works as a shared system for all research organizations in the public sector: universities, university colleges, university hospitals, and independent research institutes. This sharing allows multiple use of the same data, legitimating the costs of running CRISTIN. The driver in the development of a shared CRIS was the Norway research evaluation system. One reason is that it is necessary to identify all institutions participating in a publication in this system for validating the process and in the indicator development. Moreover, references to publications are registered only once, after which they can be used in CVs, applications to research councils, evaluations, annual reports, internal administration, bibliographies for Open Archives, links to full text, etc (SIVERTSEN, 2016b).

Considering other countries, Sweden has a CRIS, SwePub, which has not been fully developed for bibliometric analysis yet (HAMMARFELT, 2018). Italy implemented its CRIS in 2014 and in less than one year more than 66 Italian Institutions adopted the same platform. There is also an institutional Open-Access repository embedded on it. Furthermore, the rapid adoption of ORCID at a national level gave the opportunity to enhance the quality of metadata (GALIMBERTI; MORNATI, 2017). Such systems are also used, among others, in Croatia, Czech Republic, Denmark, Finland, and Poland (KULCZYCKI; KORYTKOWSKI, 2018). The main difference among the research evaluation systems among countries is if they have or not a national or regional database for recording their research outputs (SĪLE *et al*. 2017), which can be tightly integrated to CRIS or be a standalone database. These databases are especially important with regards to book records, because the coverage of books in the international databases is weak and not sufficient for evaluative purposes (GORRAIZ; PURNELL; GLÄNZEL, 2013). In the SSH, books are among the most relevant forms of output.

In order to follow these international trends concerning research evaluation, CAPES has made some improvements in the Sucupira Platform, and others have been designed for future implementations. These information system developments follow the principle that understanding the reality of research data and how they can and should be used for the needs and objectives of research evaluation is crucial. In many countries, evaluation agencies, research funders, and research organizations require to systematically keep track of research outputs and several other data and metrics, such as staff numbers and characteristics, research project funding contracts, etc. Hence, CAPES is in the process of developing and implementing a CRIS to improve the quality of data received and accessibility to end-users. The plan of CAPES for improvements is to enable the Sucupira Platform to become a CRIS (SIQUEIRA, 2019).

In that regard, CAPES has been exploring the CERIF-CRIS model of interoperability to determine the feasibility of applying it in the Sucupira Platform environment (SIQUEIRA, 2019). Figure 4 presents the project of a new concept to improve the connections of Sucupira Platform with other national systems according to Siqueira (2019). The author explains the CAPES model of data integration in detail, which is summarized as follows:

1) The figures represent researchers and faculty, CV Lattes, and ORCID organization. Nowadays, researchers and faculty fill their CV Lattes from which Sucupira collects

information. ORCID is a unique research identifier that has been recently adopted by Brazil. It provides new data collection, and authors' recognition possibilities, which are currently not feasible in CV Lattes.

2) The graduate programs provide their data to the Sucupira Platform every year. Some data from faculty and students are manually informed. The idea is to reduce these entries manually with the implementation of CRIS.

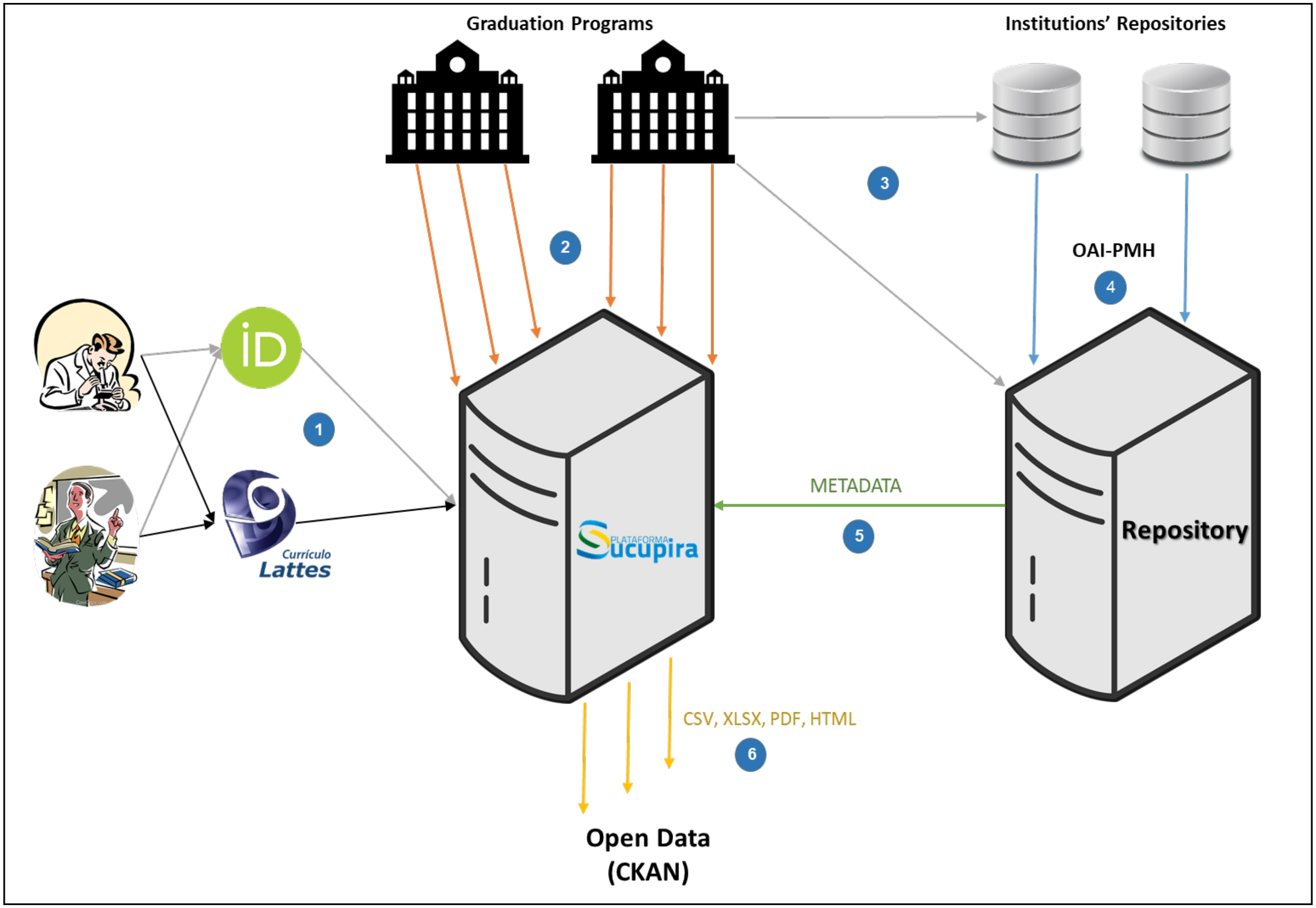
3) Nowadays, the graduate programs exchange their research data with repositories of institutions and others, before sharing them with CAPES. This flow will be inverted with CRIS; therefore, Sucupira will receive the data first.

4) It shows interoperability among the different repositories.

5) Metadata from repositories of institutions will be listed in Sucupira Platform and be related to graduate programs, therefore inverting the current flow.

6) It is the process of analyzing and checking data by Sucupira, making them available to all interested parties.

Figure 4 – Interoperability project for Sucupira Platform



Source: Siqueira (2019).

Considering the implementation needs for research data integration in CRIS, Schöpfel, Prost and Rebouillat (2017) in his overview established a list of recommendations that should cover at least six aspects:

- evaluation should not concentrate on data but on data management;
- data deposit in repositories of labelled data is preferred (expected);
- standard, generic and rich metadata must be required;
- standard persistent identifiers for data and contributors (authors), namely DOI and ORCID (Open Researcher and Contributor iD), must be required;
- open data policy should be the default, at least for public funded research;
- evaluation should include explicit measures for reporting and follow-up (no simple declaration of intention).

Among these recommendations, we highlight that CAPES has recently adopted ORCID for the registration in Sucupira Platform. ORCID provides researchers with unique persistent identifiers. It has the potential to make difference to a researcher's ability to gain full credit for their work and is a useful tool for universities as they track, evaluate and report research work. This identifier became available through a Brazilian consortium of institutions, named National Consortium of Education, Science, Technology and Innovation (CONECTI). The consortium members are:

- Three federal institutions: CAPES, CNPq, and Brazilian Institute of Information in Science and Technology (IBICT);
- One nonprofit organization: Brazilian National Council for the State Funding Agencies (CONFAP);
- One social organization: Brazilian National Research and Educational Network (RNP) that is the consortium lead;
- One publishing company – Scientific Electronic Library Online (SciELO).

The ORCID adoption represents to CAPES an important gain in agility and quality mainly in Sucupira Platform, which records data from graduate programs, in grant systems, and in international project bidding systems[6].

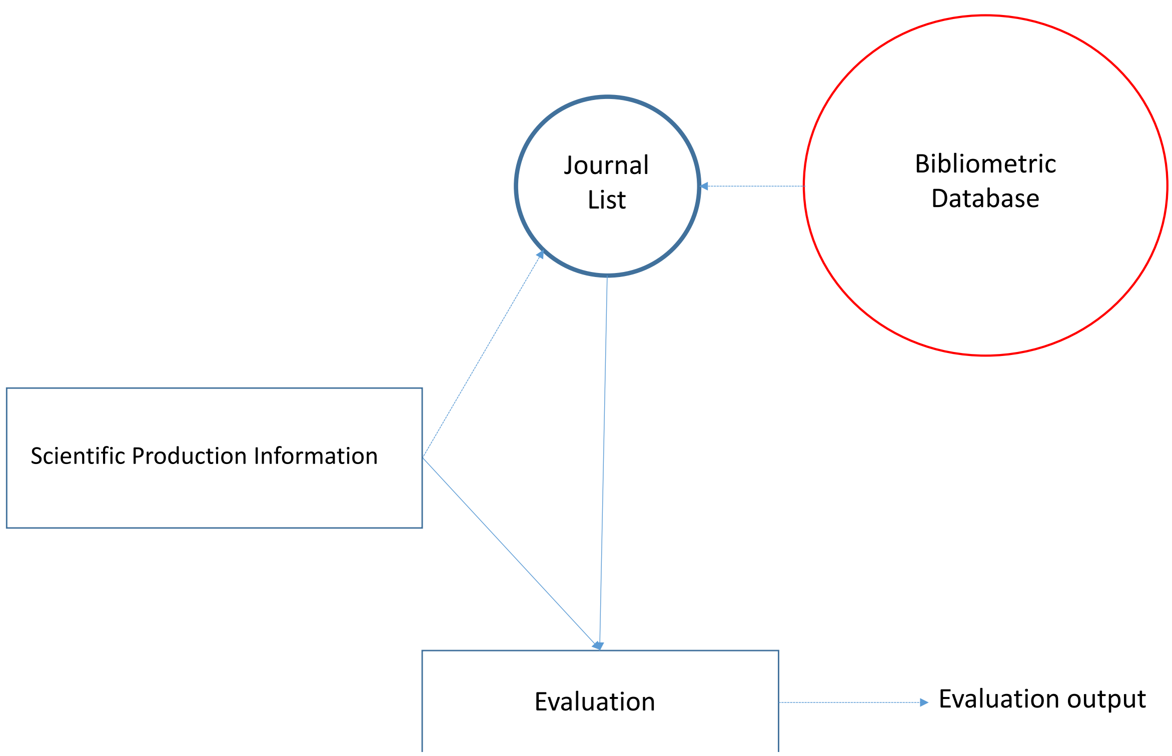
Even though Brazil is putting efforts to implement an efficient research information system, there is a great challenge forward, especially when comparing Figures 4 and 3. The adoption of the same CRIS is the first step towards stronger harmonization of procedures, which is still in the process of development by Brazil with CONECTI. Furthermore, the nationwide adoption of ORCID is beginning now in Brazil and it will take some time for its harmonization in all CRISs installations. Once it is done, adopting an architecture of data at the level of the links between authors and publications, publications and research projects, projects and investigators, and so on, will be possible. Disambiguation of researchers' names might allow an even more extensive interoperability worldwide. Regarding research evaluation procedures, it would be supported by an enormous amount of open, comparable, and interoperable data at a national level with the implementation of CRIS. Furthermore, it would enable the investigation and comprehension of publication strategies, strengths and weaknesses, main research areas, new areas, among others.

[6] Available at: https://orcid.org/blog/2018/06/18/brazilian-leadership-open-research. Accessed on: July 08, 2019.

5 MATERIAL AND METHODS

This section presents a framework for the analysis of scientific production evaluation systems based on ranking lists. The framework is based on the integration of several databases that store information about scientific production, their citation impact, among other data. Its goal is to bring to light patterns as well as potential distortions in academic production behavior induced by these evaluation systems. Figure 5 shows the overall framework for the analysis of scientific production in a system based on ranking lists. In such system, scientific production information is commonly used to identify journals in which researchers publish their works. These are the journals that are considered to form the Journal List. In the composition of these lists, bibliometric data from known databases (*e.g.* Scopus, WoS) may be used to classify journals from more to less relevant in each field. The ranking of journals does not necessarily follow bibliometric criteria, and other quantitative or qualitative parameters may also be used to decide on the relevance of each journal. Once the Journal List is established, it becomes one of the main sources of information for scientific production evaluation. Therefore, it is important to monitor evaluation output results.

Figure 5 – Framework for the analysis of scientific production in a system based on ranking lists

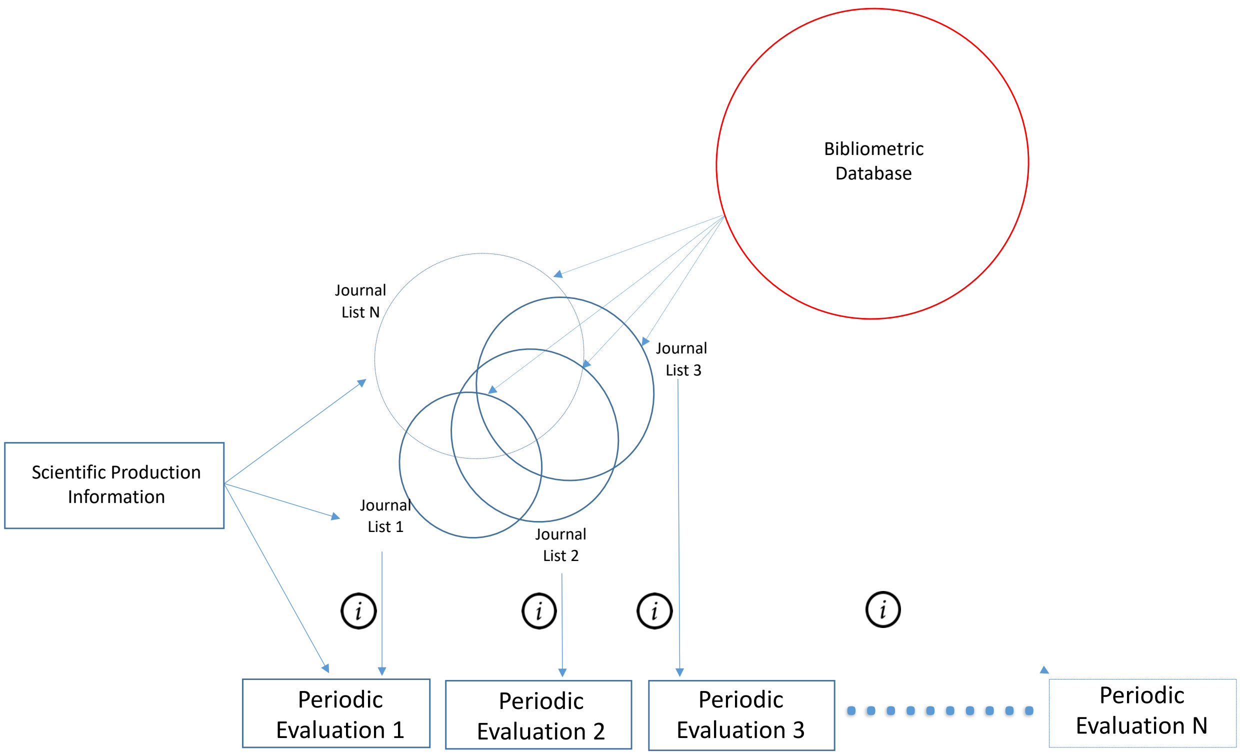


Source: The author.

.

The goal of having such a model is to identify patterns that may show positive as well as negative aspects in academic production behavior induced by these evaluation systems. However, the model should encompass historical evaluation steps as well as different Journal Lists for identifying behavioral changes, as each of these steps could potentially consider a different set of journals. Figure 6 provides this more detailed view of the framework.

Figure 6 – The detailed framework



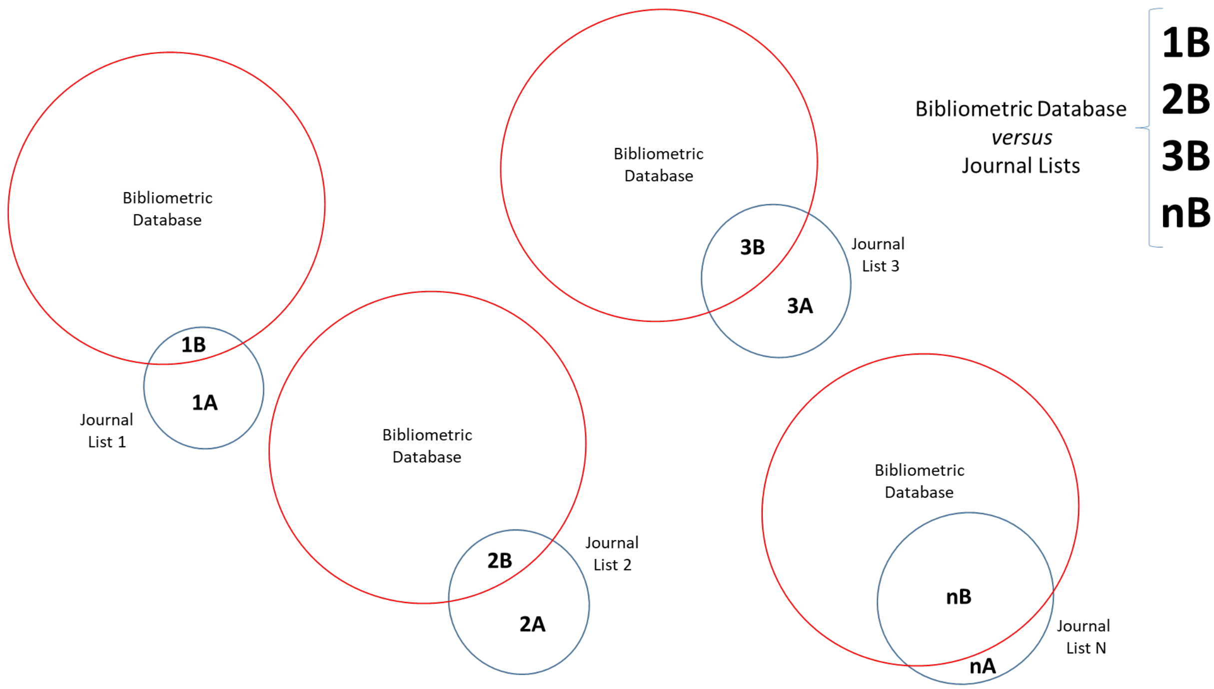
Source: The author.

Figure 6 presents N periodic evaluations, each of them considering a different Journal List. Bibliometric information is again obtained from the Bibliometric Database for each Journal List. To make the visualization easier, the framework was split into two phases.

5.1 FRAMEWORK PHASE 1

Framework Phase 1 is a comparison between Journal Lists and Bibliometric Database in each periodic evaluation (Figure 7).

Figure 7 – Journal Lists compared to Bibliometric databases



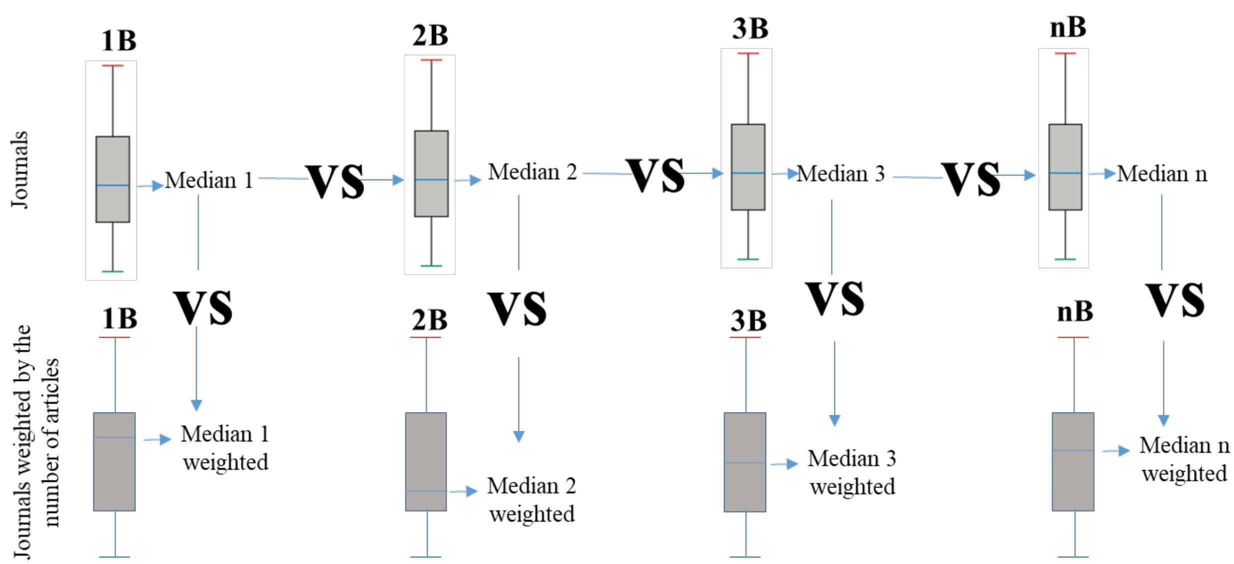
Source: The author.

This phase is composed of four steps. The first one in Phase 1 is tailored to identify whether the journals that are part of the Journal List are indexed in international bibliometric databases. The goal is to recognize patterns in scientific production as well as degrees of its coverage in the primary international databases. In Figure 7, 1B, 2B, 3B, and nB represent journals from the Journal Lists that are indexed in the leading international databases. Next, the percentage of these journals by the total journals in the Journal Lists are compared as follows:

% 1B	**vs**	**% 2B**	**vs**	**% 3B**	**vs**	**............**		**% nB**
Jornals List 1		**Jornals List 2**		**Jornals List 3**				**Jornals List N**

The second step in Phase 1 is based on the comparison of the bibliometric indicator distribution by the indexed journals of the ranking list (1B, 2B, 3B…nB), weighted and not weighted by the number of articles, between the periodic evaluations (Figure 8). One goal is to show if, in general, there are changes in the international impact of the journals chosen by faculty and students to publish their work between the analyzed periods. Another goal is to indicate if faculty and students are publishing more articles in journals with the highest or lowest international impact and how it changed over the periodic evaluations.

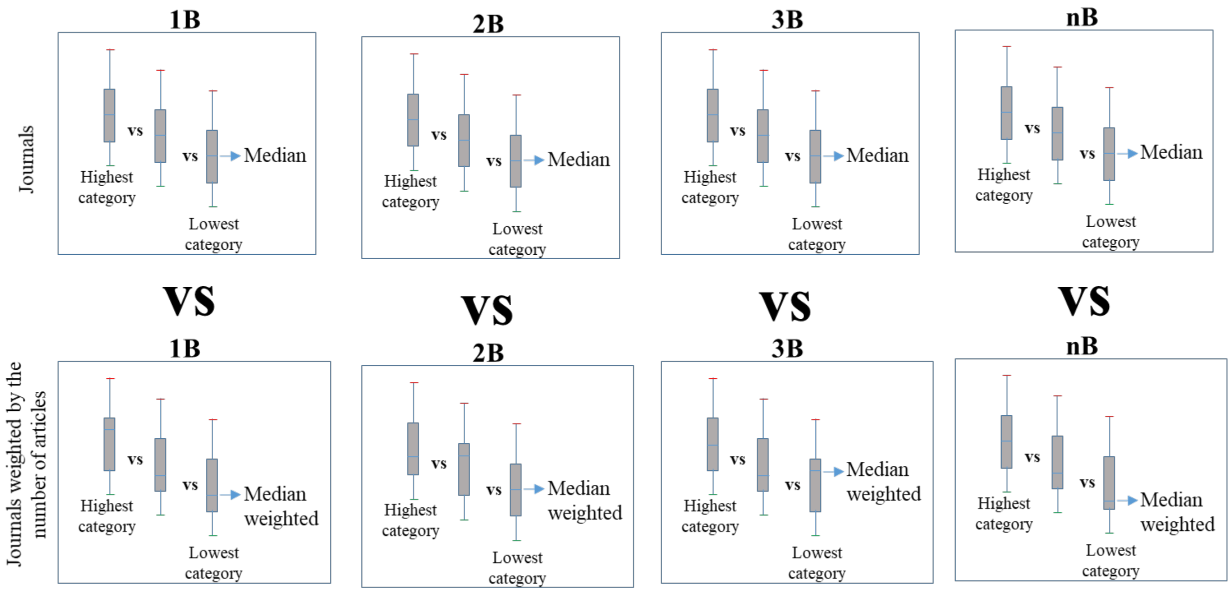
Figure 8 – Bibliometric indicator distribution by the indexed journals (1B, 2B, 3B…nB), weighted and not weighted by the number of articles, compared between the periodic evaluations



Source: The author.

The third step in Phase 1 refers to the comparison of the bibliometric indicator distribution among ranking categories of indexed journals, weighted or not by the number of articles, always considering a class and the one ranked immediately below it (Figure 9). One goal is to determine the level of agreement between bibliometric indicators and peer-review evaluation. Another goal is to check if the articles are more concentrated in journals with the highest or lowest international impact considering the same category.

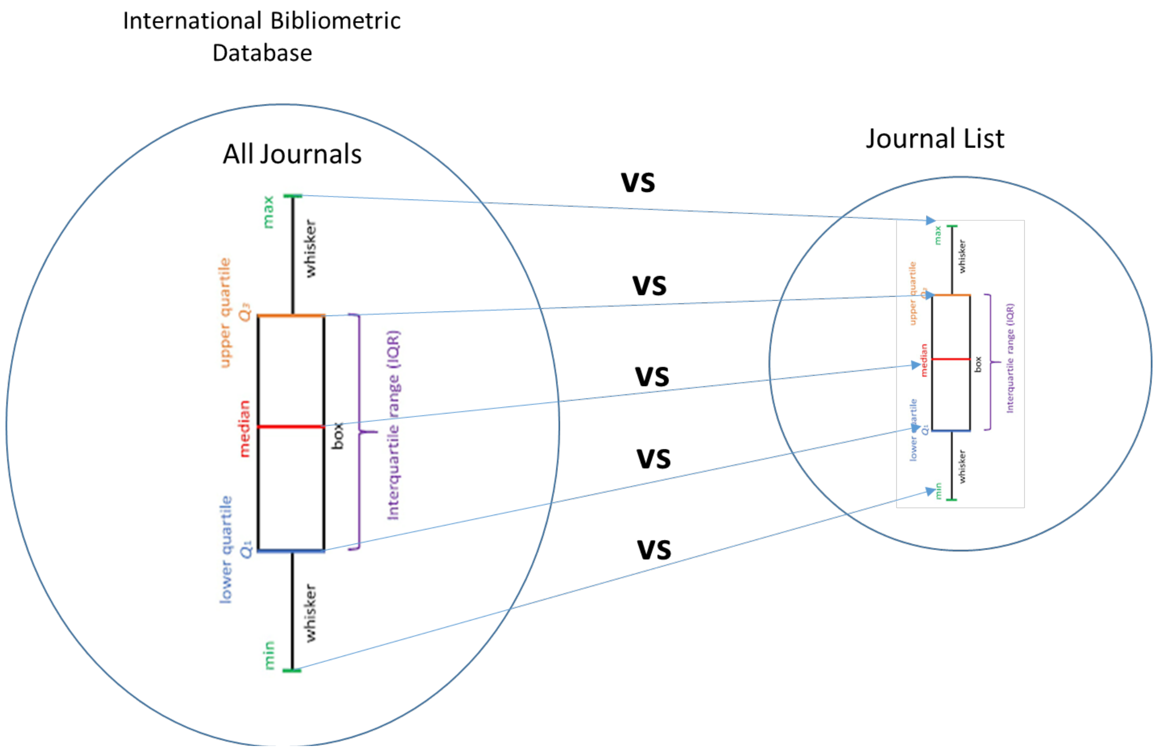
Figure 9 – Bibliometric indicator distribution among ranking categories, weighted and not by the number of articles, compared considering a class and the one ranked immediately below it



Source: The author.

