A Comparison between a Chinese Bibliometric Database and the Web of Science in terms of Authors and their Output

Fei Shu
School of Information Studies
McGill University, Montreal
August 2018

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

TABLE OF CONTENTS

ACKNOWLEDGEMENT

Doctoral research is a long journey and I therefore have many to thank for everyone helping me during my doctoral studies. First and foremost, I thank my supervisors, Charles-Antoine Julien and Vincent Larivière, for their supervision, support and advice over the years. I also thank all assistance given by other faculty members, staff, fellow students, research assistants, technicians, families, and others in the collection of materials and data, the analysis of data, and the preparation of the thesis. I have stored up all these names in my heart.

CONTRIBUTION OF AUTHORS

This thesis consists primarily of three papers, which are solely my own work and writing, however all three benefited from the review, feedback, and advice of my thesis committee: Dr. Charles-Antoine Julien, Dr. Vincent Larivière, Dr. Jamshid Beheshti and Dr. Ilja Frissen, with paper 3 receiving additional help from Wei Quan at Wuhan University and Dr. Bikun Chen at Nanjing University of Science and Technology prior to its submission for publication.

Paper 2 further benefited from comments and data collection provided by Dr. Lin Zhang at Wuhan University and Jing Zhang at National Science Library of Chinese Academy of Sciences. Paper 3 also benefited from anonymous reviewers from the Aslib Journal of Information Management in addition to Wei and Bikun. As the principle investigator of paper 3, I designed the research, analyzed the data and drafted the manuscript, while Wei and Bikun collected the data. Please note that the author order of paper 3 is based on Chinese co-authors' request that my name is in the last author position as the corresponding author. The contribution of authors for this paper has been declared in the published article as below:

> *This study is a part of Fei Shu's dissertation research, which is supported by iFellows Doctoral Scholarship provided by Andrew W. Mellon Foundation and The Fonds de recherche du Québec – Société et culture (FRQSC)(Quan, Chen, & Shu, 2017, p. 486).*
>
> *……*
>
> *Fei Shu is a PhD Candidate in the School of Information Studies at McGill University. His research interests focus on bibliometrics, informetrics, scholarly communication and research evaluation. He also works with Dr Vincent Larivière at the Université de Montréal on various research projects regarding the transformations of scholarly communications. He is the Principal Investigator of this paper. Fei Shu is the corresponding author and can be contacted at: fei.shu@mail.mcgill.ca (Quan et al., 2017, p. 502).*

This thesis research is funded by the Fonds de recherche du Québec Société et culture (FRQSC) and the iFellows Doctoral Scholarship provided by Andrew W. Mellon Foundation.

ABSTRACT

In this thesis I compare the Web of Science (WoS) with a Chinese bibliometric database in terms of authors and their output, and address the issues that could interpret the results. This is done in three chapters, constituting a comparison between two bibliometric databases, a methodological contribution, and an investigation on the monetary reward policies in China.

In the first chapter I compare Web of Science (WoS) with a Chinese bibliometric database in terms of authors and their output, demonstrate the extent of the overlap between the two groups of Chinese scientific elites in both international and Chinese bibliometric databases, and determine how different disciplines may affect this overlap. The results of this study indicate that Chinese bibliometric databases, or a combination of WoS and Chinese bibliometric databases, should be used to evaluate Chinese research performance except in few disciplines in which Chinese research performance could be assessed using WoS only.

Since WoS adopts the journal classification system while a paper classification system is applied in Chinese bibliometric databases, in the second chapter I compare the science classification systems at the journal level and paper level for the same data set using the same classification scheme. Results show almost half of papers may be misclassified in the current journal classification system, which could methodologically explain the difference found in the first chapter.

In the third chapter I investigate the China's monetary reward policies and present the landscape of the cash-per-publication reward policy in China and reveal its trend since the late 1990s. Since the monetary reward policies may influence Chinese scholars’ publication activities reported in the first chapter. The investigation could explain why the difference found in the first chapter exist by analyzing the impact of science policies on Chinese scholars’ publishing activities.

In summary, this thesis presents three original research papers that advance knowledge in bibliometrics and scholarly communication. This thesis also lays the necessary foundation for further study on research evaluation, classification of science, and science policies. The most notable implication for the future research is to investigate the impact of metrics-based science policies, which focus on outcomes such as number of publications and rankings, on research output, either in China or worldwide.

RÉSUMÉ

Cette thèse porte sur une étude de la production scientifique des chercheurs chinois en comparant les données disponibles dans le Web of Science (WoS) et dans une base de données chinoise. Dans le but d'identifier les problèmes qui pourraient survenir lors de l'interprétation des résultats, cette thèse est organisée en trois chapitres où nous détaillons les différentes caractéristiques des deux bases de données utilisées lors de notre recherche, nous précisons l'apport méthodologique de cette étude et nous analysons les politiques de récompenses monétaires octroyées pour le travail scientifique en Chine.

Le premier chapitre explique les différences entre le WoS et la base de données chinoise en utilisant les variables des auteurs et de leurs production. Nous y exposons l'étendue du chevauchement entre deux groupes de chercheurs élites et la manière dont les pratiques disciplinaires déterminent ce chevauchement. Les résultats de cette étude indiquent que l'usage des bases de données bibliométriques chinoises ou une combinaison du WoS et de bases de données chinoises constituer un moyen plus fiable pour l'évaluation de la recherche dans le contexte chinois. Des exemptions peuvent se faire dans le cas de quelques disciplines pour lesquelles l'usage du WoS semble approprié.

Dans le chapitre 2, nous nous intéressons aux différences entre les systèmes de classification scientifique du WoS (fait à partir des revues) et de la base de données chinoise (fait à partir des articles). En comparant un même ensemble de données dans l'une et l'autre base de données, nous arrivons à la conclusion que le système fondé sur la classification des revues (WoS) génère une classification erronée pour près de la moitié des articles, ce qui explique – en termes méthodologiques – les différences rapportées dans le premier chapitre.

Le troisième chapitre présente une étude des politiques des récompenses monétaires et de la politique de récompense par publication (*cash-per-publication*) et un portrait général de la politique monétaire mise en place en Chine dans les années 1990. Considérant que les récompenses monétaires peuvent déterminer les pratiques de publication des chercheurs analysées dans le premier chapitre, nous étudions les résultats obtenus lors des évaluations de l'impact des politiques scientifiques sur les pratiques de publication des chercheurs chinois.

En résumé, cette thèse présente trois articles de recherche originaux qui contribuent à l'avancement des recherches en bibliométrie et en communication savante. Cette thèse contribue également à la poursuite des études sur l'évaluation de la recherche, la classification des sciences et les politiques scientifiques. L'implication la plus notable pour la recherche future est d'étudier l'impact des politiques scientifiques basées sur les métriques qui se concentrent sur les résultats tels que le nombre de publications et les classements, ou encore sur les retombées de la recherche, en Chine ou ailleurs dans le monde.

LIST OF FIGURES

LIST OF TABLES

1. Introduction

With the significant development of China's economy and scientific activity, China's scientific publication is experiencing a period of rapid growth. Since 2009, China[1] has become the second largest source country in terms of its share of international scientific production (ISTIC, 2010), contributing 17.1% of scientific articles indexed by Clarivate's Web of Science (WoS) (ISTIC, 2017). The emergence of China as a leading nation in science has changed the balance of power among the formerly leading nations as measured by scientific production (Leydesdorff & Zhou, 2005; Zhou & Leydesdorff, 2006; Zhou, Thijs, & Glänzel, 2009a). As a result, more and more bibliometric studies focus on China and try to evaluate China's contribution to the world's scientific activity.

Previous studies indicate significant differences between international and Chinese bibliometric databases when analyzing their co-authorship network (Hennemann, Wang, & Liefner, 2011) and citation counts (Meho & Yang, 2007). Previous studies point out that it is very important to understand the extent to which Chinese authors publish their articles in international journals in order to determine the proportion of Chinese scientific production covered in international bibliometric databases (Jin et al., 1999; Moed, 2002b; Rousseau, 2015). Unfortunately, no current study addresses the question of the coverage of Chinese scientific production.

[1] In this study, China refers to mainland China, which is the geopolitical area under the direct jurisdiction of the People's Republic of China excluding Hong Kong and Macau.

The objectives of this thesis are to compare an international bibliometric database (i.e., WoS) with a Chinese bibliometric database in terms of authors and their output, to demonstrate the extent of the overlap between the two groups of Chinese scientific elites in both international and Chinese bibliometric databases, and to determine the effect of disciplines. The results of this study will indicate the extent to which international bibliometric databases can be used to evaluate Chinese national research production as a whole and in individual research disciplines.

1.1. Bibliometrics

The term *bibliometrics* was created by Paul Otlet in 1934 as the measurement of all aspects related to the publication and reading of books and documents (Otlet, 1934). It is also defined as "the statistical or quantitative description of a literature" (Nicholas & Ritchie, 1978, p. 9), "the quantification of bibliographic information for use in analysis" (Garfield, Malin, & Small, 1978, p. 180), and "the study and analysis of scientific output with the use of publication-based data" (Melkers, 1993, p. 44). Bibliometrics is frequently used in the field of library and information science, and the American Library Association (ALA) describes it as "the use of statistical methods in the analysis of a body of literature to reveal the historical development of subject fields and patterns of authorship, publication, and use" (Young & Belanger, 1983, p. 22). Finally, the Organization for Economic Co-operation Development (OECD) (2002) also states that "bibliometrics is the generic term for data on publications collecting data on numbers of scientific articles and other publications, classified by author and/or by institution, field of science, country, etc., in order to construct simple 'productivity' indicators for academic research" (p. 203-204).

Bibliometrics became more important with the appearance of the Science Citation Index (SCI) created by Eugene Garfield and his Institute for Scientific Information (ISI) in 1963. Citation indexes allow for the efficient citation tracking and journal impact evaluation based on the number of citations (Garfield, 1972, 1979), and the exploration of research networks by co-citation analysis (Small, 1973). Price (1963) used quantitative methods to describe the growth of science, which established the field of *bibliometrics* as “the science of science” (Price, 1965). Bibliometrics is now a research method that is broadly used in many research fields to explore the advancement of knowledge and the impact of research in a specific field (Melkers, 1993).

The source for bibliometrics is a database that is called the bibliometric database (Martin et al., 2010; Okubo, 1997), which should not be confused with the bibliographic database. For the purpose of research evaluation, the bibliometric database contains details of cited references, full institutional and author details as well as some indicators that could be used in bibliometric studies. On the other hand, the main purpose of the bibliographic database is not to evaluate research but help literature retrieval; many bibliographic databases lack the data needed for bibliometric studies. Although some bibliographic databases offer author affiliations and cited references, such data are not in a consistent and comparable form (Martin et al., 2010).

There are three major bibliometric databases: Web of Science (WoS), Scopus and Google Scholar. WoS includes the Science Citation Index Expanded (SCIE), the Social Science Citation Index (SSCI) and the Arts and Humanities Citation Index (AHCI), which annually indexes documents published in about 12,000 journals, covering all areas of research. WoS, developed by Clarivate (Previously Thomson Reuter) was the only bibliometric data source before the emergence of Scopus and Google Scholar; it is the only bibliometric database covering a century's worth of scientific production (Moed, 2005). Scopus also provides publication and citation data for 1996 onwards; Scopus could be an alternative to WoS (Norris & Oppenheim, 2007; Torres-Salinas, Lopez-Cózar, & Jiménez-Contreras, 2009) but some studies find that it contains more uncertainty, sources of error, and potential misinterpretations (Leydesdorff, 2012; Meho & Yang, 2007). Google Scholar is not considered a valid database to measure scholarly activity (Norris & Oppenheim, 2007) because it is an unstructured data source and Google refuses to open its source data. Some scholars argue that Google Scholar could not be used for bibliometric studies because of its poor data quality and data selection (Hicks & Wang, 2009; White, 2006). Thus, WoS was selected as the bibliometric database in this study.

Bibliometric studies typically start by choosing the databases that best represent the population under study (Okubo, 1997). This is always a challenge for bibliometric studies on China's research activities because no single database can cover all Chinese scientific literature that may be in either Chinese or English. For example, in 2016, Chinese scholars produced about 494,200 articles (mostly in Chinese) in Chinese scientific journals that are indexed by the Chinese Scientific and Technical Papers and Citations Database (CSTPCD), while they also published about 324,300 articles (most in English) in international journals that are indexed by WoS (ISTIC, 2017). Therefore, combining international bibliometric databases with local Chinese bibliometric databases and measuring their overlap is necessary in order to accurately measure China's research activities in terms of the number of publications and citations (Jin et al., 1999; Jin, Zhang, Chen, & Zhu, 2002; Liang, 2003; Moed, 2002b); this work has yet to be reported.

1.2. Structure of the Thesis

This thesis consists primarily of three papers, constituting Chapters 2, 3, and 4. These three papers have been or will be published in the following journals:

- Shu, F., Julien, C-A., & Larivière, V. (submitted). Does the Web of Science accurately represent Chinese scientific output? *Journal of the Association for Information Science and Technology*.
- Shu, F., Julien, C-A., Zhang, L., Zhang, J., & Larivière, V. (submitted). A comparison of classification system of science between journal level and paper level. *Journal of Informetrics.*
- Quan, W., Chen, B., & Shu, F. (2017). Publish or impoverish: An investigation of the monetary reward system of science in China (1999-2016). *Aslib Journal of Information Management*, *69*(5), 486-502.

We compare the WoS with the Chinese Science and Technology Periodical Citation Database (VIP) in terms of most highly productive Chinese authors as well as their output in 115 selected disciplines in Chapter 2. The comparison answers the following questions: 1) in each discipline, are the most highly productive Chinese authors the same in the WoS and in the VIP?; 2) in each discipline, are the institutional affiliations of the group of most highly productive Chinese authors in WoS different from the institutional affiliations of the group in VIP?; 3) what is the overlap between both groups of most highly productive Chinese authors different in the various disciplines?

Classifying science into disciplinary structure is an essential element in bibliometric studies, and WoS adopts the journal classification system assigning indexed journals to about

250 WoS categories while Chinese bibliometric databases classify the science at the paper level (the paper classification system) using the Chinese Library Classification Scheme. The difference, in terms of the structure of classification system, may affect the comparison between WoS and Chinese bibliometric databases, but it has never been systematically investigated. In Chapter 3, we address the issue by comparing the classification system of science between the journal level and the paper level from the same data set using the same classification scheme, which reveals the extent of possible paper misclassification between different classification systems.

Although Chapter 2 attempts to reveal the differences between WoS and Chinese bibliometric databases by answering the three research questions outlined above, it does not explain why the differences exist. In Chapter 4, we investigate the cash-per-publication reward policy in China and try to explore how science policies influence Chinese scholars' publishing activities. The results presented in this study also form a foundation for future studies that investigate the determinants and consequences of monetary reward policies.

While the three main chapters (Chapter 2, 3, 4) of this thesis build linearly upon one another, and are ordered accordingly, the conclusion section, Chapter 5, notes how the thesis objectives were met and how the contributions were made, summarises the thesis findings and limitations as well as discusses directions for future research.

1.3. Original Contributions to Knowledge

The following are the elements of this thesis that constitute original scholarship and distinct contributions to knowledge:

- Paper 1 (Chapter 2)

1. Revealing the differences between WoS and Chinese bibliometric databases at the level of authors and their publication strategies across all domains.
2. Identification of different publication patterns among Chinese scholars in different domains.
3. Identification of different publication patterns among Chinese scholars from different institutional sectors.
4. Design and creation of a methodology for author name disambiguation and institution name disambiguation.

- Paper 2 (Chapter 3)

1. First comparison of the classification system of science between the journal level and the paper level from the same data set using the same classification scheme.
2. Improving our understanding of the accuracy of the classification system of science by indicating the extent of paper misclassification in the journal classification system.
3. Identification of need for developing the paper classification system instead of the journal classification system.

- Paper 3 (Chapter 4)

1. First paper presenting the landscape of the cash-per-publication reward policy in China, which has never been systematically studied and investigated before.
2. Identification of trends of the monetary reward policy in China.
3. Design and creation of a methodology collecting, comparing and analyzing the monetary reward policies.

2. Does the Web of Science Accurately Represent Chinese Scientific Output?

2.1. Introduction

Over the last 20 years, the contribution of China to the world's scientific activity—as measured by its number of Web of Science (WoS) publications—has increased at an impressive rate (Zhou, 2013). While part of this trend might be attributed to an increase in the number of research papers written in English by Chinese researchers (Montgomery, 2013), some Chinese scholars might still prefer to publish their manuscripts in Chinese scholarly journals that are indexed by Chinese bibliometric databases only (Jin et al., 2002; Moed, 2002b). Hence, measuring China's research output remains a challenge, as no bibliometric database covers both Chinese and English scientific literature.

Many scholars have concluded that the WoS is not an appropriate tool to measure Chinese research output (Guan & He, 2005; Jin & Rousseau, 2004; Zhou & Leydesdorff, 2007), as significant differences have been found in the coverage of international and national Chinese bibliometric databases (Hennemann et al., 2011; Meho & Yang, 2007). While previous work has attempted to explain differences between WoS and Chinese bibliometric databases by looking at journal hierarchies and citation relations (Zhou & Leydesdorff, 2007), or regional publications (Liang, 2003), no research has yet analysed the discrepancies at the level of authors. For instance, little is known on the extent to which scholars from Chinese institutions publish their articles in international journals, or whether "top" Chinese authors give up publishing papers in Chinese in order to be more visible internationally. A better understanding of these trends might help to explain the differences between the international and national Chinese bibliometric databases.

The purpose of this study is to compare an international bibliometric database (i.e., WoS) with a national Chinese bibliometric database in terms of authors and their publications, demonstrate the extent of the overlap between the two groups of Chinese scientific elites in both international and Chinese bibliometric databases, and determine how different disciplines affect this overlap. The results of this study can reveal the extent to which international bibliometric databases can be used to evaluate Chinese national research production as a whole, and in individual research disciplines.

2.2. Literature Review

2.2.1. Background

Over the past two decades, China's Research and Development (R&D) expenditures have been linked to an increase of the research production of Chinese researchers (Y. Sun & Cao, 2014). As Figure 1 shows, China's scientific research inputs and outputs have exhibited constant growth between 1995 and 2015. More specifically, China's R&D expenditures increased more than 40 times over this period, from 5.23 billion USD[2] to 212.55 billion USD (National Bureau of Statistics of China, 1996-2016), while its number of international publications (indexed by WoS) increased by more than 22 times, from 12,997 to 287,374. Since 2009, China has become the

[2] China's R&D expenditures are counted in Chinese Yuan (CNY), which are converted to US dollars (USD) for reference in this study. To avoid the impact of exchange rate changes on the monetary values, a fixed exchange rate was selected for the conversion. In this study, all monetary values in Chinese Yuan (CNY) were converted to US dollars (USD) at the rate 1 CNY = 0.15 USD, which was retrieved from xe.com on October 3, 2016.

second largest producer of scholarly papers, and contributes 17.1% of scientific articles indexed by WoS (ISTIC, 2017). Although the US is still the global leader in science and technology (S&T), its global share of S&T activities is declining as China continues to close the gap in the international race for scientific supremacy (National Science Board, 2018).