The fourth step in Phase 1 has been designed to compare the bibliometric indicator distribution of the journals ranked in the national system with all the available journals in the international databases (Figure 10). The goal is to monitor the diversity of journals ranked in the national system, define in which percentiles of the international rankings these journals are more concentrated, and verify any changes related to the number of published articles in these journals.

Figure 10 – Bibliometric indicator distribution of the journals ranked in the national system versus all available journals in the international databases

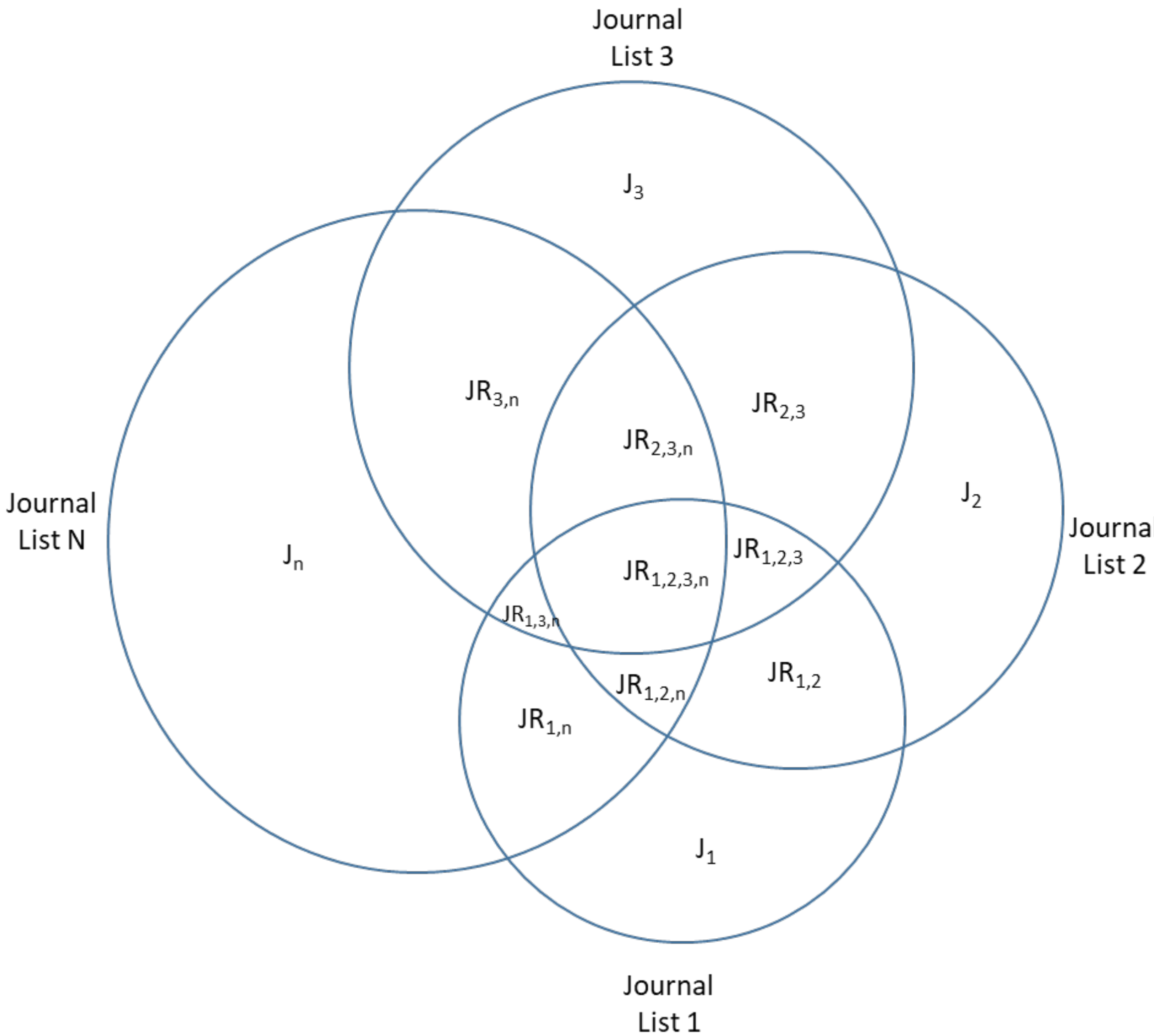


Source: The author.

5.2 FRAMEWORK PHASE 2

Framework Phase 2 is a comparison among the Journal Lists from different periodic evaluations. This phase has four steps. The first in Phase 2 is to group journals according to their frequency by periodic evaluation (Figure 11). The goal is to identify the journals that remained in the ranking lists (JR) for at least two periodic evaluations.

Figure 11 – Journals from different Journal Lists grouped according to their frequency by periodic evaluation



J_1: journals evaluated only in the first periodic evaluation; J_2: journals evaluated only in the second periodic evaluation; J_3: journals evaluated only in the third periodic evaluation; J_n: journals evaluated only in the n^{th} periodic evaluation.
$JR_{1,2}$: journals that remained in the system and were evaluated only in the first and second periodic evaluations.
$JR_{1,3}$: journals that remained in the system and were evaluated only in the first and third periodic evaluations.
$JR_{2,3}$: journals that remained in the system and were evaluated only in the second and third periodic evaluations.
$JR_{1,2,3}$: journals that remained in the system and were evaluated in all periodic evaluations.
$JR_{1,2,n}$: journals that remained in the system and were evaluated only in the first, second and $\mathbf{n}^{th}$ periodic evaluations.
$JR_{1,3,n}$: journals that remained in the system and were evaluated only in the first, third and $\mathbf{n}^{th}$ periodic evaluations
$JR_{2,3,n}$: journals that remained in the system and were evaluated only in the second, third and $\mathbf{n}^{th}$ periodic evaluations
$JR_{1,2,3,n}$: journals that remained in the system and were evaluated only in the first, second, third and $\mathbf{n}^{th}$ periodic evaluations

Source: The author.

The second step in Phase 2 is organized around a descriptive analysis of JR. The goal is to understand the profile of JR, such as nationality, insertion in the leading international databases, as well as how they are concentrating the articles over time.

The third step in Phase 2 is tailored to check changes in the category of JR and the frequency of their articles between the periodic evaluations. The goal is to verify any relationship between the maintenance or increase of QUALIS categories and changes in concentration of articles.

The fourth step in Phase 2 is carried out to compare the studied periods with the bibliometric indicator distribution of the indexed JR journals, weighted or not by the number of articles, and ranked in the highest categories in the last periodic evaluation. One goal is to test if the journal international impact increases or decreases, despite their first category being maintained or improved over periodic evaluations. Another goal is to verify improvements in the journal category with changes in their number of articles.

This complete framework for the analysis of scientific production behavior has been used to monitor Brazilian research output for the last 10 years. CAPES is responsible for the evaluation system of scientific production in Brazil, while it also manages a significant part of research funding in the country. Because the results of this evaluation are somehow linked to funds, monitoring this system is essential to ensure better results. Sections 5.4 and 5.5 detail the analysis carried out in the Brazil case.

5.3 DATA COLLECTION

For this study, we worked with data from eight different subject fields. The CAPES organization into subject fields was indicated previously in section 3.1, page 54. The criterion was to select the subject field with the highest number of graduate programs by major field. In the case of a tie, the program with the highest number of graduate courses was considered. This selection was made in August 2016, and the chosen subject fields as well as their number of graduate programs and courses are described in Table 1. Although Interdisciplinary subject field was among those selected, it was excluded from this study given its complexity.

Table 1 – Selected subject fields and their number of graduate programs and courses in 2016

		Graduate programs					Graduate courses			
Major field	Subject field	Academic master	PhD	Professional master	Academic master/PhD	Total	Academic master	PhD	Professional master	Total
Agrarian Sciences	Agrarian Science I	59	1	20	145	225	204	146	20	370
Biological Sciences II	Biological Science II	10	2	8	56	76	66	58	8	132
Health Sciences	Medicine II	17	3	15	71	106	88	74	15	177
Exact and Earth Sciences	Computer Science	33	2	11	32	78	65	34	11	110
Human Sciences	Education	56	0	47	74	177	130	74	45	249
Applied Social Sciences	Management	49	4	77	62	192	110	65	75	250
Engineering	Engineering III	41	1	29	57	128	98	58	29	185
Linugistics, Literature and Arts	Literature and Linguistics	56	1	9	91	157	147	92	9	248

Source:

The QUALIS system is the primary data source of this study. Thus, the journals used by Brazilian researchers and graduate students to publish their articles from 2007 to 2016 were considered. This period encompasses three periodic evaluations performed by CAPES: the Triennial Evaluation from 2007 to 2009; the Triennial Evaluation from 2010 to 2012; and the Quadrennial Evaluation from 2013 to 2016. In the present thesis, they will be referenced as the first, second, and third periodic evaluations, respectively. Furthermore, CAPES provided access to the journal ranking lists of the Scopus database regarding the years of 2010, 2013 and 2016.

The journals were organized in three QUALIS Lists per subject field, which were called as: QUALIS 1 – List of the first periodic evaluation, QUALIS 2, of the second and QUALIS 3, of the third. The design of each one of these lists followed the steps:

- only journals that published at least one article from a graduate program in the period were ranked[7];
- the last rank was considered when a journal received more than one QUALIS rank in the same periodic evaluation;
- the same journal had its number of articles summed inside each periodic evaluation;
- journals classified as C were excluded from the analysis.

5.4 SOURCE NORMALIZED IMPACT PER PAPER

---

[7] Some journals that did not publish any article from the graduate programs were ranked in the QUALIS 2 list. This fact was observed only for that list, which was a board decision to maintain the journals that were ranked in QUALIS 1.

The bibliometric indicator chosen for this study was SNIP, which was developed based on Scopus. The fact that Scopus is considered one of the most extensive and widespread databases in different knowledge fields (GUZ; RUSHCHITSKY, 2009) directed this choice. SNIP, introduced by Henk Moed in 2010, uses a normalized source approach to correct differences in citation practices among the scientific fields. It is defined as the ratio of the journal citation count per paper and the citation potential in its subject field. It aims to allow the direct comparison of sources in different subject fields. The strength of this approach is that it does not require a field classification system in which the boundaries of fields are explicitly defined. According to Moed (2010), SNIP is arguably the most sophisticated of all journal-level indicators, accounting for differences in citation potential between subject fields and allowing a direct comparison of journals across different subjects in an unbiased way. Other authors consider SNIP as an innovative measure, since it not only normalizes both the number of papers and field but it is also calculated based on a specific set of reference journals, instead of being defined beforehand somewhat arbitrarily (MINGERS; LEYDESDORFF, 2015b). SNIP indicators from 2010, 2013, and 2016 were retrieved from the SCImago Journal & Country Rank website for the QUALIS Lists 1, 2, and 3, respectively.

5.5 QUALIS LISTS VERSUS SCOPUS RANKING

Framework Phase 1 was applied for comparing the QUALIS Lists with Scopus ranking, considering each subject field. Regarding Step 1 in Phase 1, the journals ranked by the QUALIS system, also indexed in Scopus, considering three periodic evaluations, have been identified. For this purpose, some analyses were carried out:

- identification of journals of QUALIS indexed in Scopus;
- use of Excel version 16.3 to calculate the percentage of QUALIS journals indexed in Scopus in each periodic evaluation;
- use of Stata/IC version 15.0 software to conduct a growth curve considering the percentage results in each periodic evaluation.

Step 2 compared the distribution of SNIP values by the indexed QUALIS journals, weighted and not by the number of articles, between two periodic evaluations. The driven analyses in this step were:

- retrieving of the SNIP value for each journal from QUALIS Lists 1, 2 and 3;
- application of the Kruskal Wallis statistical test on Stata/IC version 15.0 software to compare the distribution of the indicator values in the journals among all periodic evaluations;
- application of Mann-Whitney's U Test (Wilcoxon Rank Sum Test) on Stata/IC version 15.0 software to identify which periodic evaluation had the highest SNIP ranking distribution. This test was applied once all distributions were non-normal;
- development of box plots on Stata/IC version 15.0 software. For each box plot, the upper bar is maximum observation; the lower bar is minimum observation; the middle bar is median value; the top of the box is the third quartile, and the bottom of the box is the first quartile. The outliers were excluded. The box plot shows a 95% confidence interval of the estimates.

Step 3 compared the distribution of SNIP values among the ranking categories of the indexed QUALIS journals, weighted and not by the number of articles. In this step, the performed analyses were:

- application of Mann-Whitney's U Test (Wilcoxon Rank Sum Test) on Stata/IC version 15.0 software to compare the distribution of SNIP values' between two categories. This comparison was between A1 and A2, A2 and B1, and so on. It was always between a category, and the one ranked immediately below it;
- application of Mann-Whitney's U Test (Wilcoxon Rank Sum Test) on Stata/IC version 15.0 software to compare the distribution of SNIP values in the sample, considering only the journals and the one weighted by the number of articles.

Finally, step 4 compared the distribution of SNIP values from the journals ranked in the QUALIS system, weighted or not by the number of articles, with all the available journals in Scopus. The analyses of this step were as follows:

- splitting of the Scopus journal-ranking list into four. One file by each top-level subject fields of Scopus, which are Life Sciences, Health Sciences, Physical Sciences, and Social Sciences;
- the journals considered in this analysis were those evaluated on the third periodic evaluation;
- application of Mann-Whitney's U Test (Wilcoxon Rank Sum Test) on Stata/IC version 15.0 software to compare the distribution of SNIP values of the indexed journals in the QUALIS List, weighted or not by the number of articles, with the journals of each Scopus file;
- application of Mann-Whitney's U Test (Wilcoxon Rank Sum Test) on Stata/IC version 15.0 software to compare the distribution of SNIP values of the indexed journals by A1 and A2 categories, weighted or not by the number of articles, with the journals of each Scopus file.

5.6 POTENTIAL EFFECTS LINKED TO THE QUALIS SYSTEM

Framework Phase 2 was applied for studying the potential effects of the QUALIS system. Step 1 compared QUALIS Lists 1, 2, and 3 to identify the frequency of each journal. Based on it, the journals were organized into the following groups:

- J_1: journals evaluated only in the first periodic evaluation;
- J_2: journals evaluated only in the second periodic evaluation;
- J_3, journals evaluated only in the third periodic evaluation;
- $JR_{1,2}$: journals that remained in the system and were evaluated only in the first and second periodic evaluations;
- $JR_{1,3}$: journals that remained in the system and were evaluated only in the first and third periodic evaluations;
- $JR_{2,3}$: journals that remained in the system and were evaluated only in the second and third periodic evaluations;
- $JR_{1,2,3}$: journals that remained in the system and were evaluated in all periodic evaluations.

It is noteworthy that each journal belongs to only one of these groups. In order to investigate the possible effects of lists, this study considered only journals that remained in the system for at least two consecutive periodic evaluations. However, as $JR_{1,2}$ journals appeared only in the first two periodic evaluations and, therefore, were no longer used by the academic community in the subsequent period, this set was excluded from the analyses. Hence, $JR_{2,3}$ and $JR_{1,2,3}$ were the investigated groups considered in the next steps.

Regarding step 2, descriptive analyses were performed for $JR_{2,3}$ and $JR_{1,2,3}$, such as their nationality, frequency of articles, and proportion inserted in Scopus. Step 3 investigates the shift of $JR_{1,2,3}$ and $JR_{2,3}$ journals among the QUALIS categories, as well as changes in their frequency of articles, over the periodic evaluations. Thus, firstly, we identified the initial (QUALIS 1 for $JR_{1,2,3}$ and QUALIS 2 for $JR_{2,3}$) and final (QUALIS 3) QUALIS categories. Secondly, A1 to B5 categories were coded from 7 to 1. Next, we subtracted the initial code from the latter. These differences ranged from -6 to 6, considering all the possible movements among the categories. The result -6 to -1 represents journals that had their QUALIS categories reduced over time, while 0 means no change in category, and 1 to 6 shows the journals that had an improvement of their quality according to peer review judgment. The graphs were plotted in Tableau Desktop 2018.3.

Step 4 examined only $JR_{1,2,3}$, and $JR_{2,3}$ indexed in Scopus ranked as A1, A2, or B1 in the last periodic evaluation and weighted or not by the number of articles. The SNIP value distributions of these journals, as well as their frequency of articles, were compared among the three periodic evaluations. For this test, Mann-Whitney's U Test (Wilcoxon Rank Sum Test) was applied using the Stata/IC version 15.0 software to compare the distribution of SNIP values.

6 RESULTS AND DISCUSSION

The results of this study are presented in two sections, 6.1 and 6.2. The purpose is to apply the developed framework for identifying patterns as well as potential effects of the QUALIS system on the Brazilian scientific production behavior. In the first section, this study shows an overview of the Brazilian journal evaluation system in terms of international impact from 2007 to 2016, by using Framework Phase 1. In the second section, Framework Phase 2 is applied to identify journals that remained in the system over time and study them regarding the shift in their QUALIS categories and concentration of articles over the periodic evaluations.

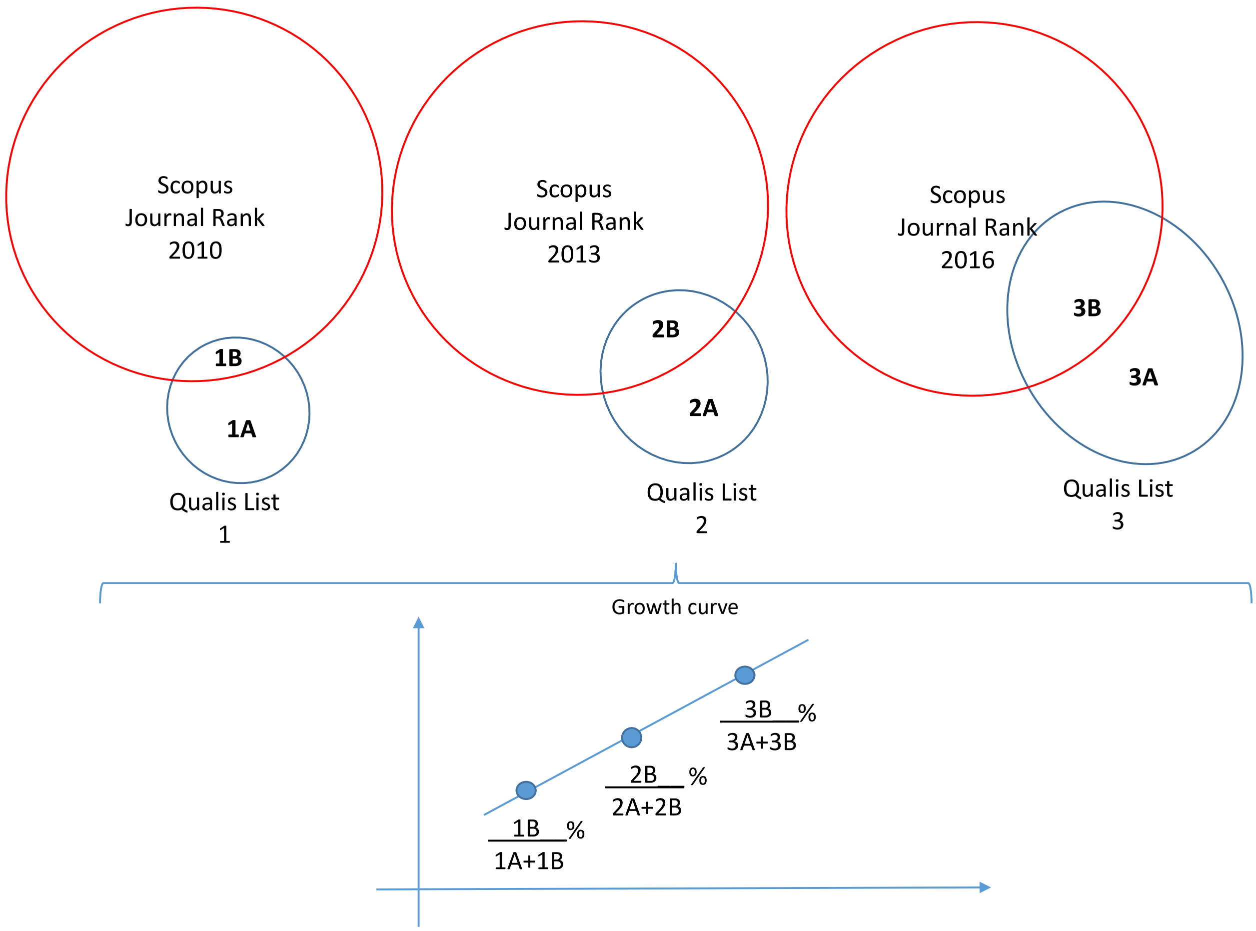
6.1 QUALIS LISTS *VERSUS* SCOPUS RANK

Framework Phase 1 was applied for this comparison. Each following subsection represents a Phase 1 step as described in Section 5.5.

6.1.1 Overview of journals ranked by QUALIS and indexed in Scopus

Step 1 of Framework Phase 1 is employed for this overview. The identification of the ranked journals in the QUALIS system, also indexed in Scopus, is represented in Figure 12, as well as a growth curve of their percentages.

Figure 12 – Step 1 in Phase 1 applied to the QUALIS system



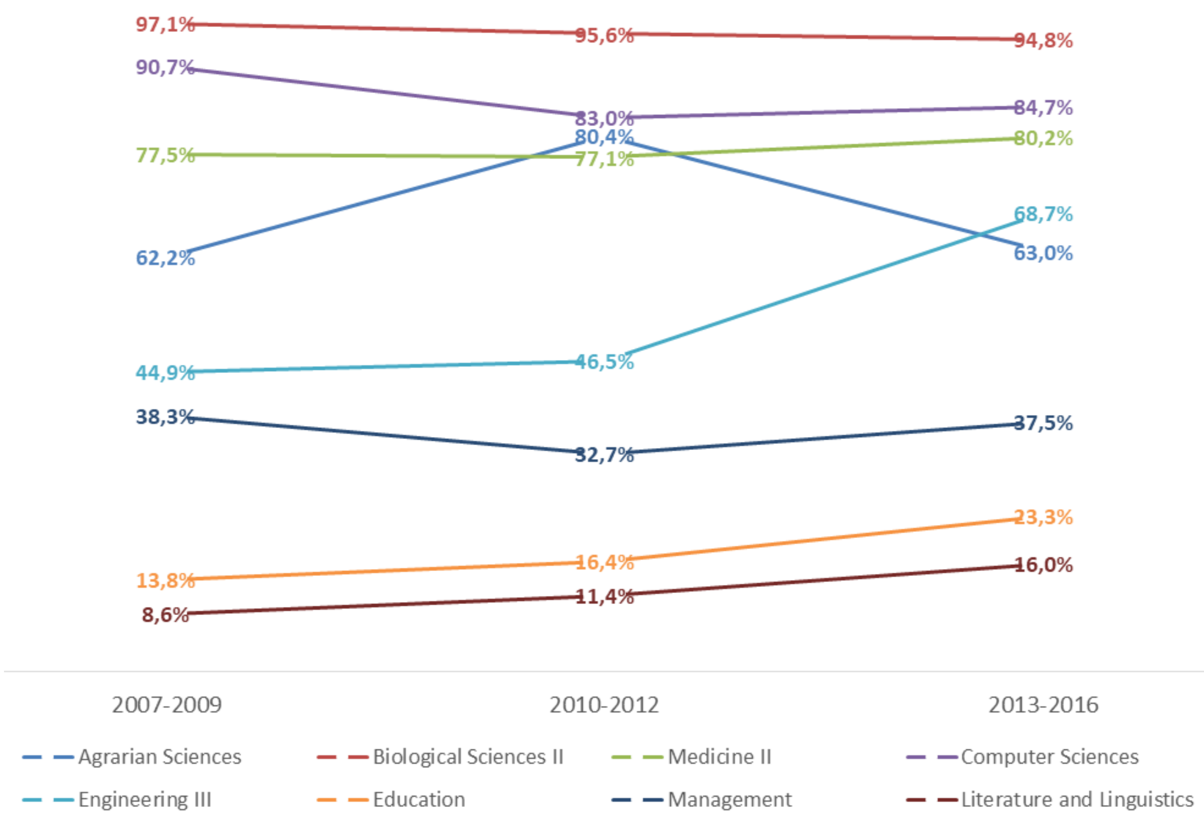
Source: The author.

Journals indexed by Scopus have greater visibility and potential for international dissemination. The presence of journals in a broader bibliographic database, such as Scopus, allows identifying which subject fields share a tradition of publication in indexed journals. However, in some of these subject fields, faculty and students do not direct a significant part of their scientific production to indexed journals as recognized in the Brazilian case (BARATA *et al*. 2014). The framework use showed that the percentage of journals adopted by the Brazilian academic community to publish its work, which was also indexed in Scopus, varied among the subject fields, as well as over ten years (2007-2016) (Figure 13). The subject fields in which research is generally more regionally and nationally engaged – Agrarian Sciences I, Management, Education and Literature, and Linguistics – displayed the lowest percentages of journals indexed in Scopus. Although Education, Literature, and Linguistics revealed an increase in their rate over the years, it is still less than 25% for both subject fields in the third periodic evaluation.

In literature review, a study about the weight of international publications to the whole Brazilian productivity found that Agriculture, Humanities, Literature and Linguistics, Health Sciences, and Social Sciences are stronger fields among researchers with national productivity. On the other hand, Biology, Engineering, and Earth and Exact Sciences are more representatives among those with international productivity (LEITE; MUGNAINI; LETA, 2011). These findings are consistent with the present results. Likewise, a previous study about the Slovenian National System for the Evaluation of Research (SICRIS) from 1996 to 2011 revealed a continuous growth in agricultural, medical (and health), natural, engineering and technology sciences in Scopus and WoS. The total number of articles was the highest in natural sciences, indicating specific publishing patterns in this scientific field, and the agricultural sciences presented the lowest figures. On the other hand, social sciences displayed accelerated growth after 2006 in both databases (BARTOL *et al*., 2014).

Additionally, a previous research in Norway analyzed data coverage from 2015 and 2016 publications by domain. It showed that 48% for Social Sciences and 27% for Humanities of all peer-reviewed scientific papers in the CRISTIN were found in Scopus (SIVERTSEN, 2018). The same study states that these deficiencies in coverage of the SSH happen mainly due to incomplete coverage of the international journals, limited or no coverage of national scholarly journals, and minimal coverage of peer-reviewed academic books. Although this Norwegian study focuses on counted paper, and in our research, journals are considered (Figure 13), both countries presented percentages in some way similar.

Figure 13 – Percentage of journals classified by the QUALIS system and indexed in Scopus per evaluation periods and subject field

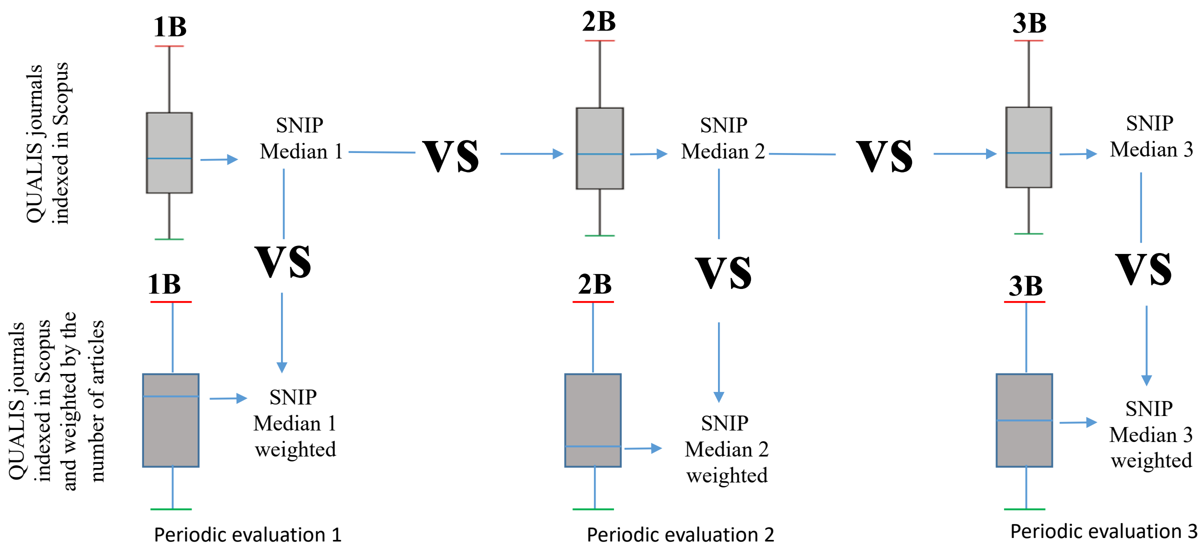


Source: The author.

6.1.2 Distribution of the SNIP values by journals and articles of QUALIS

Once the specific publishing patterns of subject fields are noted, as well as their percentage of journals indexed in Scopus, the next step was to investigate the impact of these journals on the periodic evaluations. Hence, the SNIP indicator was used for evaluating it. Thus, step 2 in Phase 1, as arranged in Figure 14, was employed to analyze the distribution of SNIP values by the indexed journals of QUALIS, weighted or not by the number of articles, compared between two periodic evaluations.

Figure 14 – Step 2 in Phase 1 applied to the QUALIS system

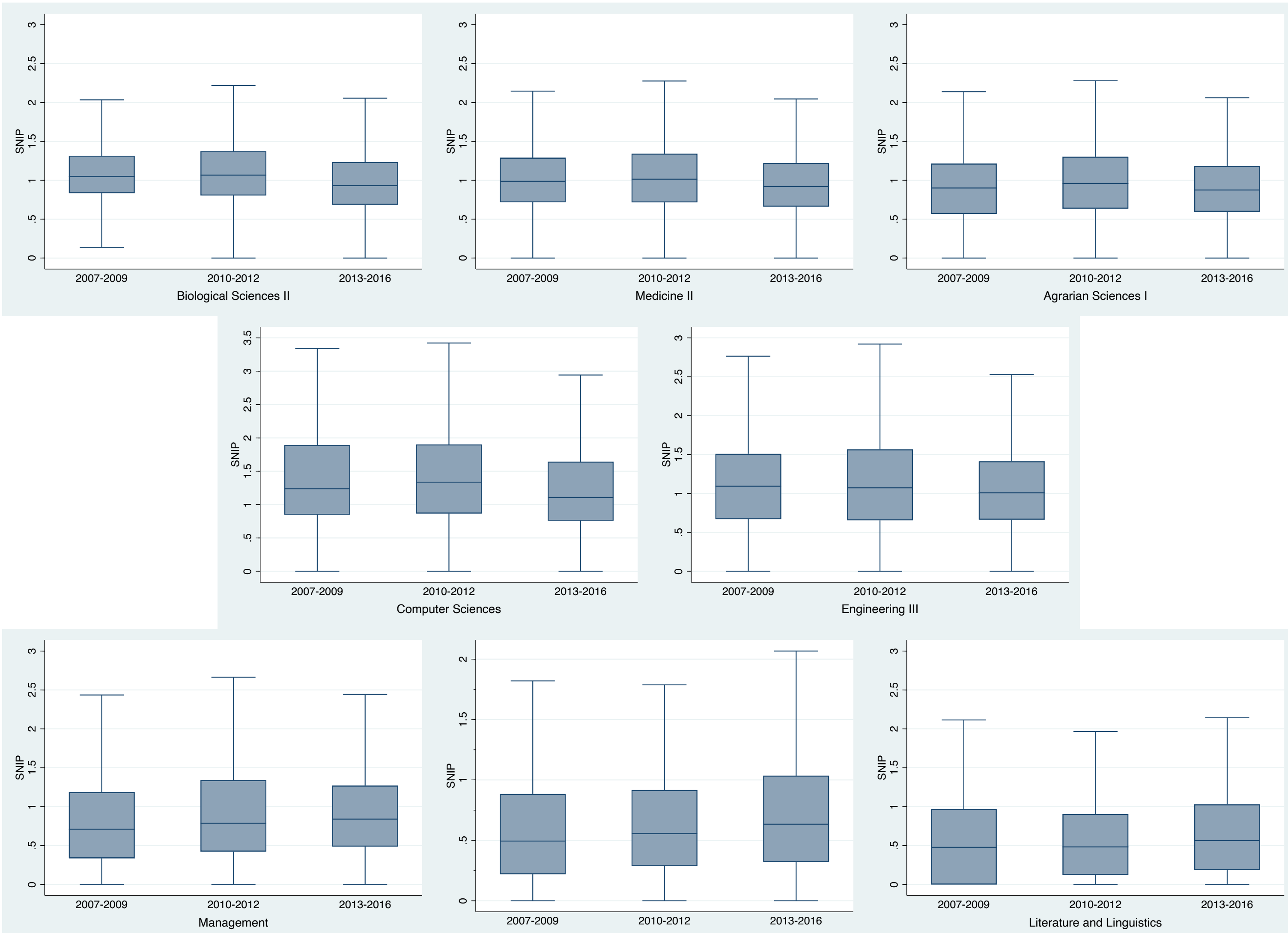


Source: The author.

The use of Step 2 produced Figure 15, which shows the distribution of the SNIP values by journals, periodic evaluations, and subject fields of QUALIS. The Kruskal-Wallis test demonstrated statistically significant differences between the three periodic evaluations in all subject fields, except for Literature and Linguistics ($p=0.196$). Additionally, the Wilcoxon-Mann-Whitney test presented that these differences were statistically lower in the third periodic evaluation compared to the first for Biological Sciences II ($p<0.001$), Medicine II ($p<0.001$), Engineering III ($p=0.02$), and Computer Sciences ($p<0.001$). However, they were higher for Management ($p<0.01$) and Education ($p<0.01$). In general, therefore, these results indicate that there was a decline or no change in the impact of journals chosen by faculty and students to publish their work from 2007 to 2016.

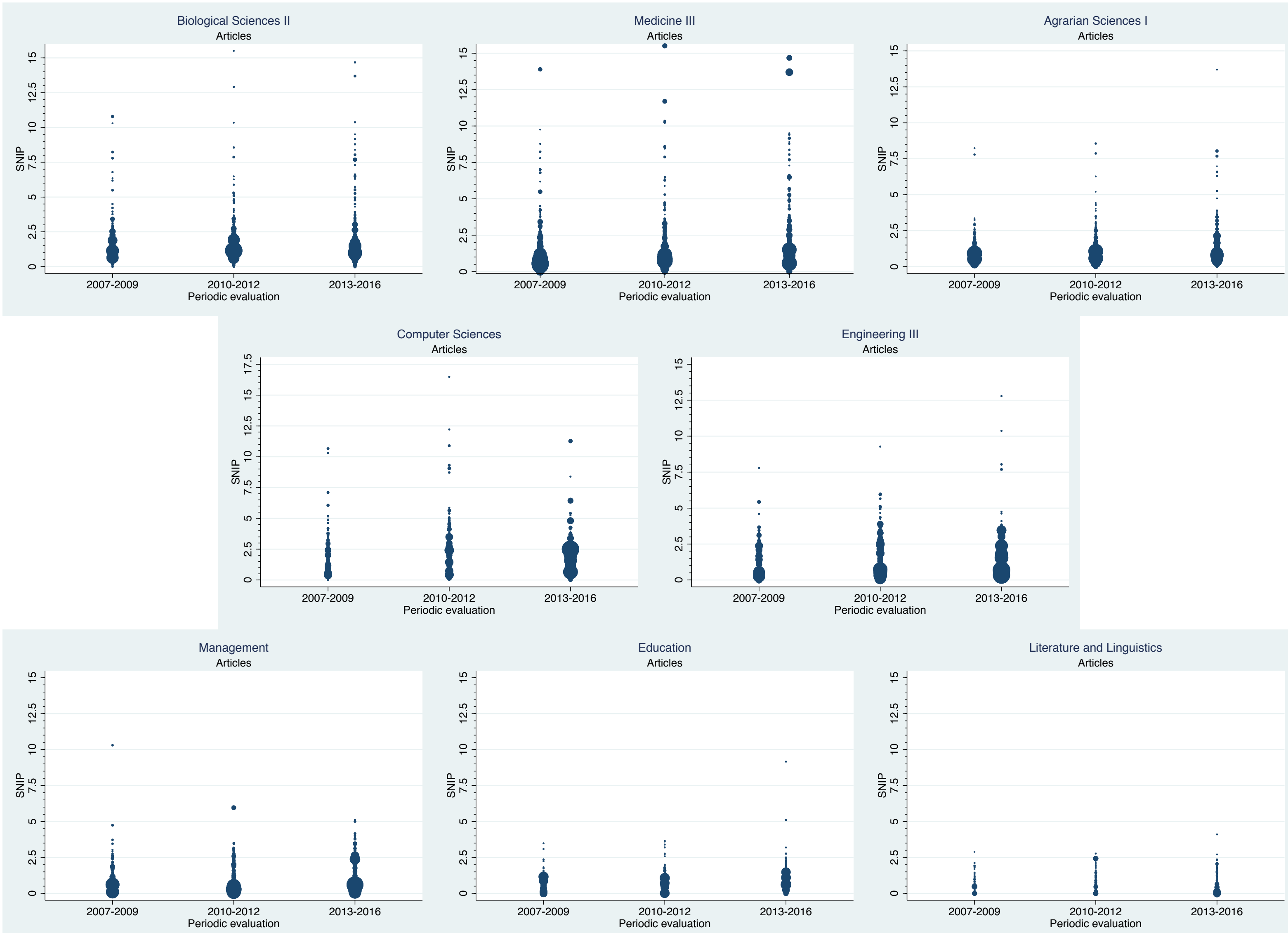
Figure 16 presents the distribution of SNIP values by the articles, periodic evaluation, and subject fields of QUALIS. After comparing the distribution of SNIP values in Figures 15 and 16, Agrarian Sciences I, Management, Literature and Linguistics showed a statistically significant higher concentration of articles in those journals with the lowest SNIP values in all periodic evaluations. Although journals from Computer Sciences presented a decrease in the distribution of their SNIP values in the third periodic evaluation compared to the first (Figure 15), the concentration of articles was statistically significant and higher in journals with the highest impact in the second and third evaluation periods (Figure 16). No significant difference was found for the other subject fields. These outcomes indicate that in the more regionally and nationally engaged subject fields, faculty and students have, in general, published in journals with lower impact in all periodic evaluations.

Figure 15 – Distribution of the SNIP values by journals, periodic evaluations, and subject fields of QUALIS



Source: The author.

Figure 16 – Distribution of the SNIP values by articles, periodic evaluations, and subject fields of QUALIS

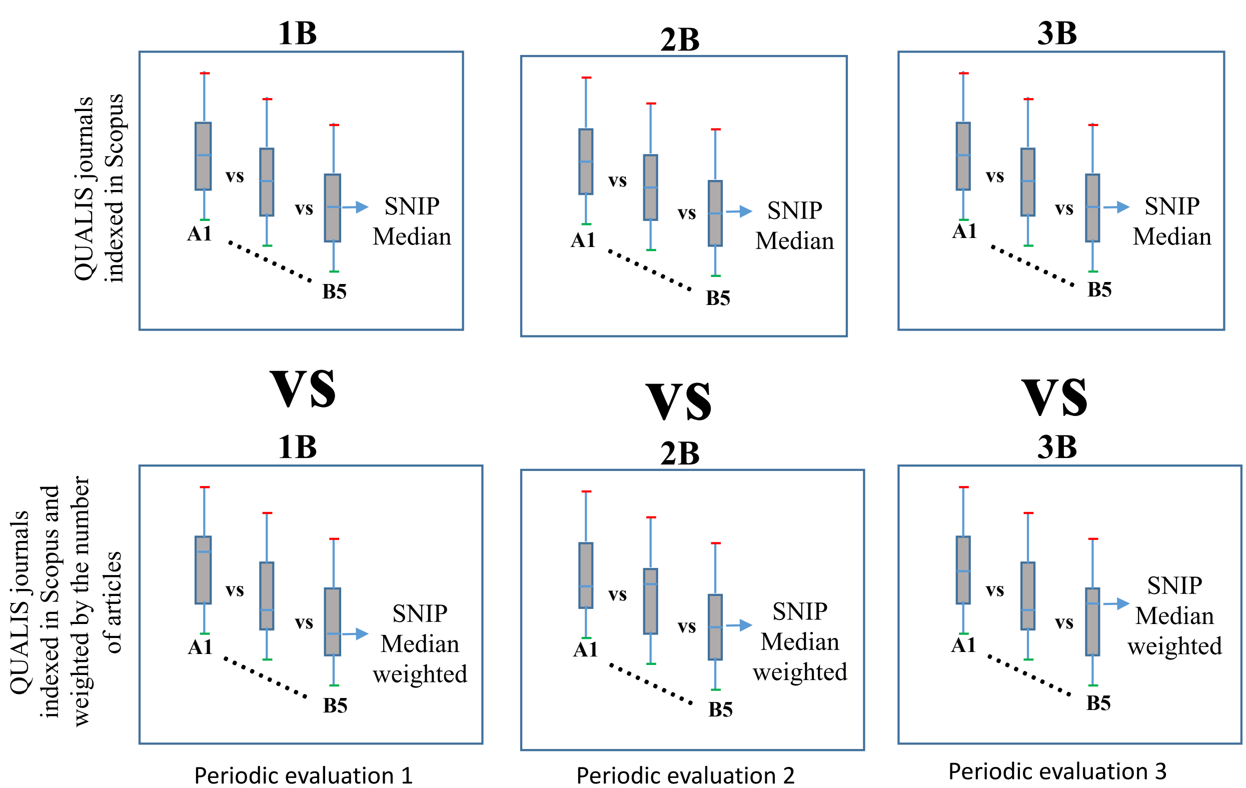


Source: The author.

6.1.3 Distribution of the SNIP values by journals and articles of QUALIS, considering the rank categories from A1 to B5

The analyses that led to the results reported in this subsection were described in Step 3 of Phase 1. As introduced in Figure 17, the distribution of SNIP values among the ranking categories of the indexed journals of QUALIS, weighted or not by the number of articles, has been compared.

Figure 17 – Step 3 in Phase 1 applied to the QUALIS system



Source: The author.

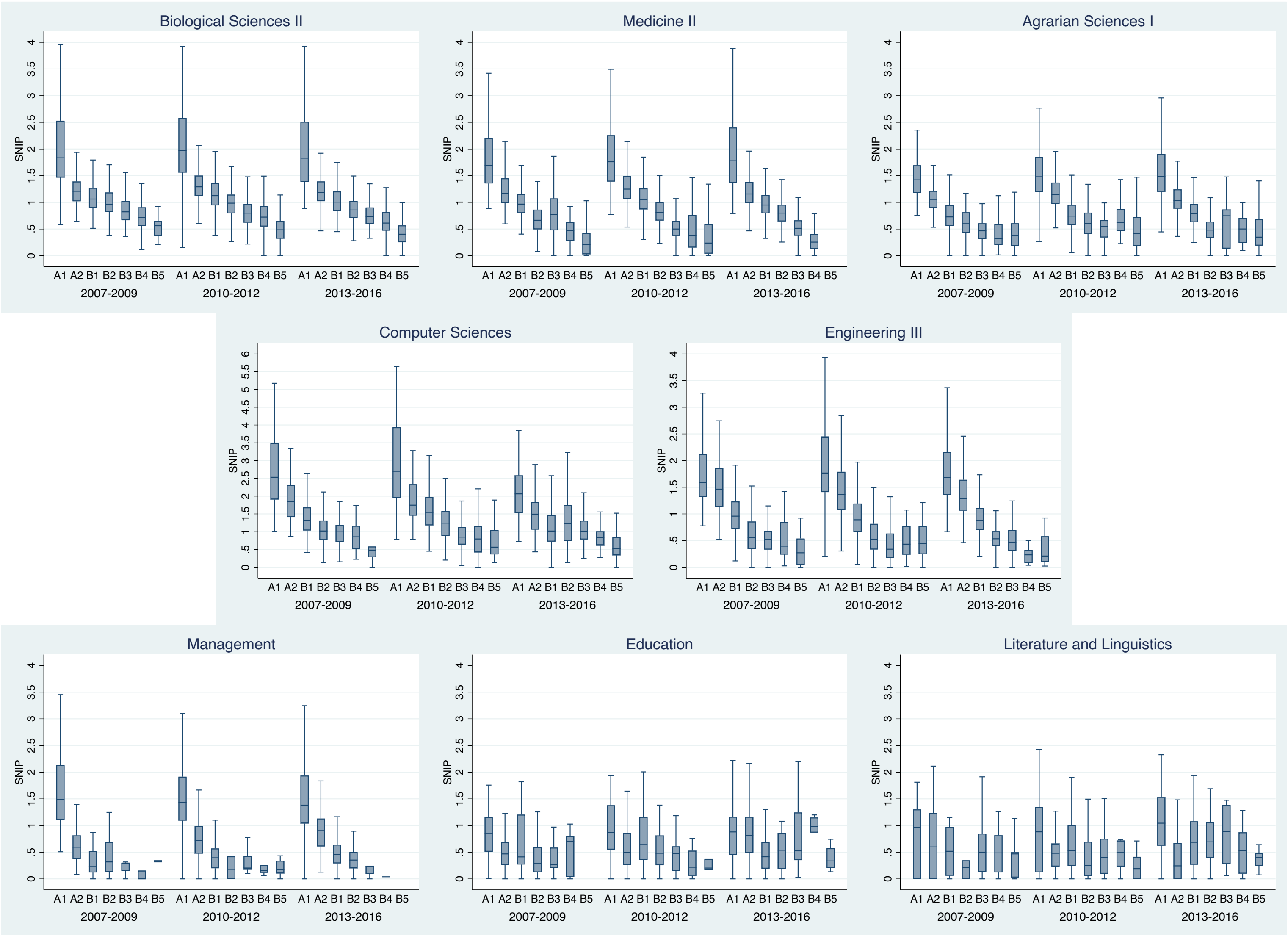
The use of Step 3 provided Figures 18 and 19. Figure 18 presents the distributions of the SNIP values by the journals, periodic evaluations, and subject fields of QUALIS, considering the rank categories from A1 to B5. A comparison of the SNIP values between A1 and A2 categories, A2 and B1, and so on, always between a category and the one ranked immediately below it, showed that only Biological Sciences II had a significant difference in the comparisons in all periodic evaluations. Medicine II displayed a similar result, except in the first periodic evaluation. Furthermore, although those two subject fields did not adopt a normalized indicator in their evaluation processes, the committees could normalize all the journals efficiently based on their subfields. On the other hand, the same investigation for Education, Literature and Linguistics did not follow a pattern concerning SNIP indicator values and the categories in QUALIS. In regard to the other subject fields, what most drew our attention was the presence of high-impact journals classified from B2 to B5 categories, whose impact is sometimes equal or higher than those classified from A1 to B1. In such situation, Computer Sciences stands out in the third periodic evaluation, with a significantly higher SNIP median in B2 compared with B1.

Regarding these results, CAPES implemented a rule in the evaluation process, in which the number of journals in A1 and A2 categories needs to be lower than 25%, considering all journals evaluated in each subject field and A1 and A2 categories plus B1 category, smaller than 50%. This rule could contribute to the observed outcomes, as each subject field selects the 50% most relevant journals to it, ranking them from A1 to B1. By following these standards, the evaluators can leave out from the upper strata those journals with high impact but that are not considered pertinent to the subject field as the chosen ones. This fact may bring significant consequences for subject fields in the frontier of knowledge as these journals may not be well evaluated by peer reviewer. From these results, most often there is no direct relationship between peer-review evaluation and international impact indicators in the Brazilian case. This also demonstrates different ways of evaluating the journals by subject fields. For example, Education, Literature and Linguistics do not take any international impact indicators into account. Dissimilarly, Biological Sciences II only values publications in the main databases. Furthermore, it was possible to infer that peer review has been one the major determinants in the QUALIS ranking. At the end, peer review decides which journals should be ranked in the highest and lowest categories, regardless of their international impact, whereas this is more evident in some subject fields.

In most of the subject fields, the A1 category (Figure 18) displayed the strongest skewness with a long tail on the right in all periodic evaluations. Skewness and long right tails are a notorious feature of bibliometric indicators in sciences, particularly for individual scientists or articles (SEGLEN, 1992) and across journals (STERN, 2013; BERTOCCHI *et al*., 2015). This result suggests that most journals are concentrated towards the lower limit in the A1 category. No changes were observed concerning the distribution of SNIP values in the A1 category from 2007-2016.

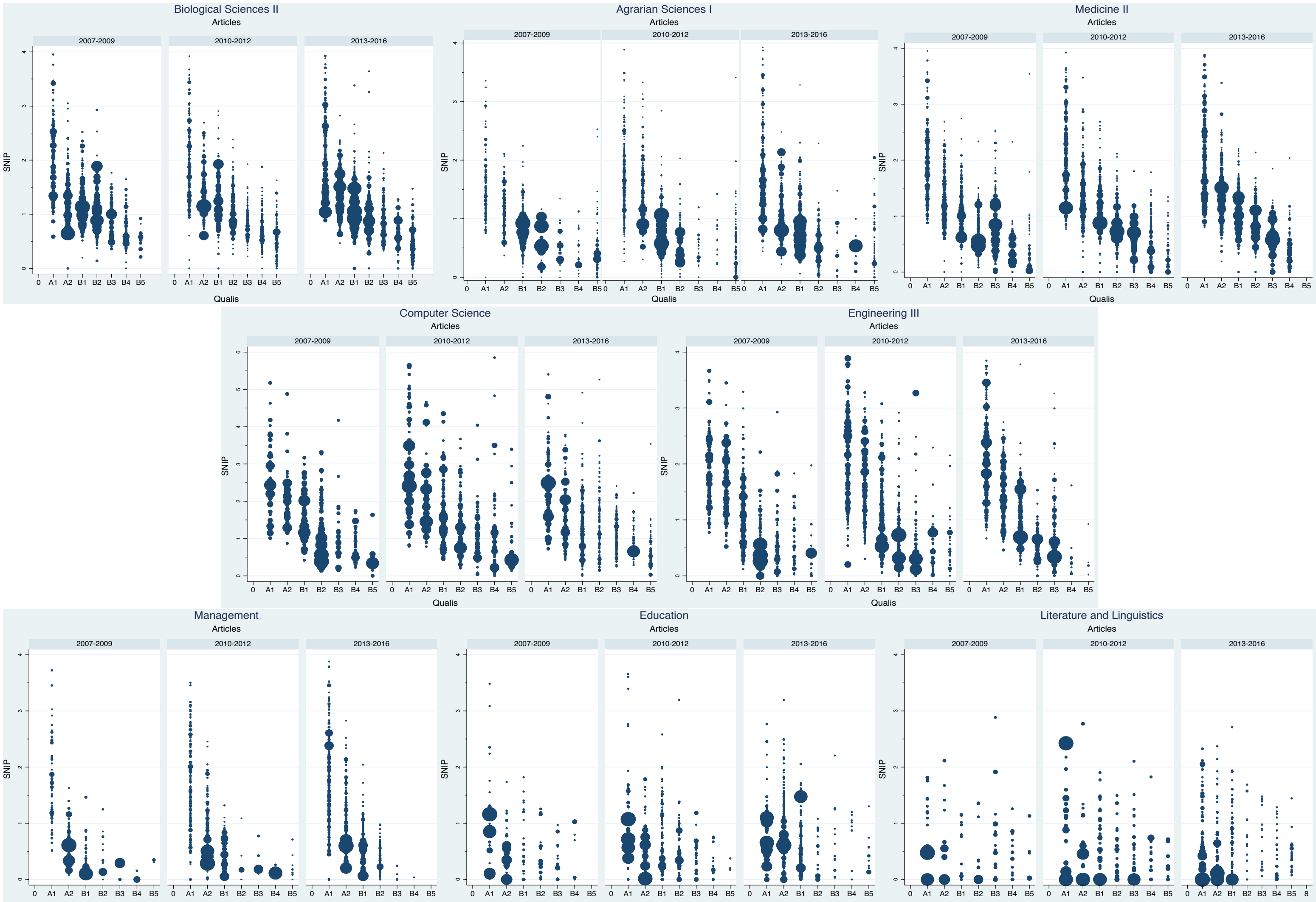
Figure 19 presents the distribution of SNIP values by the articles, periodic evaluations, and subject fields of QUALIS, considering the rank categories from A1 to B5. It also provides a visualization of the concentration of articles in each rank category and periodic evaluation. Given the high number of analyzed subject fields as well as rank categories, comparisons between Figures 18 and 19 were made only considering journals ranked in A1 and A2 categories in the third periodic evaluation. Thus, Agrarian Sciences I, Literature and Linguistics were the only subject fields to show a statistically significant higher concentration of articles in those journals with the lowest SNIP values in both categories. In contrast, Engineering III was the only one to concentrate most of its articles in journals with the highest SNIP values in both categories. Considering only A1, there were more articles on journals with the lowest SNIP values for Biological Sciences II, and with the highest SNIP values for Management. No significant difference was seen for the other fields. In regard to only A2, Management and Education showed more concentration of articles in journals with the lowest SNIP values; Medicine III in journals with the highest values; and non-significant difference was found in the other subject fields.

Figure 18 – Distribution of the SNIP values by the journals, periodic evaluations and subject fields of QUALIS, considering the rank categories from A1 to B5



Source: The author.

Figure 19 – Distribution of the SNIP values by articles, periodic evaluations and subject fields of QUALIS, considering the rank categories from A1 to B5

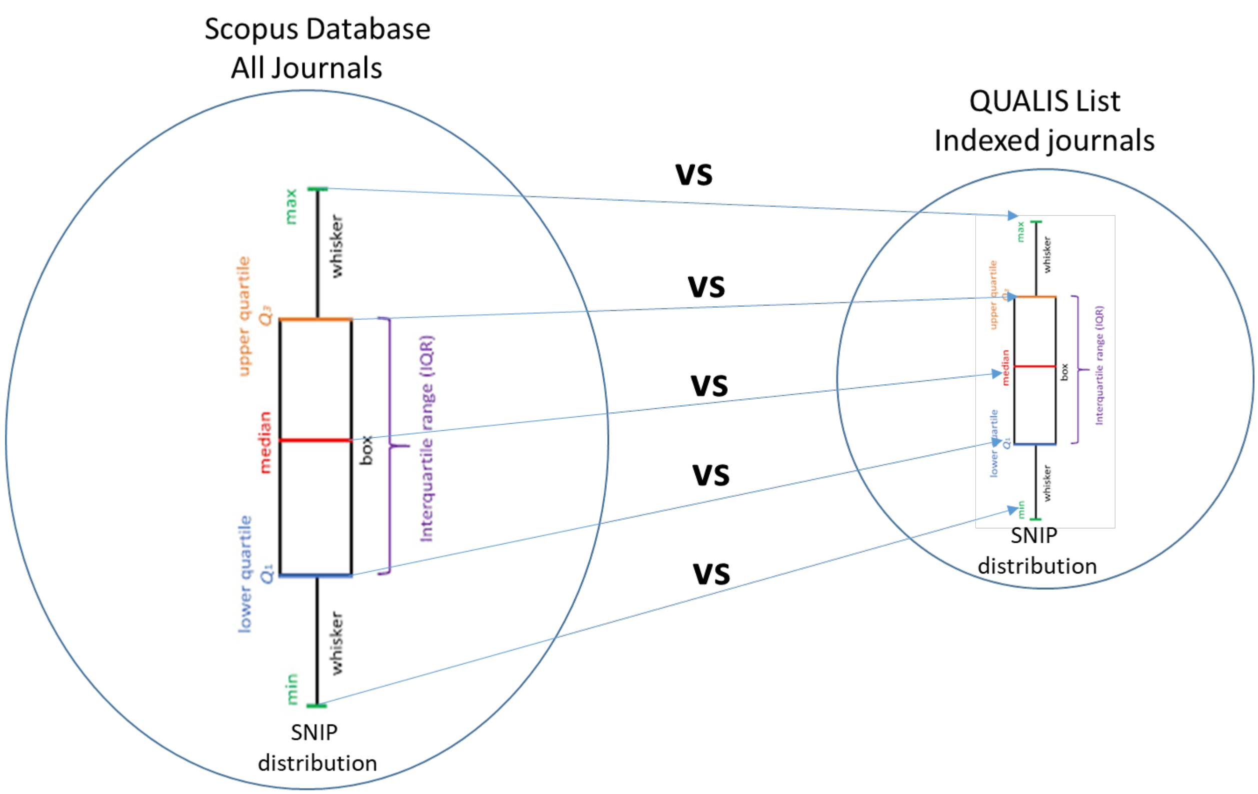


Source: The author.