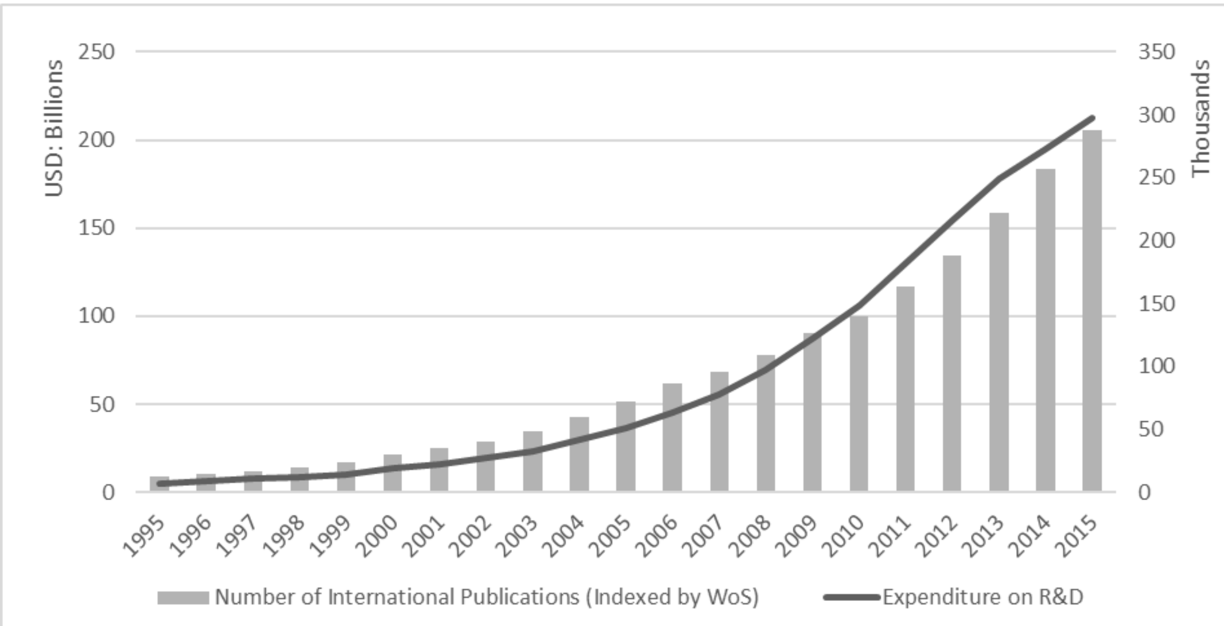

Figure 1. Research Inputs and Outputs in China (1995-2015)

In order to increase the international visibility of Chinese research, the Chinese government, Chinese universities, as well as Chinese scientific institutes[3] offer preferential policies and monetary rewards to encourage scholars to publish in international journals (Cao, Li, Li, & Liu, 2013; Peng, 2011; Quan et al., 2017). The preferential policies give more weight to international publications, specifically the number of WoS papers, when evaluating the research

[3] In China, scientific institutes refer to independent scientific institutes that purely focus on research and development (e.g., Chinese Academy of Science) as opposed to higher education institutions (e.g., universities) which are engaged in both teaching and research. These scientific institutes are administered by the Ministry of Science and Technology or other Ministries while all higher education institutions are administered by Ministry of Education.

output (Cao et al., 2013; Gong & Qu, 2010). After investigating the landscape of cash-per-publication policy in China, Quan et al. (2017) revealed that Chinese scholars could be awarded up to 165,000 USD for a single paper published in WoS journals with high Impact Factors. The pressure to publish in WoS journals is sizeable; WoS papers are central to research funding, faculty appointments and promotions (Peng, 2011; J. Qiu, 2010), in addition to monetary awards. Some PhD programs even require that their students publish a WoS paper as a condition of graduation (Cao et al., 2013; Yuan, Xu, & Hu, 2013). However, these preferential policies and monetary rewards vary by institute and geographic region, which may influence Chinese scholars' publishing behaviour (Quan et al., 2017).

Writing in English can, however, be challenging for Chinese scholars. It may require significant effort and money (Cargill, Oconnor, & Li, 2012; Ge, 2015; Montgomery, 2013), and Chinese authors have to bear a higher rejection rate than English-speaking scholars (Lee, Sugimoto, Zhang, & Cronin, 2013). Some Chinese scholars may also question their English abilities and be unwilling to submit their papers to international journals (Montgomery, 2013). For these reasons, many Chinese scholars still prefer to publish in Chinese scholarly journals (Jin et al., 2002; Moed, 2002b). According to the National Bureau of Statistics of China (1996-2016), the number of papers published in Chinese journals also increased nearly fivefold from 107,924 in 1995 to 497,849 in 2014, contributing to around two-thirds of all scholarly papers published by Chinese authors in 2014. Thus, some have argued that publishing in an English journal reflects the author's English language proficiency more than their research abilities (Cargill et al., 2012; Montgomery, 2013), and that WoS papers do not accurately represent China's research output (Guan & He, 2005; Jin & Rousseau, 2004; Zhou & Leydesdorff, 2007).

2.2.2. Bibliometric Studies on China

Partly as a result of its preferential policies and monetary incentives, China's international publication count has increased at an exponential rate of 20% annually since the 1990s (Kostoff, Briggs, Rushenberg, Eowles, et al., 2007). This makes China the fastest growing source of scholarly article publications and the second largest source country in terms of the number of articles published in WoS (National Science Board, 2018). A number of studies have attempted to characterize Chinese research achievements beyond the number of scholarly publications. Wu et al. (1991) is one of the first bibliometric studies to provide a brief overview of China's research activities, and it revealed that an international bibliometric database (i.e. WoS) could be a good supplement to national Chinese bibliometric databases when evaluating Chinese research output. Table 1 shows that several international and Chinese scholars have tracked China's publication records, measured China's collaborative networks, evaluated their academic journals and bibliometric databases, assessed Chinese university rankings, and analyzed their international visibility.

The results obtained by bibliometric studies depend on the chosen dataset. Table 1 also presents the various databases used in these bibliometric studies; it shows that 52 out of 63 articles have used international bibliometric databases (i.e., WoS, Scopus) while 24 out of 63 articles used national Chinese bibliometric databases (i.e., Chinese Scientific and Technical Papers and Citations Database (CSTPCD); Chinese Science Citation Database (CSCD); China Academic Journal Full-text Database (CJFD); Chinese Science and Technology Periodical Citation Database (VIP); Chinese Social Science Citation Index (CSSCI)) to evaluate Chinese research production, and that 13 studies (bold in Table 1) selected data from both international and national Chinese sources. This shows that international databases, especially WoS (used by 46 out of 63 articles), are still the major data sources for bibliometric studies on China. However, the extent to which they are representative of the Chinese output is not known. In other words, it is unclear whether the trends observed in international databases differ from those found in national Chinese databases. This study attempts to address this issue.

Table 1. Overview of Articles Using Bibliometric Databases in the Context of China

		General Evaluation of Chinese Research Output	**Collaboration Network**	**Journal & Database**	**University Ranking**	**International Visibility**
International Bibliometric Database	**WoS**	Gao and Guan (2009); Guan and He (2005); Jin and Rousseau (2004, 2005); Kostoff, Briggs, Rushenberg, Bowles, et al. (2007); Kostoff, Briggs, Rushenberg, Eowles, et al. (2007); Leydesdorff and Zhou (2005); Liang (2003); Liang, Havemann, Heinz, and Wagner-Döbler (2006); Liu, Tang, Gu, and Hu (2015); Mely, El Kader, Dudognon, and Okubo (1998); Meng, Hu, and Liu (2006); Moed (2002b); Zhi and Meng (2016); Zhou and Leydesdorff (2006); Zhou et al. (2009a); Zhou, Thijs, and Glänzel (2009b)	**Hennemann et al. (2011)**; He (2009); J. Li and Li (2015); Niu and Qiu (2014); Park, Yoon, and Leydesdorff (2016); L. L. Wang and Wang (2017); X. Wang, Xu, Wang, Peng, and Wang (2013); H. Zhang and Guo (1997); Zheng et al. (2012); Zhou and Glänzel (2010)	**Basu (2010); Leydesdorff and Jin (2005); Liang (2003)**; Ren and Rousseau (2002); Shelton, Foland, and Gorelskyy (2009); S. Wang, Wang, and Weldon (2007); S. Wang and Weldon (2006); **Zhou and Leydesdorff (2007)**	Cheng and Liu (2008); Fu and Ho (2013); **Liang, Wu, and Li (2001); Meho and Yang (2007)**; J. P. Qiu, Yang, and Zhao (2010)	**Basu (2010);** Fu, Chuang, Wang, and Ho (2011); Leydesdorff and Jin (2005); J. P. Qiu et al. (2010); Ren and Rousseau (2002); Shu and Larivière (2015)
	Scopus	L. L. Wang (2016)	Royle, Coles, Williams, and Evans (2007);	**Basu (2010);** Ding,	**Meho and Yang (2007);** Zhu, Hassan,	**Basu (2010); Shu and Lariviere**

			W. Wang, Wu, and Pan (2014)	Zheng, and Wu (2012)	Mirza, and Xie (2014)	**(2015);** L. L. Wang (2016)
National Chinese Bibliometric Database	**CSCD**	**Jin and Rousseau (2004); Liang (2003); Moed (2002)**	Liang and Zhu (2002)	**Leydesdorff and Jin (2005); Liang (2003);** Jin and Wang (1999); Rousseau, Jin, and Yang (2001)	**Liang, Wu, and Li (2001)**	**Leydesdorff and Jin (2005);** Rousseau et al. (2001)
	CSTPCD	**Liang, Havemann, Heinz, and Wagner-Döbler (2006); Liang (2003); Guan and He (2005); Z. Wang, Li, Li, and Li (2012)**	Yan Wang, Wu, Pan, Ma, and Rousseau (2005)	Wu et al. (2004); **Zhou and Leydesdorff (2007);**	**Liang, Wu, and Li (2001)**	
	CJFD	Hu, Guo, and Hou (2017); **Z. Wang et al. (2012)**				Yang, Ma, Song, and Qiu (2010)
	VIP	**Z. Wang et al. (2012)**	**Hennemann et al. (2011)**			
	CSSCI	Song, Ma, and Yang (2015)	Yan, Ding, and Zhu (2010)			

Note: 13 studies (bold) selected data from both international and national Chinese sources.

2.2.3. Problem Statement

No bibliometric database contains all the published literature. However, bibliometric databases must be representative of the population of researchers they aim at studying (Okubo, 1997). Several studies have shown that the WoS may not adequately represent China's research activities (Guan & He, 2005; Jin & Rousseau, 2004; Jin et al., 2002; Liang et al., 2001; Moed, 2002b; Zhou & Leydesdorff, 2007), as more than 97% of Chinese language scholarly journals are excluded from its coverage (ISTIC, 2014). Moed (2002b) and Liang (2003) have suggested using national Chinese bibliometric databases to assess Chinese research output. Some researchers have combined the WoS with national Chinese bibliometric databases and report significant differences between the databases when investigating co-authorship networks (Hennemann et al., 2011), regional publications (Liang, 2003) as well as citation analysis (Meho & Yang, 2007). Taken as a whole, these results suggest that the WoS and national Chinese bibliometric databases tell different stories about Chinese research, although it is not clear how much they differ.

Critical factors to consider when analyzing data from different bibliometric databases are their coverage and comparability, which determine the study's validity and reliability (Hennemann et al., 2011). Previous studies show that coverage differences between WoS and national Chinese bibliometric databases will lead to different results (Liang, 2003; Zhou & Leydesdorff, 2007). It is not known to what extent all differences can be attributed to differences in coverage. This study addresses the lack of current research comparing coverage between WoS and a national bibliometric Chinese database at the level of individual authors.

2.2.4. Scientific Elites

According to Merton (1957), science could be regarded as a social institution, with values, norms, and organization, which can reward its members (scientists) for their performance. Merton (1973) also points out that rewards for scientific achievement can be given only if others recognize it. As a result, scientists are eager to present their achievements via publications because "a scientist is rewarded through recognition for producing results which are seen as new, important and true" (Gilbert, 1977, p. 116). Scholars typically diffuse their research findings in the form of peer-reviewed journal articles or monographs, and the authorship of a publication assigns credit to its creator(s) and formally establishes the responsibility for the published work (Cole & Cole, 1967). Author productivity in terms of the number of publications is a basic descriptive indicator of performance in bibliometrics. The most highly productive authors are defined as scientific elites by Zuckerman (1977) because they contribute the majority of literature produced in their disciplines (Larivière, Macaluso, Archambault, & Gingras, 2010; Price, 1963).

In addition, since scholars in different disciplines have different traditions and habits of publication, publication activities vary significantly by discipline (Glänzel, 2003; Larivière, Archambault, Gingras, & Vignola-Gagnè, 2006; Okubo, 1997). Scholars in the Social Sciences and Humanities publish their works in books or monographs in addition to journals, through which scholars in the Natural Sciences diffuse most of their research findings. It is difficult to make comparisons of different disciplines due to the disciplinary variation (Okubo, 1997). Thus, this study compares two groups of Chinese scientific elites between WoS and a Chinese bibliometric database discipline by discipline.

2.2.5. Research Questions

The purpose of this study is to compare the overlap in the scientific elite of Chinese scholars found in WoS and in a national Chinese bibliometric database, and describe the differences observed according to disciplines. It will answer the following research questions:

1. In a given discipline, are the researchers who represent the scientific elite (in terms of numbers of publications) the same in WoS and in a Chinese database?

2. In a given discipline, are the institutional affiliations of the group of Chinese scientific elites in WoS different from the institutional affiliations of the group in a Chinese database?
3. How does the overlap between the scientific elites differ by discipline?

On the whole, this analysis aims to improve our understanding of the differences between WoS and national Chinese bibliometric database, at the level of authors and their publication strategies, and provides insights into the coverage of the Chinese research activity obtained through the WoS. More specifically, it contributes to the knowledge of 1) the extent to which scholars from Chinese institutions publish their articles in international journals; 2) whether Chinese scientific elites give up publishing papers in English; and 3) whether these two factors explain the differences between WoS and a Chinese database.

2.3. Methodology

2.3.1. Research Design

2.3.1.1. Database

In this study, the Web of Science (WoS) and the Chinese Science and Technology Periodical Citation Database (VIP) are used as data sources because of their coverage and representation. WoS is the only bibliometric database covering a century of citation-based indicators for all disciplines, as well as, since 1973, all authors and their institutional affiliations (Moed, 2005). Indeed, most previous bibliometric studies on China are based on WoS (Zhou & Leydesdorff, 2007). Although there are five major bibliometric databases in China (see Table 2), VIP has the largest coverage and offers author rankings in terms of publications and citations that are not provided by other databases (Zhao, Lei, Ma, & Qiu, 2008).

The Chinese Science and Technology Periodical Citation Database (VIP) was established by the CQVIP Corporation in 1994. VIP indexes about 14,000 academic journals covering all disciplines, more than any other Chinese bibliometric database. VIP offers bibliometric indicators that measure Chinese scientific research output in terms of the number of publications and citations by authors, institutions, journals or topics. which are not provided in other databases.

Table 2. Comparisons of Five Chinese Bibliometric Databases

	Chinese Science Citation Database	**Chinese Science and Technology Paper and Citation Database**	**Chinese Social Science Citation Index**	**China Academic Journals Full-Text Database**	**Chinese Science and Technology Periodical Citation Database**
Appreciation	CSCD	CSTPCD	CSSCI	CJFD	VIP
Chinese Name	中国科学引文数据库	中国科技论文与引文数据库	中文社会科学引文索引	中国学术期刊全文数据库	中文科技期刊引文数据库
URL	http://sciencechina.cn	http://www.istic.ac.cn	http://cssci.nju.edu.cn/	http://oversea.cnki.net/	http://www.cqvip.com/
Coverage (in 2017)	1,195 journals	2,054 journals	533 journals	10,324 journals	14,352 journals
Established	1989	1987	1998	1994	1994
Update Frequency	Yearly	Yearly	Yearly	Monthly	Quarterly

2.3.1.2. Disciplinary Classification

WoS and VIP use different disciplinary classification systems. WoS assigns journals to 232 subject categories while the VIP classifies Chinese literature into 35 fields and 457 subfields. Equivalences between the WoS and VIP disciplinary classification systems were first established based on the descriptions of each subject category. This produced 116 obvious one-to-one matches. *Dance* was removed from the list since no Chinese publication was found in this WoS category. Therefore, 115 disciplines with equivalent classes across WoS and VIP were compared in this study (See Appendix I), which account for 66.08 % of Chinese publications (959,728 of 1,452,380) in WoS and 65.15% of literature (19,472,497 of 29,889,566) in VIP. This list includes 83, 21 and 12 disciplines[4] in Natural Sciences, Social Sciences, and Arts and Humanities respectively.

Some inconsistencies between the WoS and VIP classification systems were also found. WoS adopts the journal classification system assigning indexed journals to roughly 250 WoS categories while VIP classifies the discipline at the paper level (the paper classification system) using the Chinese Library Classification Scheme. An inclusive classification is applied to both databases; in other words, journals or papers may be assigned to one or multiple disciplines, which produces 1,240,677 and 22,727,318 assignments in WoS and VIP respectively.

2.3.2. Data

2.3.2.1. Data Collection

Data used in this study were collected for the period between 2008 and 2015 considering WoS only offers bibliometric data distinguishing authors' first name and their institutions since 2008. All papers with a Chinese address (CU = Peoples R China) published between 2008 and 2015 (n=1,452,380) as well as their bibliographic information were retrieved from WoS and assigned to relevant disciplines. In the 115 selected disciplines, Chinese authors contributed the most papers in *Chemistry, Physics* (92,342), followed by *Engineering, Electrical & Electronic*

[4] History is classified as discipline under both Social Science and Arts and Humanities.

(70,318) and *Optics* (49,038) while they only contributed 2, 5 and 6 papers in *Folklore*, *Literary Theory & Criticism* and *Film, Radio & Television* respectively. On the other hand, 29,940,090 Chinese papers published between 2008 and 2015 were indexed by VIP under 457 subfields (disciplines), ranging from 1,667 papers in *Physics, Condensed Matter* to 4,223,457 papers in *Education & Educational Research* In the 115 selected disciplines. No correlation was found between WoS and VIP in terms of the number of publication among these 115 disciplines.

In each discipline, Chinese authors were ranked by their number of published papers during the period of 2008-2015 in both WoS and VIP dataset. The top 100 (and tied) authors in the 115 disciplines were retrieved and formed 115 pairs of author groups, for a total of 26,969 records in the two databases.

2.3.2.2. Author Name Disambiguation

Author name ambiguity is a significant issue when conducting bibliometric analysis at the level of individual researchers (Moed, 2002a). This is even more evident in studies that investigate Chinese and Korean names (Strotmann & Zhao, 2012). Although WoS indexes the complete first name of the authors from 2008 onwards, author name ambiguity remains an issue in WoS, especially since different Chinese names can be transliterated to a single Romanized form name. The issue of author name ambiguity is less important in the VIP data, as full author names are recorded using Chinese characters. However, there remain cases where different Chinese authors share the same Chinese name.

Both automatic and manual validation were performed to disambiguate author names in the WoS and the VIP data. A combination of the author's full name and her/his primary institutional affiliation was used for automatic validation. A pilot test with fully manual validation was conducted based on data from 10 selected disciplines (Shu, Larivière, & Julien, 2016), and the results indicated that the automatic validation allows to disambiguate about 97% of WoS data and almost all VIP data. Exceptional cases were caused by two or more Chinese authors that share the same (Chinese or English) name, and who were active within the same institution or the same discipline. In addition to the automatic validation, a thorough manual validation (that lasted about 6 months) was performed to disambiguate these exceptions. In each discipline, the same name affiliated to different institutions was validated as either an author having multiple affiliations or different authors sharing the same name. Incomplete entries and inconsistent formats were also corrected. The manual validation disambiguated 120,953 ambiguous records regarding Chinese author names.

Moreover, in addition to typos and incomplete entries, serious institutional name ambiguity was also found in WoS data. For example, *JINAN-UNIV* refers to Jinan University located at city of Guangzhou in the province of Guangdong while *UNIV-JINAN* refers University of Jinan located in the city of Jinan in the province of Shandong; *BEIJING-UNIV-TECHNOL* (Beijing University of Technology) and *BEIJING-INST-TECHNOLOGY* (Beijing Institute of Technology) are two different institutions while both *BEIJING-INST-CHEM-TECHNOL* and *BEIJING-UNIV-CHEM-TECHNOL* refer to the same Beijing University of Chemical Technology (formerly Beijing Institute of Chemical Technology). Both *CHINESE-ACAD-MED-SCI* and *PEKING-UNION-MED-COLL* refer to the same institution with two different names (Chinese Academy of Medical Science and Beijing Union Medical College). Institution name disambiguation was conducted manually at the same time as the author name disambiguation was performed, and clarified 1,398 ambiguous records regarding Chinese institution names.

2.3.2.3. Classification into institutional sectors

All institutional affiliations were classified into six sectors: universities, scientific institutes, government, enterprises, hospitals not affiliated with universities, and other sectors. Since universities play the dominant role in China's scientific research output, contributing 82.8% of monographs and 73.4% of journal articles including 83.0% of WoS papers (National Bureau of Statistics of China, 2015), all Chinese universities were further classified into two sub-categories: elite universities that are defined by Ministry of Education of China (2016), including but not limited to universities within the two national research programs (Project 211 and Project 985 (Quan et al., 2017)), and non-elite universities, which refers to the remainder.