6.1.4 The percentiles of QUALIS compared to those of SCImago Journal Rank

The last analyses allowed comparisons among the journals evaluated by QUALIS, considering the different rank categories and periodic evaluations. In general, the results showed that the evaluation system seems to evidence no significant changes in terms of the impact of journals chosen by faculty and students to publish their work in Brazil. This scenario was worse in some subject fields. Hence, these findings led us to two questions: In which percentiles of Scopus are these journals selected by Brazilian researchers to disseminate their work? Does this improve when the journals ranked in the highest categories of QUALIS (A1 and A2) are considered? In order to answer the last questions, step 4 of Phase 1 was applied (Figure 20). Thus, the SNIP distributions for all indexed journals, from lists of QUALIS in the third periodic evaluation, were compared to all journals from Scopus by percentiles.

Figure 20 – Step 4 in Phase 1 applied to the QUALIS system



Source: The author.

In order to follow the analyses in Step 4, the journals were split by top-level subject fields of Scopus, which are Life Sciences, Health Sciences, Physical Sciences, and Social Sciences. The results are established in Table 2, which presents the distributions of SNIP for all journals in the QUALIS Lists, Scopus by top-level subject fields, and QUALIS A1 plus A2 categories, providing descriptive measures based on percentiles. This table also presents the journals of QUALIS weighted by the number of articles. Overall, a low diversity of journals was observed when comparing the ones ranked by QUALIS with the variety of journals in Scopus, according to N values. This scenario is even more evident for subject fields from SSH.

Nevertheless, all SNIP values of QUALIS journals are distributed in all Scopus rank. This result means that despite the low diversity of QUALIS journals, the evaluation policy has led to an international showcase of the Brazilian scientific production, even in subject fields from SSH. Considering the weighted journals, the percentile limits are lower than those of Scopus rank for Agrarian Sciences I, Management, Education, Literature and Linguistics. Thus, once more, the articles are focused on journals with the lowest international impact in these fields.

Table 2 – Distribution of SNIP by journals of QUALIS and Scopus journals

		Journals or Articles	N	IQR	p25	p50	p75	p90	p99
Scopus top-level subject fields	Life Sciences	All journals	4797	0,626	0,401	0,716	1,027	1,427	3,631
Brazilian subject fields	Biological Sciences II	All journals	2595	0,549	0,686	0,932	1,235	1,676	4,763
		All weighted journals	19957	0,408	0,823	0,977	1,231	1,508	2,941
		A1+A2 journals	663	0,773	1,131	1,383	1,904	2,685	8,791
		A1+A2 weighted journals	6168	0,459	1,103	1,299	1,562	1,942	4,894
	Agrarian Sciences I	All journals	1926	0,587	0,596	0,875	1,183	1,631	3,405
		All weighted journals	29334	0,436	0,568	0,802	1,004	1,516	2,621
		A1+A2 journals	780	0,617	0,968	1,220	1,585	2,101	4,742
		A1+A2 weighted journals	10065	0,669	0,886	1,139	1,555	1,885	3,454
Scopus top-level subject fields	Health Sciences	All journals	6911	0,750	0,262	0,642	1,012	1,403	3,307
Brazilian subject field	Medicine II	All journals	3093	0,559	0,663	0,921	1,222	1,669	4,309
		All weighted journals	23494	0,550	0,707	0,967	1,257	1,589	4,959
		A1+A2 journals	920	0,707	1,091	1,346	1,798	2,487	8,039
		A1+A2 weighted journals	7537	0,568	1,120	1,339	1,688	2,482	13,698
Scopus top-level subject fields	Physical Sciences	All journals	7705	0,730	0,373	0,702	1,103	1,684	3,847
Brazilian subject field	Computer Sciences	All journals	1070	0,885	0,758	1,107	1,643	2,239	4,101
		All weighted journals	5338	1,133	0,906	1,393	2,039	2,525	4,807
		A1+A2 journals	301	0,998	1,239	1,725	2,237	3,099	5,404
		A1+A2 weighted journals	3037	1,204	1,288	1,842	2,492	3,102	6,438
	Engineering III	All journals	1819	0,749	0,664	1,008	1,413	1,921	3,332
		All weighted journals	10418	0,990	0,655	1,119	1,645	2,091	3,454
		A1+A2 journals	706	0,743	1,173	1,483	1,916	2,405	3,847
		A1+A2 weighted journals	10418	0,746	1,305	1,676	2,051	2,393	3,454
Scopus top-level subject fields	Social Sciences	All journals	8967	0,780	0,217	0,567	0,997	1,483	3,181
Brazilian subject field	Management	All journals	940	0,783	0,487	0,840	1,270	1,868	3,454
		All weighted journals	4687	0,637	0,429	0,609	1,066	1,837	3,080
		A1+A2 journals	635	0,771	0,752	1,075	1,523	2,123	3,521
		A1+A2 weighted journals	3221	0,745	0,599	0,749	1,344	2,152	3,213
	Education	All journals	482	0,713	0,322	0,634	1,035	1,401	2,492
		All weighted journals	2597	0,556	0,384	0,618	0,940	1,100	1,832
		A1+A2 journals	272	0,682	0,483	0,836	1,165	1,530	2,767
		A1+A2 weighted journals	1930	0,479	0,526	0,657	1,005	1,100	1,905
	Literature and Linguistics	All journals	258	0,845	0,184	0,564	1,029	1,532	2,372
		All weighted journals	1175	0,505	0,000	0,112	0,505	1,005	2,051
		A1+A2 journals	135	0,978	0,132	0,584	1,110	1,742	2,372
		A1+A2 weighted journals	918	0,421	0,000	0,112	0,421	0,819	2,051

Source: The author.

Regarding the journals in A1 and A2 categories, 25% of them in Biological Sciences II and Agrarian Sciences I were among the Top 10% journals of Scopus in Life Sciences. This pattern was maintained as well for the weighted journals. The same was observed for Medicine II when compared to Scopus journals in Health Sciences, and for Computer Sciences and Engineering III once matched with those in Physical Sciences. Considering the subject fields in Social Sciences, 25% of the journals ranked as A1 and A2 were among the top 10% of Scopus only for Management. Once the journals weighted by articles were reviewed, the pattern was different for that subject field, showing fewer articles in that percentile as demonstrated by the SNIP value decrease. For Education, Literature and Linguistics, only 10% of the journals ranked as A1 and A2 were among the top 10% of Scopus, but less than 10% of the articles were focused on those journals. Overall, even considering the journals in the highest categories of QUALIS, the studied subject fields of SSH displayed fewer journals and articles in the top percentiles of Scopus (10 and 1%).

After comparing the percentile limit values from A1 and A2 journals, the journals, when weighted by articles, had a decrease in their SNIP value for all fields, except for Computer Sciences and Engineering III. Therefore, these results demonstrated that, in general, the articles concentrated more on journals with the lowest international impact. In addition, if considering the representativeness of indexed journals among those ranked as A1 and A2, the impact of journals was even lower in some subject fields. Thus, only the indexed journals account for 25% of all journals evaluated in Biological Sciences II and Medicine II as A1 and A2, 27% in Computer Sciences, 38% in Engineering III, 40% in Agrarian Sciences, 52% in Literature and Linguistics, 56% in Education, and 66% in Management. The subject fields of SSH presented less than 40% of their journals indexed in Scopus (Figure 13), in which around 50% are A1 or A2. From these perspectives, the publication patterns of SSH subject fields in international journals in Brazil once more seem to worsen.

Regarding international visibility or "internationalization" in SSH, there is a general belief that research quality can be promoted in SSH through more publications in a restricted number of international journals selected for indexing in the major databases (SIVERTSEN, 2016a). In Brazil, many efforts have been made to internationalize the Brazilian journals, specially to publish more research in international collaboration. Following this path, it was essential to improve the impact of Brazilian journals, giving more emphasis to those already indexed journals (PACKER, 2011). Moreover, the same research project may well contribute with different parts to both national and international dimensions. SSH research cannot be mainly communicated in international journals that are only read by peers abroad, because it will lose their raison d'être by disconnecting from the surrounding culture and society. At the same time, publishing in those specialized journals at the international level is required to be confronted with and inspired by the scholarly standards, critical discussions, and new developments among other experts in the field (SIVERTSEN, 2016a).

In terms of research evaluation exercises, SSH are well known to have heterogeneous publication patterns and not only articles in internationally oriented journals. Furthermore, although publication patterns differ between the disciplines of SSH, they are similar across countries within the areas (VAN LEEUWEN, 2006; SIVERTSEN, 2016b). In an effort to use research information systems to reflect more this richness of SSH research, several countries in Europe have set up alternative data sources that are national bibliographic databases (VERLEYSEN; GHESQUIÈRE; ENGELS, 2014; STOJANOVSKI, 1999; SIVERTSEN, 2016a; MOSKALEVA *et al*., 2018). The purpose behind these databases is to provide more comprehensive bibliographic data, accounting for the diversity of research output types in SSH, as pointed out by either the Leiden Manifesto (HICKS *et al*., 2015) or the San Francisco Declaration on Research Assessment (DORA, 2012). Thus, a study about the comprehensiveness of 13 European national bibliographic databases showed that data from some national bibliographic databases were more comprehensive. Therefore, they may be a better fit for bibliometric analyses when compared to data from commercial databases such as WoS and Scopus (SĪLE *et al*., 2018).

In some cases, these national bibliographic databases are already integrated into the national information systems. As an example, Norway was the first country to establish a CRIS with complete quality-assured bibliographic data covering all peer-reviewed scholarly publishing in the total higher education sector. Publication data from the Norwegian model include both aggregated data at the institutional and field levels, and publication data at the individual level. They also produce their own publication indicators based on bibliographic data that are applied in their performance-based funding formula (SCHNEIDER, 2009; SIVERTSEN, 2010; AHLGREN; COLLIANDER; PERSSON, 2012; AAGAARD; BLOCH; SCHNEIDER, 2015). The information system called CRISTIN (CRIS in Norway) has been expanded beyond the higher education sector. Sivertsen (2016a) has studied internationalization patterns and research evaluation criteria in SSH in Norway. The author observed that even in the category of international journals used by Norwegian researchers, the coverage of articles in WoS and Scopus was below 50% and has been decreasing since 2005. Although there is an expansion of these databases in SSH, it has not kept up with the rapid development of new international and specialized journals in those fields.

Furthermore, CRISs do not allow international comparison or benchmarking and they lack data on citations. Thus, citation analysis and international comparison are possible if the data are matched to data from Scopus and WoS. This match is almost automatically in the Norwegian CRISTIN system because bibliographical records from the two external sources are imported into CRISTIN. In addition, they are validated by CRISTIN to facilitate the researchers' registration of the publications (SIVERTSEN, 2018).

In sum, the results from the Framework Phase 1 application indicated different publication patterns among the studied Brazilian subject fields. The ones with researches more regionally and nationally engaged, especially the ones of SSH, usually publish less in indexed journals, and among the last ones, those with lower impact are preferable. This scenario seems to get worsen after analyzing some consecutive periodic evaluations. A limitation of this study is that only one impact indicator was applied in the study. This indicator was based on citations from an international database, which is one of the dimensions of scientific or scholarly quality. Additionally, this indicator type is usually inappropriate for evaluating scientific production in SSH fields. Considering the majority of other subject fields, the impact of the chosen journals did not change or declined over the years. Additionally, there was a low diversity of journals ranked in QUALIS compared to the available variety in Scopus. This diversity is even more reduced for A1 and A2 journals. For these highest categories, all subject fields produced more articles in journals with the lowest impact comparing the percentiles of QUALIS and Scopus rankings. Based on all these results, this section hypothesis is that these identified patterns arise as a potential effect of the own QUALIS system structure and operation over the last years. These possible effects are the main topic viewed and discussed in the next section.

6.2 POTENTIAL EFFECTS LINKED TO THE QUALIS SYSTEM

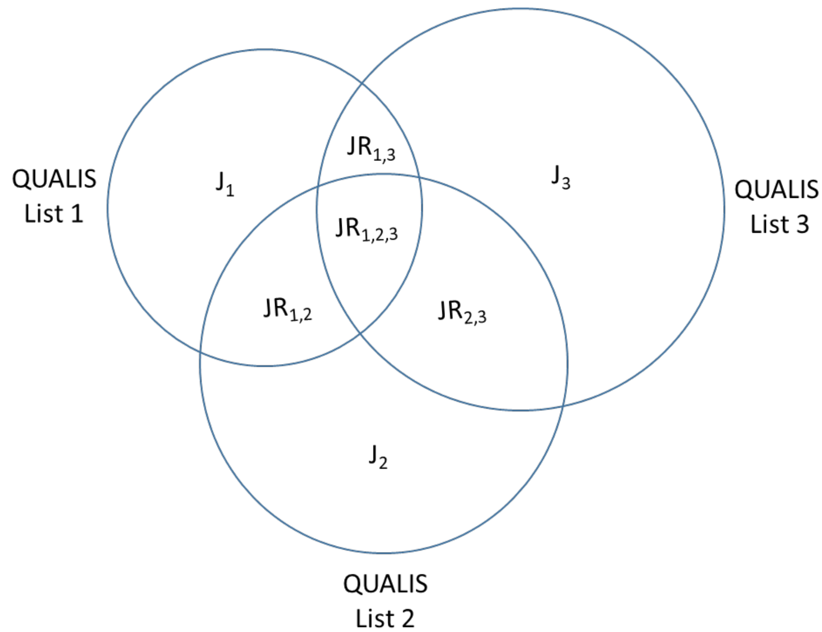
Framework Phase 2 was applied for verifying the potential effects of QUALIS system. This part of the study aims to check potential systematic effects linked to it. Two intrinsic aspects of QUALIS could have influenced specific publication patterns observed in Section 6.1. One is that the QUALIS system provides some journal quality lists as a result. The logic of the journals' list is 'one size fits all.' As such, an effect is "to endorse and cultivate a research monoculture in which particular criteria, favored by a given list, assume the status of a universal benchmark of performance ('research quality')" (MINGERS; WILLMOTT, 2013). Another is that the QUALIS system has been working as a PRFS in the national scenario. Previous research states an apparent lack of formal considerations from the designers of these systems on how they affect institutional practices and individual behavior, although there have been abundant and heated debates on the intended and unintended consequences for both individuals and systems in the public and academic levels (AAGAARD, 2015).

Examining the effects of PRFS is an extremely challenging task, and robust evidence on these effects at the individual level still lacks (GLÄSER, 2007; BUTLER *et al.*, 2010). A better empirical understanding of how this kind of system affects individuals through direct and indirect channels across levels, fields, and institutions could ground future policy decisions (SAUDER; ESPELAND, 2009). Hence, this study is an attempt to scrutinize the potential effects of QUALIS List on scientific production behavior to better support Brazilian government decisions about the implementation of research evaluation information systems in the country.

6.2.1 Description by groups of all journals ranked in the QUALIS system from 2007 to 2016

Step 1 in Framework Phase 2 compared QUALIS List 1, QUALIS List 2 and QUALIS List 3 to identify each journal frequency. Based on them, the journals were organized into groups according to Figure 21.

Figure 21 – Step 1 in Phase 2 applied to the QUALIS system



J_1: journals evaluated only in the first periodic evaluation.
J_2: journals evaluated only in the second periodic evaluation.
J_3, journals evaluated only in the third periodic evaluation.
$JR_{1,2}$: journals that remained in the system and were evaluated only in the first and second periodic evaluations.
$JR_{1,3}$: journals that remained in the system and were evaluated only in the first and third periodic evaluations.
$JR_{2,3}$: journals that remained in the system and were evaluated only in the second and third periodic evaluations.
$JR_{1,2,3}$: journals that remained in the system and were evaluated in all periodic evaluations.

Source: The author.

Table 3 describes the number of journals ranked in the QUALIS system in each periodic evaluation, the journals organized into groups, and the total of unique journals evaluated from 2007 to 2016.

Table 3 – Description by groups considering all journals ranked in the QUALIS system from 2007 to 2016

Subject field	2007-2009	2010-2012	2013-2016	J_1	J_2	J3	$JR_{1,2}$	$JR_{1,3}$	$JR_{2,3}$	$JR_{1,2,3}$	Total of unique journals (2007-2016)
Biological Sciences II	1,348	2,121	2,753	205	509	1,106	142	177	646	824	3609
Agrarian Sciences I	1,124	2,165	3,081	218	550	1,481	117	102	811	687	3966
Medicine II	2,336	3,445	3,894	485	910	1,453	310	216	900	1,325	5599
Computer Sciences	472	849	1,271	106	302	680	36	80	261	250	1715
Engineering III	1,236	2,303	2,668	329	891	1,299	182	139	644	586	4070
Management	521	1,574	2,544	130	631	1,618	76	59	611	256	3381
Education	922	1,728	2,076	252	616	1,055	158	67	509	445	3102
Literature and Linguistics	1,071	1,476	1,609	422	506	747	195	87	408	367	2732

J_1: journals evaluated only in the first periodic evaluation.
J_2: journals evaluated only in the second periodic evaluation.
J_3, journals evaluated only in the third periodic evaluation.
$JR_{1,2}$: journals that remained in the system and were evaluated only in the first and second periodic evaluations.
$JR_{1,3}$: journals that remained in the system and were evaluated only in the first and third periodic evaluations.
$JR_{2,3}$: journals that remained in the system and were evaluated only in the second and third periodic evaluations.
$JR_{1,2,3}$: journals that remained in the system and were evaluated in all periodic evaluations.
Source: The author.

The number of new journals selected by the Brazilian community increased over the years. These journals were represented by J_1, J_2, and J_3. The last group constitutes more than 40% of all journals evaluated in the third period. Table 4, however, shows that in each periodic evaluation, on average, the articles were more concentrated on journals that remained in the system for at least two consecutive periodic evaluations – $JR_{2,3}$ and $JR_{1,2,3}$. This result demonstrates a high concentration of articles in a few journals that have remained in the system over the years.

Table 4 – Average of articles by groups of journals

Subject field	J1	J2	J3	JR1,2	JR1,3	JR2,3	JR1,2,3
Biological Sciences II	4%	6%	17%	4%	3%	18%	71%
Agrarian Sciences I	3%	4%	12%	5%	1%	19%	75%
Medicine II	6%	7%	15%	6%	2%	15%	74%
Computer Sciences	11%	17%	20%	5%	5%	23%	57%
Engineering III	10%	19%	20%	12%	3%	22%	56%
Management	6%	10%	19%	7%	2%	42%	45%
Education	7%	10%	19%	8%	1%	29%	58%
Literature and Linguistics	10%	9%	12%	12%	3%	27%	60%

J_1: journals evaluated only in the first periodic evaluation.
J_2: journals evaluated only in the second periodic evaluation.
J_3, journals evaluated only in the third periodic evaluation.
$JR_{1,2}$: journals that remained in the system and were evaluated only in the first and second periodic evaluations.
$JR_{1,3}$: journals that remained in the system and were evaluated only in the first and third periodic evaluations.
$JR_{2,3}$: journals that remained in the system and were evaluated only in the second and third periodic evaluations.
$JR_{1,2,3}$: journals that remained in the system and were evaluated in all periodic evaluations.
Source: The author.

Thus, if the QUALIS List of a year or period is used as a reference by faculty and students for choosing in which journal they publish in the next period, this higher concentration in those restrict groups of journals is a reasonable effect. $JR_{1,2,3}$ and $JR_{2,3}$ concentrated together more than 80% of the published articles, on average, considering all fields (Table 3), although they represent less than 40% of Total unique journals ranked by QUALIS from 2007 to 2016 (Table 3). This result also bears some relation to Bradford Law, which states that despite the existence of many journals, scientific information is concentrated in a minority of them that publish most articles (BRADFORD, 1934). Hence, it is essential to know more about these groups of journals to understand the main publication patterns in Brazil and the possible connections with the evaluation system process.

6.2.2 Overview of the $JR_{1,2,3}$ and $JR_{2,3}$ journals and their articles

Step 2 in Framework Phase 2 was applied for this overview, in which descriptive analyses were done for $JR_{2,3}$ and $JR_{1,2,3}$, such as their nationality, frequency of articles, and proportion inserted in Scopus. Table 5 describes the number of $JR_{1,2,3}$ and $JR_{2,3}$ journals and the concentration of their articles, besides indicating among them those indexed in Scopus. The majority of $JR_{1,2,3}$ and $JR_{2,3}$ journals were indexed in Scopus, except for Management, Education, Literature and Linguistics. Table 5 shows a decrease in the number of indexed journals considering all $JR_{1,2,3}$ and $JR_{2,3}$ journals in all fields. This decline is even sharper in SSH, especially considering the percentage of published articles (% indexed articles). In sum, Life Sciences, Health Sciences, and Physical Sciences journals of choice were mostly indexed in Scopus, while in Humanities and Social Sciences, these numbers were considerably lower.

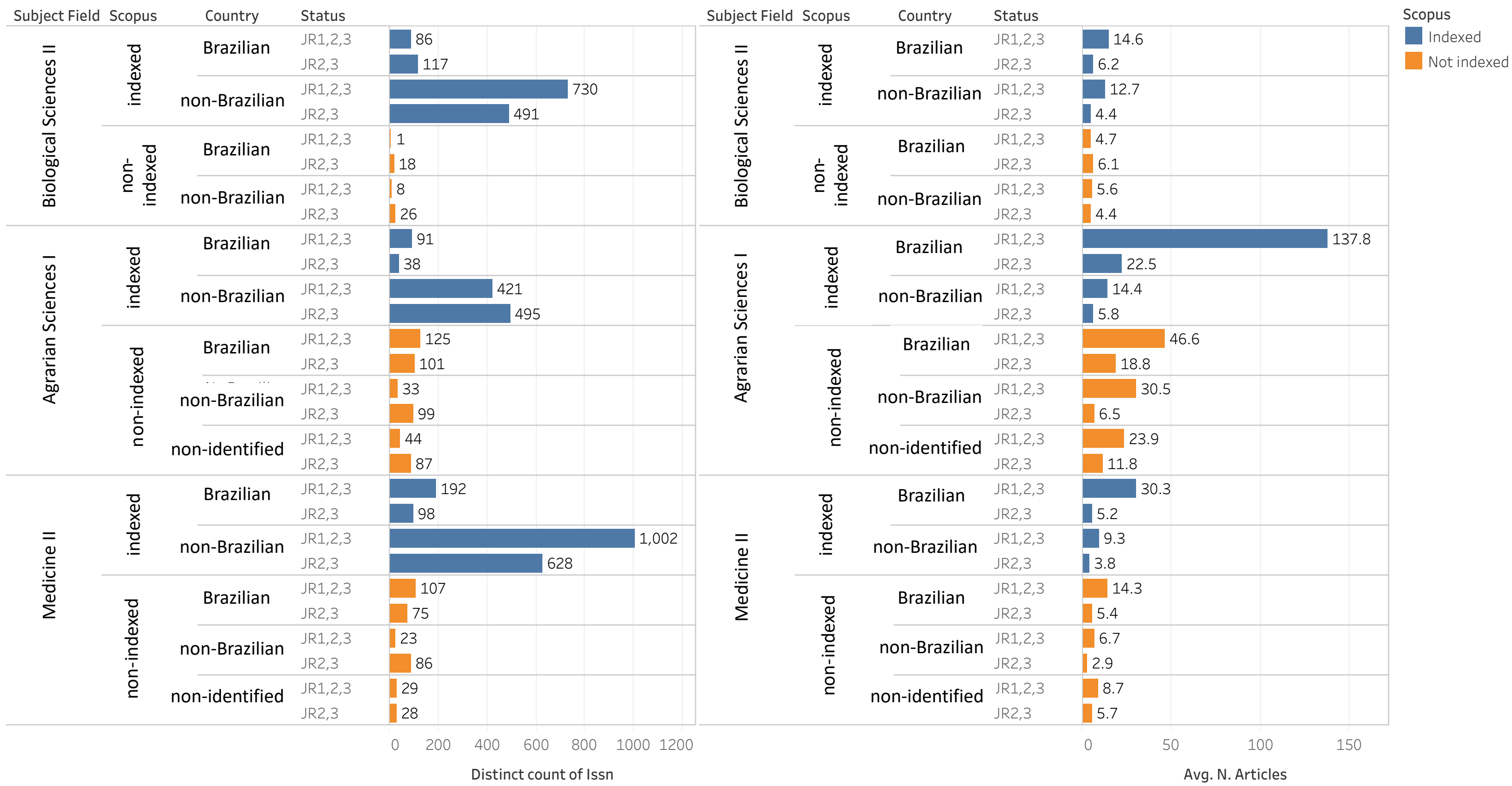
Table 5 – Description of $JR_{1,2,3}$ and $JR_{2,3}$ journals

Subject field	JR1,2,3 + JR2,3			
	Number of journals	% of articles	Number of indexed journals	% of indexed articles
Biological Sciences II	1470	85.2%	1418	83.8%
Agrarian Sciences I	1498	89.3%	1035	60.6%
Medicine II	2225	84.9%	1909	74.9%
Computer Sciences	511	76.4%	453	63.3%
Engineering III	1230	74.1%	879	53.8%
Management	867	81.7%	318	14.1%
Education	954	81.2%	155	10.1%
Literature and Linguistics	775	81.2%	85	4.1%

Source: The author.

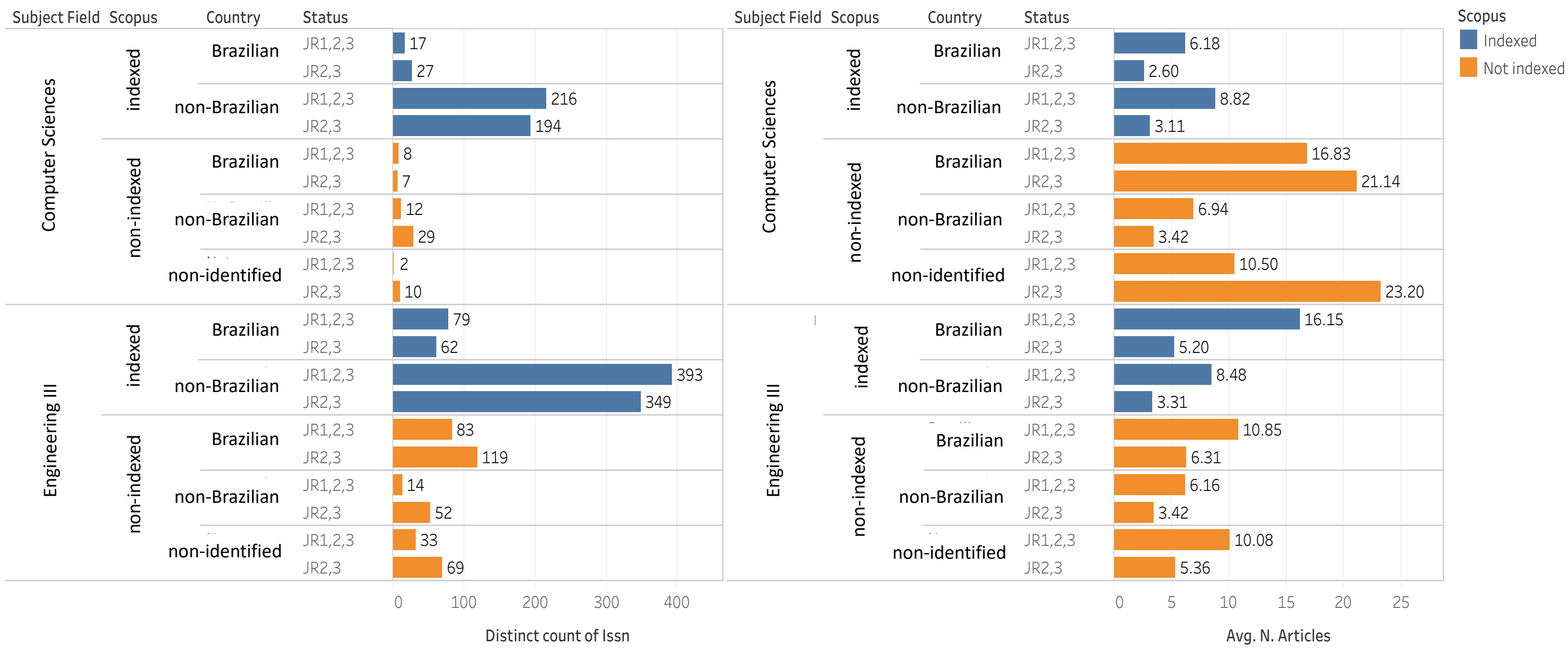
Figures 22, 23, and 24 show indexed and not indexed journals grouped according to their nationality. The non-Brazilian indexed journals were the majority for Biological Sciences II, Medicine II, Computer Sciences, and Engineering III, besides showing the highest average of articles. Although the non-Brazilian indexed journals were the majority in Agrarian Sciences, a group of 91 Brazilian indexed journals displayed a much higher ratio of articles. Considering the few indexed journals in Education, Literature and Linguistics, the majority are Brazilians. Although mostly indexed journals in Management were not Brazilians, 44 Brazilian indexed journals demonstrated a much higher average of articles.

Figure 22 – Number of JR1,2,3 and JR2,3 grouped by indexed and not indexed, Brazilian and non-Brazilian, as well as the average of articles in each group for Biological Sciences II, Agrarian Sciences I, and Medicine II



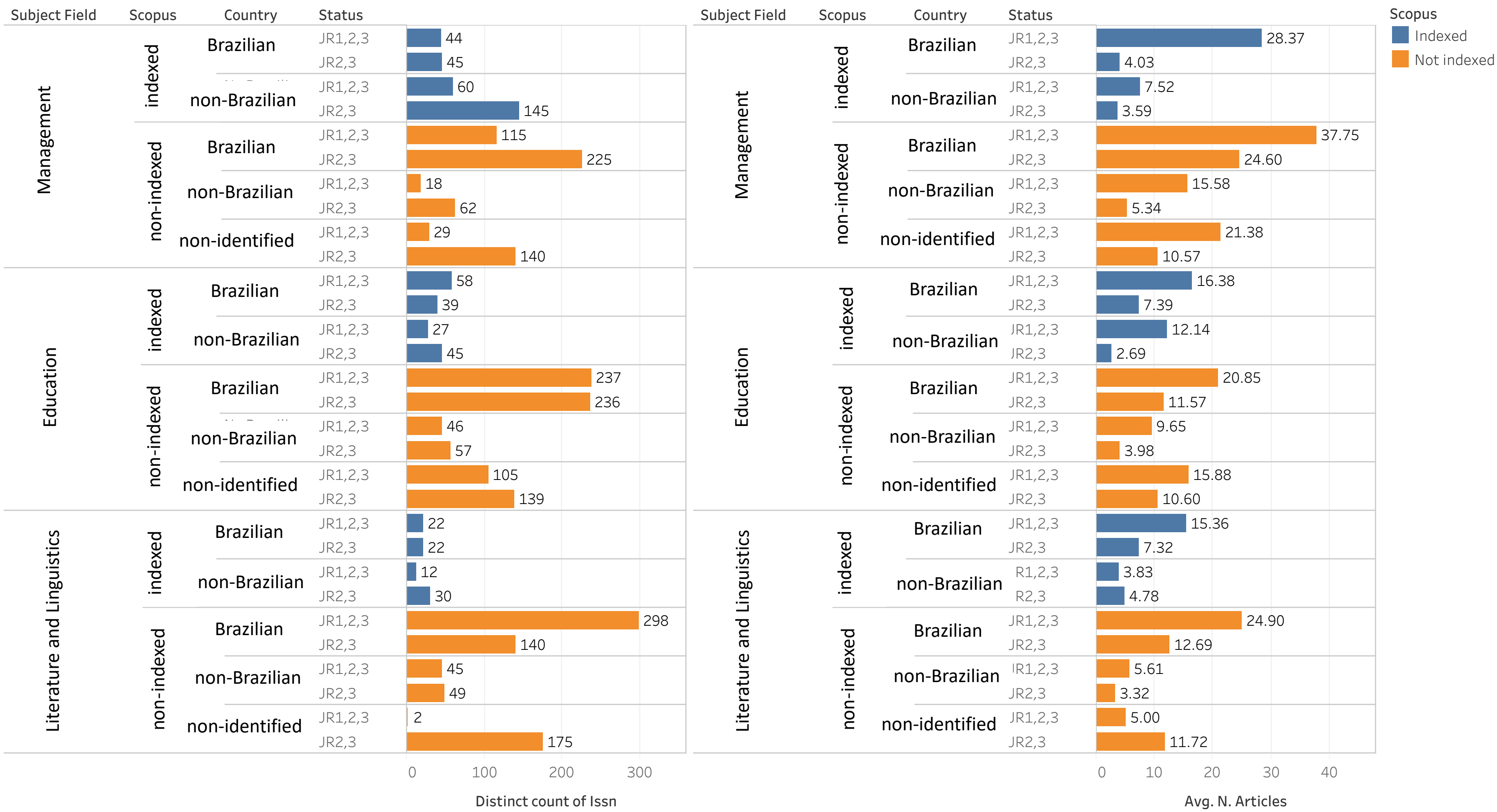
Source: The author.

Figure 23 – Number of JR1,2,3 and JR2,3, grouped by indexed and not indexed, Brazilian and non-Brazilian, as well as the average of articles in each group for Computer Sciences and Engineering III



Source: The author.

Figure 24 – Number of JR1,2,3 and JR2,3 grouped by indexed and not indexed, Brazilian and non-Brazilian, as well as the average of articles in each group for Management, Education, Literature and Linguistics



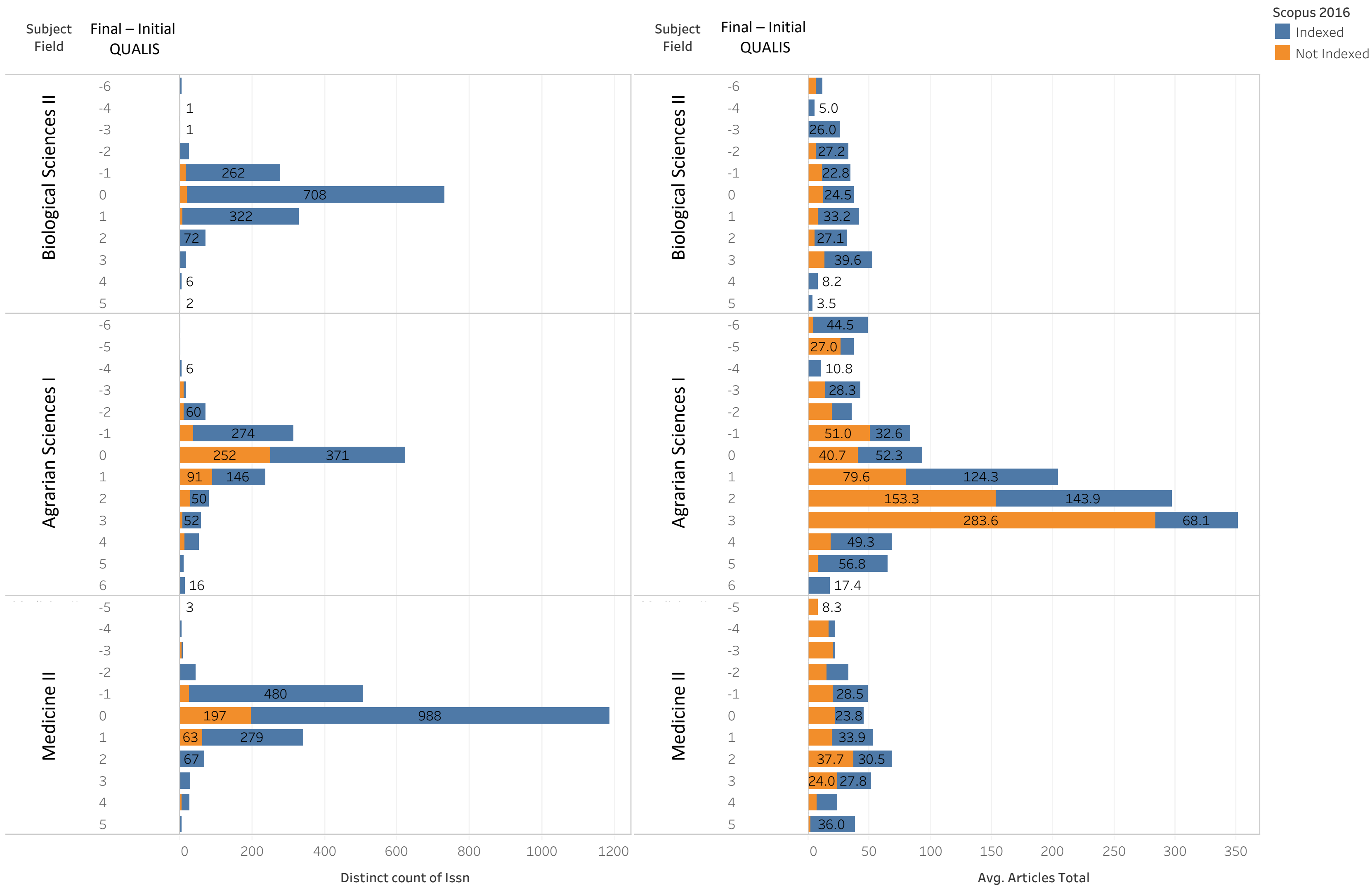
Source: The author.

6.2.3 Shifts in the categories of $JR_{1,2,3}$ and $JR_{2,3}$ journals, considering changes in frequency of their articles, between the periodic evaluations

Step 3 in Framework Phase 2 was applied to investigate the shift of $JR_{1,2,3}$ and $JR_{2,3}$ journals among the categories of QUALIS, as well as changes in their frequency of articles, over the periodic evaluations. Firstly, the initial (QUALIS 1 for $JR_{1,2,3}$ and QUALIS 2 for $JR_{2,3}$) and the final (QUALIS 3 for both $JR_{1,2,3}$ and $JR_{2,3}$) QUALIS categories were identified. Secondly, A1 to B5 categories were coded from 7 to 1. Next, the initial code was subtracted from the last one. These differences ranged from -6 to 6, considering all the possible movements among the categories. The result -6 to -1 represents journals that had their QUALIS categories reduced over time, while 0 means no change in category, and 1 to 6 shows the journals that had an improvement of their quality according to peer-review judgment.

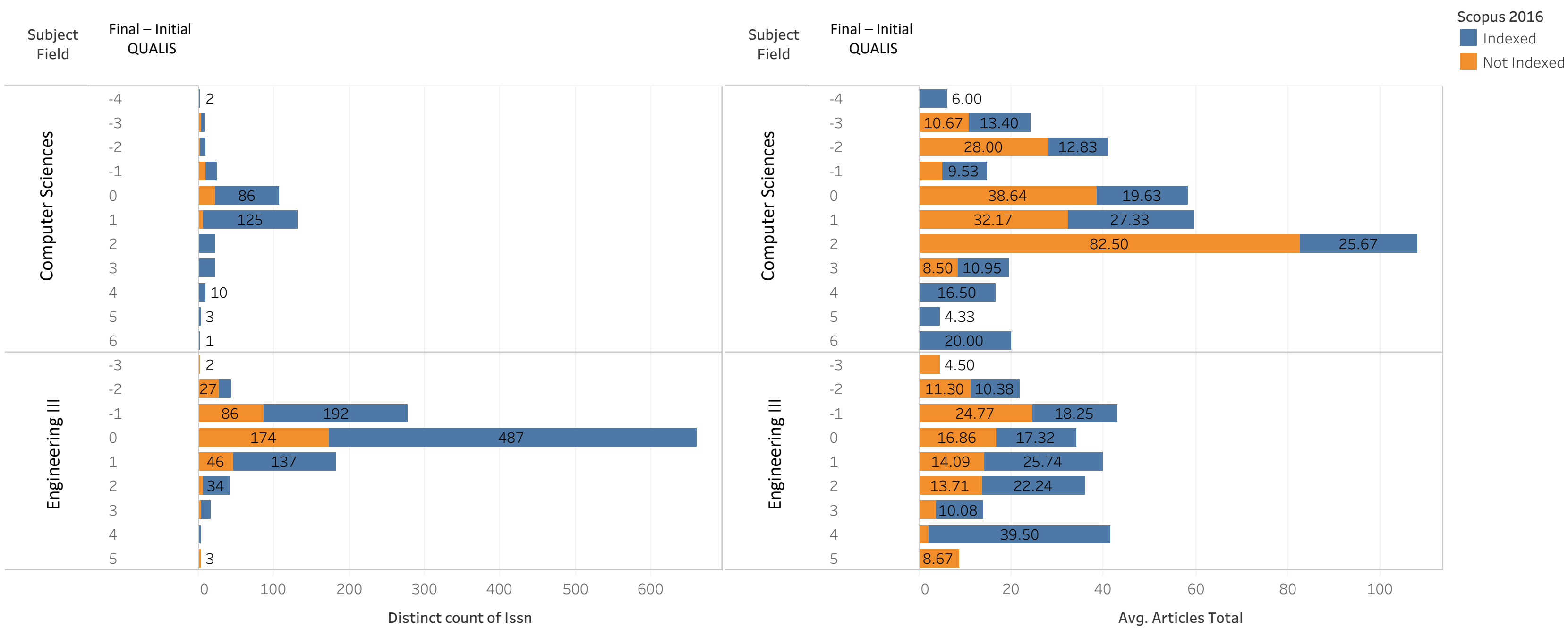
In this regard, Figures 25, 26 and 27 illustrate how the remaining journals are distributed according to the difference between their final (F) and initial (I) QUALIS, their insertion in the Scopus database, and their average of articles. More than 70% of the $JR_{1,2,3}$ and $JR_{2,3}$ journals had their initial QUALIS preserved (difference equal 0) or improved (difference ranging from 1 to 6) over the periodic evaluations in all subject fields. Furthermore, this preserved or improved QUALIS subset concentrated more than 70% of the articles published by $JR_{1,2,3}$ and $JR_{2,3.}$ In view of the insertion in Scopus, the $JR_{1,2,3}$ and $JR_{2,3}$ not indexed journals concentrated a great average of articles in most of the subject fields, especially among the preserved or improved QUALIS subset. Likewise, these not indexed journals represented the majority for the subject fields of SSH. All these results provide evidence of a relationship between the maintenance or increase of QUALIS categories and a high concentration of articles in a restrict group of journals, mainly in those not indexed at Scopus.

Figure 25 – Shifts in the categories of JR1,2,3 and JR2,3 journals between the periodic evaluations, and average of their articles, for Biological Sciences II, Agrarian Sciences I and Medicine II



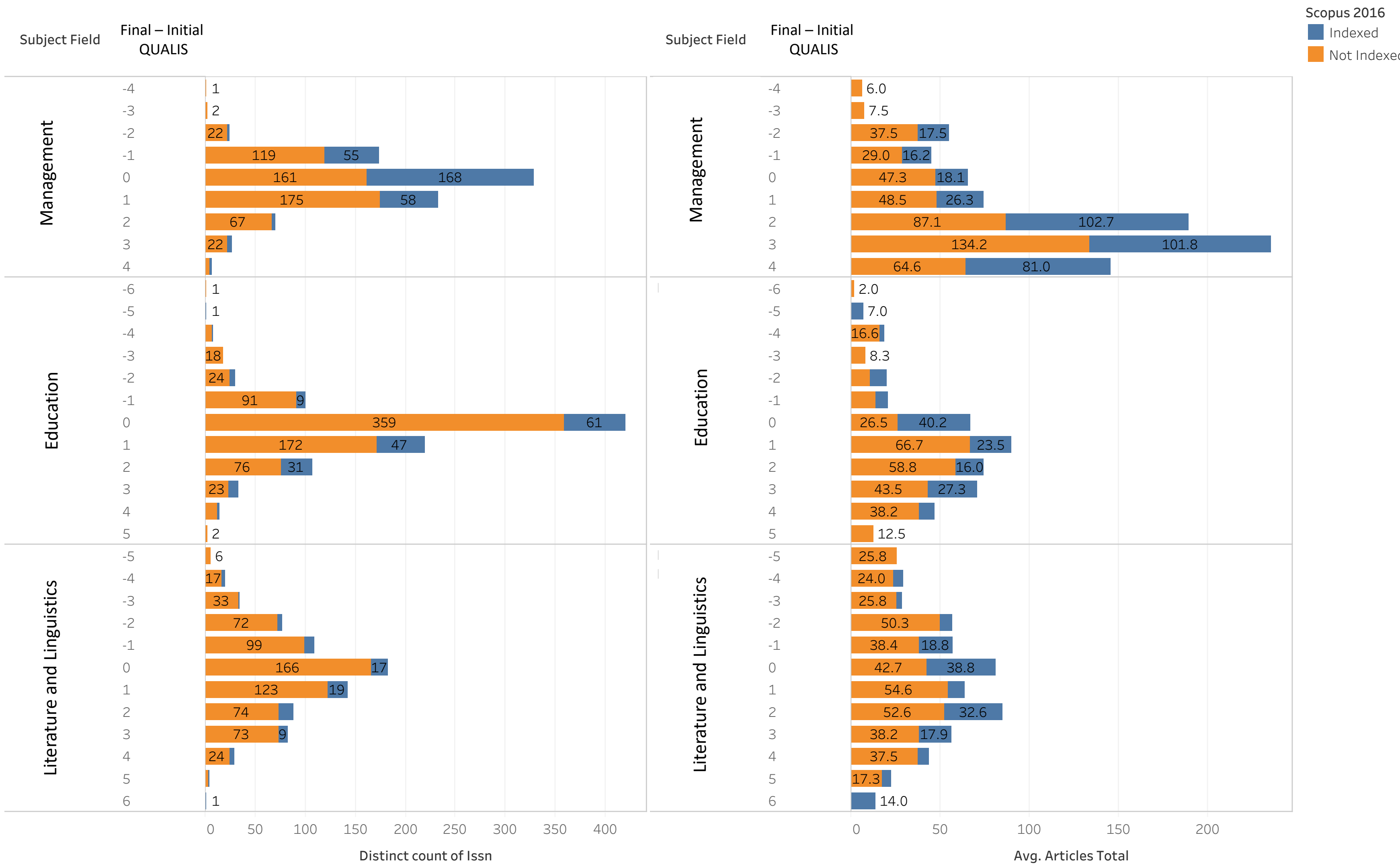
Source: The author.

Figure 26 – Shifts in the categories of JR1,2,3 and JR2,3 journals between the periodic evaluations, and average of their articles for Computer Sciences and Engineering III



Source: The author.

Figure 27 – Shifts in the categories of JR1,2,3 and JR2,3 journals between the periodic evaluations, and average of their articles, for Management, Education, Literature and Linguistics

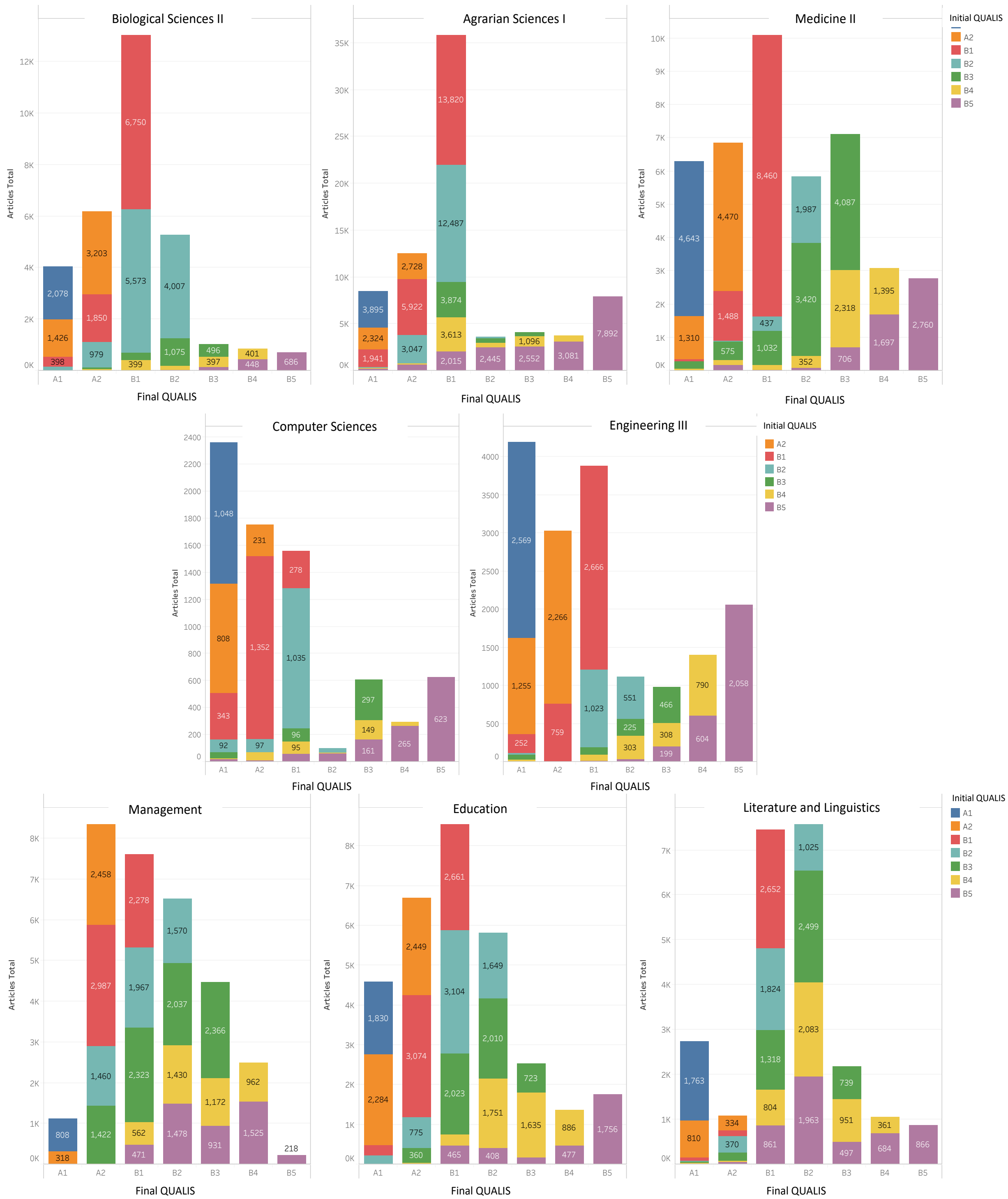


Source: The author.

Figure 28 shows the number of articles considering the final QUALIS in the x-axis and the initial QUALIS represented by the colors in the bars. Taking this preserved or improved QUALIS journals subset of $JR_{1,2,3}$ and $JR_{2,3}$ into account, more than 50% of their articles were in the journals ranked as A1, A2, and B1 in the final QUALIS in all subject fields. In some subject fields such as Biological Sciences II, Agrarian Sciences I and Computer Sciences, this percentage was even higher. It is noteworthy that publishing in these three categories guarantees more points in the final grade of the graduate program.

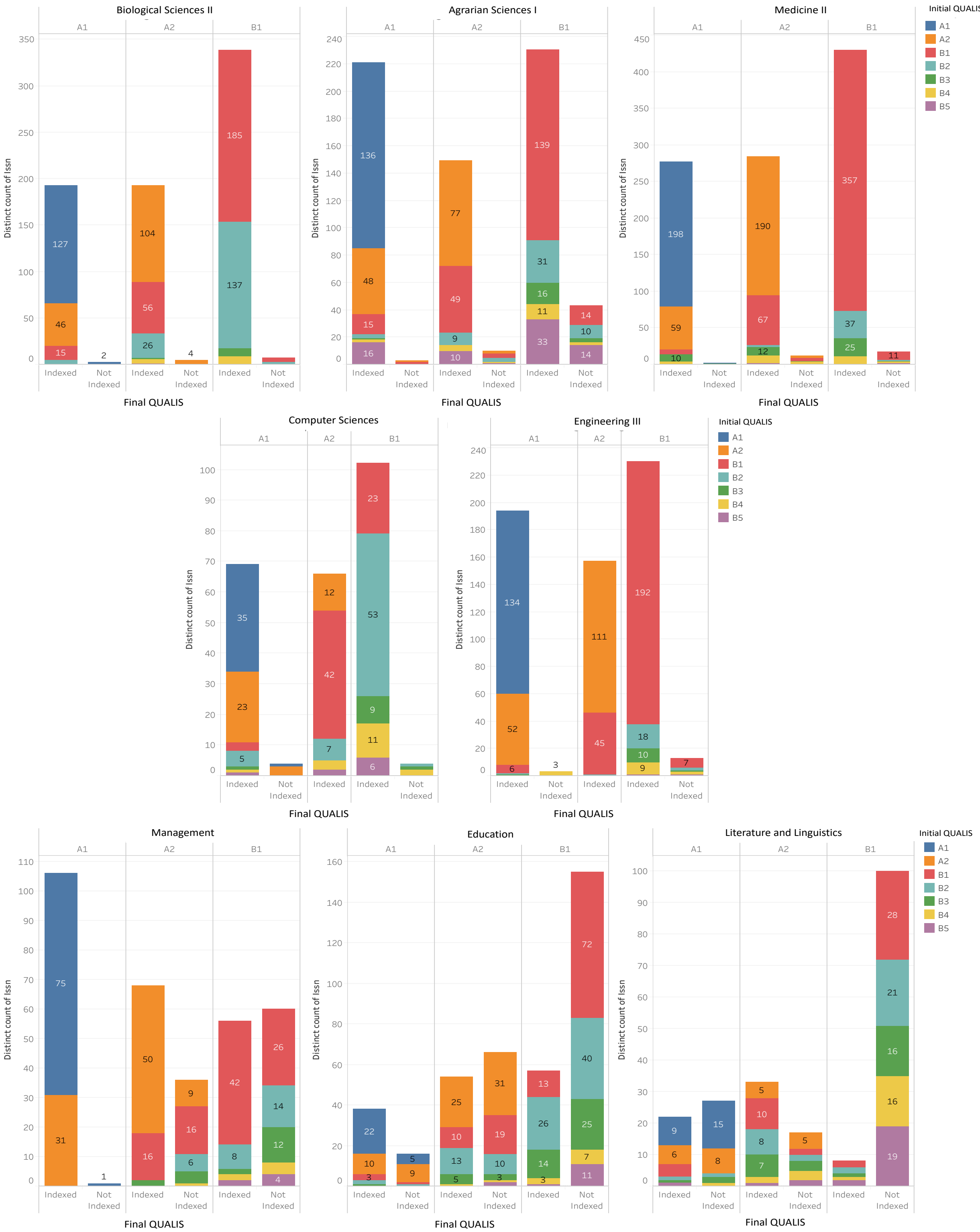
Figure 29 represents the journals ranked as A1, A2, and B1 in the final QUALIS divided into indexed and not indexed journals, in which the final QUALIS is in the x-axis and the colors in the bars represent the initial QUALIS. Most of the indexed journals ranked as A1 in the final QUALIS had an equal initial QUALIS in all subject fields. On the other hand, in general, these A1 journals do not include most of the articles in all subject fields (Figure 28). Moreover, not indexed journals are the majority of these set of journals only in Education and Literature and Linguistics.

Figure 28 – Number of articles considering final and initial QUALIS



Source: The author.