In addition, China consists of 31 provincial-level divisions that were traditionally grouped into seven geographical regions: North, Northeast, Northwest, Center, East, Southwest, and South. Economic development in different regions differs significantly. The average university budgets in the developed regions (i.e. the North, East and South) are much higher than those in the developing regions (i.e. the Northeast, Northwest, Center, and Southwest) (Ministry of Education of China, 2015; National Bureau of Statistics of China, 2015). Considering economic development in different regions differs significantly in China, the provinces in which the affiliated institutions located were also analyzed. Please note that some universities, scientific institutes and hospitals are owned by China's People's Liberation Army (PLA), which are not administrated by Ministry of Education or any local governments.

2.3.2.4. Indicators

In this study, Chinese authors were defined as those whose primary affiliated institution is in China, regardless of their citizenship. In WoS, all articles with a Chinese address were selected.

The authors of these selected articles were likely to qualify as Chinese authors, but co-author(s) whose affiliated institution was not located in China were excluded. Authors with multiple affiliated institutions were manually validated for their eligibility.

The Chinese scientific elites are defined as the most highly productive authors in terms of the number of publications they have produced in their disciplines. For each of the 115 disciplines chosen, the number of papers per author was compiled in order to produce ranked lists of top Chinese authors in WoS and VIP. The top 100 (and tied) authors in terms of the number of publications produced between 2008 and 2015 in the 115 identified disciplines formed 115 pairs of Chinese scientific elite researchers. The amount of overlap between each of these 115 sets of researchers indicated whether the Chinese scientific elites found in the WoS is the same as the one found in the VIP. For each discipline, the overlap between those researchers who are among the top 100 in WoS and the top 100 in VIP (hereafter referred to as the overlap rate) was calculated based on the formula,

$$\text{Overlap rate} = \frac{\text{N}}{(\text{Tv} + \text{Tw})/2}$$

where N=number of Chinese most productive authors exist in both databases, T_V=number of top 100 and tied VIP authors, and T_W=number of top 100 and tied WoS authors.

For example, the overlap rate is 20% when 22 shared authors are found between 105 authors in WoS and 115 authors in VIP (considering that the number of top 100 authors may equal more than 100 when ties are included).

The publication counts presented in this paper were based on the number of articles, notes, and review articles but exclude editorials, book reviews, letters to the editor and meeting abstracts that are not generally considered original contributions to scholarly knowledge (Moed, 1996). In China, not all co-authorship credits are assigned based on an individual's scientific contribution but on the basis of seniority (Shen, 2016). However, Chinese bibliometric databases, including VIP, give full credit to all co-authors when counting the number of publications. This study applied the same approach regardless of the argument on whether a full count or divided count is better to measure the co-authorship.

In addition to the overlap rate, eight indicators were also compiled for each discipline for the purpose of data analysis, as shown in Table 3.

Table 3. List of Indicators Used in Data Analysis

Indicator	Description
The Overlap Rate	The share of Chinese scientific elites found in both databases
The number of VIP papers	The number of papers that were published between 2008 and 2015 and indexed by VIP
The number of VIP authors	The number of Chinese scholars who published at least one paper indexed by VIP between 2008 and 2015
The number of Chinese WoS papers	The number of papers that were published by Chinese scholars between 2008 and 2015 and indexed by WoS
The number of Chinese WoS authors	The number of Chinese scholars who published at least one paper indexed by WoS between 2008 and 2015
The number of WoS papers	The number of papers that were published between 2008 and 2015 and indexed by WoS
The ratio of Chinese WoS papers to all WoS papers ($Ratio_{c2w}$)	The share of Chinese WoS papers to all WoS papers
The ratio of Chinese WoS papers to all Chinese papers ($Ratio_{w2c}$)	The share of Chinese WoS papers to all Chinese papers including both WoS papers and VIP papers
The ratio of Chinese WoS authors to VIP authors ($Ratio_{w2v}$)	The ratio of the number of Chinese WoS authors to the number of VIP authors

2.4. Results

Among the 26,969 records retrieved from WoS and VIP (14,911 records from WoS and 12,058 records from VIP), 12,270 and 11,066 Chinese elite researchers as well as their primary affiliated institutions were identified from WoS and VIP, respectively, across the 115 selected disciplines. As noted above, Chinese scientific elites in multiple disciplines tied for the top 100 ranking. In addition, the total numbers of Chinese scientific elites in 7 disciplines in WoS and 3 disciplines in VIP totaled fewer than 100 because fewer than 100 Chinese authors published papers in these disciplines between 2008 and 2015.

2.4.1. Overlap

As Figure 3 and Figure 2 show, the average overlap rate between the two groups of Chinese scientific elites was 10.52% ranging from 0% to 33.98% across the 115 disciplines. The overlaps in the Natural Sciences including Life Sciences & Biomedicine, Physical Sciences, and Technology[5] were higher than those in the Social Sciences and Humanities. Although the size of discipline in terms of the total number of publications varies, no correlation was found between the size of discipline and the overlap rate.

[5] WoS classifies research areas into five domains including Arts and Humanities, Social Science, Life Science & Biomedicine, Physical Science, and Technology, the last three of which constitute Natural Sciences. All WoS categories could correspond to research areas and be assigned to these five domains using a conversion table.

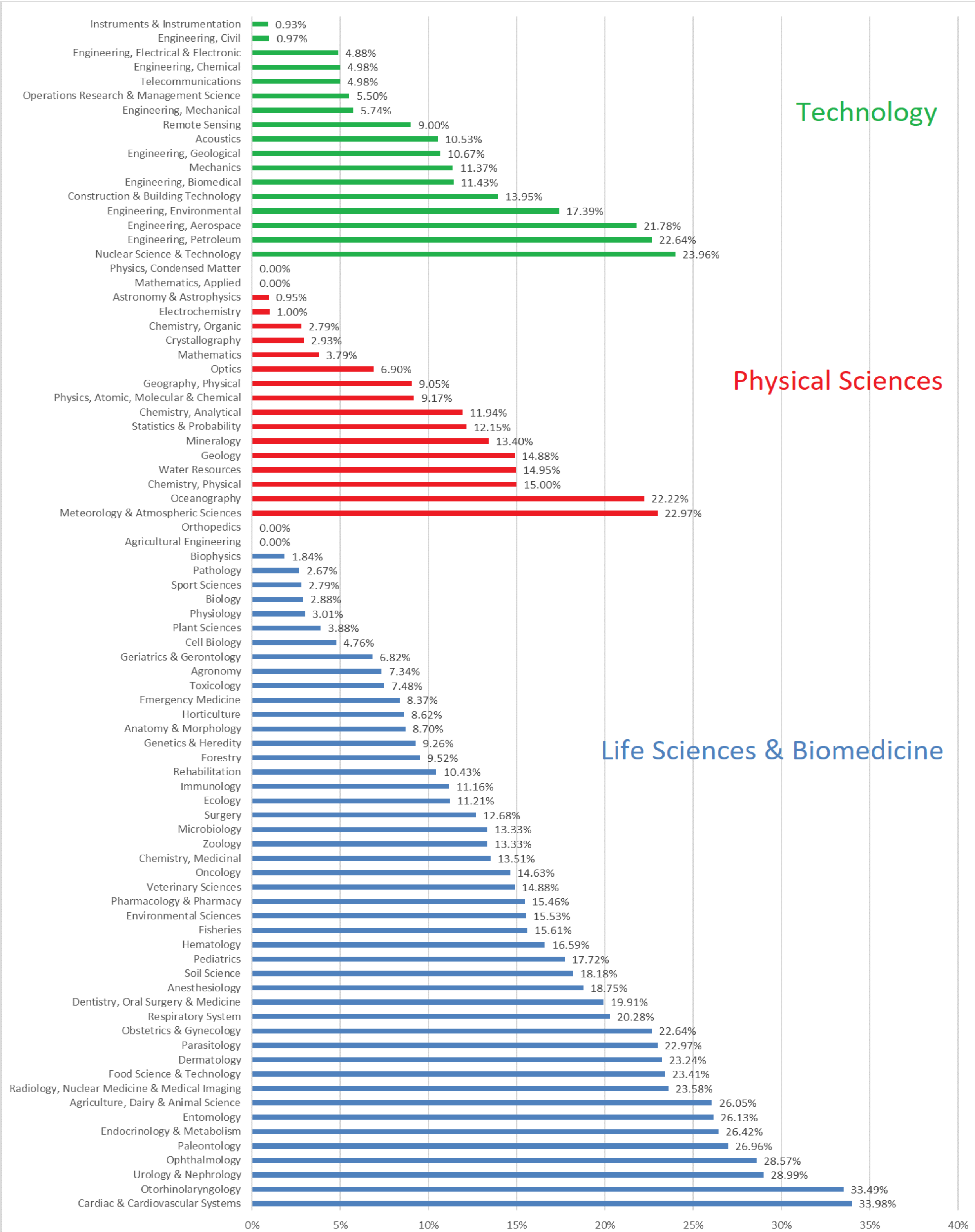

Figure 2. The Overlap Rate between Groups of Chinese Scientific Elites in Natural Sciences

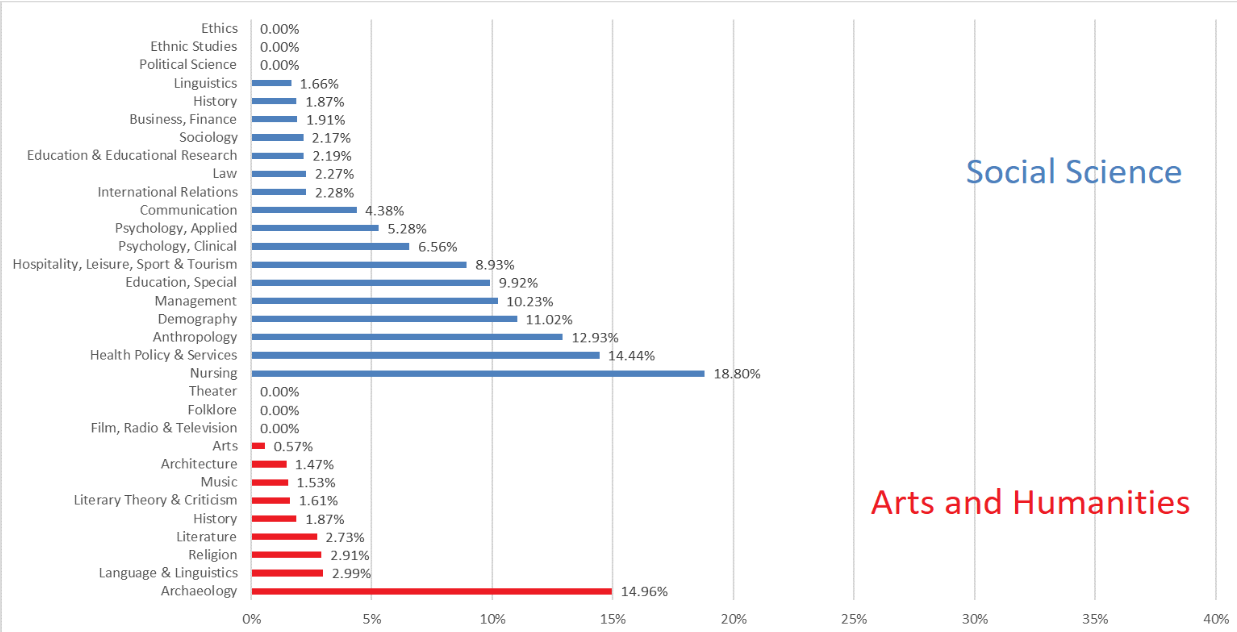

Figure 3. The Overlap Rate between Groups of Chinese Scientific Elites in Social Sciences and Humanities

2.4.1.1. Social Sciences and Humanities

Among the 13 disciplines in Arts and Humanities selected for this study, all the overlap rates were lower than 3% except for *Archaeology*, where 19 authors were presented in both elite groups in WoS (148) and VIP (106), contributing to an overlap rate of 14.96%. Indeed, during the period of 2008-2015, Chinese scholars only published 3,929 WoS papers in these 13 disciplines, ranging from 2 papers in *Folklore* to 1,203 papers in *Literature*. However, while Chinese scholars' contribution to WoS literature in Arts and Humanities remains marginal, they published 4,330,239 Chinese papers indexed by VIP in these 13 disciplines. The ratio of Chinese WoS papers to all Chinese papers ($Ratio_{w2c}$) in these 13 disciplines was 0.09%. Few Chinese scholars published WoS papers in these disciplines, ranging from 2 authors in *Folklore* to 796 authors in *Archaeology*. Additionally, less than 100 Chinese authors published WoS papers during the period of 2008-2015 in 6 out of these 13 disciplines.

The overlaps were a little higher (6.10% in average) among the 21 disciplines in the Social Sciences, and ranged from 0% in *Ethics*, *Ethnic Studies*, and *Political Science* to 18.80% in *Nursing* which is related to Medical Sciences but classified as a Social Science discipline in WoS. Indeed, the top 3 Social Science disciplines in terms of the overlap (*Nursing*, *Health Policy & Services* and *Anthropology*) were all related to Health.

Compared to Chinese scholars in Arts and Humanities, Chinese scholars in Social Sciences published more WoS papers (20,507 across 21 disciplines), contributing 2.91% of international scientific production. However, the ratio of Chinese WoS papers to all Chinese papers ($Ratio_{w2c}$) in Social Sciences remains very low (0.23%); in other words, more than 99% of Chinese papers in Social Sciences are published in national Chinese journals. The only exception was *Psychology, Applied,* where Chinese scholars published 948 WoS papers and 7,048 VIP papers respectively, but it is a small discipline considering that only 24,938 papers were indexed by WoS over the eight years. In addition, the number of Chinese authors in WoS's Social Sciences was also low, ranging from 21 in *Ethnic Studies* to 3,688 in *Management*.

2.4.1.2. Natural Sciences

The overlap rates (12.68% in average) were higher in Natural Sciences than those in Social Sciences and Humanities. The overlap rates varied across the 83 disciplines, which could be classified into three broad categories in WoS: Life Sciences & Biomedicine, Physical Sciences, and Technology. In Life Sciences & Biomedicine, the average overlap rate was 14.74% ranging from 0% in *Orthopedics* and *Agricultural Engineering* to 33.98% in *Cardiac & Cardiovascular Systems, which* was the highest among all 115 disciplines. The average overlap rate was 9.12% in Physical Sciences ranging from 0% in *Physics*, *Condensed Matter* and *Mathematics*, *Applied* to 22.97% in *Meteorology & Atmospheric Sciences;* and the average overlap rate was 10.63% in Technology ranging from 0.93% in *Instruments & Instrumentation* to 23.96% in *Nuclear Science & Technology*.

The share of Chinese WoS papers to all WoS papers ($Ratio_{c2w}$) was higher in Natural Sciences than Social Sciences and Humanities. Chinese scholars contributed 14.25% of WoS papers during the period of 2008-2015 among these 83 disciplines in Natural Sciences (9.82% in Life Sciences & Biomedicine, 20.32% in Physical Sciences, and 18.64% in Technology) ranging from 2.22% in *Sports Science* to 31.00% in *Crystallography*. The correlation between the overlap rate and China's share in international scientific publications ($Ratio_{c2w}$) seemed to be negative (r=-.4094) as shown in Figure 4. The overlap rates were not higher among those disciplines in which Chinese scholars contributed more in international scientific literature.

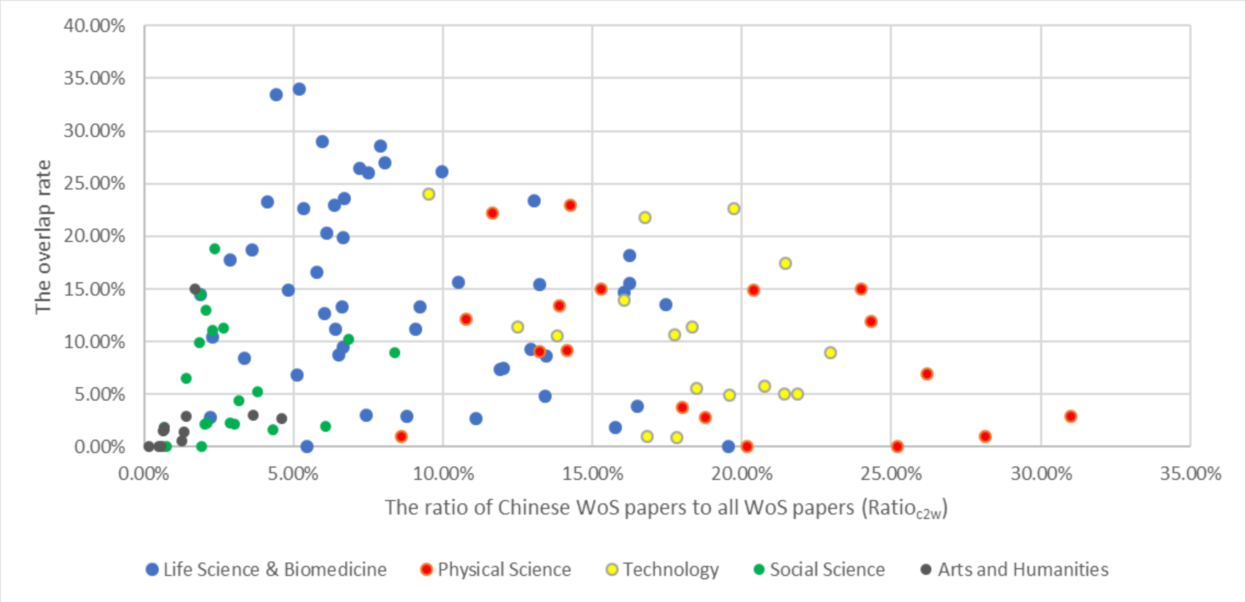

Figure 4. Scatterplot of the Overlap Rate and China's Share in International Scientific Literature

A similar negative correlation (r=-.4305) was also found between the overlap rate and the ratio of Chinese WoS papers to all Chinese papers ($Ratio_{w2c}$) in Natural Sciences. As shown in Figure 5, the overlap rates were low among those disciplines in which Chinese scholars published more international papers indexed by WoS. Indeed, the average $Ratio_{w2c}$ was 12.16% among 83 disciplines in Natural Sciences (7.74% in Life Sciences & Biomedicine, 35.43% in Physical Sciences, and 7.35% in Technology respectively) ranging from 0.35% in *Sports Science* to 96.29% in *Physics*, *Condensed Matter*.

An interesting pattern was revealed when we investigated the relationship between the overlap rate and the combination of $Ratio_{c2w}$ and $Ratio_{w2c.}$ As Figure 6 shows, the overlap rates were less than 15% in all disciplines in which the $Ratio_{w2c}$ was over 30%. When the threshold was increased to $Ratio_{w2c}$ > 40% and $Ratio_{c2w}$ > 10%, the overlap rates in 11 out of 13 disciplines within this section were less than 10%. It means that the share of Chinese scientific elites is low among those disciplines in which Chinese scholars published a lot of international articles.