Figure 29 – Number of journals ranked as A1, A2, and B1 in the final QUALIS divided into indexed and not indexed journals

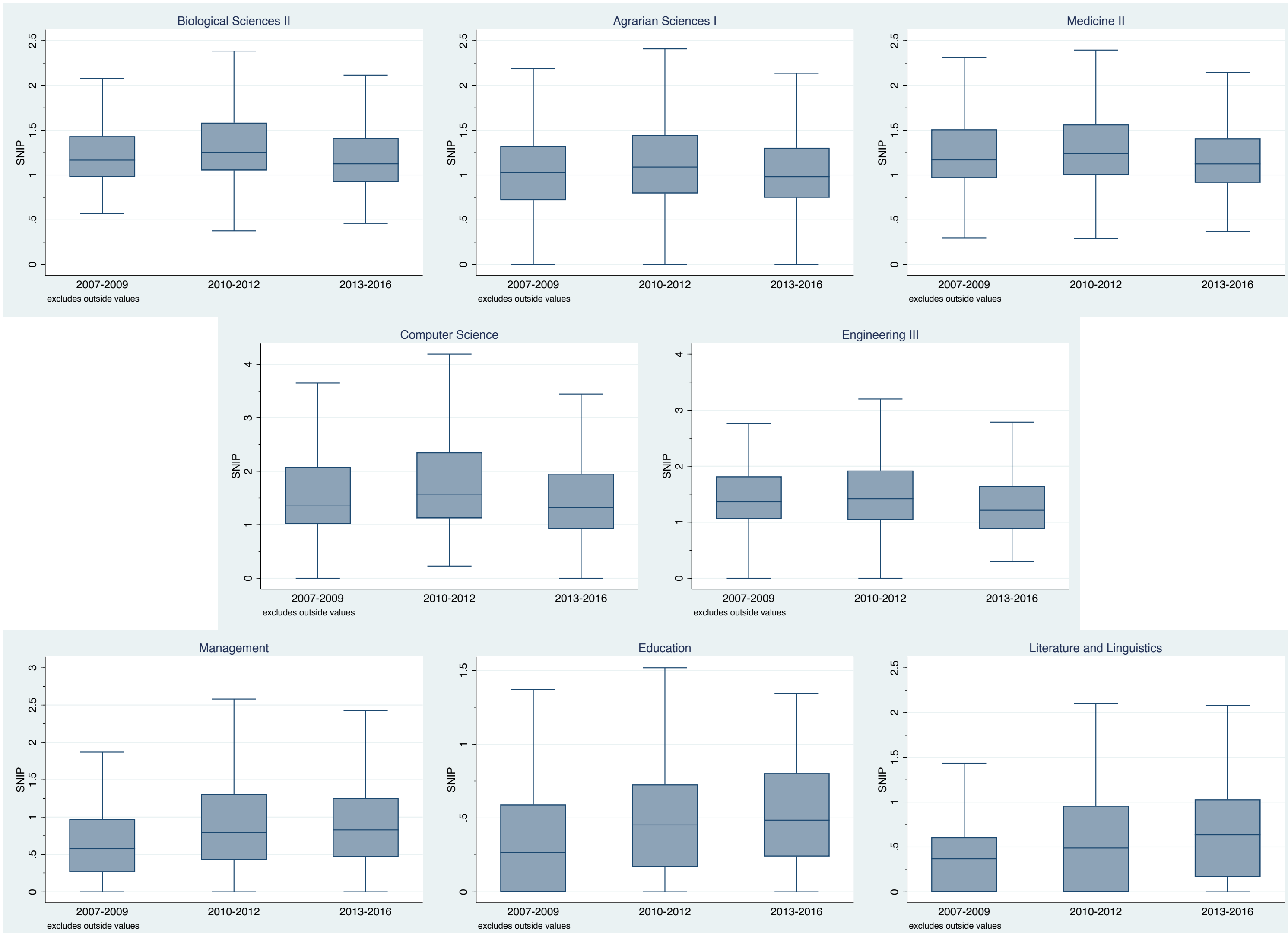


Source: The author.

6.2.4 Analysis of SNIP values of the indexed $JR_{1,2,3}$ and $JR_{2,3}$ journals ranked as A1, A2, and B1 in the final QUALIS

In this subsection, Step 4 of Framework Phase 2 was applied to examine only $JR_{1,2,3}$ and $JR_{2,3}$ indexed in Scopus ranked as A1, A2, or B1 in the last periodic evaluation and weighted or not by the number of articles. The SNIP value distributions of these journals, as well as their frequency of articles, were compared among the three periodic evaluations (Figure 30). The results showed that the distribution of SNIP values from this journal set was lower in the third periodic evaluation than in the second one for Biological Sciences II, Agrarian Sciences I, Medicine II, Computer Sciences, and Engineering III. This distribution was, however, higher in the third periodic evaluation when compared to the first one for Management, Education, Literature and Linguistics. This observation indicates that regarding international impact there was a decrease in most of the subject fields, although the QUALIS categories of this set were maintained or improved during the periodic evaluations.

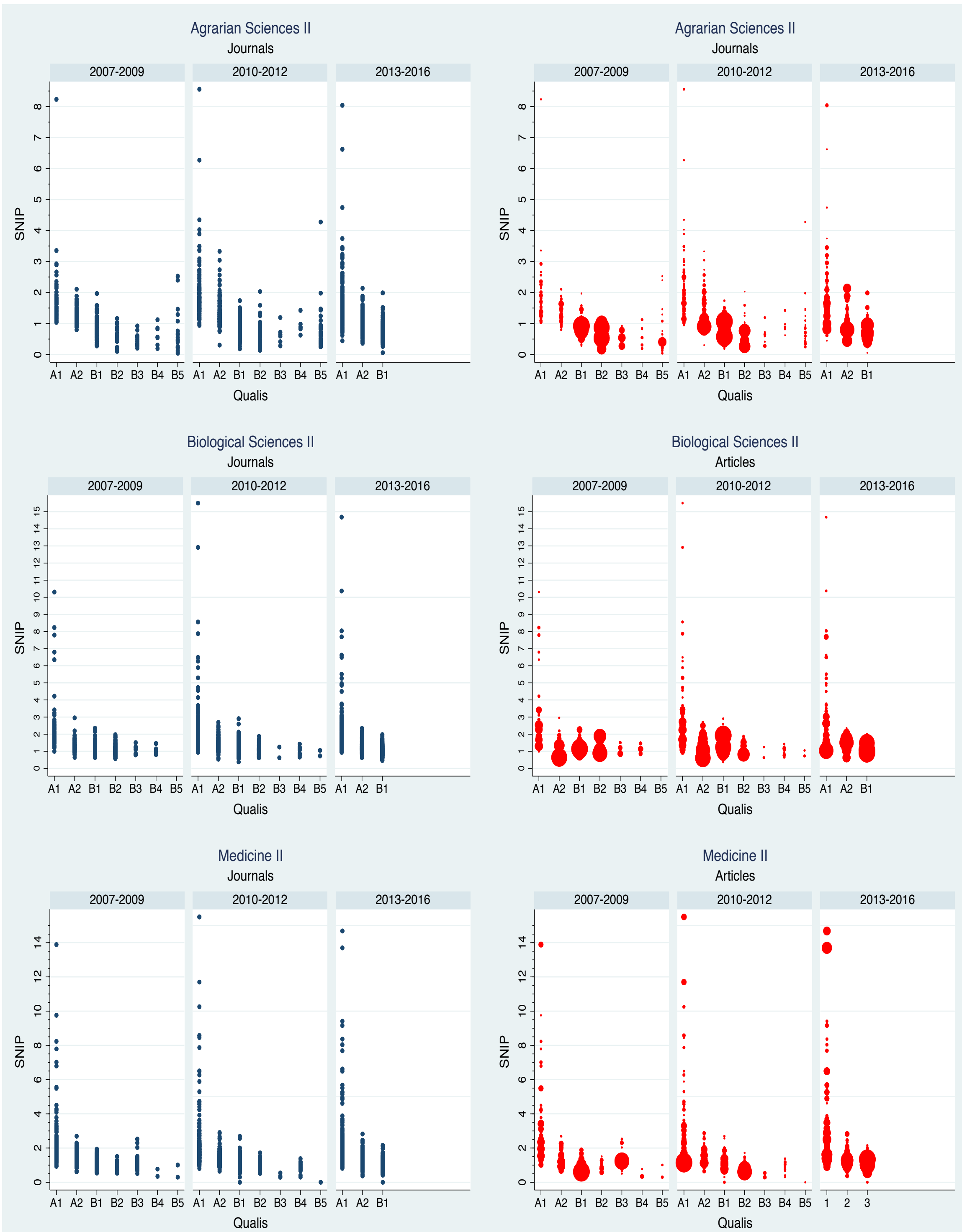
Figure 30 - JR1,2,3 and JR2,3 indexed in Scopus ranked as A1, A2, or B1, distributed by their SNIP values in each periodic evaluation and subject field



Source: The author.

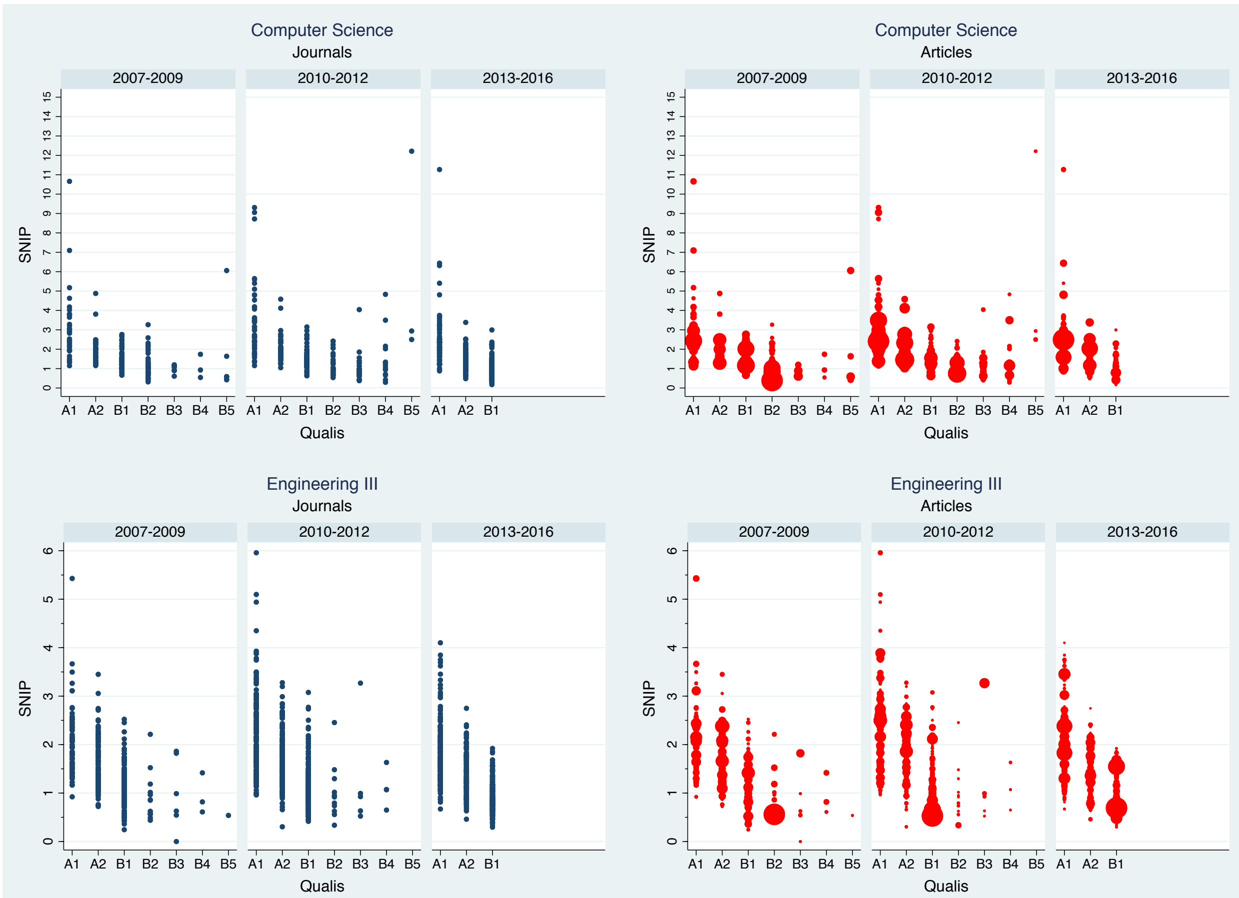
Figures 31, 32, and 33 show the shifting of these journals as well as of the same journals weighted by the number of articles. Lower-impact journals are moving to the highest categories of QUALIS, and this shifting happens simultaneously with the increase in the number of articles in journals with a lower international impact in all fields. These results suggest that some journals, despite their lower impact, have become more qualified according to peer review from distinct subject fields during the periodic evaluations. At the same time, there is an increase of articles in low-impact journals. These results might be due to inappropriate use of the QUALIS lists. Thus, faculty and students may have been choosing low-impact journals from QUALIS lists ranked in the highest categories. Another possible explanation is the presence of low-impact journals that become more qualified over time, thus drawing more articles. Since QUALIS works as a PRFS, publishing in the highest categories is more important than the journal impact, and in the end, it will guarantee the same funds to the program.

Figure 31 – Shifts in the categories of JR1,2,3 and JR2,3 journals, considering only journals ranked as A1, A2, and B1 in the final QUALIS, besides the average of journal articles for Agrarian Sciences I, Biological Sciences II, and Medicine II



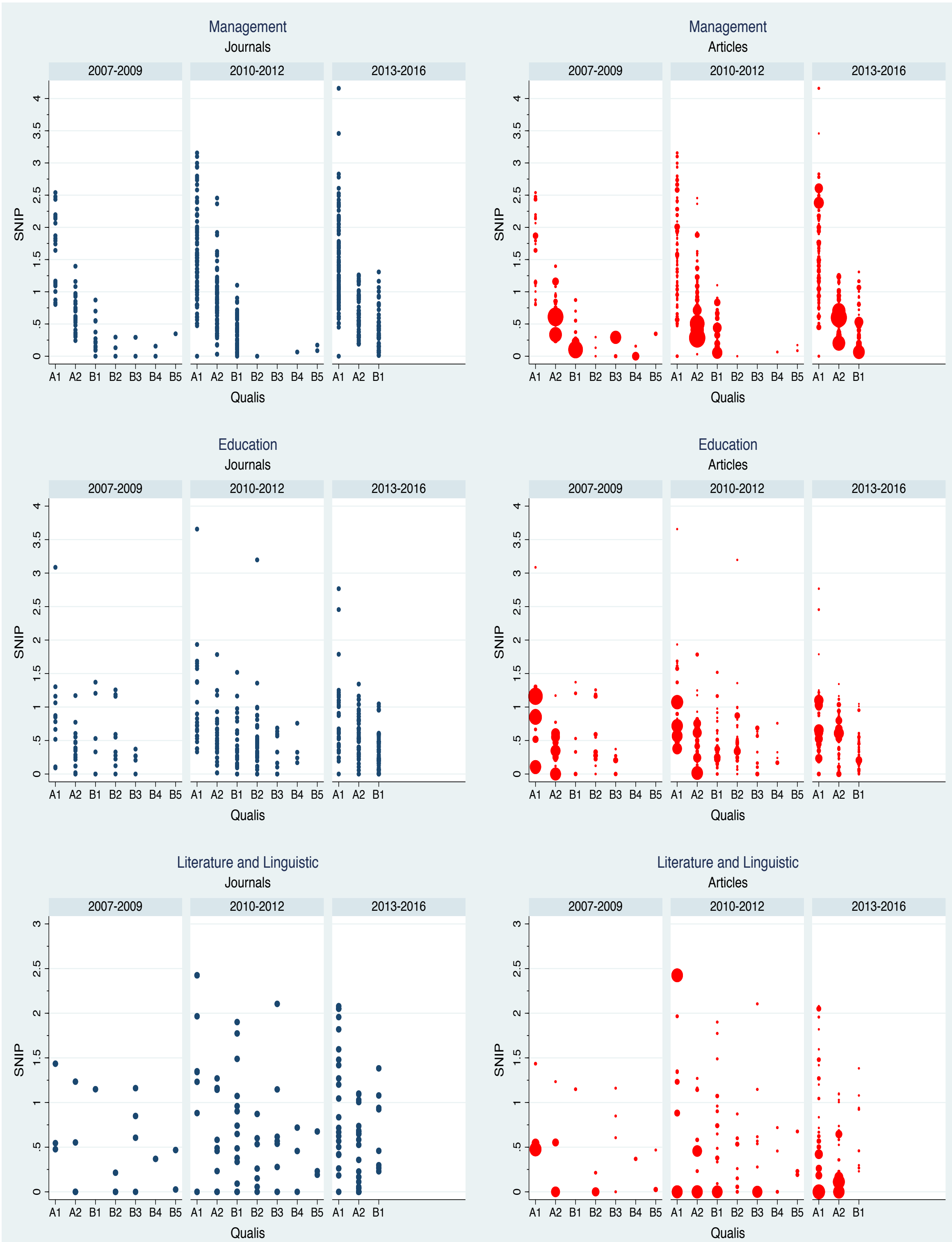
Source: The author.

Figure 32 – Shifts in the categories of JR1,2,3 and JR2,3 journals, considering only the journals ranked as A1, A2 and B1 in the Final QUALIS, as well as the average of journal articles for Computer Sciences and Engineering III



Source: The author.

Figure 33 - Shifts in the categories of JR1,2,3 and JR2,3 journals, considering only the journals ranked as A1, A2, and B1 in the final QUALIS, as well as the average of journal articles for Management, Education, Literature and Linguistics



Source: The author.

6.3 DISCUSSION

The development of a clear framework is essential to guide monitoring and evaluation. The proposed framework can provide a foundation for monitoring the progress of a national research evaluation system based on a journal list and for determining if this system is on course to achieve its intended results. OECD (2002, p. 30) defined performance monitoring as "a continuous process of collecting and analyzing data to compare how well a project, program, or policy is being implemented against the expected results." All results obtained by applying the proposed framework are of prime importance for CAPES, since the foundation has been rethinking the Brazilian research evaluation system, as well as the development of a national research information system. Thus, empirical knowledge or evidence of the possible aggregated effects caused by the previous model are essential for designing a new one. Additionally, the designers of the new QUALIS model, as well as the national research information system, need to be aware of possible effects of evaluation exercises, 'gaming' of indicators, and strategic responses by scientific communities and other players for the requirements in research evaluations (RIJCKE *et al*., 2016).

As reported by the results, the QUALIS system does not seem to encourage publication in journals indexed in the major international databases in some subject fields and especially in high-impact journals. Taking into account that the QUALIS list intends to be a national quality indicator, it is expected that journals in the same category have the same quality. Each category showed journals ranging from high to low impact. Not indexed journals in Scopus were also present in all categories. Moreover, QUALIS is composed by a funding formula that at the end is linked to funds. Given that publications in the same QUALIS category have identical rewards to the graduate program, faculty and students may be choosing to publish more in quantity in low-impact journals or even those not indexed in the leading international databases to publish their work. Furthermore, the articles were mostly concentrated in a restrict group of journals that remained in the system in the last periodic evaluations. Most of these journals either maintained the same QUALIS category or improved it over the periodic evaluations, although, in general, there was a significant decrease in their impact. Furthermore, this movement of journals from the lowest to the highest categories coincided with the increase in the average of articles in low-impact journals. Similarly, QUALIS does not seem to counter the 'perverse' publication effects, in which researchers seek to publish more, but with less effort.

The works developed by Butler are in line with our results. The author conducted previous studies about the research evaluation system in Australia and has found similar outcomes (BUTLER, 2002, 2003a, 2003b, 2004). Analogous to QUALIS, the Australian model was targeted at the institutional level and based on counting approaches. One of Butler's analyses on the Australian university publications clearly showed how academics quickly reacted to the funding formulae that rewarded quantity rather than quality. As reported by the author, with no differentiation between the quality or impact of the publications, there was little incentive to strive for placement in a prestigious journal in the Australian system. Thus, the rewards for a publication in Nature or in a low-impact journal were identical.

As a consequence, journal publication productivity has significantly increased in a decade, but its impact has declined (BUTLER, 2002, 2003a). Furthermore, Butler (2003b) demonstrated that the most significant increase in productivity had been in those journals at the lower end of the impact scale. The Australia funding formulae were based on a method encapsulating some performance measures – graduate student numbers or completion rates, research income, and publications, but the last had three times more weight. Another feature of the Australian model was that it did not differentiate between publication types and outlets, which according to Butler (2004) has presumably led to more activity but with 'less effort,' resulting in lower national impact.

Based on the Australian case, Gläser and Laudel (2007, p. 138) argue that universities tend to mirror the national formula internally to maximize income – even in situations where there is as an inappropriate measure of research quality. In Brazil, considering that almost all research takes place in the universities and the majority of funding is based on CAPES evaluation, it is expected that universities reinforce the signals of CAPES by using similar or identical measures to distribute their funds. Several reports have shown that evaluation systems that affect money or reputation are based on peer review or indicators and will tend to influence researchers' behavior, who play an active role in this context. Goal displacement is one of these behaviors. The goal becomes scoring high on the established criteria rather than as a means of evaluating if specific objectives (or performance levels) have been met. Another behavior is the scientific or scholarly process transformation due to the evaluation criteria that may be more difficult to recognize, such as avoiding risks in selecting research topics (RIJCKE *et al*., 2016).

In the UK, a longitudinal bibliometric study of its publication patterns between 1985 and 2003 suggested that specific publication patterns emerged years before three Research Assessment

Exercises (RAE) that took place in such period (1992, 1996, 2001), depending on whether the RAE were aimed at quantity or quality of publications (MOED, 2008). Another strategic response by the research community in the UK was reported by Harley (2002), based on a survey of academic staff in social sciences and business-related disciplines. It was carried out immediately after the 2001 RAE results, which he called 'playing the RAE game.' According to the author, UK academics continued to legitimate the unequal distribution of research funds despite recognizing the mechanism to be fundamentally flawed.

In Spain, unlike the UK and Australia systems, the focus is clearly on the individual rather than the institution. Thus, salary bonuses for publishing in prestigious journals, mainly papers published in a relatively high position (approximately the top one third) in the Institute of Scientific Information (ISI)'s Journal Citation Report lists by subject category, were the most common reward. The effect of this policy on Spanish publication output demonstrated clearly that Spanish researchers have also responded to funding stimuli by increasing their production well above the long-term trend line for Spanish publications in the ISI indices. Thus, the National Commission for the Evaluation of Research Activity (CNEAI) achieved its goal, *i.e.* increase productivity and internationalization of Spanish research (JIMÉNEZ-CONTRERAS; MOYA ANEGÓN; LÓPEZ-CÓZAR, 2003).

Osuna, Cruz-Casto and Sanz-Menéndez (2011) are particularly critical of attempts to argue for a causal relationship between the introduction of evaluation systems and the rise in the number of publications at a national level (see, for instance, BUTLER, 2003b; JIMÉNEZ-CONTRERAS; MOYA ANEGÓN; LOPEZ-COZAR, 2003). After analysis of the Spanish case, they argue that there is a range of other explanations, such as the maturation of Spanish science, the rise of R&D budget, and the number of researchers. Moreover, their attempt to isolate the effect of the introduction of a new evaluation system in Spanish science in 1989 does not find any apparent effects.

In sum, most of these systems are designed to allocate funding at aggregated levels: either institutional or department level. In this regard, insufficient consideration has been given to their link to individual behavior. Furthermore, there is an apparent lack of formal concerns from the system designers on how these systems affect institutional practices and individual behavior. On the other hand, public and academic debates have been abundant and heated on the intended and unintended consequences for both individuals and systems; however, they have been more anecdotal than evidence-based (BUTLER *et al.*, 2010; AAGAARD, 2015).

In addition, QUALIS system intended to provide a journal quality list. Some studies point out more significance and influence of these lists, as there is increasing competition between institutions for resources. Therefore, these lists come to shape nature, structure, and academic work conditions (ESPELAND; SAUDER, 2007; SAUDER; ESPELAND, 2009). A previous study regarding the 'Journal Guide' from the Association of Business Schools (ABS) in the UK argued about the effect of the 'one size fits all' logic of journal lists. According to them, these lists endorse and cultivate a research monoculture in which specific criteria, favored by a given list, assume the status of a universal benchmark of performance ('research quality'). Moreover, they demonstrated that a list of journals could come to dominate and define the focus and trajectory of a research field, with detrimental consequences for knowledge development (MINGERS; WILLMOTT, 2013).

Thus, the Arts and Humanities Research Council in the UK opposed a project to implement rankings across various disciplines, further advised against the use of the [ranking] outcomes as the basis for assessing individual candidates for employment or funding (ARTS AND HUMANITIES RESEARCH COUNCIL, 2011). In 2012, the San Francisco Declaration warned against the use of journal rating for the evaluation of individuals. In the same line, the Research Excellence Framework (REF) in the UK proclaimed not to use journal rating. Authoritative sources have pleaded for an integrated use of bibliometrics and peer review, warning against the exclusive use of bibliometric indicators for the assessment of individual researchers, in the context of recruitment, internal analysis, or promotion (WILSDON *et al*., 2015).

The policy instruments developed by public authorities to measure scientific performance and selectively allocate resources led to the transformation of the research production process within higher education institutions. Moreover, they rely on peer-review processes, reinforcing existing practices and traditions, and an academic elite (MUSSELIN, 2013).

> Such performative effects are, of course, higher when they weaken or marginalize alternative criteria and evaluation processes. Examining the use and impact of journal lists is, therefore, important not merely for better understanding, refining how such metrics are devised (see TRUEX *et al*., 2011 for a critical review), but also, and more significantly, for appreciating and questioning their constitutive role in defining and policing the focus and direction of research activity (MINGERS; WILLMOTT, 2013, p. 1052).

According to Musselin (2013), all evaluation processes that link funding to performance relates to the rise of what he calls an “incentivizing” state. Thus, governments, rather than prescribing how things should be done, develop “rules of the game,” which require compliant behaviors if one wants access to funding. Regarding “rules of the game”, while in many countries, researchers are pressured to frequently publish in high-ranking journals (LAWRENCE, 2003), in Brazil faculty and students have been pushed to post in the QUALIS highest categories as a way to assure funding of their graduate program by CAPES as well as by other national agencies.

In an attempt to specify the “rules of the game” practices for research metrics, the Leiden Manifesto (HICKS, 2015) proposed using “ten principles to guide research evaluation,” and also warned against “morphing the instrument into the goal.” In this regard, Oliveira and Amaral (2017) compared the practices of the Brazilian evaluation process with the principles established in that manifesto as well as in the San Francisco Declaration on Research Assessment (DORA, 2012) and in the Metric Tide (WILSDON *et al*.,2015) on the best use of metrics. The authors reported that the processes and practices in the Brazilian evaluation model mainly follow the principles established in those international documents. Contrary to what the authors advocated, since QUALIS is intended to be a national indicator of quality in a list format, it has become a new indicator or metric, whereby the community morphed the instrument into the goal, as cautioned by the Leiden Manifesto. The low diversity in journals used by the faculty and students to publish their work and their lower impact, as observed in our work, might be potential effects of the QUALIS system.

Such mechanisms can be conceptualized as of ‘reactivity,’ in which the basic idea is that individuals alter their behavior in reaction to being evaluated, observed, or measured (ESPELAND; SAUDER, 2007). Scientific elites, with access to more resources and established reputations within academic hierarchies, are expected to learn the rules and norms of an evaluation system quickly and then strategically and tactically manipulate them to maximize their advantages (MARQUES *et al*., 2017). This behavior shifting based on changing rules of the game can be understood as a form of ‘reverse engineering,’ which in general encourages an attitude of focusing on the number rather than what it is supposed to measure (ESPELAND, 2016).

Hence, implementing more advanced quantitative analysis as well as in-depth qualitative research on how quantitative performance measurements influence organizational behavior would benefit the understanding of the potential effects of those systems (SAUDER; ESPELAND, 2009;

RIJCKE *et al*., 2016). The proposed framework based on quantitative analyses is an attempt to monitor those potential effects, allowing research funding institutions to verify if the system is leading to desired outcomes and impacts. Hence, the rationalities of this system need to be well defined by CAPES, along with other research funding institutions in the country. Thus, the percentage of funding associated with this research evaluation system by each funding institution is an essential aspect to consider when evaluating the achievement of the intended results and to detect the unintended ones. Therefore, differently from the CAPES statement, the role of QUALIS is not exclusively to evaluate the scientific output of graduate programs. Additionally, other uses of QUALIS beyond the scope of graduate programs are surely the responsibility of CAPES, because almost all the research in Brazil is performed in universities inside the graduate programs. These are evaluated by CAPES that link program performance to fund. Moreover, the delineation of rationalities cannot disregard the different scientific output patterns among the distinct subject fields, as well as their purpose.