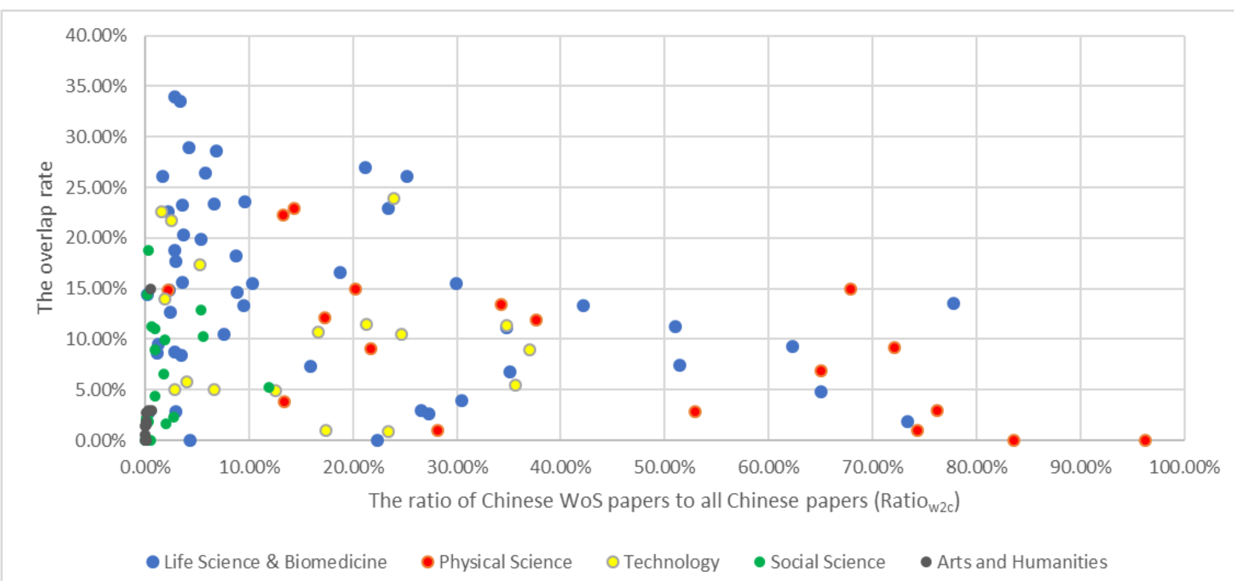

Figure 5. Scatterplot of the Overlap Rate and the Proportion of WoS Papers to China's National Scientific Literature

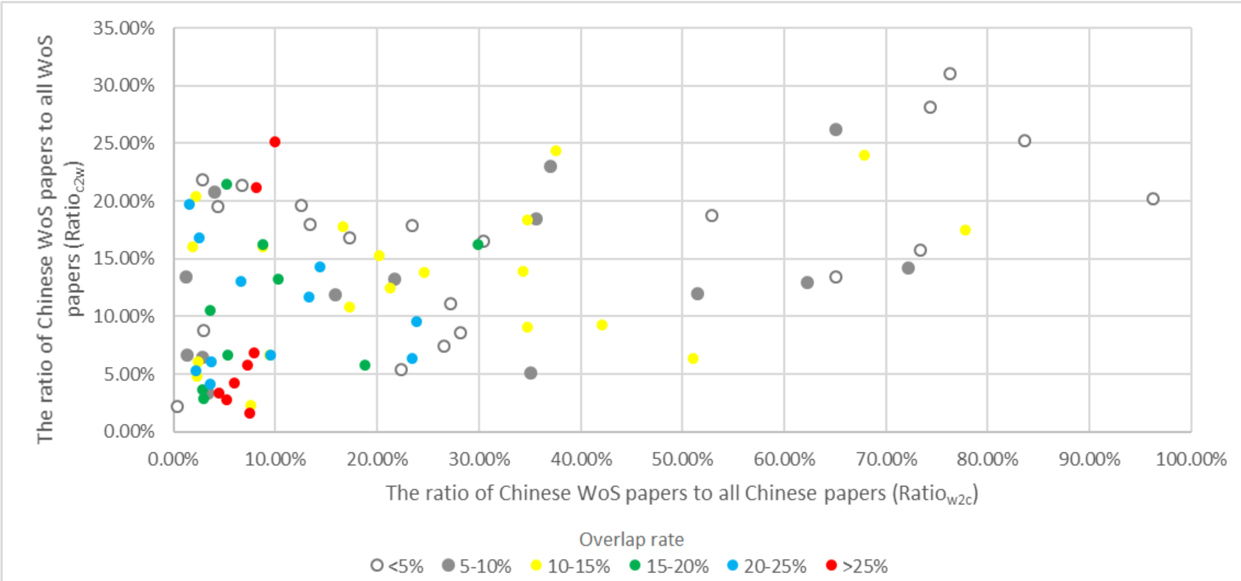

Figure 6. Correlation between the Overlap Rate and the Proportion of China's WoS papers to International and National Scientific Literature

2.4.2. Affiliated Institution

2.4.2.1. Type of Affiliated Institutions

As shown in Figure 7, the 11,066 Chinese scientific elite researchers identified in VIP are from different sectors: universities contributed the most scientific elites (73.78%) including 40.44% from elite universities and 33.35% from non-elite universities, followed by scientific institutes (12.05%), non-affiliated hospitals (7.15%), Other sections (3.24%), Government (3.15%) and Enterprises (0.62%).

The distribution of Chinese scientific elites differed among different disciplines, but not significantly. Unsurprisingly, Table 4 shows that universities contributed more scientific elites in Social Sciences and Humanities (81.62% and 78.55%, respectively) compared to other sectors. In the Natural Sciences, the share of scientific elites from universities was lower, at 69.23% in Life Sciences & Biomedicine, as 15.15% of scientific elites were from non-affiliated hospitals. The proportions of scientific elites between elite universities and non-elite universities were close across these major disciplines except for Technology (51.60% vs. 23.91%) and Social Sciences (49.06% vs. 29.50%).

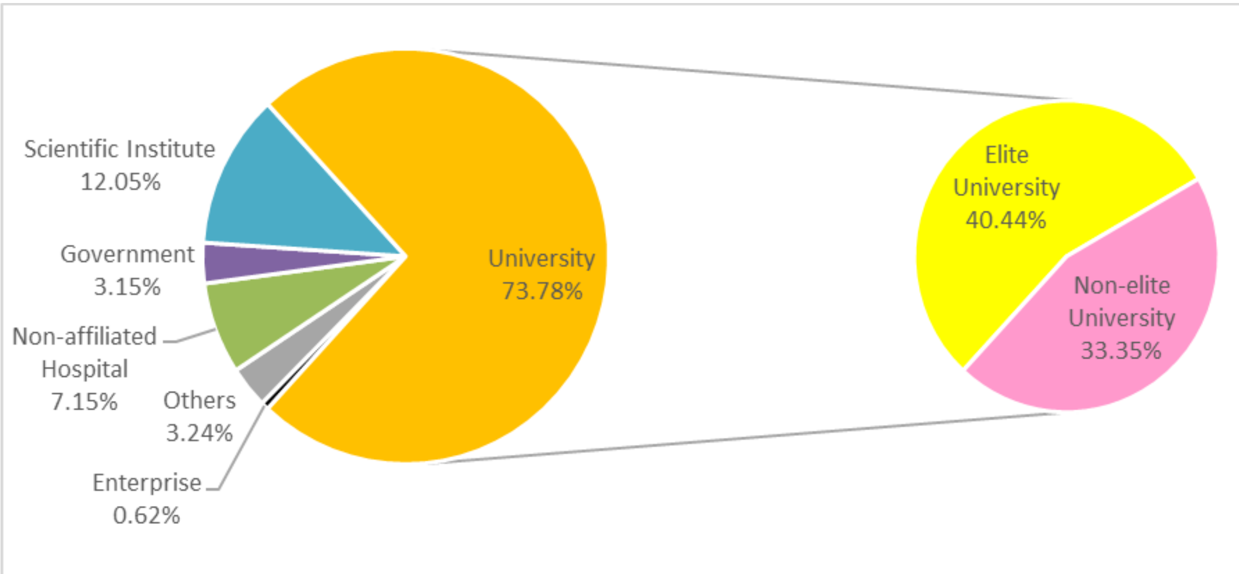

Figure 7. Distribution of Chinese Scientific Elites (VIP) by Type of Affiliated Institutions

Table 4. Distribution of Chinese Scientific Elites in VIP (upper) and WoS (lower) by Discipline in Percentage

	Arts and Humanities	Social Sciences	Natural Sciences: Life Sciences & Biomedicine	Natural Sciences: Physical Sciences	Natural Sciences: Technology	Natural Sciences: All Natural Sciences	Total
Enterprises	0.17 1.56	0.23 1.01	0.21 0.13	0.39 0.06	2.67 0.32	0.76 0.15	0.62 0.59
Others	6.18 4.90	7.69 1.26	1.06 0.11	0.50 0.00	3.71 0.00	1.48 0.06	3.24 0.99
Non-affiliated Hospitals	0.00 0.00	2.39 4.04	15.15 5.41	0.17 0.00	2.09 0.19	9.05 3.26	7.15 3.06
Government	2.09 0.60	4.10 1.28	2.02 0.44	6.25 1.51	2.26 0.13	2.96 0.61	3.15 0.80
Scientific institutes	9.93 8.55	7.04 6.56	12.32 14.25	17.27 30.05	13.76 14.08	13.58 17.55	12.05 12.92
Universities	81.62 84.39	78.55 85.84	69.23 79.66	75.43 68.38	75.51 85.29	72.18 78.36	73.78 81.64
Elite Universities	44.77 61.00	49.06 70.57	33.92 60.70	37.08 56.63	51.60 78.44	38.57 63.35	40.44 65.22
Non-elite Universities	36.85 23.39	29.50 15.28	35.31 18.96	38.35 11.75	23.91 6.85	33.61 15.01	33.35 16.42

On the other hand, the distribution of Chinese scientific elites identified in WoS was slightly different; as shown in Figure 8, scientific institute contributed a similar share of scientific elites as 12.92%; the university contributed 81.64% of scientific elites while the contribution of non-affiliated hospital (3.06%), other sections (0.99%), government (0.80%) and enterprise (0.59%) were less than 5%. Indeed, 65.22% of scientific elites came from elite universities while 16.42% of them were from non-elite universities, which was significantly different from the ratio found in VIP.

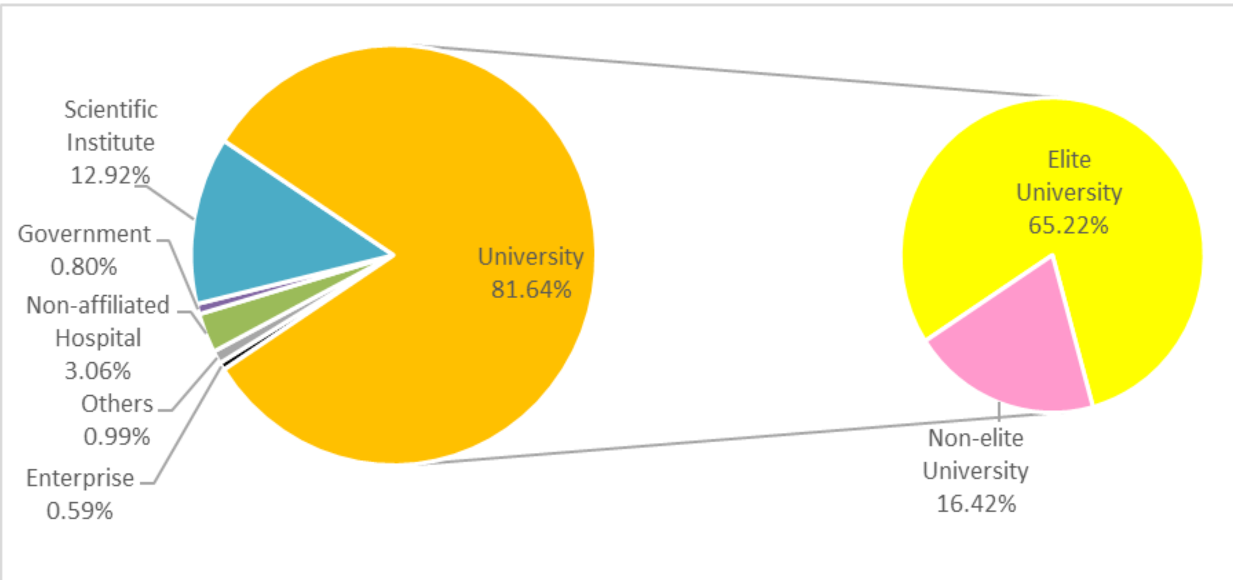

Figure 8. Distribution of Chinese Scientific Elites (WoS) by Type of Affiliated Institutions

Similar to the results from VIP, the distribution of scientific elites in WoS differed slightly among different disciplines, as shown in Table 4. The university contributed more scientific elites in Social Sciences and Humanities but fewer in Physical Sciences in which scientific institutes contributed more. However, there were two differences in terms of the distribution between VIP and WoS. The scientific institute's share of scientific elites in WoS (17.55%) was higher than the share in VIP (13.58%) in Natural Sciences, especially in Physical

Sciences (30.05% vs. 17.27%). A difference between elite universities and non-elite universities in terms of the proportion of scientific elites was found across all disciplines.

2.4.2.2. Geographic Area

Few differences were found between VIP and WoS in terms of the distribution of Chinese scientific elites by geographic area. As shown in Table 5, most of Chinese scientific elites came from two major developed areas in China: The North (Beijing-centered) and the East (Shanghai-centered) that contributed more than 50% and 60% of scientific elites across all disciplines in VIP and WoS, respectively. In addition, 5.93% and 3.28% of Chinese scientific elites, in VIP and WoS respectively, were affiliated to the Chinese People's Liberation Army (PLA), which has many affiliated scientific institutes and hospitals across the country. The North was the largest contributor to Chinese scientific elites in all disciplines in VIP and most disciplines in WoS except for Arts & Humanities and Life Sciences & Biomedicine in WoS in which the East contributed the most scientific elites.

Table 5. Distribution of Chinese Scientific Elites in VIP (upper) and WoS (lower) by Geographic Area in Percentage

	Arts and Humanities	Social Sciences	Natural Sciences				Total
			Life Sciences & Biomedicine	Physical Sciences	Technology	All Natural Sciences	
PLA	0.26 0.12	2.99 1.37	9.32 6.66	2.27 0.64	7.60 3.93	7.38 4.71	5.93 3.28
Center	10.98 10.65	12.52 6.75	8.68 6.55	7.97 8.69	9.17 6.53	8.74 6.95	9.56 7.41
East	29.44 37.14	26.23 30.50	24.58 32.49	23.57 27.45	21.42 29.30	23.69 30.51	24.76 31.45
North	30.66 29.90	32.40 37.17	26.77 30.13	31.60 45.17	32.10 36.08	28.93 34.76	29.77 34.71
Northeast	7.93 4.49	5.66 3.85	7.75 5.08	8.97 6.49	7.08 8.88	7.78 6.11	7.37 5.31
Northwest	6.27 5.92	4.60 5.06	6.07 3.20	10.02 3.94	9.05 7.17	7.53 4.13	6.89 4.66

South	6.36 7.66	5.80 9.16	8.68 8.39	5.31 3.07	2.90 2.16	6.76 6.00	6.51 7.07
Southwest	7.93 4.13	9.71 6.15	8.13 7.51	10.18 4.57	10.68 5.96	9.16 6.83	9.15 6.12

2.5. Discussion

2.5.1. Exceptional Disciplines

This study indicated that disciplines in Natural Sciences including Medical Science and Engineering exhibited a much higher level of overlap than those in Social Sciences and Humanities. It is unsurprising that scholars could easily disseminate knowledge internationally in the Natural Sciences in which all scholars share the same paradigm (Kuhn, 2012). On the other hand, scholars in Social Sciences and Humanities have to apply multi-paradigmatic approaches to understand complex social or human behaviour, making it a more difficult task to publish research in different languages in these disciplines (Cole, 1975; Delanty, 2005).

Although Chinese scholars contributed more international publications in Natural Sciences, we found an unexpected negative correlation between the overlap rate and both the ratio of Chinese WoS papers to all WoS papers ($Ratio_{c2w}$) and the ratio of Chinese WoS papers to all Chinese papers ($Ratio_{w2c}$) in those disciplines. Low overlap rates of Chinese scientific elites between WoS and VIP were found in disciplines that are most international in scope. As shown in Figure 6 above, the overlap rates were below 10% in 11 out of 13 disciplines in which the ratio of Chinese WoS papers to all WoS papers ($Ratio_{c2w}$) was over 10% and the ratio of Chinese WoS papers to all Chinese papers ($Ratio_{w2c}$) was over 40%. The overlap rate declined to less than 3% in four disciplines in which the two ratios were increased to 20% and 70%, respectively, as shown Table 6. Indeed, in those disciplines that were most international in scope, most Chinese scholars preferred diffusing their research results in international journals to publishing in national Chinese journals. For example, in *Physics, Condensed Matter*, 97,483 Chinese scholars published 43,319 WoS articles while 3,568 Chinese scholars published 1,667 articles in Chinese journals during the same period; Chinese scholars had almost abandoned publishing in national Chinese journals as 96.29% of their publications were in WoS journals. Thus, although the overlap rates were low, Chinese WoS papers could still represent Chinese research activities in those disciplines in which international publication was dominant.

Table 6. Top Four Disciplines that are Most International in Scope

Discipline	**# VIP Papers**	**# VIP Authors**	**# WoS Papers**	**# WoS Authors**	**$Ratio_{c2w}$**	**$Ratio_{w2c}$**	**$Ratio_{w2v}$**	**Overlap Rate**
Physics, Condensed Matter	1,667	3,568	43,319	97,483	20.18%	96.29%	27.32	0.00%
Mathematics, Applied	9,311	14,423	47,499	23,896	25.20%	83.61%	1.66	0.00%
Crystallography	6,570	13,772	21,102	40,742	31.00%	76.26%	2.96	2.93%
Electrochemistry	9,206	10,792	26,621	60,194	28.14%	74.30%	5.58	1.00%

2.5.2. Publication Patterns

Being a productive author in both English and Chinese publication is not easy. These Chinese scholars have to allocate their manuscripts in two directions; some are sent to national journals while others are submitted to international journals[6]. They also need to balance the number of submissions between national and international publications to compete with scholars who may only focus on publishing nationally or internationally. Thus, the different publication patterns of

[6] In practice, some Chinese scholars submitted the same manuscript to both Chinese journals and English journals in different languages, which should be explored in future research.

Chinese scholars result in the low overlap rates between the two groups of Chinese scientific elites found in this study.

For example, 355,387 Chinese authors published 321,875 VIP papers while 123,839 Chinese authors published 36,836 WoS papers in *Pharmacology & Pharmacy*[7] between 2008 and 2015. 103 Chinese authors who published 73 or more VIP papers and 104 Chinese authors who published 42 or more WoS papers were identified as Chinese scientific elites while 16 scholars were included in both groups. 87 out of 103 VIP scientific elites (84.47%) also published WoS papers while 101 out of 104 WoS scientific elites (97.12%) also published VIP papers. Although most of Chinese scientific elites published papers in both national journals and international journals, they have different publication patterns as shown in Figure 9. Some scholars (red nodes in Figure 9) published most of their papers in international (WoS) journals; some (blue nodes in Figure 9) preferred to diffuse most of their research results in national Chinese journals; 16 scholars (green nodes in Figure 9) could keep the balance and published their manuscripts in both international and national journals. It is therefore difficult to evaluate China's research output based on a single database.

[7] *Pharmacology & Pharmacy* is selected because its indicators, such as Ratioc2w (13.22%), Ratiow2c (10.27%), Ratiow2v (0.3485), and the overlap rate (15.46%) are close to the average of all disciplines in Natural Science (14.25%, 12.16%, 0.3733, and 12.68%, respectively).

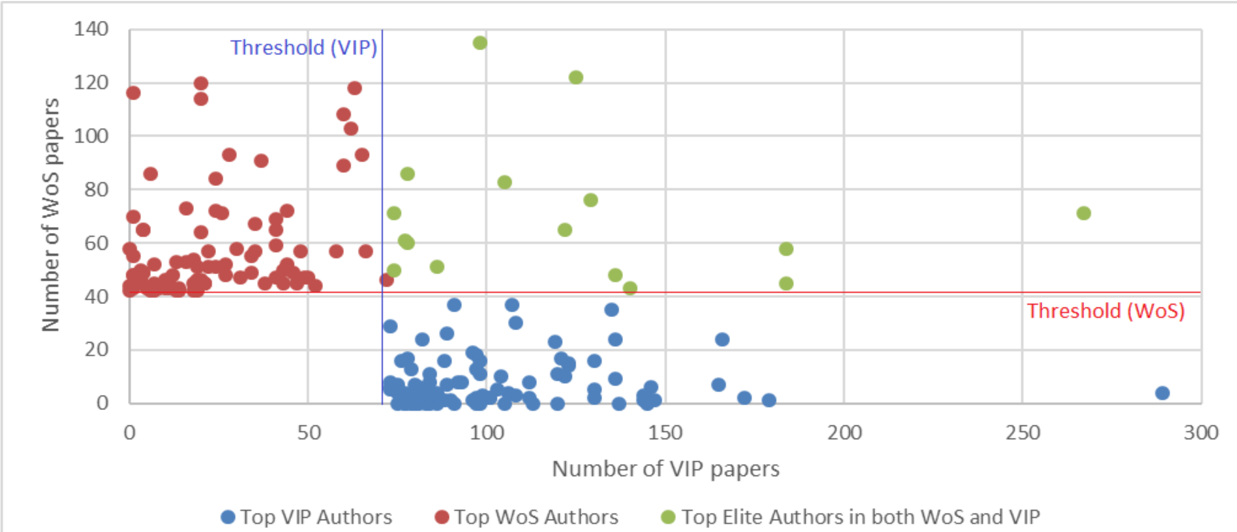

Figure 9. Publishing Pattern of Chinese Scientific Elites in Pharmacology & Pharmacy

2.5.3. Science Policy

The publication patterns of Chinese scholars are also influenced by China's science policies that promote international publication (Quan et al., 2017). Since the 1980s, the number of WoS papers has been used to evaluate the research output in China of both institutions and individuals (Cao et al., 2013; Gong & Qu, 2010) to increase the international visibility of Chinese research. Chinese scholars are required to publish WoS papers to obtain promotion, while their affiliated institutions need the number of WoS papers for ranking and funding applications (Cao et al., 2013; YJ Wang & Li, 2015). Chinese research institutions even offer the monetary rewards to their scholars who publish internationally (Quan et al., 2017). These science policies create a negative goal displacement effect (Cao et al., 2013; Frey, Osterloh, & Homberg, 2013; Osterloh & Frey, 2014), the result of which is that, for Chinese scholars, the purpose of publishing their works is not only to advance knowledge but also to fulfill promotion requirements and earn money (X. Sun & Zhang, 2010; L. Wang, 2016).

In China, international publication is a mandatory requirement for tenure and promotion at most elite universities, but is only an optional requirement at non-elite universities in which Chinese scholars could use national publications as alternatives (Cao et al., 2013). In order to fulfill the requirement, Chinese scholars from elite universities mostly publish papers in international (WoS) journals while those from non-elite universities prefer to diffuse their research results in national journals in consideration of the language barrier and high rejection rate of WoS journals. As a result, the proportion of Chinese scientific elites from elite universities is much higher than those from non-elite universities in WoS, but non-elite universities contribute the similar number of scientific elites to elite universities in VIP.

2.5.4. Limitations

Classifying science into a disciplinary structure is one of the basic preconditions in bibliometrics (Glänzel & Schubert, 2003), and most classification systems of science (e.g., WoS) are established at the level of journals (journal classification system) but all Chinese bibliometric databases (e.g., VIP) classify science at the paper level (paper classification system) using the Chinese Library Classification Scheme. In this study, WoS offers a strictly journal classification system while the paper classification system is applied in VIP; thus, the comparison of the two groups of Chinese scientific elites in both databases had to be conducted across different classification systems. The effects of this will be discussed in Chpater 3.

A combination of an author's full name and her/his primary affiliated institution is used in this study for name disambiguation. Although this method can disambiguate about 97% of WoS data and almost all VIP data, it cannot disambiguate scholars who were affiliated to different institutions because of academic mobility. In other words, 3% of Chinese scholars are ranked by the number of publications with only one of her/his affiliated institutions. In addition, this study defines Chinese scientific elites as those who have produced the greatest number of publications in their discipline, but the impact of their research (i.e. citation) is not considered. Future work may need to investigate the impact of the Chinese scientific elite according to their average number of citations per paper and total number of citations.

2.6. Conclusion

This study indicates that Chinese scholars do not have homogeneous publication patterns. While some Chinese scientific elites mostly publish in international (WoS) journals, others prefer to diffuse their research results in national Chinese journals. Unsurprisingly, disciplines that are most international in scope such as those of the Natural Sciences exhibit a much higher level of overlap than those of the Social Sciences and Humanities. On the whole, these results suggest that the WoS does not accurately represent Chinese research activities, which confirms the findings of previous research (Guan & He, 2005; Jin & Rousseau, 2004; Jin et al., 2002; Liang et al., 2001; Moed, 2002b; Zhou & Leydesdorff, 2007). However, this study also finds a relative overlap with the Chinese national scientific literature in the Natural Sciences including Life Sciences & Biomedicine, Physical Sciences, and Technology, in which WoS may be used to evaluate Chinese research output.

1. In Social Sciences and Humanities, in which Chinese scholars publish few WoS papers compared to the large number of publications in national journals, WoS does not represent Chinese research activities. Instead, Chinese bibliometric databases should be

used to evaluate Chinese research output, as suggested by Moed (2002b) and Liang (2003).

2. In Natural Sciences including Life Sciences & Biomedicine, Physical Sciences, and Technology, in which Chinese scholars diffuse their research results in both international journals and national journals, Chinese research output could be evaluated using a combination of WoS and national Chinese bibliometric databases.

3. Exceptionally, in some disciplines in which Chinese scholars publish few papers in national journals compared to the large number of WoS papers, national publications cannot represent Chinese research activities when international publications become dominant. In such cases, WoS could be used to evaluate Chinese research output.

This study also reveals different publication patterns among Chinese scholars: those from elite universities prefer publishing in international journals indexed by WoS, while those from non-elite universities publish more papers in national journals. Although the difference could be partly attributed to the impact of China's science policies that promote international publication, the detailed relationship between publication patterns and science policies should be investigated in future work.

Transition I

In the previous chapter we compared Web of Science (WoS) with Chinese Science and Technology Periodical Citation Database (VIP) in terms of most productive authors and their research output, and demonstrated the extent of the overlap between the two groups of Chinese scientific elites. Since WoS adopts journal classification system while VIP adopts paper classification system, the comparison of the two groups of Chinese scientific elites in both databases had to be conducted across different classification systems, which may have the effect on the results of this comparison. Thus, in the next chapter we compared the classification system of science between journal level and paper level, the need for which was identified and discussed in the Chapter 2.

3. A comparison of the classification system of science between journal level and paper level

3.1. Introduction

Classification system of science plays an important role in bibliometric studies that provide quantitative analyses of scientific literature (De Bellis, 2009; Glänzel & Schubert, 2003). It assigns scientific literature to research areas or disciplines to describe the structure and historical development of scientific disciplines (Young & Belanger, 1983), and research production (Melkers, 1993). The classification system of science is usually established at the level of journals (hereafter referred to as *journal classification*). This has well-known limitations since, for example, papers published in multidisciplinary journals cannot be properly classified at the journal level. Previous studies try to construct a classification system of science at the paper level (hereafter referred to as *paper classification*) (Boyack & Klavans, 2010; Boyack et al., 2011; Klavans & Boyack, 2017; Waltman & Eck, 2012), but its accuracy is difficult to be evaluated without a "golden standard" (Waltman & Eck, 2012, p. 2390). Although the accuracy of the classification system of science is questioned at both journal- and paper levels, no study isolates and compares these classification systems of science between both levels. This study begins to address this gap by using a single dataset that contains both journal and paper classifications to specifically reveal their respective impacts on the structure of science.

3.2. Background

Classifying science into a disciplinary structure is one of the basic preconditions in bibliometrics (Glänzel & Schubert, 2003). A classification system of science is used to assign journals or individual publications to scientific disciplines or research areas in which scientific literature is studied and evaluated (De Bellis, 2009; Melkers, 1993), either within a single discipline or across multiple disciplines (Leydesdorff & Bornmann, 2016; Porter & Rafols, 2009; Porter, Roessner, & Heberger, 2008; L. Zhang, Rousseau, & Glänzel, 2016).

Classification systems of science are widely used in bibliometric studies, not only for information retrieval, but also as a critical grouping factor to normalize scientific evaluations across the varying scholarly practices (e.g., The Times Higher Education World University Rankings are based on the Web of Science [WoS] subject classification scheme). Normalizing citation impact based on journal classification systems has become an established practice in evaluative bibliometrics (L. Zhang, Janssens, Liang, & Glänzel, 2010); however, these systems have known drawbacks. Leydesdorff and Bornmann (2016) indicate that WoS categories do not provide sufficient analytical clarity to carry bibliometric normalization in evaluation practices because of *indexer effects,* which refer to the categories added by an indexer which may generate relations among other unrelated journals. Janssens, Zhang, De Moor, and Glänzel (2009) also find that some Essential Science Indicators (ESI) subject areas are not coherent enough in terms of the cross-citation and textual evaluation, and produce a list of inappropriate journal classifications or misclassification. The issue is more serious when measuring interdisciplinarity, which is often operationalized as a measure of diversity in disciplines assigned to the article's references (L. Zhang et al., 2016); an inaccurate disciplinary assignment or reference misclassification within the classification system of science may cause bias, especially if there is a significant proportion of multidisciplinary journals in the reference list (L. Zhang et al., 2010).

3.2.1. Classification System of Science

Traditional classification systems of science are established at the journal level, which has been adopted by major international bibliometric databases such as WoS and Scopus. Using a journal classification system, papers published in the same journal are classified as the same discipline as their journal's classification(s) except that papers in multidisciplinary journals could be reclassified with their most relevant disciplines in practice. On the other hand, Chinese bibliometric databases classify science at the paper level using the Chinese Library Classification Scheme, which allows for a greater level of classification specificity since papers published in the same journal can be classified into different disciplines.

3.2.1.1. Journal Classification System

The most well-known journal classification system is the WoS Categories that assigns indexed journals to about 250 categories representing the full spectrum of scientific research (Leydesdorff & Rafols, 2009). Based on the citation pattern and selection by experts, journals are assigned to one or multiple WoS categories (Pudovkin & Garfield, 2002) but the detailed methodology has never been published. In addition to WoS categories, WoS also offers two other classification systems: WoS Research Area (SU) consisting of around 150 research areas (hereafter referred to as WoS SU) and Essential Science Indicators (ESI) including 22 subject areas. Alternatively, Scopus offers a 2-level journal classification system consisting of 27 major disciplines and 304 minor disciplines but does not reveal any information regarding how this system is built.

There are other journal classification systems in addition to WoS and Scopus. The National Science Foundation (NSF) classification system is also a 2-level journal classification system consisting of 14 broad fields and 144 fine fields, but it assigns each individual journal into only one single field (Javitz et al., 2010). This exclusive classification is also adopted by the Science-Metrix classification system, which is a 3-level journal classification system including 6 domains, 22 fields and 176 subfields (Archambault, Beauchesne, & Caruso, 2011). Glänzel and Schubert (2003) have designed a 2-level Leuven-Budapest (ECOOM) subject-classification scheme including 15 fields and 64 subfields[8] and have reassigned WoS journals into this classification system. Börner et al. (2012) have mapped the science using both WoS and Scopus data and have established the UCSD (University of California, San Diego) classification system that consists of 13 disciplines and 554 subdisciplines. In addition, some countries maintain their own journal classification system (e.g., the Australian and New Zealand Standard Research Classification (ANZSRC)[9], the Chinese Journal Classification[10], etc.) mainly for national research evaluation.

3.2.1.2. Other Classification Systems

Some studies try to construct a paper classification system in which publications are clustered into disciplines based on citation analysis techniques such as direct citation, bibliographic

[8] The updated version of ECOOM consists of 16 fields and 75 subfields.
[9] http://www.arc.gov.au/australian-and-new-zealand-standard-research-classification-anzsrc.
[10] http://clc.nlc.cn/ztfbb.jsp.

coupling, co-citation and hybrid methods (Shu, Dinneen, Asadi, & Julien, 2017). These provide measures of document similarity where documents judged adequately similar (e.g., 95% similar) are grouped to form a structure of science (Griffith, Small, Stonehill, & Dey, 1974; Small & Griffith, 1974). For example, the ECOOM classification system assigns multidisciplinary journal papers to specific subfields based on their references (Glänzel & Schubert, 2003). Based on the UCSD classification system, Klavans and Boyack (2010) have assigned over 5.68 million papers into more than 84,000 research areas based on their reference distributions. On the basis of 97.6 million citations received by 10.2 million publications, Waltman and Eck (2012) have developed a methodology to construct a 3-level paper classification system that consists of 20, 672 and 22,412 research areas. However, no study constructs the classification systems at both journal level and paper level using the same methodology and compares the difference.

Library classification also groups scientific literature in books and monographs, but it is rarely applied at the journal level or paper level. Although the U.S. Library of Congress classification (LCC)[11] is the most widely-accepted classification system (Klavans & Boyack, 2009), it is used predominantly in research and academic libraries to classify physical books and monographs into a single discipline for the purpose of establishing a unique address on a shelf for that book, rather than classifying journal articles. Shu et al. (2017) present a methodology that classifies the scientific literature into disciplines of the Library of Congress Subject Headings (LCSH), but LCSH is not applied at the paper level. Medical Subject Headings (MeSH)[12] is a candidate classification since it is applied to both journal articles and books, but

[11] https://www.loc.gov/catdir/cpso/lcc.html.
[12] https://www.nlm.nih.gov/mesh/.

findings based on this single domain classification are difficult to be generalized to disciplines beyond the medical sciences. The most adequate dataset found is the Chinese Library Classification (CLC)[13], which is used to construct the classification system of science, at both the journal level and the paper level. This offers an opportunity to compare the classification system of science between the journal level and the paper level from the same data set using the same classification scheme, which has never been done in previous studies.

3.2.2. Chinese Library Classification

The Chinese Library Classification (CLC) is the national library classification scheme of China (*Zhongguo Tushuguan Fenleifa [Chinese Library Classification]*, 2010) that has been used by the education system, research institutions, as well as public libraries since 1973. It is also used by publishers in China to classify all publications including books, monographs, and journals. CLC is designed to produce an alphanumeric representation of the publication's main topic as well as the discipline to which it belongs. CLC should not be confused with the Disciplinary Classification and Code (DCC)[14] and the China Subject Categories by State Council of China (CSSC)[15], which are designed for the purpose of administering other parts of the education system.

3.2.2.1. Chinese Library Classification Code (CLC Code)

CLC is analogic to Library of Congress Classification (LCC) in the United States, in that, it also provides a tree structure consisting of a small group of top categories (22 in the case of CLC) and

[13] http://www.ztflh.com/.
[14] http://dean.pku.edu.cn/urtpku/yjxk.html.
[15] http://www.cdgdc.edu.cn/xwyyjsjyxx/sy/glmd/272726.shtml.

their sub-categories. Each category or sub-category is represented by its name or label, as well as, an alpha-numeric CLC Code (i.e., letters of the Roman alphabet and Arabic numeral). The CLC code is composed of one or two uppercase letters and whole number composed of up to three digits (1-999) that can have decimal extensions. A hyphen can be used to represent further discipline specificity.

The 22 top categories are denoted by single capital letters, and a combination of letters and numbers is used to express sub-categories. Figure 10 presents an example of a 4-level CLC tree structure indicating that *Scientometrics* (G301) is the fourth level category when following the succession of *Culture, Science, Education & Sports* (G) – *Science & Science Studies* (G3) – *Theory of Science Studies* (G30) – *Scientometrics* (G301).

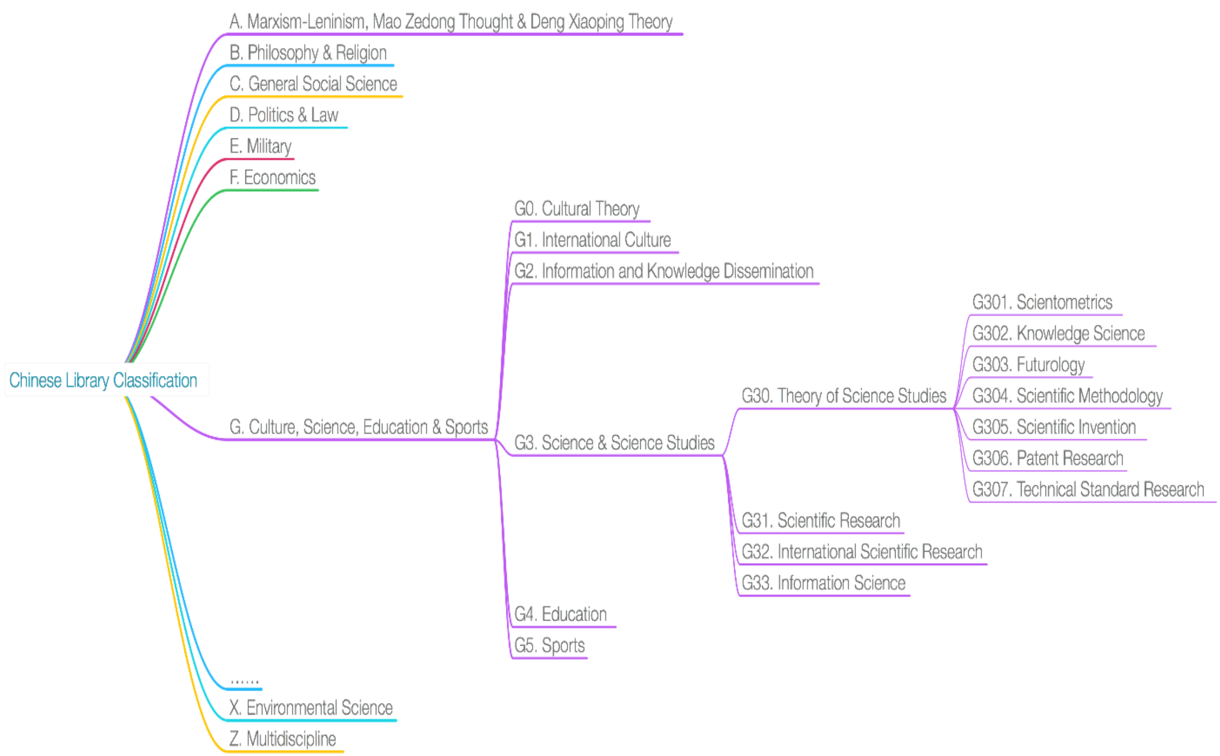

Figure 10. Example of CLC tree structure

3.2.2.2. CLC Code Assignment

In China, CLC is used to classify all legal publications, not only books and monographs but also newspapers and journals. A Chinese Serial Number (CN) is legally required for each academic journal published in China[16]. A CN is assigned by State Administration of Press, Publication, Radio, Film and Television of The People's Republic of China (SAPPRFT) upon request. A CN consists of the registration number and the classification code, which is the CLC classification code, separated by slash "/" (General Administration of Press and Publication of China, 1989).

Figure 11 presents an example where the ISSN (1001-7143) and the CN (11-2684/G3) of the *Chinese Journal of Scientific and Technical Periodicals* are printed in the top right corner of the cover page. In this example, 11-2684 is the registration number while G3 is its classification code corresponding to *Science & Science Studies* in the CLC. Thus, each journal could be classified into different disciplines in a journal classification system of science based on its CLC code.

[16] An International Standard Serial Number (ISSN) is also required if the publication is published internationally.

Figure 11. Example of journal CN

The CLC code is also applied to each journal article. This is generally done by asking authors to provide the CLC code when submitting their manuscript(s). Authors are meant to self-identify the major topic of the manuscript and assign the appropriate CLC code based on the CLC instructions. In some cases, the editor may modify the CLC code provided by the authors if she or he believes this improves the classification. Assigning CLC codes at the paper level is a normal practice in China's academic publishing process, and the CLC codes are assumed to represent the paper topic's discipline(s).

Figure 12 shows an example of a paper's CLC code, which is included in its bibliographic information. The CLC code G301 means that this paper published in journal

Studies in Science of Science is identified by the author as a *Scientometrics* (G301) paper. This dual and independent classification of the journal and its individual articles entails that articles can be classified in disciplines that differ from the journal's CLC code. Note that the CLC's alphanumeric structure permits classification at different abstraction levels. For example, the paper above could also be identified as a *Culture, Science, Education & Sports* (G) paper, or a *Science & Science Studies* (G3) paper, or a *Theory of Science Studies* (G30) paper.


第35卷 第8期 科 学 学 研 究 Vol. 35 No. 8
2017年8月 Studies in Science of Science Aug. 2017

文章编号: 1003 – 2053(2017) 08 – 1121 – 09

科学学的研究进路暨前瞻
——基于贝尔纳奖的分析视角

和 钰[1,2],陈 悦[1],崔银河[1],刘则渊[1]
(1. 大连理工大学人文与社会科学学部,辽宁大连 116024; 2. 内蒙古科技大学艺术与设计学院,内蒙古包头 014030)

摘 要: 贝尔纳奖以其广泛的影响力和学术权威性成为科学学研究领域的代表性奖项。在对贝尔纳奖获奖者及获奖成果(1981 – 2015) 进行统计分析的研究中发现,贝尔纳奖获奖者具有研究地域集中,研究内容多路演进,研究领域存在较为广泛的跨学科性,获奖学者年龄普遍老龄化,取得重大科研成果以及科研成果获奖周期长等现象。透过这些现象,结合学科发展背景,可以较为客观地捋清科学学领域发展的历史沿革、研究进路以及科学学研究发展中形成的多种特点,有助于从科学史、科学与技术的联系、研究策略等多角度思考科学学研究的前瞻延展。

关键词: 贝尔纳奖; 科学学; 科学计量学; 科学社会学; 科学、技术与社会

中图分类号: G301 文献标识码: A

DOI:10.16192/j.cnki.1003-2053.2017.08.001

Figure 12. Example CLC Code of journal article

CLC codes provide a classification system of science at both the journal and the paper level, for the same set of bibliographic records. It is noteworthy that all Chinese bibliometric databases classify science at the paper level, which is in sharp contrast with major international bibliometric databases, such as WoS and Scopus, who classify science at the journal level. Paper level CLCs classify individual papers into author-specified disciplines that can differ from the CLCs of their publishing journals. This means that two papers published in one journal may belong to different disciplines, and papers within the same discipline may be published by journals from different disciplines. This has the potential to produce a more specific and representative description of the structure of science.