7 CONCLUSIONS AND FUTURE WORKS

This Thesis set out a study of a feasible framework for analyzing the scientific production behavior on research evaluation systems based on national ranking lists. The proposed framework sets the basis for the implementation of a computational system that could support the monitoring of the progress of a national research evaluation system. Besides, it was efficient to reveal some patterns as well as potential effects of the Brazilian journal evaluation system in academic production behavior during the last years. Some of these patterns refer to those subject fields more regionally and nationally engaged, especially the ones from SSH. In Brazil, the studied subject fields in SSH publish less in indexed journals. This Brazilian pattern for SSH is also worldwide. Fields in SSH are well known to usually publish less in journals indexed in the leading international databases for many reasons. They publish in a variety of output types that go beyond journals, which are not covered by those international databases. Additionally, many of their publications are in the country's language, considering that the purpose of SSH fields is overall more related to societal and cultural issues, thus influencing a diverse public.

What stood out in the SSH patterns in Brazil's case was that beyond publishing more in lower impact journals among the few ones indexed, the SSH subject fields also published proportionally less in these indexed journals over the years. Therefore, although there was an increase in the used indexed journals curve in SSH, there was a decrease in the same curve, considering the articles published on them. In sum, Brazil produces fewer and fewer articles in the primary international databases in SSH when considering the production of their articles over the years. If for measuring international visibility one considers the presence of an article in the leading international databases, it is possible to conclude that the overall scientific production in the SSH studied subject fields have become internationally less visible over the years. If the objective of QUALIS system was to improve the internationalization of scientific production based on the last criterion, the goal has not been achieved at least regarding the ratio to overall output in the SSH fields. On the other hand, if the QUALIS goal was to induce more quantity of articles in in highly ranked journals according to the system, it has been achieved. In this direction, the monitoring process of a system should take into account, considering the specificities of each field.

Nevertheless, literature has demonstrated that commercial indexing databases such as Scopus and WoS have shown no capability of keeping up with the rapid development of new international and specialized journals in these fields, consequently inadequately for evaluating internationalization in SSH (CURRY; LILLIS, 2010, p. 6; SIVERTSEN, 2016a). Moreover, more than 80% of all articles registered in these databases are from North America (the United States and Canada) and Europe together; thus, these countries have an advantage in the global research environment. As aforementioned, there is also a diversity in publication patterns in SSH both between and within fields. Thus, in Brazil's case, for example, the expected level of internationalization from Literature and Linguistic cannot be the same from Education or Management, albeit all of them are SSH fields. All these SSH characteristics led some countries to improve their output coverage by building national databases and repositories, which became alternatives for evaluation purposes.

Concerning the other subject fields, a found pattern was that, in the majority, the impact of the chosen journals has not changed or declined over the years. Moreover, comparing the ranking percentiles of QUALIS and Scopus, all subject fields, in general, produced more articles in A1 and A2 journals with the lowest impact by each percentile. Finally, these patterns along with the last one showed that the QUALIS system did not encourage publication in journals indexed in major international databases in some subject fields, and overall in high-impact journals.

Therefore, these patterns of specific publications, as well as the other ones found in this work, may be considered potential effects of QUALIS. It is noteworthy that no causality relation was established in this work since we were unable to isolate the possible consequences of other national and international factors. Hence, two QUALIS intrinsic characteristics may be related to these potential effects. One of them is that QUALIS generates some journal lists that are deemed national quality indicators of scientific production in Brazil. Second, these lists are somehow connected to research funding. Although CAPES considers other elements in its evaluation process, it has been the main one. Other research fund institutions and graduate programs have also been using QUALIS to distribute their resources in the national level.

In order to check that connection, the framework enabled to compare the three QUALIS lists from different periodic evaluations regarding their concentration of journals and articles. These lists, when shared, exhibited a few journals that remained in the system for at least two periodic assessments. Considering all ten years (2007-2016), this group concentrated the great majority of articles produced during that time. Among these journals, only a few ones were indexed in the SSH subject fields. Most of these journals either maintained the same QUALIS category or improved it over the periodic evaluations, although, in general, there was a significant decrease in their impact. Additionally, this movement of journals from the lowest to the highest categories matches with the increase in the average of articles in lower-impact journals in those categories in all subject fields. All these results are in some extent similar to those found by Butler in 2003, as discussed before. In that moment, Butler attributed the results to the PRFS developed in Australia. On other hand, it is important to highlight that PRFS do not operate in isolation, and other funding mechanisms as well as unrelated government policies can be involved in the consequences of evaluation (OECD, 2010, p. 15).

In brief, considering that QUALIS generates a list that is used at the national level as a quality indicator, Brazilian researchers may have been using them as a reference to choose journals where to publish. Therefore, faculty and students are probably selecting journals in the highest categories not indexed or with lower impact to game the system. By doing so, they will guarantee the same points and funds to their graduate programs in the end, thus publishing more, but with less effort. CAPES nowadays has the majority of federal funds to research in the country; therefore, the unrestricted economic importance of QUALIS as a national indicator of research quality may be seen as a way of maximizing the potential adverse effects on academic research behavior. Hence, monitoring research evaluation systems based on journal lists is very important, seeking to check constantly shifts in publication activity, their constitutive role in defining and policing the focus and direction of research activity. The developed framework enables any country that makes use of journal lists in its research evaluation to reproduce it. In the Brazilian case, the use of QUALIS should be monitored to verify whether its goals are being achieved, according to the government research agenda. This requires interdisciplinary collaboration between science policy studies, sociology of science, and bibliometrics, which can all contribute to the necessary analytical toolbox.

Future investigations are necessary to validate the kinds of conclusions that can be drawn from this study. Considering that Brazil has different sources of research funding at national and state levels, studies should consider tracking the percentage of direct and indirect financing linked to QUALIS on both levels. This action will enable CAPES to comprehend its strength to induce scientific production behaviors. Furthermore, more research is needed to verify the changes in terms of scientific production behavior before and after 2008, when CAPES funds become higher than CNPq.

In addition, surveys about the use of QUALIS ranking among researchers in the university and university college sector, as well as among rectors, deans, and heads of departments in the university and university college sector, might extend the explanations of the observed potential effects of that system. Besides, other bibliometric analysis of the QUALIS impact in the national and international context and its properties can be useful to address intended political goals. In the same direction, further studies on how significant differences in the intellectual and social organizations of CAPES scientific fields are likely to affect and be affected by QUALIS could be of high relevance.

REFERENCES

AAGAARD, Kaare. How incentives trickle down: Local use of a national bibliometric indicator system. **Science and Public Policy**, Oxford, v. 42, n. 5, p. 725-737, 2015.

AAGAARD, Kaare; BLOCH, Carter; SCHNEIDER, Jesper W. Impacts of performance-based research funding systems: The case of the Norwegian Publication Indicator. **Research Evaluation**, Oxford, v. 24, n. 2, p. 106-117, 2015.

ABRAMO, Giovanni; D'ANGELO, Ciriaco Andrea; DI COSTA, Flavia. National research assessment exercises: a comparison of peer review and bibliometrics rankings. **Scientometrics**, Berlin, v. 89, n. 3, p. 929-941, 2011.

ABRAMO, Giovanni; D'ANGELO, Ciriaco Andrea. Evaluating research: from informed peer review to bibliometrics. **Scientometrics**, Berlin, v. 87, n. 3, p. 499-514, 2011.

______. The VQR, I taly's second national research assessment: Methodological failures and ranking distortions. **Journal of the Association for Information Science and Technology**, Hoboken, v. 66, n. 11, p. 2202-2214, 2015.

ADLER, Nancy J.; HARZING, Anne-Wil. When knowledge wins: Transcending the sense and nonsense of academic rankings**. Academy of Management Learning & Education**, New York, v. 8, n. 1, p. 72-95, 2009.

AGENZIA NAZIONALE DI VALUTAZIONE DEL SISTEMA UNIVERSITARIO E DELLA RICERCA – ANVUR. **Valutazione della Qualità della ricerca 2004-2010**. Roma, 2011. Avaiable in https://www.anvur.it/wp-content/uploads/2011/11/bando_vqr_def_07_11.pdf. Accessed on: Jan. 28, 2020.

AHLGREN, Per; COLLIANDER, Cristian; PERSSON, Olle. Field normalized citation rates, field normalized journal impact and Norwegian weights for allocation of university research funds. **Scientometrics**, Berlin, v. 92, n. 3, p. 767-780, 2012.

AKSNES, Dag W.; TAXT, Randi Elisabeth. Peer reviews and bibliometric indicators: a comparative study at a Norwegian university. **Research evaluation**, Oxford, v. 13, n. 1, p. 33-41, 2004.

ALEXANDER, F. King. The changing face of accountability: Monitoring and assessing institutional performance in higher education. **The Journal of Higher Education**, Columbus, v. 71, n. 4, p. 411-431, 2000.

ALVES, Alexandre D.; YANASSE, Horacio H.; SOMA, Nei Y. LattesMiner: a multilingual DSL for information extraction from lattes platform. *In:* LOPES, Crista Videira (Ed.). **Proceedings of the compilation of the co-located workshops on DSM'11, TMC'11, AGERE! 2011, AOOPES'11, NEAT'11, & VMIL'11**. New York: Association for Computing Machinery, 2011. p. 85-92.

ALPERIN, Juan Pablo; FISCHMAN, Gustavo (eds.). **Made in Latin America**: open access, scholarly journals, and regional innovations. Ciudad Autónoma de Buenos Aires: CLACSO, 2015.

ALTBACH, Philip G. **Global perspectives on higher education**. Baltimore: JHU Press, 2016.

AMIN Mayur; MABE Michael. Impact Factors: Use and Abuse. **Medicina**, Buenos Aires, n. 63, p. 347-354, 2003.

ANCAIANI, Alessio *et al*. Evaluating scientific research in Italy: The 2004–10 research evaluation exercise. **Research Evaluation**, Oxford, v. 24, n. 3, p. 242-255, 2015.

ANFOSSI, Alberto *et al*. Large-scale assessment of research outputs through a weighted combination of bibliometric indicators. **Scientometrics**, Berlin, v. 107, n. 2, p. 671-683, 2016.

ARAUJO, Sonia M. La evaluación y la Universidad en Argentina: políticas, enfoques y prácticas. **Revista de La Educación Superior**, Cidade do México, v. 43, n. 172, p. 57-77, 2014.

ARCHAMBAULT, Éric; LARIVIÈRE, Vincent. History of the journal impact factor: Contingencies and consequences. **Scientometrics**, Berlin, v. 79, n. 3, p. 635-649, 2009.

ARTS AND HUMANITIES RESEARCH COUNCIL. **The Impact of AHRC Research 2010/11**. Wiltshire: Arts and Humanities Research Council Polaris House, 2011. Available at http://www.ahrc.ac.uk/About/Publications/Documents/AHRCImpactReport2011.pdf. Accessed on: Jan. 28, 2020.

BALANCIERI, Renato *et al*. A análise de redes de colaboração científica sob as novas tecnologias de informação e comunicação: um estudo na Plataforma Lattes. **Ciência da Informação**, Brazilia, v. 34, n. 1, p. 64-77, 2005.

BALL, Rafael; TUNGER, Dirk. Bibliometric analysis-A new business area for information professionals in libraries?. **Scientometrics**, Berlin, v. 66, n. 3, p. 561-577, 2006.

BARATA, Rita Barradas. Necessary changes in the evaluation of graduate programs in Brazil. **Interface, comunicação, saúde e educação**, Botucatu, v. 23, p.1-6, 2019.

BARATA, Rita Barradas *et al*. The configuration of the Brazilian scientific field. **Anais da Academia Brasileira de Ciências**, Rio de Janeiro, v. 86, n. 1, p. 505-521, 2014.

BARATA, Rita Barradas. Ten things you should know about the Qualis. **Revista Brasileira de Pós-Graduação**, Brazilia, v. 13, n. 30, p. 13-43, 2016.

BARTOL, Tomaz *et al*. Assessment of research fields in Scopus and Web of Science in the view of national research evaluation in Slovenia. **Scientometrics**, Berlin, v. 98, n. 2, p. 1491-1504, 2014.

BASTOW, Simon; DUNLEAVY, Patrick; TINKLER, Jane. **The impact of the social sciences**: How academics and their research make a difference. London: Sage, 2014.

BEDEIAN, Arthur G. The manuscript review process: The proper roles of authors, referees, and editors. **Journal of Management Inquiry**, Thousand Oaks, v. 12, n. 4, p. 331-338, 2003.

BERGSTROM, Carl. Eigenfactor: Measuring the value and prestige of scholarly journals. **College & Research Libraries News**, Chicago, v. 68, n. 5, p. 314-316, 2007.

BERTOCCHI, Graziella *et al*. Bibliometric evaluation vs. informed peer review: Evidence from Italy. **Research Policy**, Amsterdam, v. 44, n. 2, p. 451-466, 2015.

BESLEY, Tina A. C; PETERS, Michael A. Neoliberalism, performance and the assessment of educational research quality: Comparing United Kingdom, Australia & New Zealand. *In*: BESLEY, Tina A.C. **Assessing the Quality of Educational Research in Higher Education**. Boston: Brill Sense, 2009. p. 25-48.

BLOCH, Carter; SCHNEIDER, Jesper W. Performance-based funding models and researcher behavior: An analysis of the influence of the Norwegian Publication Indicator at the individual level. **Research Evaluation**, Oxford, v. 25, n. 4, p. 371-382, 2016.

BONACCORSI, Andrea; CICERO, Tindaro; FERRARA, Antonio. Journal ratings as predictors of articles quality in Arts, Humanities and Social Sciences: an analysis based on the Italian Research Evaluation Exercise. **F1000Research**, London, v. 4, n. 196, p. 1-18, 2015.

BORNMANN, Lutz. Peer review and bibliometric: potentials and problems. *In:* SHIN, Jungcheol; TEICHLER, Ulrich; TOUTKOUSHIAN, Rob. **University rankings**: Theoretical Basis, Methodology and Impacts on Global Higher Education. Dordrecht: Springer Netherlands, 2011a. p. 145-164.

______. Scientific peer review. **Annual Review of Information Science and Technology**, Maryland, v. 45, n. 1, p. 197-245, 2011b.

BRADFORD, Samuel C. Sources of information on specific subjects. **Engineering**, London, v. 137, p. 85-86, 1934.

BRAZIL. MINISTÉRIO DA EDUCAÇÃO - MEC. **Plano Nacional De Pós-Graduação (PNPG) 2005-2010**. Brazilia, 2005. Available at: https://www.capes.gov.br/images/stories/download/editais/PNPG_2005_2010.pdf. Accessed on: Jan. 28, 2020.

______. **Plano Nacional De Pós-Graduação (PNPG) 2011-2020**. Brazília, 2010. Available at: https://www.capes.gov.br/images/stories/download/Livros-PNPG-Volume-I-Mont.pdf. Accessed on: Jan. 28, 2020.

BURTON, Hilary D. Use of a virtual information system for bibliometric analysis**. Information Processing & Management**, Amsterdam, v. 24, n. 1, p. 39-44, 1988.

BUTLER, Linda. A list of published papers is no measure of value. **Nature**, London, v. 419, n. 6910, p. 877, 2002.

______. Explaining Australia's increased share of ISI publications—the effects of a funding formula based on publication counts. **Research Policy**, Amsterdam, v. 32, n. 1, p. 143-155, 2003a.

______. Modifying publication practices in response to funding formulas. **Research Evaluation**, Oxford, v. 12, n. 1, p. 39-46, 2003b

______. 'What Happens When Funding is Linked to Publication Counts?'. *In:* MOED, H. F.; GLANZEL, W; SCHMOCH, U (eds.). **Handbook of Quantitative Science and Technology Research**. Netherlands: Kluwer Academic Publishers, 2004. p. 389-405.

BUTLER, Linda *et al*. **Impacts of performance-based research funding systems**: A review of the concerns and the evidence. Paris: OECD Publishing, 2010.

BUTLER, Linda; MCALLISTER, Ian. Metrics or peer review? Evaluating the 2001 UK research assessment exercise in political science. **Political Studies Review**, California, v. 7, n. 1, p. 3-17, 2009.

CAMPANARIO, Juan Miguel. Empirical study of journal impact factors obtained using the classical two-year citation window versus a five-year citation window. **Scientometrics**, Berlin, v. 87, n. 1, p. 189-204, 2011.

CASTRO, Cláudio de Moura; SOARES, Gláucio AD. As avaliações da Capes. Pesquisa Universitária em Questão. Campinas: UNICAMP, 1986. p. 173-189.

CASTRO, Pablo de. The role of current research information systems (CRIS) in supporting open science implementation: the case of Strathclyde. **Informačné technológie a knižnice (ITLib)**, Bratislava, p. 21-30, nov. 2018.

______. **Practical application of CRIS/IR interoperability@ Uni Strathclyde**. *In:* INTERNATIONAL CONFERENCE ON OPEN REPOSITORIES, 14., 2019. Hamburg. **Annals** […]. Hamburg: OR2019, 2019.

CHEN, Xiaotian. Google Scholar's dramatic coverage improvement five years after debut. **Serials Review**, Amsterdam, v. 36, n. 4, p. 221-226, 2010.

CLERIDES, Sofronis; PASHARDES, Panos; POLYCARPOU, Alexandros. Peer review vs metric-based assessment: testing for bias in the RAE ratings of UK Economics Departments. **Economica**, London, v. 78, n. 311, p. 565-583, 2011.

CONSELHO NACIONAL DE DESENVOLVIMENTO CINETÍFICO E TECNOLOÓICO – CNPQ. **Plataforma Lattes**. Brazilia: CNPq, 2016.

COOPER, Simon; POLETTI, Anna The New ERA of Journal Ranking: The Consequences of Australia's Fraught Encounter with "Quality". **Australian Universities' Review**, Melbourne, v. 53, n. 1, p. 57-65, 2011.

COORDENAÇ ÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR – CAPES. **60 Anos – Revista Comemorativa**. Brazilia: CAPES, 2011.

COUTINHO, Marília. Ecology and enviromental sciences in Brazilian higher education. Graduate programs, research and intellectual identity. **Documentos de Trabalho**, Universidade de São Paulo, São Paulo, n. 6, 1996.

COUTINHO, Renato X *et al*. Brazilian scientific production in science education. **Scientometrics**, Berlin, v. 92, n. 3, p. 697-710, 2012.

CRONIN, Blaise; SUGIMOTO, Cassidy R (eds.). **Beyond bibliometrics:** Harnessing multidimensional indicators of scholarly impact. Massachusetts: MIT Press, 2014.

CRUZ-CASTRO, Laura; SANZ-MENÉNDEZ, Luis. Research evaluation in transition. Individual versus Organisational Assessment in Spain. *In:* WHITLEY, Richard; GLÄSER, Jochen (eds.). **The Changing Governance of the Sciences**. The Advent of the Research Evaluation Systems. Dordrecht: Springer, 2007. p. 205-223.

CURRIE, Jan. Globalization practices and the professoriate in Anglo-Pacific and North American universities. **Comparative Education Review**, Chicago, v. 42, n. 1, p. 15-29, 1998.

CURRY, Mary Jane; LILLIS, Theresa M. Academic research networks: Accessing resources for English-medium publishing. **English for Specific Purposes**, Amsterdam, v. 29, n. 4, p. 281-295, 2010.

CURRY, Stephen. Let's move beyond the rhetoric: it's time to change how we judge research. **Nature**, London, v. 554, n. 7691, 2018.

DAY, Theodore Eugene. The big consequences of small biases: A simulation of peer review. Research Policy, Amsterdã, v. 44, n. 6, p. 1266-1270, 2015.

DE SOUZA, Rosali Fernandez. Organização e representação de áreas do conhecimento em ciência e tecnologia: princípios de agregação em grandes áreas segundo diferentes contextos de produção e uso de informação. Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação, v. 11, n. 1, p. 27-41, 2006.

DECLARATION ON RESEARCH ASSESSMENT - DORA. **San Francisco Declaration on Research Assessment**. San Francisco, 2012. Available at: https://sfdora.org/read/. Accessed on: Jan. 30, 2020.

DERRICK, Gemma Elizabeth *et al*. The association between four citation metrics and peer rankings of research influence of Australian researchers in six fields of public health. **PLoS One**, San Francisco, v. 6, n. 4, 2011.

ELSEVIER. **‘Scopus Content Coverage Guide’**, Amsterdam, 2013. Available at: http://files.sciverse.com/documents/pdf/ContentCoverageGuide-jan-2013.pdf. Accessed on: Feb. 2, 2020.

ELTON, Lewis. The UK research assessment exercise: unintended consequences. **Higher Education Quarterly**, United Kingdom, v. 54, n. 3, p. 274-283, 2000.

ENGELS, Tim C. E.; OSSENBLOK, Truyken L. B.; SPRUYT, Eric H. J. Changing publication patterns in the social sciences and humanities, 2000–2009. **Scientometrics**, Berlin, v. 93, n. 2, p. 373-390, 2012.

ESPELAND, Wendy. Reverse engineering and emotional attachments as mechanisms mediating the effects of quantification. **Historical Social Research/Historische Sozialforschung**, Illinois, v. 41, n. 2, p. 280-304, 2016.

ESPELAND, Wendy; SAUDER, Michael. Rankings and reactivity: How public measures recreate social worlds. **American Journal of Sociology**, Chicago, v. 113, n. 1, p. 1-40, 2007.

FERRARA, Antonio; BONACCORSI, Andrea. How robust is journal rating in Humanities and Social Sciences? Evidence from a large-scale, multi-method exercise. **Research Evaluation**, Oxford, v. 25, n. 3, p. 279-291, 2016.

FERREIRA, Marieta de Moraes; MOREIRA, Regina da Luz. **CAPES 50 anos**: depoimentos ao CPDOC/FGV. Brazilia: FGV/CPDOC/CAPES, 2002.

FRIEND, Frederick J. Google Scholar: Potentially good for users of academic information. **Journal of Electronic Publishing**, Michigan, v. 9, n. 1, 2006.

GALIMBERTI, P.; MORNATI, S. The Italian model of distributed research information management systems: a case study. **Procedia Computer Science**, Amsterdam, n. 106, p. 183-195, 2017.

GALLO, Stephen A.; SULLIVAN, Joanne H.; GLISSON, Scott R. The influence of peer reviewer expertise on the evaluation of research funding applications. **PLoS One**, San Francisco, v. 11, n. 10, p. e0165147, 2016.

GANS, Joshua S.; SHEPHERD, George B. How are the mighty fallen: Rejected classic articles by leading economists. **Journal of Economic Perspectives**, Nashville, v. 8, n. 1, p. 165-179, 1994.

GARFIELD, Eugene. Citation analysis as a tool in journal evaluation. **Science**, Pennsylvania, v. 178, n. 4060, p. 471-479, 1972.

GEHANNO, Jean-François; ROLLIN, Laetitia; DARMONI, Stefan. Is the coverage of Google Scholar enough to be used alone for systematic reviews. **BMC Medical Informatics and Decision Making**, London, v. 13, n. 1, p. 7, 2013.

GEISLER, Eliezer; ABDALLAH, Wagdy. **The metrics of science and technology**. Santa Barbara/California: Greenwood Publishing Group, 2000.

GEUNA, Aldo; MARTIN, Ben R. University research evaluation and funding: An international comparison. **Minerva**, Berlin, v. 41, n. 4, p. 277-304, 2003.

GEUNA, Aldo; PIOLATTO, Matteo. Research assessment in the UK and Italy: Costly and difficult, but probably worth it (at least for a while). **Research Policy**, Amsterdam, v. 45, n. 1, p. 260-271, 2016.

GIMÉNEZ-TOLEDO, Elea *et al*. Taking scholarly books into account, part II: A comparison of 19 European countries in evaluation and funding. **Scientometrics**, Berlin, v. 118, n. 1, p. 233-251, 2019.

GLÄNZEL, Wolfgang. A bibliometric approach to social sciences. National research performances in 6 selected social science areas, 1990–1992. **Scientometrics**, Berlin, v. 35, n. 3, p. 291-307, 1996.

GLÄNZEL, Wolfgang; MOED, Henk. Journal impact measures in bibliometric research. **Scientometrics**, Berlin, v. 53, n. 2, p. 171-193, 2002.

GLÄSER, Jochen; LAUDEL, Grit. Evaluation without evaluators. *In:* WHITLEY, Richard; GLÄSER, Jochen (eds.). **The Changing Governance of the Sciences**. The Advent of the Research Evaluation Systems. Dordrecht: Springer, 2007. p. 127-151.

GOERGEN, Pedro. Educação superior na perspectiva do sistema e do plano nacional de educação. **Educação & Sociedade**, Campinas, v. 31, n. 112, p. 895-917, jul./set. 2010.

GONZÁLEZ-PEREIRA, Borja; GUERRERO-BOTE, Vicente P.; MOYA-ANEGÓN, Félix. A new approach to the metric of journals' scientific prestige: The SJR indicator. **Journal of Informetrics**, Amsterdam, v. 4, n. 3, p. 379-391, 2010.

GORRAIZ, Juan; PURNELL, Philip J.; GLÄNZEL, Wolfgang. Opportunities for and limitations of the Book Citation Index. **Journal of the American Society for Information Science and Technology**, Hoboken, v. 64, n. 7, p. 1388-1398, 2013.

GUERRERO-BOTE, Vicente P.; MOYA-ANEGÓN, Félix. A further step forward in measuring journals' scientific prestige: The SJR2 indicator. **Journal of Informetrics**, Amsterdam, v. 6, n. 4, p. 674-688, 2012.

GUETZKOW, Joshua; LAMONT, Michèle; MALLARD, Grégoire. What is Originality in the Humanities and the Social Sciences?. **American Sociological Review**, Thousand Oaks v. 69, n. 2, p. 190-212, 2004.

GUZ, Alexander N.; RUSHCHITSKY, Jeremiah. J. Scopus: A system for the evaluation of scientific journals. **International Applied Mechanics**, Berlin, v. 45, n. 4, p. 351, 2009.

HADDOW, Gaby; GENONI, Paul. Citation analysis and peer ranking of Australian social science journals. **Scientometrics**, Berlin, v. 85, n. 2, p. 471-487, 2010.

HAMANN, Julian. The visible hand of research performance assessment. **Higher Education**, New York, v. 72, n. 6, p. 761-779, 2016.

HAMMARFELT, Björn; DE RIJCKE, Sarah. Accountability in context: Effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the faculty of Arts at Uppsala University. **Research Evaluation**, Oxford, v. 24, n. 1, p. 63-77, 2015.

HAMMARFELT, Björn. Interdisciplinarity and the intellectual base of literature studies: Citation analysis of highly cited monographs. **Scientometrics**, Berlin, v. 86, n. 3, p. 705-725, 2011.

HAMMARFELT, Björn. Taking comfort in points: The appeal of the Norwegian Model in Sweden. **Journal of Data and Information Science**, Beijing, v. 3, n. 4, p. 84-94, 2018.

HARLEY, Sandra. The impact of research selectivity on academic work and identity in UK universities. **Studies in Higher Education**, Abingdon-on-Thames, v. 27, n. 2, p. 187-205, 2002.

HARZING, Anne-Wil. A preliminary test of Google Scholar as a source for citation data: a longitudinal study of Nobel prize winners. **Scientometrics**, Berlin, v. 94, n. 3, p. 1057-1075, 2013.

_____. **Publish or Perish**. London, 2007. Available at: http://www.harzing.com/pop.htm. Accessed on: Jan. 30, 2020

HARZING, Anne-Wil; VAN DER WAL, Ron. A Google Scholar h-index for journals: An alternative metric to measure journal impact in economics and business. **Journal of the American Society for Information Science and technology**, Hoboken, v. 60, n. 1, p. 41-46, 2009.

_____. Google Scholar as a new source for citation analysis. **Ethics in Science and Environmental Politics**, v. 8, n. 1, p. 61-73, 2008.

HASLAM, Nick; KOVAL, Peter. Possible research area bias in the Excellence in Research for Australia (ERA) draft journal rankings. **Australian Journal of Psychology**, Melbourne, v. 62, n. 2, p. 112-114, 2010.

HEINRICH, Carolyn J.; MARSCHKE, Gerald. Incentives and their dynamics in public sector performance management systems. **Journal of Policy Analysis and Management**, Washington, v. 29, n. 1, p. 183-208, 2010.