3.2.3. Chinese Science Citation Database (CSCD)

Journal/Paper classification is also offered by the Chinese Science Citation Database (CSCD), which was developed by the Chinese Academy of Sciences (CAS) in 1989. It covers more than one thousand core journals in the natural sciences, engineering and medical sciences; including mathematics, physics, chemistry, life science, earth science, agricultural science, medicine, industrial technology, and environmental sciences. It offers information retrieval features, journal selection and academic performance evaluation that are comparable to Science Citation Index (SCI) (Jin & Wang, 1999). In 2007, the CAS cooperated with Clarivate (known then as Thomson Reuters) to provide access to CSCD from WoS.

Similar to other Chinese bibliometric database, CSCD adopts the paper classification system that each paper can be assigned different disciplines based on their CLC codes. According to the agreement with WoS, these papers also could be grouped into different WoS SUs using a conversion table (provided by Clarivate) that maps CLC codes to their corresponding WoS SU(s). CSCD users can strictly access data classified by WoS SU while the CLC-based paper classification system is only for internal reference.

3.2.3.1. CSCD Journal Classification

In addition to the CLC-based paper classification system, a CLC-based journal classification system is also internally used for journal selection and evaluation in CSCD. This provides an opportunity to compare the paper classification system with the journal classification system from the same data set. According to the National Science Library of Chinese Academy of Science (2017), this journal classification system is established on the basis of CLC top categories (level-1 disciplines) as well as their direct sub-categories (level-2 disciplines). Based on direct citation analysis and clustering techniques, some disciplines merge and form a journal classification system with 12 level-1 disciplines and 66 level-2 disciplines[17] as shown in Table 7. Most journals are assigned to a single level-2 discipline except for about 4% of journals that are assigned to more than one level-2 discipline (National Science Library of Chinese Academy of Science, 2015).

Table 7. List of disciplines in CSCD journal classification system

Level-1 Discipline	Level-2 Discipline
General Social Science[18]	
General Natural Science	
Physical Science and Chemistry	Physical Science and Chemistry – General Topics, Mathematics, Mechanics, Physics, Chemistry
Astronomy & Earth Science	Astronomy & Earth Science – General Topics, Astronomy, Topography, Geophysics, Meteorology, Geology, Oceanology, Physical Geography
Biology	Biology – General Topics, Biology Principle & Theory, Paleontology, Microbiology, Botany, Zoology & Anthropology, Entomology
Medical Science & Health	Medical Science & Health – General Topics, Disease Prevention & Hygiology, Chinese Medicine, Preclinical Medicine, Clinical Medicine, Internal Medicine, Surgery, Obstetrics & Gynecology, Pediatrics, Oncology, Neurology &

[17] Since no level-2 discipline is under *General Social Science, General Natural Science, Transportation, Aviation & Aerospace,* and *Multidiscipline,* these five level-1 disciplines will also be investigated as level-2 disciplines.
[18] Although CSCD is a science citation database, three journals are classified as *General Social Science* journals.

	Psychiatry, Dermatology & Venereology, Otorhinolaryngology, Ophthalmology, Dentistry, Special Medicine, Pharmaceutical Science
Agricultural Science	Agricultural Science – General Topics, Agricultural Basic Science, Agricultural Engineering, Agronomy, Plant Protection, Horticulture, Forestry, Animal Science & Veterinary Medicine, Fisheries
Engineering & Technology	Engineering & Technology – General Topics, General Engineering & Technology, Mining Engineering, Petroleum Engineering, Metallurgical Engineering, Metal Science, Machinery & Instruments, Military Science, Energy & Power Engineering, Nuclear Science & Technology, Electrical Engineering, Electronic Technology & Telecommunications, Automation & Computer Technology, Chemical Engineering, Light Industry, Architecture, Hydraulic Engineering
Transportation	
Aviation & Aerospace	
Environmental & Safety Science	Environmental & Safety Science – General Topics, Safety Science
Multidiscipline	

3.2.3.2. WoS Research Area (WoS SU)

WoS SUs is a subject classification scheme that is shared by all WoS product databases. As a result, users can identify, retrieve and analyze documents from multiple databases that pertain to the same subject. For example, journals covered by the WoS Core Collection are assigned at least one WoS category, and each WoS category can be mapped to one WoS SU.

WoS SUs include 151 research areas that are classified into five major research domains: *Arts Humanities*, *Social Sciences*, *Life Sciences Biomedicine*, *Physical Sciences*, and *Technology*. In CSCD, 683 CLC categories in four levels could be matched to 89 out of 151 WoS SUs (11/14 in *Arts Humanities*, 10/24 *in Social Sciences*, 41/75 in *Life Sciences Biomedicine*, 12/18 in *Physical Sciences*, and 15/20 in *Technology*) through a conversion table provided by Clarivate. Thus, the comparison between the paper classification system and the journal classification system can also be conducted under the WoS SU since CSCD indexed papers and journals can be matched with corresponding WoS SUs.

Since both the paper classification system and the journal classification system are available in CSCD, each CSCD indexed paper can be assigned to the corresponding discipline(s) from either classification system. Thus, we can compare the two classification systems and investigate the difference within the same CSCD data set. Since CSCD is also available from WoS using their own classification (WoS SU), both at the journal- and paper level, this study could compare journal and paper classification for the same data, across two different classification schemas (i.e., CSCD CLC, and WoS SU).

3.3. Research Questions

The objective of this paper is to compare journal classification systems and paper classification systems, which should be done using one data set that offers both levels of classification. This study improves our understanding in the specificity of the classification system of science by answering the following research questions:

1. In a journal classification system, what percentage of journal articles, in a given discipline, are contributed to by other disciplines?
2. In a paper classification system, what percentage of papers in a given discipline are published in journals in other disciplines?
3. Do these percentages vary by discipline?

Answering these questions will reveal possible paper misclassification in current journal classification systems, which has never been systematically investigated and may lead to restructuring the classification system of science used in bibliometric studies.

3.4. Data and Method

The following sections describe the datasets and their treatment, followed by the measures used to answer the research questions.

3.4.1. Data

All CSCD data was provided by the National Science Library of the Chinese Academy of Science. The number of journals indexed by CSCD between 2008 and 2015 was 1,480, from which 869 were indexed throughout that period. The raw data analyzed by this study was 1,830,307 CSCD records of papers published in these 869 journals between 2008 and 2015, of which 0.1% (2,035/1,830,307) were excluded because their CLC codes were not available. Note that 6.2% of these papers (113,917/1,830,307) contained multiple CLC codes; this means that those papers might be classified into multiple disciplines in the paper classification system.

The 869 journals were assigned to 12 level-1 disciplines and 66 level-2 disciplines as shown in Table 7. Of these journals, 31 (3.6%) were assigned to 2 or 3 level-2 disciplines while 838 journals (96.4%) were assigned to a single level-2 discipline. No journal was assigned to more than one level-1 discipline, and 6.2% (54/869) of journals were English language journals, while the rest were published in Chinese.

3.4.2. Method

Based on the CLC-codes, each journal is classified into the corresponding discipline(s) from the journal classification system while each paper is assigned to the corresponding discipline(s) from the paper classification system. For journal classification, each paper is assigned to the same discipline(s) of its journal, which is compared to the higher specificity offered by paper level classification to reveal differences and similarities.

3.4.2.1. CLC-based System

As described above, all journals, as well as their papers, were assigned to 12 level-1 disciplines and 66 level-2 disciplines in the journal classification system. The less controlled nature of paper level classification, that is performed by authors, is highlighted by the fact that there is no requirement for author(s) to provide CLC-codes at a specific CLC level of abstraction. Indeed, paper CLC codes abstraction levels vary: among the 1,830,307 papers investigated, 24.1% (441,383) contain a CLC code representing a level-1 or level-2 CLC level, while 75.9% (1,388,924) include a CLC code at level 3 or below.

To compare journal- and paper level classification, the structure of the journal classification system (12 level-1 disciplines and 66 level-2 disciplines) was adopted as the *gold standard* to compare with the paper level classification. As a first step, papers with a CLC code at level 3 or below were re-assigned to their parent or grandparent level-2 CLC disciplines; for example, a paper with a CLC code G301 was assigned to the level-2 discipline *Science & Science Studies* (G3) instead of the original assignment to the level 4 discipline *Scientometrics* (G301). Secondly, all papers assigned to level-2 disciplines beyond the selected 66 level-2 journal level disciplines were re-assigned to their parent level-1 discipline(s). For example, the paper assigned to the level-2 discipline *Science & Science Studies* (G3) was re-assigned to the level-1 discipline *Culture, Science, Education & Sports* (G), since the former is not included in the list of 66 level-2 disciplines. This method of reassignment to broader or more general abstraction levels, has been used in library classification mapping where its robustness has been confirmed (Shu et al., 2017). Finally, 11 level-1 disciplines in social science and humanities were merged to two level-1 disciplines: *General Social Science* corresponding to CLC codes starting with B, C, D, E, F, G and *Arts and Humanities*[19] corresponding to CLC codes starting with A, H, I, J, K. For example, all papers assigned to *Culture, Science, Education & Sports* (G) were moved to *General Social Science.*

As a result, papers classified at the journal- or paper levels were modified to fit in the same structure of 12 level-1 disciplines and 66 level-2 disciplines. The outcome of the data treatment is the same set of papers assigned to the same disciplines in two different ways: the paper level based on their own CLC codes, and the journal level based on the publishing journals' CLC codes.

3.4.2.2. WoS SU-based System

Since the CLC categories can be converted to the corresponding WoS SUs via the conversion table (see the sample as Table 8), we can conduct the same journal- vs. paper level comparison with the WoS SU structure. To achieve this, all CSCD indexed journals and papers were respectively assigned to corresponding categories in the WoS SUs with some modifications

[19] No *Arts and Humanities* Journal in CSCD but some papers are identified by their authors as *Arts and Humanities* papers.

made to render the disciplinary structures consistent across both classification systems (i.e., CLC and WoS SU).

Table 8. The sample Chinese Library Classification - WoS Research Area conversion table

CLC Code	CLC Category	WoS Research Domain	WoS Research Area (SU)
A	Philosophy	Arts & Humanities	Philosophy
A1	Philosophy	Arts & Humanities	Philosophy
A2	Philosophy	Arts & Humanities	Philosophy
A3	Philosophy	Arts & Humanities	Philosophy
A4	Philosophy	Arts & Humanities	Philosophy
A49	Philosophy	Arts & Humanities	Philosophy
A5	Philosophy	Arts & Humanities	Philosophy
A7	Philosophy	Arts & Humanities	Philosophy
A8	Philosophy	Arts & Humanities	Philosophy
B	Humanities multidisciplinary	Arts & Humanities	Arts & Humanities - Other Topics
…	…	…	…
X7	Environmental sciences	Life Sciences & Biomedicine	Environmental Sciences & Ecology
X8	Environmental sciences	Life Sciences & Biomedicine	Environmental Sciences & Ecology
X9	Public environmental occupational health	Life Sciences & Biomedicine	Public, Environmental & Occupational Health

Using the conversion table, 869 journals were assigned to 42 WoS SUs in the journal classification system, while 1,830,307 papers were assigned to 86 WoS SUs in the paper classification system, except for 51 papers which CLC-codes were not convertible. Lastly, 41 WoS SUs were found in both the classification systems. Three journals were assigned to *Arts & Humanities - Other Topics*[20] but no paper was assigned to this WoS SU; on the other hand, 309,361 papers (16.9%) were assigned to 45 WoS SUs in which no journal was assigned.

[20] In the WoS SU, "Other Topics" in research area names means "General Topics". For example, *Arts & Humanities - Other Topics* covers general topics in Arts and Humanities.

A modification was made to keep the same structure in both classification systems for the comparison. First, these 309,361 papers were re-assigned to WoS SUs included in both classification systems based on their CLC codes at upper level(s). For example, a paper with a CLC code of TU4 was originally assigned to *Construction & Building Technology* that is not included in the journal classification system; we re-assigned it to *Architecture* according to its upper level CLC code as TU. Secondly, we added *Arts & Humanities - Other Topics* in the paper classification system and moved 581 papers originally assigned to seven Arts and Humanities SUs (i.e., *Religion, Literature, Philosophy, History, Music, Art* and *Film, Radio & Television*) to it. Eventually, all journals and papers were respectively assigned to 42 WoS SUs which are included in both the journal classification system and the paper classification system.

3.4.3. Measurement

Papers could be classified into a given discipline in either the journal classification system or the paper classification system and form two corpuses. As shown in Figure 13, the circle J and P represent all papers classified as the discipline X in the journal classification system and the paper classification system, respectively. The overlap area O represents all papers classified as discipline X in both systems. The area A (A = J \ O) represents papers classified as discipline X using journal classification but self-identified by their authors as concerning another discipline(s). The area B (B = P \ O) represents papers in corpus B that are self-identified by their authors as disciplines X but published in journals from another discipline or disciplines. Both A and B are articles misclassified by the journal classification system.

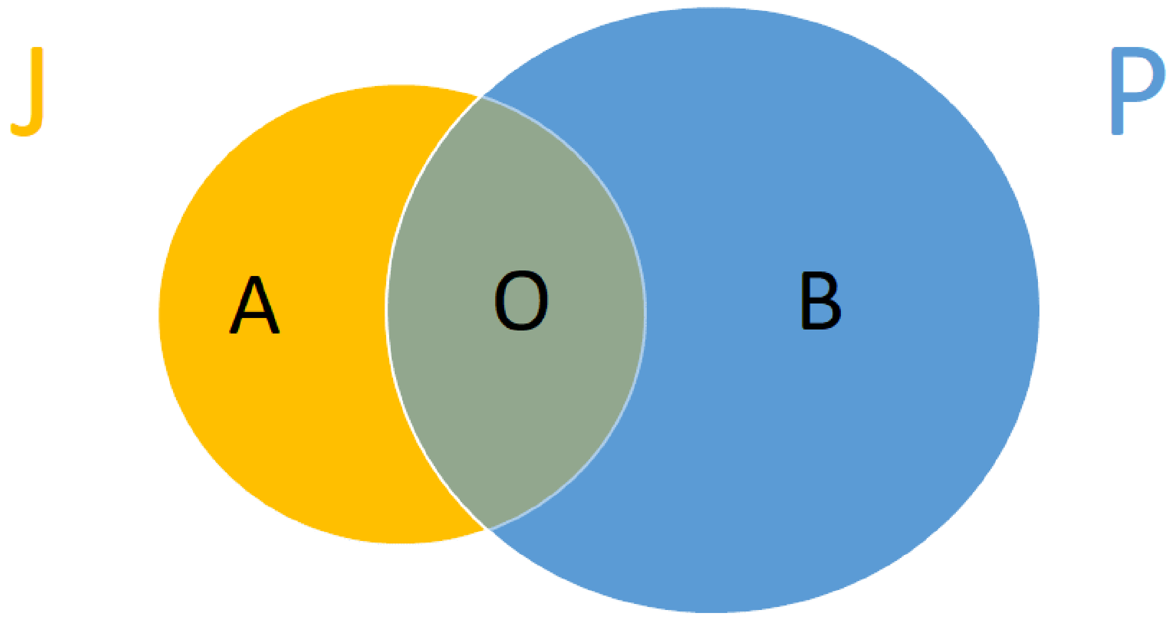

Figure 13. Example measurement of paper misclassification in a given discipline X between journal classification (J) and paper classification (P)

Since the size of corpus J, P as well as the overlap O may vary by discipline, two ratios measure the misclassification as shown in Table 9. In a given discipline X, the ratio of papers contributed from other disciplines ($Ratio_A$ = A / J) represents the percentage of papers published in journals in discipline X but self-identified by their authors as other disciplines than X. The ratio of papers contributed to other disciplines ($Ratio_B$ = B / P) represents the percentage of papers self-identified by their authors as discipline X but published in journals in other disciplines than X.

Table 9. Measurements used by this study

Research Question	**CSCD Level-1**	**CSCD Level-2**	**WoS SU**
RQ1	$Ratio_A$	$Ratio_A$	$Ratio_A$
RQ2	$Ratio_B$	$Ratio_B$	$Ratio_B$
RQ3	$Ratio_A$, $Ratio_B$	$Ratio_A$, $Ratio_B$	$Ratio_A$, $Ratio_B$

3.5. Results

Between 2008 and 2015, 1,830,307 papers were published in 869 CSCD indexed journals. The average number of papers per journal is 2,106, ranging from 124 in *Progress in Physics* (物理学进展) to 20,623 in *Chinese Journal of Gerontology* (中国老年学杂志). Most papers were published in Chinese except that 69,118 English papers (3.8%) were published in 54 English journals (6.2%) administrated by Chinese publishers.

3.5.1. CLC-based Comparison

In the journal classification system, 869 journals as well as their 1,830,307 papers were classified into 12 level-1 disciplines and 66 level-2 disciplines (See Appendix II and Appendix III). *Engineering & Technology*, *Medical Science & Health*, *Physical Science and Chemistry* were the top 3 level-1 disciplines contributing more than 60% of journals and 70% of papers. *Geology* was the largest level-2 discipline in terms of the number of indexed journals (45) while *Automation & Computer Technology* contributed the most papers (125,023). Although CSCD is a science citation database, it still indexes three *General Social Science* (level-1) journals that include 5,281 papers. In addition to 45 *Multidiscipline* journals[21], three journals are classified as *General Natural Science* (level-1).

These 1,830,307 papers were also assigned to 12 level-1 disciplines and 66 level-2 disciplines in the paper classification system (See Appendix IV). *Engineering & Technology*

[21] CSCD does not provide any details regarding the criteria assigning a journal to *Multidiscipline*.

(671,417) and *Automation & Computer Technology* (173,865) are the largest level-1 and level-2 discipline respectively in terms of the number of indexed papers.

3.5.1.1. Disciplinary Distribution of Journal Articles

Comparing assignments between the journal classification system and the paper classification system shows that 21.3% of journal articles were contributed by other level-1 disciplines ($Ratio_A$) than their journal's discipline, ranging from 4.4% in *Medical Science & Health* to 47.6% in *Biology* (See Table 10). The $Ratio_A$ increased to 46.0% ranging from 5.3% in *Horticulture* to 81.8% in *Agricultural Engineering* when the 59 level-2 disciplines[22] were investigated (See Table 11). Although $Ratio_A$ varied in different Level-2 disciplines, the largest contributions were still from their own disciplines except for *Metallurgical Engineering* which had 42.0% of papers identified by authors as *Metal Science* papers while only 19.5% of papers were identified as *Metallurgical Engineering* papers.

Note that papers with general topics might affect the results. For example, 29.1% of papers in *Neurology & Psychiatry* journals were identified by authors as *Medical Science & Health – General Topics*, and *General Social Science* papers account for 19.5 % of papers in *Physical Geography* journals. The share of these papers with general topics could increase $Ratio_A$ although some of them might partly relate to their journals' disciplines.

[22] Seven Level-2 disciplines (*Multidiscipline, Engineering & Technology – General Topics, Agricultural Science – General Topics, Medical Science & Health – General Topics, Biology – General Topics, General Natural Science*, and *General Social Science*) were excluded because most journals under these disciplines are multidisciplinary journals.

Table 10. Ratio of papers contributed from other disciplines (Ratio$_A$) by level-1 discipline presented by the heat map

J \ P	AA	AE	AH	AS	BI	ES	ET	GN	GS	MH	PC	TR	Ratio A
AA	54.8%	2.2%	0.0%	0.0%	0.0%	0.1%	37.1%	0.2%	0.5%	0.1%	3.9%	0.9%	45.2%
AE	0.1%	66.7%	0.1%	3.7%	4.9%	6.4%	12.2%	0.1%	2.9%	0.3%	1.8%	0.9%	33.3%
AS	0.0%	2.1%	0.0%	66.8%	10.7%	5.9%	9.0%	0.0%	3.1%	0.9%	1.1%	0.5%	33.2%
BI	0.0%	0.8%	0.2%	20.8%	52.4%	4.3%	2.1%	0.0%	1.0%	18.0%	0.4%	0.0%	47.6%
ES	0.1%	6.6%	0.0%	4.1%	3.3%	68.9%	10.0%	0.1%	4.5%	0.4%	1.3%	0.7%	31.1%
ET	1.4%	2.0%	0.0%	0.8%	1.0%	2.1%	81.1%	0.4%	1.3%	1.0%	7.2%	1.7%	18.9%
GN	0.1%	7.5%	4.8%	0.9%	1.7%	0.2%	10.7%	54.1%	5.0%	2.8%	11.0%	1.2%	45.9%
GS	0.0%	0.0%	0.3%	0.1%	0.2%	0.1%	0.6%	0.0%	88.2%	10.3%	0.2%	0.0%	11.8%
MD	1.3%	5.0%	0.1%	2.2%	6.0%	4.5%	48.1%	0.8%	4.3%	3.2%	20.0%	4.5%	N/A
MH	0.0%	0.0%	0.0%	0.5%	1.9%	0.1%	0.7%	0.0%	0.8%	95.6%	0.3%	0.0%	4.4%
PC	1.3%	0.7%	0.0%	0.9%	1.1%	1.1%	24.8%	0.2%	1.1%	1.8%	66.0%	1.0%	34.0%
TR	1.2%	1.5%	0.0%	0.2%	0.1%	1.3%	25.3%	0.1%	2.7%	0.1%	2.7%	64.7%	35.3%

AA: Aviation & Aerospace
AE: Astronomy & Earth Science
AH: Arts and Humanities (paper only)
AS: Agricultural Science
BI: Biology
ES: Environmental & Safety Science
ET: Engineering & Technology
GN: General Natural Science
GS: General Social Science
MD: Multidiscipline (journal only)
MH: Medical Science & Health
PC: Physical Science and Chemistry
TR: Transportation

Table 11. List of Level-2 disciplines (journal level) in terms of the ratio of papers contributed from other disciplines (Ratio$_A$)

Rank	Discipline	Ratio$_A$
1	Agricultural Engineering	81.8%
2	Metallurgical Engineering	80.5%
3	General Engineering & Technology	80.3%
4	Physical Geography	79.8%
5	Special Medicine	78.2%
6	Clinical Medicine	73.4%
7	Machinery & Instruments	72.9%
8	Agricultural Basic Science	70.7%
9	Mining Engineering	68.9%
10	Military Science	67.9%
…	…	…
50	Automation & Computer Technology	24.1%
51	Dentistry	22.6%
52	Ophthalmology	21.0%
53	Mathematics	20.2%
54	Chinese Medicine	17.0%
55	Oncology	16.7%
56	Otorhinolaryngology	16.1%
57	Metal Science	15.7%
58	Entomology	7.4%
59	Horticulture	5.3%

3.5.1.2. Disciplinary Distribution of Paper Publication

Generally, it is assumed that scholars would like to submit their papers to journals within the same discipline or multidisciplinary journals, but we found that 24.5% of papers were published in journals from other Level-1 disciplines ($Ratio_B$), ranging from 6.2% in Medical Science & Health to 86.9% in General Social Science (See Table 12). The $Ratio_B$ increased to 51.1% when Level-2 discipline were analyzed: ranging from 20.0% in Light Industry to 77.0% in Oncology (See Table 13)

Table 12. Ratio of papers contributed to other disciplines ($Ratio_B$) by level-1 discipline presented by the heat map

J \ P	AA	AE	AH	AS	BI	ES	ET	GN	GS	MH	PC	TR
AA	55.3%	0.6%	0.1%	0.0%	0.0%	0.0%	1.6%	1.4%	0.4%	0.0%	0.6%	0.8%
AE	0.4%	72.8%	10.3%	2.8%	5.9%	8.5%	2.0%	1.6%	9.3%	0.1%	1.1%	3.0%
AS	0.2%	3.5%	3.0%	76.8%	19.9%	11.9%	2.3%	1.3%	15.1%	0.3%	0.9%	2.6%
BI	0.0%	0.7%	19.5%	11.6%	47.3%	4.2%	0.3%	0.7%	2.3%	3.1%	0.2%	0.0%
ES	0.1%	4.1%	1.2%	1.7%	2.3%	51.6%	0.9%	1.7%	8.3%	0.0%	0.4%	1.3%
ET	30.9%	12.8%	15.7%	3.3%	7.3%	16.4%	79.1%	50.6%	24.0%	1.4%	24.4%	34.1%
GN	0.0%	0.1%	10.7%	0.0%	0.0%	0.0%	0.0%	20.1%	0.2%	0.0%	0.1%	0.1%
GS	0.0%	0.0%	1.8%	0.0%	0.0%	0.0%	0.0%	0.0%	13.1%	0.1%	0.0%	0.0%
MD	3.4%	3.7%	11.0%	1.1%	5.0%	4.1%	5.5%	14.0%	9.4%	0.5%	7.9%	10.6%
MH	0.4%	0.1%	21.4%	1.5%	9.9%	0.6%	0.5%	0.7%	10.6%	93.8%	0.7%	0.3%
PC	8.4%	1.3%	4.3%	1.1%	2.3%	2.4%	6.9%	7.3%	5.6%	0.7%	63.4%	5.6%
TR	0.9%	0.3%	1.0%	0.0%	0.0%	0.3%	0.8%	0.6%	1.6%	0.0%	0.3%	41.5%
Ratio B	**44.7%**	**27.2%**	**N/A**	**23.2%**	**52.7%**	**48.4%**	**20.9%**	**79.9%**	**86.9%**	**6.2%**	**36.6%**	**58.5%**

AA: Aviation & Aerospace
AE: Astronomy & Earth Science
AH: Arts and Humanities (paper only)
AS: Agricultural Science
BI: Biology
ES: Environmental & Safety Science
ET: Engineering & Technology
GN: General Natural Science
GS: General Social Science
MD: Multidiscipline (journal only)
MH: Medical Science & Health
PC: Physical Science and Chemistry
TR: Transportation

Table 13. List of Level-2 disciplines (paper level) in terms of the ratio of papers contributed to other disciplines (RatioB)

Rank	Discipline	Ratio
1	Oncology	77.0%
2	Clinical Medicine	76.8%
3	Horticulture	76.5%

4	Mechanics	73.5%
5	Plant Protection	73.4%
6	General Engineering & Technology	73.0%
7	Zoology & Anthropology	73.0%
8	Botany	69.3%
9	Agronomy	68.7%
10	Neurology & Psychiatry	68.6%
…	…	…
50	Oceanology	38.8%
51	Biology Principle & Theory	38.7%
52	Otorhinolaryngology	38.7%
53	Forestry	38.6%
54	Geology	38.4%
55	Ophthalmology	35.6%
56	Electrical Engineering	32.3%
57	Pharmaceutical Science	29.4%
58	Physics	27.7%
59	Light Industry	20.0%

It is noteworthy that 25.9% of author-identified *Clinical Medicin*e papers were published in *Special Medicine* journals while only 23.2% of *Clinical Medicin*e papers were published in that journal's discipline. In addition, *Mathematics* papers made the largest contribution to *Multidiscipline* journals accounting for 24.6% of total papers, followed by *General Natural Science* papers (14.0%), *Arts and Humanities* papers (11.0%), *Transportation* papers (10.6%) and *General Social Science* papers (9.4%).

3.5.2. WoS SU-based Comparison

Similar results were found when all 1,830,307 papers and 869 journals were assigned to WoS SUs in both the journal classification system and the paper classification system using the conversion table (See Appendix V and Appendix VI). Among the 39 analysed SUs[23], 39.6% of journal articles were contributed from other SUs than their journal's SU(s) ($Ratio_A$), ranging from 7.5% in *Entomology* to 76.2% in *Biotechnology & Applied Microbiology* (See Table 14). In 38 out of these 39 SUs, most papers were contributed from the same SU as their journal's. The exception is *Neurosciences & Neurology* where 58.3% of papers stemmed from *General & Internal Medicine* as compared with 30.7% of papers contributed by *Neurosciences & Neurology* itself.

A similar $Ratio_B$ was found in papers contributed to other SUs: 39.2% of papers were published in journals classified in other SUs ranging from 24.7% in *Agriculture* to 76.2% in *Genetics & Heredity* (See Table 15), and papers in 6/39 SUs were published more in journals classified in another SU than journals classified in its own SU. Only 6.4% of *Genetics & Heredity* paper were published in *Genetics & Heredity* journals while 23.8% of these papers contributed to *Agriculture* journals; most *Mechanics* papers were not published in their own *Mechanics* journals (26.9%) but in *Engineering* journals (37.5%); 44.1% of *Paleontology* papers appear in *Geology* journals compared to 40.7% of papers in *Paleontology* journals. *General & Internal Medicine* journals are popular since most papers from *Neurosciences & Neurology* (54.7%), *Oncology* (53.1%), *Public, Environmental & Occupational Health* (46.7%) were published in *General & Internal Medicine* journals rather than journals in their own SUs (30.7%, 23.1%, 28.8% respectively).

---

[23] *Science & Technology - Other Topics, Life Sciences & Biomedicine - Other Topics, Arts & Humanities - Other Topics* were not included because most journals assigned to this three SUs are multidisciplinary journals.

Table 14. List of WoS SUs (journal level) in terms of the ratio of papers contributed from other WoS SUs (Ratio$_A$)

Rank	SU	Ratio$_A$
1	Biotechnology & Applied Microbiology	76.2%
2	Mechanics	65.8%
3	Mining & Mineral Processing	64.4%
4	Oceanography	63.6%
5	Genetics & Heredity	62.4%
6	Geology	60.3%
7	Plant Sciences	59.9%
8	Neurosciences & Neurology	58.3%
9	Microbiology	55.6%
10	Anthropology	55.6%
…	…	…
30	Paleontology	26.0%
31	Pediatrics	25.8%
32	Automation & Control Systems	23.9%
33	Metallurgy & Metallurgical Engineering	23.4%
34	Dentistry, Oral Surgery & Medicine	22.7%
35	Ophthalmology	21.0%
36	Mathematics	19.4%
37	Oncology	16.7%
38	Otorhinolaryngology	16.1%
39	Entomology	7.5%

Table 15. List of WoS SUs (paper level) in terms of the ratio of papers contributed to other WoS SUs (Ratio$_B$)

Rank	SU	Ratio$_B$
1	Genetics & Heredity	76.2%
2	Zoology	72.8%
3	Anthropology	71.1%
4	Plant Sciences	69.0%
5	Microbiology	66.2%
6	Biotechnology & Applied Microbiology	66.1%
7	Fisheries	63.5%
8	Mechanics	62.5%
9	Transportation	57.2%
10	Paleontology	56.0%
…	…	…
30	Oceanography	38.7%

31	Otorhinolaryngology	38.6%
32	Forestry	38.4%
33	Ophthalmology	35.6%
34	Pharmacology & Pharmacy	31.4%
35	General & Internal Medicine	29.0%
36	Physics	28.2%
37	Geology	28.1%
38	Engineering	27.9%
39	Agriculture	24.7%

3.6. Discussions

Classifying publications into research areas or disciplines is a basic element of bibliometric studies. Traditionally, the classification is based on the journal's discipline(s), but its accuracy has not been systematically investigated. This study shows a classification mismatch between journal and paper classification systems, which raises further questions concerning the accuracy of the dominant classification system of science.

3.6.1. Accuracy of Journal Classification System

Traditionally in bibliometric studies, it is assumed that research papers belong to the same discipline(s) of their journal. This study shows that 46.0% of journal articles, on average, come from other disciplines than that of their journal's. It means that the journal classification system returns inaccurate results for research. This study also confirms the question raised by previous studies regarding the inaccuracy of the journal classification system.

For example, between 2008 and 2015, 17,701 papers were published in four *Agriculture Engineering* journals indexed by CSCD. Of these 17,701 papers, 3,217 (18.2%) were identified by their authors as *Agricultural Engineering* papers, followed by *Agricultural Basic Science* (2,540, 14.4%), *Automation & Computer Technology* (1,775, 10.0%), and *Agronomy* (1,398, 7.9%). As Figure 14 shows, although *Agricultural Engineering* papers made the largest contribution to *Agricultural Engineering* journals, there were 14,484 papers (81.8%) that were contributed from 54 other disciplines. If we simply classified all these 17,701 papers as *Agricultural Engineering* papers using the journal classification system, 14,484 papers ($Ratio_A$ = 81.8%) would be misclassified. Meanwhile, 3,402 out of 6,619 *Agricultural Engineering* papers ($Ratio_B$ = 51.4%) published in journals in other disciplines were not classified as *Agricultural Engineering* in the journal classification system.

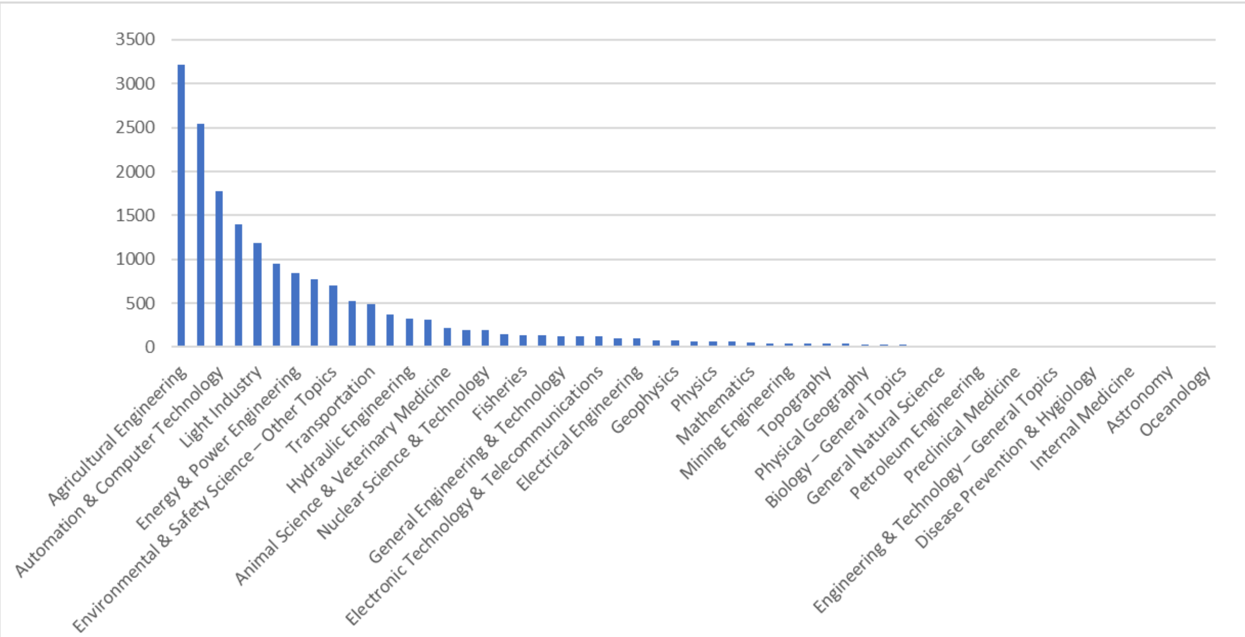

Figure 14. Distribution of papers in Agricultural Engineering journals by discipline

Journals in some disciplines such as *Horticulture* published very few papers (242/4,590) from other disciplines. Even so, 18,508 papers were identified by their authors as *Horticulture* papers but only 4,348 were published in *Horticulture* Journals, which entails that 76.5% ($Ratio_B$) of *Horticulture* papers (14,160 /18,508) were classified as other disciplines in the journal classification system. Analysis of level-2 disciplines shows similar examples where journal articles stem from, on average, 48.9 source disciplines including the journal's own, ranging from

14 source disciplines in *Horticulture* to 66 source disciplines in *Chemistry*. Figure 15 presents the diversity of between-discipline contributions in level-2 disciplines, where each point shows the number of source disciplines and the ratio of papers contributed from other disciplines ($Ratio_A$). This represents the diversity of source disciplines contributing to journal articles in each level-2 discipline. These points are grouped by their parent level-1 disciplines using different colours. Figure 15 presents the huge diversity of between-discipline contributions in most scientific disciplines, in which 5.3% to 80.5% of journal articles are contributed from other disciplines. Thus, it reveals the extent of misclassification of journal articles if we simply classify all papers into the same discipline as their journal in the journal classification system.

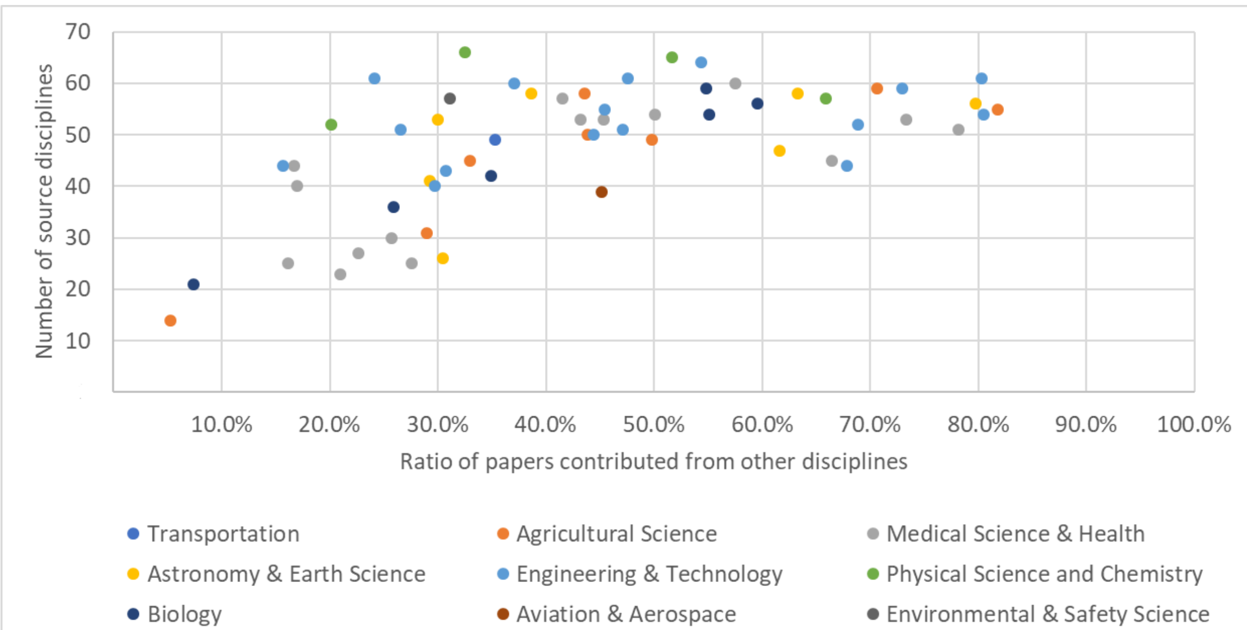

Figure 15. Diversity of between-discipline contributions in CSCD in level-2 disciplines

3.6.2. System Settings

In addition to the CSCD classification system, many journal classification systems (WoS Categories, Scopus All Science Journal Classification, National Science Foundation (NSF) system, etc.) are available but they vary by their number of disciplines, level of classification, and journal classification methodology. This study reveals that such system settings could influence the accuracy of the journal classification system.

3.6.2.1. Inclusive or Exclusive Classification

Both the inclusive and the exclusive classifications are used in the journal classification system when assigning journals to disciplines. Both WoS and Scopus assign journals to one or more disciplines while NSF and other systems (e.g. Science-Metrix) exclusively assign each journal to a single discipline. Both methods are reasonable, but the former may produce more misclassifications. For example, the journal *Rare Metal* was assigned to CSCD Level-2 disciplines *Metallurgical Engineering* and *Metal Science*; from 2008 to 2015, this journal published 988 papers that were classified by authors into 12 Level-2 disciplines as shown in Figure 16. It shows that 550 papers (55.7%) were classified as *Metal Science* and 230 papers (23.3%) were classified as *Metallurgical Engineering* papers, followed by 68 papers (6.9%) in *Engineering & Technology – General Topics*, 37 papers (3.7%) in *Chemistry* and 33 papers (3.3%) in *Chemical Engineering*. As Table 17 shows, with inclusive classification, the 988 papers were indexed as 988 *Metal Science* papers and 988 *Metallurgical Engineering* papers, which produced 438 misclassifications (44.3%) as *Metal Science* papers and 758 misclassifications (76.7%) as *Metallurgical Engineering* papers. On the other hand, with the exclusive classification, this journal would be assigned to *Metal Science* only and produce only 438 misclassifications (44.3%) as *Metal Science* papers.