HEY, Anthony; TANSLEY, Stewart; TOLLE, Kristin M. Jim Gray on eScience: a transformed scientific method. *In:* HEY, Tony. **The Fourth Paradigm**. Washington: Microsoft Research, 2009. p. 16-31

HICKS, Diana *et al*. Bibliometrics: the Leiden Manifesto for research metrics. **Nature**, London, News, v. 520, n. 7548, p. 429, 2015.

HICKS, Diana. Performance-based university research funding systems. **Research Policy**, Amsterdam, v. 41, n. 2, p. 251-261, 2012.

______. The four literatures of social science. *In:* MOED, H. F.; GLANZEL, W; SCHMOCH, U (eds.). **Handbook of Quantitative Science and Technology Research**. Netherlands: Kluwer Academic Publishers, 2004. p. 473-496.

HICKS, Diana; WANG, Jian. Coverage and overlap of the new social sciences and humanities journal lists. **Journal of the American Society for Information Science and Technology**, Hoboken, v. 62, n. 2, p. 284-294, 2011.

HOOD, William; WILSON, Concepción. Informetric studies using databases: Opportunities and challenges. **Scientometrics**, Berlin, v. 58, n. 3, p. 587-608, 2003.

______. The literature of bibliometrics, scientometrics, and informetrics. **Scientometrics**, Berlin, v. 52, n. 2, p. 291-314, 2001.

HORNBOSTEL, Stefan. **From CRIS to CRIS**: integration and interoperability. Leuven: University Press, 2006.

HORROBIN, David F. The philosophical basis of peer review and the suppression of innovation. **Jama**, Illinois, v. 263, n. 10, p. 1438-1441, 1990.

HOSTINS, Regina Célia Linhares. Os Planos Nacionais de Pós-graduação (PNPG) e suas repercussões na Pós-graduação brasileira. **Perspectiva**, Florianopolis, v. 24, n. 1, p. 133-160, 2006.

IGLESIAS, Juan E.; PECHARROMÁN, Carlos. Scaling the h-index for different scientific ISI fields. **Scientometrics**, Berlin, v. 73, n. 3, p. 303-320, 2007.

INGWERSEN, Peter; LARSEN, Birger. Influence of a performance indicator on Danish research production and citation impact 2000–12. **Scientometrics**, Berlin, v. 101, n. 2, p. 1325-1344, 2014.

JACSO, Peter. As we may search—comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. **Current Science**, Karnataka, v. 89, n. 9, p. 1537-1547, 2005a.

______. **Comparison and analysis of the citedness scores in Web of Science and Google Scholar**. *In:* International Conference on Asian Digital Libraries. Berlin: Springer, 2005b. p. 360-369.

______. Google Scholar: the pros and the cons. **Online Information Review**, Bingley, v. 29, n. 2, p. 208-214, apr. 2005c.

JAYASINGHE, Upali W.; MARSH, Herbert W.; BOND, Nigel. Peer review in the funding of research in higher education: The Australian experience**. Educational Evaluation and Policy Analysis**, Thousand Oaks, v. 23, n. 4, p. 343-364, 2001.

JEFFERY, Keith; ASSERSON, Anne. CERIF-CRIS for the European e-Infrastructure. **Data Science Journal**, France, v. 9, p. 1004260203-1004260203, 2010.

JEFFERY, Keith G. CRIS in 2020. *In:* JEFFERY, Keith G.; DVOŘÁK, Jan (eds.), "**e-Infrastructures for Research and Innovation Linking Information Systems to Improve Scientific Knowledge Production"**: Proceedings of the 11th International Conference on Current Research Information Systems. Prague: euroCRIS2012, 2012. p. 333-342

JEFFERY, Keith *et al*. Research information management: the CERIF approach. **International Journal of Metadata, Semantics and Ontologies**, Genèva, v. 1, p. 5-14, 2014.

JIMÉNEZ-CONTRERAS, E.; MOYA ANEGÓN, F de.; LÓPEZ-CÓZAR, E. D. The evolution of research activity in Spain: The impact of the National Commission for the Evaluation of Research Activity (CNEAI). **Research Policy**, Amsterdam, v. 32, n. 1, p. 123-142, 2003.

JÖRG, Brigitte. CERIF: The common European research information format model. **Data Science Journal**, France, v.9/special, p. 1006280236-1006280236, 2010.

JÖRG, Brigitte; HÖLLRIGL, Thorsten; SICILIA, Miguel Angek. Entities and Identities. *In:* JEFFERY, Keith G.; DVOŘÁK, Jan (eds.), "**e-Infrastructures for Research and Innovation Linking Information Systems to Improve Scientific Knowledge Production"**: Proceedings of the 11th International Conference on Current Research Information Systems. Prague: euroCRIS2012, 2012. p. 185-194.

KHABSA, Madian; GILES, C. Lee. The number of scholarly documents on the public web. **PloS One**, San Francisco, v. 9, n. 5, 2014.

KEKÄLE, Jouni. Conceptions of quality in four different disciplines. **Tertiary Education and Management**, Canberra, v. 8, n. 1, p. 65-80, 2002.

KOGAN, Maurice; HANNEY, Stephen. **Reforming higher education**. London: Jessica Kingsley Publishers, 2000.

KUENZER, Acacia Zeneida; MORAES, Maria Célia Marcondes de. Temas e tramas na pós-graduação em educação. **Educação & Sociedade**, Campinas, v. 26, n. 93, p. 1341-1362, 2005.

KULCZYCKI, Emanuel *et al*. Publication patterns in the social sciences and humanities: evidence from eight European countries. **Scientometrics**, Berlin, v. 116, n. 1, p. 463-486, 2018.

KULCZYCKI, Emanuel; KORZEŃ, Marcin; KORYTKOWSKI, Przemysław. Toward an excellence-based research funding system: Evidence from Poland. **Journal of Informetrics**, Amsterdam, v. 11, n. 1, p. 282-298, 2017.

KULCZYCKI, Emanuel; KORYTKOWSKI, Przemysław. Redesigning the model of book evaluation in the Polish performance-based research funding system. **Journal of Data and Information Science**, Beijing, v. 3, n. 4, p. 60-72, 2018.

KURMIS, Andrew P. Understanding the limitations of the journal impact factor. **The Journal of Bone and Joint Surgery**, Evanston, v. 85, n. 12, p. 2449-2454, 2003.

LANCHO-BARRANTES, Bárbara S.; GUERRERO-BOTE, Vicente P.; MOYA-ANEGÓN, Félix. What lies behind the averages and significance of citation indicators in different disciplines?. **Journal of Information Science**, Thousand Oaks, v. 36, n. 3, p. 371-382, 2010.

LANE, Julia. Let's make science metrics more scientific. **Nature**, London, v. 464, n. 7288, p. 488-489, 2010.

LANE, Julia; LARGENT, Mark; ROSEN, Rebecca. Science Metrics and Science Policy CRONIN, Blaise; SUGIMOTO, Cassidy R (eds.). **Beyond bibliometrics**: Harnessing multidimensional indicators of scholarly impact. Massachusetts: MIT Press, 2014. p. 397.

LANGFELDT, Liv. The decision-making constraints and processes of grant peer review, and their effects on the review outcome. **Social Studies of Science**, Thousand Oaks, v. 31, n. 6, p. 820-841, 2001.

LARIVIÈRE, Vincent; GINGRAS, Yves; ARCHAMBAULT, Éric. Canadian collaboration networks: A comparative analysis of the natural sciences, social sciences and the humanities. **Scientometrics**, Berlin, v. 68, n. 3, p. 519-533, 2006.

LARIVIÈRE, Vincent; MACALUSO, Benoit. 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 and Technology**, Hoboken, v. 62, n. 12, p. 2437-2442, 2011.

LAWRENCE, Peter A. The politics of publication. **Nature**, London, v. 422, n. 6929, p. 259-261, 2003.

LAUDEL, Grit. The art of getting funded: how scientists adapt to their funding conditions. **Science and Public Policy**, Oxford, v. 33, n. 7, p. 489-504, 2006.

LEE, Carole J *et al*. Bias in peer review. **Journal of the American Society for Information Science and Technology**, Hoboken, v. 64, n. 1, p. 2-17, 2013.

LEITE, Paula; MUGNAINI, Rogério; LETA, Jacqueline. A new indicator for international visibility: exploring Brazilian scientific community. **Scientometrics**, Berlin, v. 88, n. 1, p. 311, 2011.

LEIVA-MEDEROS, Amed *et al*. Working framework of semantic interoperability for CRIS with heterogeneous data sources. **Journal of Documentation**, West Yorkshire, v. 73, n. 3, 2017.

LIEFNER, Ingo. Funding, resource allocation, and performance in higher education systems. **Higher Education**, New York, v. 46, n. 4, p. 469-489, 2003.

LOWRY, Paul Benjamin *et al*. Evaluating journal quality and the association for information systems senior scholars' journal basket via bibliometric measures: Do expert journal assessments add value?. **MIS Quarterly**, Minnesota, v. 37, n. 4, p. 993-1012, dec.2013.

MARQUES, Marcelo *et al*. How does research evaluation impact educational research? Exploring intended and unintended consequences of research assessment in the United Kingdom, 1986–2014. **European Educational Research Journal**, Thousand Oaks, v. 16, n. 6, p. 820-842, 2017.

MARSH, Herbert W.; JAYASINGHE, Upali W.; BOND, Nigel W. Improving the peer-review process for grant applications: reliability, validity, bias, and generalizability. **American Psychologist**, Washington, v. 63, n. 3, p. 160, 2008.

MARTIN, B.R.; WHITLEY, R. The UK Research Assessment Exercise: a case of regulatory capture? *In*: WHITLEY, R.; GLÄSER, J.; ENGWALL, L (eds.). **Reconfiguring Knowledge Production**: Changing Authority Relationships in the Sciences and Their Consequences for Intellectual Innovation. Oxford: Oxford University Press, 2010. p. 51-80.

MARTÍN-MARTÍN, Alberto *et al*. Does Google Scholar contain all highly cited documents (1950-2013)?. arXiv preprint arXiv:1410.8464, 2014.

MARTIN-SARDESAI, Ann *et al*. Government research evaluations and academic freedom: A UK and Australian comparison. **Higher Education Research & Development**, v. 36, n. 2, p. 372-385, 2017.

MCCARTHY, Greg; SONG, Xianlin; JAYASURIYA, Kanishka. The proletarianisation of academic labour in Australia. **Higher Education Research & Development**, v. 36, n. 5, p. 1017-1030, 2017.

MEEK, V.; SUWANWELA, Charas (Ed.). **Higher education, research, and knowledge in the Asia-Pacific region**. Dordrecht: Springer, 2007.

MEIER, John J.; CONKLING, Thomas W. Google Scholar's coverage of the engineering literature: an empirical study. **The Journal of Academic Librarianship**, v. 34, n. 3, p. 196-201, 2008.

MINELLI, Eliana; REBORA, Gianfranco; TURRI, Matteo. How can evaluation fail? The case of Italian universities. **Quality in Higher Education**, Abingdon, v. 14, n. 2, p. 157-173, 2008.

MINGERS, John; LEYDESDORFF, Loet. A review of theory and practice in scientometrics. **European Journal of Operational Research**, v. 246, n. 1, p. 1-19, 2015a.

______. Identifying research fields within business and management: A journal cross-citation analysis. **Journal of the Operational Research Society**, v. 66, n. 8, p. 1370-1384, 2015b.

MINGERS, John; YANG, Liying. Evaluating journal quality: A review of journal citation indicators and ranking in business and management. **European Journal of Operational Research**, v. 257, n. 1, p. 323-337, 2017.

MINGERS, John; WILLMOTT, Hugh. Taylorizing business school research: On the 'one best way'performative effects of journal ranking lists. **Human Relations**, v. 66, n. 8, p. 1051-1073, 2013.

MIRANDA, Elaine Cristina; MUGNAINI, Rogerio. Estudo exploratório dos critérios de classificação de livros nas áreas de Ciências Sociais Aplicadas, Humanas e Linguística, Letras e Artes. Anais, 2014.

MOED, Henk F. Bibliometric indicators reflect publication and management strategies. **Scientometrics**, Berlin, v. 47, n. 2, p. 323-346, 2000.

______. Citation analysis of scientific journals and journal impact measures. **Current Science**, Karnataka, p. 1990-1996, 2005.

______. UK Research Assessment Exercises: Informed judgments on research quality or quantity? **Scientometrics**, Berlin, v. 74, n. 1, p. 153-161, 2008.

______. The Source-Normalized Impact per Paper (SNIP) is a valid and sophisticated indicator of journal citation impact. arXiv preprint arXiv:1005.4906, 2010.

MOLAS-GALLART, Jordi. Research governance and the role of evaluation: A comparative study. **American Journal of Evaluation**, v. 33, n. 4, p. 583-598, 2012.

MOSBAH-NATANSON, Sébastien; GINGRAS, Yves. The globalization of social sciences? Evidence from a quantitative analysis of 30 years of production, collaboration and citations in the social sciences (1980–2009). **Current Sociology**, v. 62, n. 5, p. 626-646, 2014.

MOSKALEVA, Olga *et al*. Russian index of Science citation: Overview and review. **Scientometrics**, Berlin, v. 116, n. 1, p. 449-462, 2018.

MUGNAINI, Rogério. Anos de bibliometria no Brasil: da bibliografia estatística à avaliação da produção científica nacional. *In:* HAYASHI, Maria Cristina Piumbato Innocentini; LETA Jacqueline (eds.). **Bibliometria e cientometria**: reflexões teóricas e interfaces. São Carlos: Pedro e João, 2013. p. 37-58.

_____. **Ciclo avaliativo de periódicos no Brasil**: caminho virtuoso ou colcha de retalhos? *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO. 16., 2015. João Pessoa. **Annals** [...]. Paraíba: ANCIB, 2015.

MUSSELIN, Christine. How peer review empowers the academic profession and university managers: Changes in relationships between the state, universities and the professoriate. **Research Policy**, Amsterdam, v. 42, n. 5, p. 1165-1173, 2013.

NEDERHOF, Anton J. Bibliometric monitoring of research performance in the social sciences and the humanities: A review. **Scientometrics**, Berlin, v. 66, n. 1, p. 81-100, 2006.

NORRIS, Michael; OPPENHEIM, Charles. Citation counts and the research assessment exercise V. **Journal of Documentation**, West Yorkshire,2003.

_____. Peer review and the h-index: Two studies. **Journal of Informetrics**, v. 4, n. 3, p. 221-232, 2010.

NOTESS, G. R. Scholarly web searching: Google Scholar and Scirus. Online, 29 (4), 39-41. Reed, KL (1995). Citation analysis of faculty publications: Beyond Science Citation Index and Social Science [sic] Citation Index. **Bulletin of the Medical Library Association**, v. 83, n. 4, p. 503-508, 2005.

OANCEA, Alis; PRING, Richard. The importance of being thorough: on systematic accumulations of 'what works' in education research. **Journal of Philosophy of Education**, Hoboken, v. 42, p. 15-39, 2008.

OLIVEIRA, Talita Moreira de; AMARAL, Livio. Políticas Públicas em Ciência e Tecnologia no Brasil: desafios e propostas para utilização de indicadores na avaliação. *In:* MUNAINI, Rogério; FUJINO, Asa; KOBASH, Nair Yumiko (eds.). **Bibliometria e Cientometria no Brasil**: infraestrutura para avaliação da pesquisa científica na Era do Big Data. São Paulo: Escola de Comunicação e Artes USP, 2017. p. 157-184.

ORDUÑA-MALEA, Enrique *et al*. Methods for estimating the size of Google Scholar. **Scientometrics**, Berlin, v. 104, n. 3, p. 931-949, 2015.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT – OECD. **Glossary of Key Terms in Evaluation and Results Based Management**. OECD Publishing: Paris, 2002.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT – OECD. **Performance-based Funding for Public Research in Tertiary Education Institutions**: Workshop Proceedings, OECD Publishing: Paris, 2010. https://doi.org/10.1787/9789264094611-en.

OSUNA, Carmen; CRUZ-CASTRO, Laura; SANZ-MENÉNDEZ, Luis. Overturning some assumptions about the effects of evaluation systems on publication performance. **Scientometrics**, Berlin, v. 86, n. 3, p. 575-592, 2011.

OSSENBLOK, Truyken. **Scientific communication in the social sciences and humanities: Analysis of publication and collaboration patterns in Flanders**. Doctoral Thesis University of Antwerp, 2016.

PACKER, Abel L. Os periódicos brasileiros e a comunicação da pesquisa nacional. Revista USP, n. 89, p. 26-61, 2011.

PARKER, Jonathan. Comparing research and teaching in university promotion criteria. **Higher Education Quarterly**, United Kingdom, v. 62, n. 3, p. 237-251, 2008.

PENDLEBURY, David A. **White Paper Using Bibliometrics in Evaluating Research**. Philadelphia: Thomson Reuters,
2008.

PRITCHARD, Alan *et al*. Statistical bibliography or bibliometrics**. Journal of Documentation**, West Yorkshire, v. 25, n. 4, p. 348-349, 1969.

PUUSKA, Hanna-Mari. Scholarly Publishing Patterns in Finland-A comparison of disciplinary groups. 2014.

QIU, Jane. Publish or perish in China: the pressure to rack up publications in high-impact journals could encourage misconduct, some say. **Nature**, London, v. 463, n. 7278, p. 142-144, 2010.

REALE, Emanuela; BARBARA, Anna; COSTANTINI, Antonio. Peer review for the evaluation of academic research: lessons from the Italian experience. **Research Evaluation**, Oxford, v. 16, n. 3, p. 216-228, 2007.

REALE, Emanuela *et al*. The social construction of indicators for evaluation: Internationalization of Funding Agencies. **Research Evaluation**, Oxford, v. 21, n. 4, p. 245-256, 2012.

REBORA, Gianfranco; TURRI, Matteo. The UK and Italian research assessment exercises face to face. **Research policy**, Amsterdam, v. 42, n. 9, p. 1657-1666, 2013.

RIBEIRO, Renato Janine. Quality in Research–Capes'point of View1. *In:* AUDY, Jorge Luis Nicolas; MOROSINI, Marília Costa (eds.). **Innovation and Quality in the University**. Porto Alegre, 2008. p. 127-178.

RIJCKE, Sara de *et al*. Evaluation practices and effects of indicator use—a literature review. **Research Evaluation**, Oxford, v. 25, n. 2, p. 161-169, 2016.

ROBERTS, Peter. Performativity, measurement and research: A critique of performance-based research funding in New Zealand. *In:* OZGA, Jenny; POPKEWITZ, Thomas; SEDDON, Terry.

World yearbook in education: Education research and policy (eds.). London: Routledge Falmer, 2006. p. 188–195.

ROUSSEAU, Ronald. Journal evaluation: Technical and practical issues. **Library Trends**, Illinois, v. 50, n. 3, p. 418-439, 2002.

RUSSELL, Jane M.; ROUSSEAU, Ronald. Bibliometrics and institutional evaluation. Encyclopedia of life support systems (EOLSS). **Part**, Oxford, v. 19, p. 42-64, 2010.

SÁ BARRETO, Francisco César de; DOMINGUES, Ivan; BORGES, Mário Neto. The Brazilian National Graduate Program, Past, Present and Future: a short review. **Policy Futures in Education**, Thousand Oaks, v. 12, n. 5, p. 695-706, 2014.

SAPIRO, Gisèle; SEILER-JUILLERET, Hélène. Disseminating the Social Sciences and Humanities. 2016. ffhalshs-01659501f

SAUDER, Michael; ESPELAND, Wendy Nelson. The discipline of rankings: Tight coupling and organizational change. **American Sociological Review**, Thousand Oaks, v. 74, n. 1, p. 63-82, 2009.

SCHWARTZMAN, Simon. Uses and abuses of education assessment in Brazil. **Prospects**, London, v. 43, n. 3, p. 269-288, 2013.

SCHNEIDER, Jesper W. An outline of the bibliometric indicator used for performance-based funding of research institutions in Norway. **European Political Science**, London v. 8, n. 3, p. 364-378, 2009.

SCHÖPFEL, Joachim; PROST, Hélène; REBOUILLAT, Violaine. Research data in current research information systems. **Procedia Computer Science**, Amsterdam, n. 106, p. 305-320, 2017.

SCHOTTEN, Michiel; EL AISATI, M'hamed. The Rise of National Research Assessments, and the Tools and Data That Make Them Work. **Elsevier Connect**, Amsterdam, v. 17, 2014.

SEGLEN, Per Ottar. From bad to worse: evaluation by Journal Impact. **Trends in biochemical Sciences**, Cambridge, v. 14, n. 8, p. 326-327, 1989.

______. The skewness of science. **Journal of the American Society for Information Science**, Hoboken, v. 43, n. 9, p. 628-638, 1992.

SĪLE, Linda *et al*. **European databases and repositories for social sciences and humanities research output**: report. Antwerp: ECOOM & ENRESSH, 2017.

SĪLE, Linda *et al*. Comprehensiveness of national bibliographic databases for social sciences and humanities: Findings from a European survey. **Research Evaluation**, Oxford, v. 27, n. 4, p. 310-322, 2018.

SIQUEIRA, Manoel Brod. Sucupira-A Platform for the Evaluation of Graduate Education in Brazil. **Procedia Computer Science**, Amsterdam, v. 146, p. 247-255, 2019.

SINDICATO NACIONAL DOS DOCENTES DAS INSTITUIÇÕES DE ENSINO SUPERIOR - ANDES-SN. **Crise de financiamento das Universidades Federais e da Ciência e Tecnologia**: Recursos existem, mas vão para o pagamento da dívida pública. Brazilia: ANDES, 2018.

SIVERTSEN, Gunnar; AAGAARD, Kaare. The effects of performance-based research funding systems. **R-QUEST Policy Brief**, Oslo, n. 2, p. 1-4, 2017.

SIVERTSEN, Gunnar. A performance indicator based on complete data for the scientific publication output at research institutions. **ISSI Newsletter**, Leuven, v. 6, n. 1, p. 22-28, 2010.

______. Patterns of internationalization and criteria for research assessment in the social sciences and humanities. **Scientometrics**, Berlin, v. 107, n. 2, p. 357-368, 2016a.

______. Publication-based funding: The Norwegian model. *In:* OCHSNER, Michael; HUG, Sven E.; DANIEL, Hans-Dieter (eds.). **Research Assessment in the Humanities**. Cham: Springer, 2016b. p. 79-90.

______. Publication patterns in all fields. *In:* ÅSTRÖM, Fredrik *et al. (*eds.). **Celebrating scholarly communication studies**: A Festschrift for Olle Persson at his 60th birthday, 2009. p. 55-60.

______. The norwegian model in Norway. **Journal of Data and Information Science**, Beijing, v. 3, n. 4, p. 2-18, 2018.

______. Understanding and evaluating research and scholarly publishing in the Social Sciences and Humanities (SSH). **Data and Information Management**, Warsaw, v. 1, n. ahead-of-print, 2019.

SOUZA, Elaine Pereira de; PAULA, Maria Carlota de Souza. Qualis: a base de qualificação dos periódicos científicos utilizada na avaliação CAPES. **InfoCAPES Boletim Informativo**, Brazilia, v. 10, n. 2, 2002.

SPORN, Barbara. Convergence **or Divergence in International Higher Education Policy**: Lessons from Europe. Cambridge, MA: Ford Policy Forum, 2003. p. 31–44.

STERN, David I. Uncertainty measures for economics journal impact factors. **Journal of Economic Literature**, Tennessee, v. 51, n. 1, p. 173-89, 2013.

STOJANOVSKI, Jadranka. Bibliography in the network environment: Croatian Scientific Bibliography (CROSBI). *In:* INTERNATIONAL CONFERENCE ON INFORMATION TECHNOLOGY INTERFACES. 21., 1999. **Annals** [...]. Zagreb: ITI'99, 1999. p. 359-364.

TAPPER, T.; SALTER, B. Interpreting the process of change in higher education: The case of the research assessment exercises. **Higher Education Quarterly**, United Kingdom, v. 57, n. 1, p. 4-23, 2003.

TAYLOR, Jim. The assessment of research quality in UK universities: peer review or metrics?. **British Journal of Management**, Hoboken, v. 22, n. 2, p. 202-217, 2011.

THOMAS, P. R.; WATKINS, David S. Institutional research rankings via bibliometric analysis and direct peer review: A comparative case study with policy implications. **Scientometrics**, Berlin, v. 41, n. 3, p. 335-355, 1998.

TRAAG, Vincent A *et al*. Systematic analysis of agreement between metrics and peer review in the UK REF. **Palgrave Communications**, London, v. 5, 2019.

VANCLAY, Jerome K. An evaluation of the Australian Research Council's journal ranking. **Journal of Informetrics**, Amsterdam, v. 5, n. 2, p. 265-274, 2011.

VAN DEN BESSELAAR, Peter; HEYMAN, Ulf; SANDSTRÖM, Ulf. Perverse effects of output-based research funding? Butler's Australian case revisited. **Journal of Informetrics**, Amsterdam, v. 11, n. 3, p. 905-918, 2017.

VAN LEEUWEN, Thed. The application of bibliometric analyses in the evaluation of social science research. Who benefits from it, and why it is still feasible. **Scientometrics**, Berlin, v. 66, n. 1, p. 133-154, 2006.

VEUGELERS, R *et al*. **Internationalisation of R&D**: Trends, Issues and Implications for S&T policies. *In:* Background report for the OECD forum on the internationalization of R&D. Brussels: Organisation for Economic Co-operation and Development, 2005.

VERHINE, Robert E.; DANTAS, Lys MV. Reflexões sobre o sistema de avaliação da capes a partir do V Plano Nacional de Pós-graduação. **Revista de Educação Pública**, Cuiabá, v. 18, n. 37, p. 295-310, 2009.

VERLEYSEN, Frederik; GHESQUIÈRE, Pol; ENGELS, Tim. The objectives, design and selection process of the Flemish Academic Bibliographic Database for the Social Sciences and Humanities (VABB-SHW). 2014.

VIEIRA, Elizabeth S.; GOMES, José ANF. A comparison of Scopus and Web of Science for a typical university. **Scientometrics**, Berlin, v. 81, n. 2, p. 587, 2009.

WALTMAN, Ludo *et al*. Some modifications to the SNIP journal impact indicator. **Journal of Informetrics**, Amsterdam, v. 7, n. 2, p. 272-285, 2013.

WALTMAN, Ludo. A review of the literature on citation impact indicators. **Journal of Informetrics**, Amsterdam, v. 10, n. 2, p. 365-391, 2016.

WATERMEYER, Richard. Issues in the articulation of 'impact': the responses of UK academics to 'impact'as a new measure of research assessment. **Studies in Higher Education**, Abingdon-on-Thames, v. 39, n. 2, p. 359-377, 2014.

WHITLEY, Richard. Competition and pluralism in the public sciences: the impact of institutional frameworks on the organisation of academic science. **Research Policy**, Amsterdam, v. 32, n. 6, p. 1015-1029, 2003.

WHITLEY, Richard *et al*. **The intellectual and social organization of the sciences**. Oxford: University Press on Demand, 2000.

WHITLEY, Richard; GLÄSER, Jochen. The changing governance of the sciences. **Sociology of the sciences yearbook,** London, v. 26, 2007.

WILHITE, Allen W.; FONG, Eric A. Coercive citation in academic publishing. **Science**, Pennsylvania, v. 335, n. 6068, p. 542-543, 2012.

WILSDON, James *et al*. **The Metric Tide**: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. 2015. DOI: 10.13140/RG.2.1.4929.1363

WOUTERS, Paul *et al*. **The metric tide literature review, Supplementary Report I to the Independent Review of the Role of Metrics in Research Assessment and Management**. London: HEFCE, 2015.

YOKOYAMA, Keiko. The effect of the research assessment exercise on organisational culture in English universities: collegiality versus managerialism. **Tertiary Education & Management**, Camberra, v. 12, n. 4, p. 311-322, 2006.

YOUNG, Suzanne *et al*. The impact of journal ranking fetishism on Australian policy-related research. **Australian Universities' Review**, Melbourne, v. 53, n. 2, p. 77, 2011.

ZIMMERMAN, E.; JEFFERY, Keith G. The Need for a Standardised Current Research Information System (CRIS): A Call to Arms. *In:* INTERNATIONAL CONFERENCE ON CURRENT RESEARCH INFORMATION SYSTEMS. 2004. Antwerp. **Annals** […]. Antwerp: ICIS, 20014. p. 153-160.