The purpose of inclusive classification is to allow journals to represent more related disciplines. However, since the journal classification system cannot differentiate journal articles by discipline, inclusively classifying journal articles into two or more disciplines of their journal will produce more paper misclassifications as compared with exclusive classification.

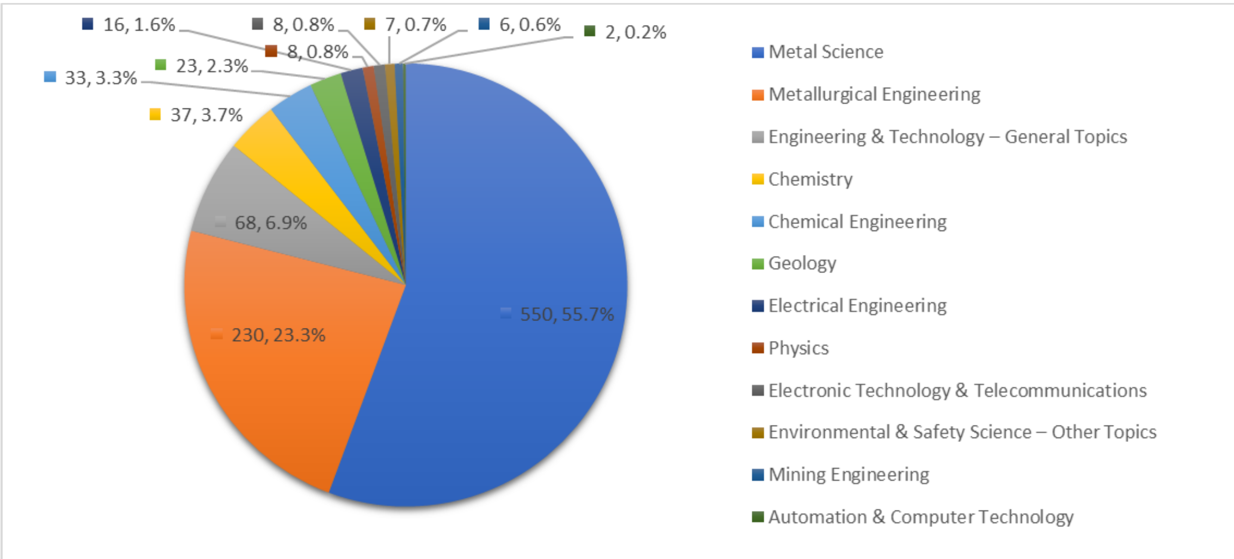

Figure 16. Distribution of papers published in the journal Rare Metal by discipline (2008-2015)

3.6.2.2. Number of Disciplines

The number of classified disciplines varies in different journal classification systems, and this has an effect on the accuracy of the journal classification system. This study made three comparisons between the journal classification system and the paper classification system in CLC Level-1 disciplines, CLC Level-2 disciplines and WoS SUs, which has a different number of disciplines in the classification system. As Table 16 shows, the ratio of papers contributed from other disciplines ($Ratio_A$) increases with the number of disciplines in the journal classification system. The more disciplines a journal classification system has, the higher the $Ratio_A$ is. A strong correlation between the number of disciplines in the classification system and the $Ratio_A$ (r=-.9847) is found.

Table 16. Comparison of the results based on three different classification systems

Classification System	**# of disciplines[24]**	**Ratio$_A$**		
		Min.	**Avg.**	**Max.**
CLC Level-1	11	4.4%	21.3%	47.6%
WoS SU	39	7.5%	39.6%	76.2%
CLC Level-2	59	5.3%	46.0%	81.8%

For another example of the journal *Rare Metal:* if we create a new discipline *Metal Science and Engineering* that covers both *Metallurgical Engineering* and *Metal Science;* as Table 17 shows, only 208 out of 988 papers are contributed from disciplines other than the new discipline *Metal Science and Engineering;* the ratio of misclassification (Ratio$_A$) will decrease from 44.3% (438/988) to 21.0% (208/988). It means that a journal classification system including only broad or high-level disciplines could reduce the paper misclassifications comparing to a detailed journal classification system including many disciplines.

Table 17. The comparison of misclassification between inclusive and exclusive classification

	Inclusive classification			**Exclusive classification**		
	# of papers	# of misclassification	% of misclassification	# of papers	# of misclassification	% of misclassification
Metal Science	988	438	44.3%	988	438	44.3%
Metallurgical Engineering	988	758	76.7%	N/A		
Metal Science and Engineering	N/A			988	208	21.0%

3.6.2.3. Multidisciplinary and Interdisciplinary

Some academic journals cover literature from multiple scientific disciplines. These are either classified into multidisciplinary categories (e.g. *Agriculture, Multidisciplinary; Multidisciplinary*

[24] Multidisciplinary level-2 disciplines and WoS SUs (e.g. *Multidiscipline, General Social Science, Biology – Other Topics*, etc.) are excluded here.

Sciences; Physics, Multidisciplinary; etc.) or inclusively assigned to multiple related disciplines in the major bibliometric databases (i.e., WoS and Scopus). As analyzed above, inclusively classifying a journal that publishes papers from diverse disciplines into multiple specific disciplines will produce more paper misclassifications than classifiying it into a multidisciplinary category. The methodologies that determine the boundary of these multidisciplinary disciplines affect the accuracy of a journal classification system; unfortunately, there is a death of publicly available information describing the specifics of these methodologies.

The Herfindahl-Hirschman index (HHI)[25], which is commonly used to measure market concentration in business and has been applied in bibliometric studies for measuring the concentration of authors, papers, journals, citations, references as well as other indicators (Chi, 2016; Huang, Fang, & Chang, 2011; Keathley-Herring et al., 2016; Tseng & Tsay, 2013). In this study, the HHI and the share of the largest contributor (Level-2 discipline) were used to measure the concentration of disciplines for all 869 CSCD journals. The HHI ranges from 0 to 1 as a journal covers all scientific disciplines down to a single discipline.

As Figure 17 **s**hows, the 869 journals were grouped into four categories: *Multidiscipline* (45 journals classified into the Level-1 category as *Multidiscipline*), *General Discipline* (78 journals classified into Level-2 multidisciplinary categories such as *Biology – Other Topics, General Social Science, etc***.**), *Cross Discipline* (31 journals classified into multiple Level-2

---

[25] The formula of HHI is
$H = \sum_{i=1}^{n} Si^2$
where Si is the market share of firm i in the market, and N is the number of firms

disciplines), and *Single Discipline* (the rest of 715 journals). These categories are identified by the color of the data points in Figure 17.

Figure 17 shows that the journal's discipline concentration increases with the growth of its share of the largest contributor to the journal. *Multidiscipline* journals are less concentrated with the HHI ranging from 4.3% to 40.6% while almost half *Single Discipline* journals (337/715) concentrated more on a single discipline with the HHI over 50%. The concentration measurement of *General Discipline* journals (18.0% on average, between 5.2% and 74.6%) and *Cross Discipline* journals (34.6% on average, between 9.7% and 87.9) are between the *Multidiscipline* (13.0% on average, between 4.3% and 40.6%) and *Single Discipline* (50.9% on average, between 6.7% and 99.7%). However, there is no obvious boundary between them; some *Single Discipline* journals are less concentrated than *Multidiscipline* journals but are still classified into a single discipline.

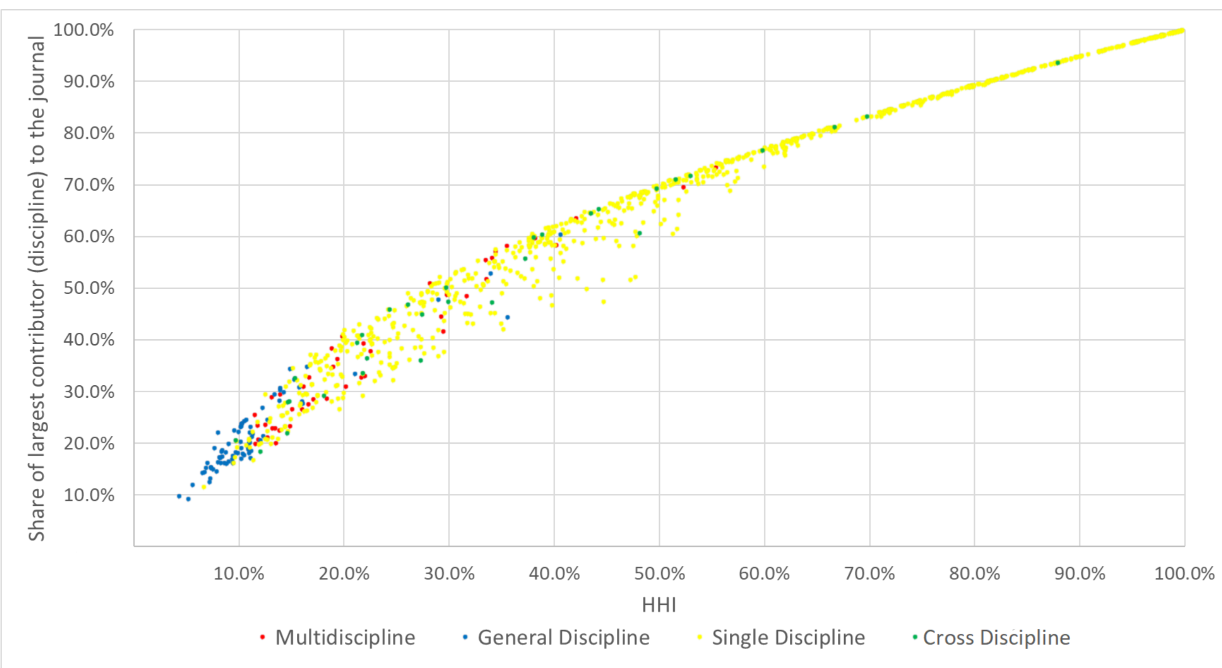

Figure 17. Diversity of source discipline in CSCD journals

Figure 18 presents data from 2008 to 2015, during which 1,114 papers were published in the *Journal of Fujian Agriculture and Forestry University (Natural Science Edition)*, which was classified as a *Forestry* journal. However, this journal does not concentrate on *Forestry* research (111) but published papers are assigned to *Animal Science & Veterinary Medicine* (129), *Agronomy* (111), *Plant Protection* (110), *Horticulture* (104) and another 36 disciplines. Based on its coverage, this journal should be classified as *Agricultural Science – Other Topics* (General Discipline) instead of *Forestry* (Single Discipline). The misclassification of this journal also leads to its papers being misclassified since 90.0% (1,003/1,114) are from disciplines other than *Forestry*.

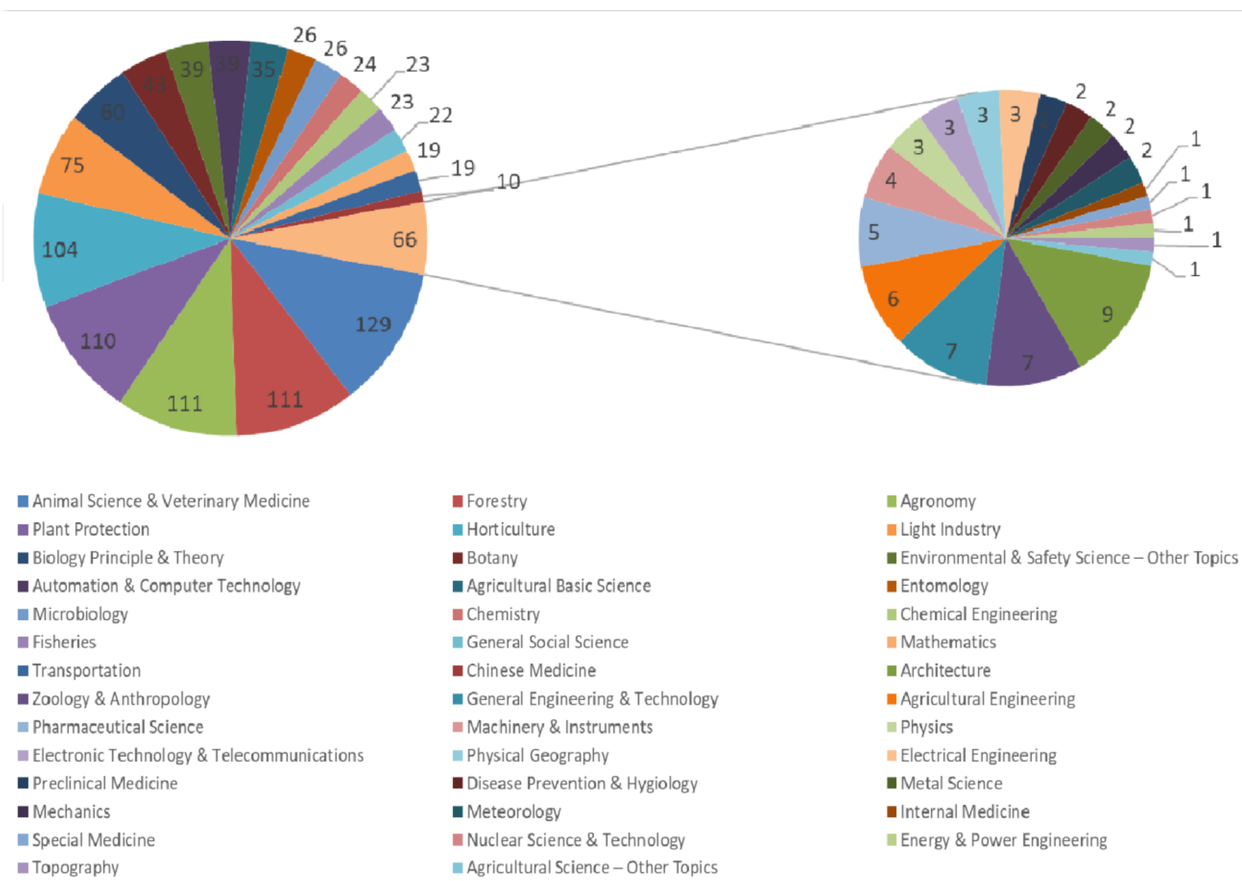

Figure 18. Distribution of papers published in Journal of Fujian Agriculture and Forestry University (Natural Science Edition) by discipline (2008-2015)

3.6.3. Limitations

It is assumed that the CLC code of a paper represents its major research area (discipline) since it is chosen by their author(s). In practice, some scholars may choose inaccurate or inconsistent CLC codes for their papers since they are not classification experts. Most authors are required to provide only one CLC code for each paper, but this is not adequate to represent the research area(s) of an interdisciplinary research paper. In addition, although the CSCD journal classification system is based on the CLC categories, some disciplines are merged or modified using direct citation analysis and clustering methods that are not detailed, which is also a limitation of this study.

3.7. Conclusions

After comparing the classification system of science between the journal level and the paper level, based on the same CSCD data, this study revealed the extent of paper misclassification in the journal classification system of science based on the CSCD data. The results of this study indicate that papers are misclassified by a journal classification system since:

- in the journal classification system, 46.0% of articles were *not* contributed from their journal's discipline(s), and
- in a paper classification system, 51.1% of papers were *not* published in journals of the same discipline(s), but
- the ratio of papers contributed from, or to, other discipline(s) varies by discipline and the methodology constructing the classification system.

This study reveals the problem of paper misclassifications in the journal classification system using the CSCD data. This makes it increasingly essential to develop the paper classification system instead of the journal classification system in bibliometric databases. In addition to the CLC that is only used in China, both the Library of Congress Subject Headings (LCSH) and the Medical Subject Headings (MeSH) could be used to classify papers into disciplines, if authors are required to provide the subject headings. Since the MeSH assignments have been applied to medical research papers, a paper classification system of medical science could be established based on the MeSH; and a comparison between the journal classification system and the paper classification system in medical science, should be proposed in future research.

Overall, this study improves our understanding of the accuracy of the classification system of science and forms a foundation for future studies investigating the difference between the journal classification system and the paper classification system.

Transition II

In the previous chapter I compared the classification system of science between journal level and paper level, which are adopted by Web of Science (WoS) and Chinese Science and Technology Periodical Citation Database (VIP) respectively. The comparison could be used to interpret the difference between WoS and VIP in terms of most productive authors and their research output from a methodological point of view. Also, chapter 2 indicated that Chinese scholars have different publication patterns that could be influenced by science policies. Thus, in the next chapter we investigated the cash-per-publication policy in China, which may influence Chinese scholars' publication activities that were reported in Chapter 2. The investigation could explain why the differences exist by analyzing the impact of science policies on Chinese scholars' publishing activities.

4. Publish or impoverish: An investigation of the monetary reward system of science in China (1999-2016)

4.1. Introduction

Although monetary rewards have been used for recognizing scientific achievement since the eighteenth century, it is not regarded as the major reward system in science (Merton, 1973), in which scientists try to publish their works and receive the recognition of their peers as the reward. Since academic prizes consisting of cash rewards are awarded only to very few scientific elites, they are considered as the metaphors of *prestige* rather than simply large sums of money (Zuckerman, 1992). However, the reward system in science changed when the monetary reward incentive for publication was introduced in 1980s in UK. It is reported that this incentive can promote research productivity (Franzoni, Scellato, & Stephan, 2011) but might create a negative goal displacement effect (Frey et al., 2013; Osterloh & Frey, 2014).

Since the early 1990s, Chinese research institutions have initiated the cash-per-publication reward polices in which Chinese scholars could get cash for each eligible publication. The purpose of publishing their works is not only to advance knowledge and win recognition, but also to earn money (X. Sun & Zhang, 2010; L. Wang, 2016). Since these cash-per-publication reward policies vary by institution and some policies are internal or confidential, they have never been systematically investigated except in some case studies. The purpose of this study is to present the landscape of the cash-per-publication reward policy in China and reveal its trend since the late 1990s.

4.2. China's Scientific Activity

With the significant development of China's economy, China's scientific activity is experiencing a period of rapid growth. As Figure 1 shows above (p. 13) , China's scientific research inputs and outputs have exhibited a consistent growth pattern over the past 20 years; from 1995 to 2015, China's expenditures on Research and Development (R&D) increased almost 40 times from 5.23 billion USD to 212.55 billion USD, while its number of international publications (indexed by Web of Science) increased about 22 times from 12,997 to 287,374 (National Bureau of Statistics of China, 1996-2016). Indeed, China, the second largest share of international scientific production per country since 2009, contributes 17.1% of scientific articles indexed by Web of Science (hereafter referred to as WoS) (ISTIC, 2017).

4.2.1. Universities in China

In China, although the universities, the research institutions, the enterprises, the hospitals not affiliated with universities, and other sectors are all involved in scientific activity, the universities play the dominant role in China's scientific research output, contributing 82.8% of monographs and 73.4% of journal articles, including 83.0% of WoS papers (National Bureau of Statistics of China, 2015). There are 2,595 higher education institutions in China, including 1,236 universities offering undergraduate programs (Ministry of Education of China, 2016). Traditionally these universities vary by ownership, speciality, and region; however, they also can be classified into three tiers by two national research programs: *Project 211* and *Project 985*.

Project 211 was initiated in 1995 by China's Ministry of Education. The objective of this project was to construct 100 world-class universities in the beginning of the twenty-first century

(Ministry of Education of China, 2000). The Chinese government offers preferential policies and financial support to designated universities who are admitted to this project, and has contributed around 2.7 billion USD to it (Tang & Yang, 2008). Eventually, 116 universities were admitted to Project 211, forming an elite group of universities occupying 70% of national research funding and supervising 80% of doctoral students (Tang & Yang, 2008). Today, 112 universities are still included in Project 211, even after 4 universities merged.

Project 985 was first announced by Zemin Jiang, former Chairman of the People's Republic of China, on May 4, 1998 to promote the development of a Chinese equivalent of the US Ivy League (Chen, 2006). This *Chinese Ivy League* started with 9 universities in 2009 and accepted another 30 universities in the following two years. These 39 universities are all Project 211 universities, but receive more government funding than other Project 211 universities (Mohrman, 2005). Both Project 985 and Project 211 ceased admission in 2011, grouping Chinese universities into a 3-tier pyramid hierarchy as shown in Figure 19: 39 universities within Project 985 (hereafter referred to as *Tier 1* universities), 73 universities within Project 211 but excluded from Project 985 (hereafter referred to as *Tier 2* universities), and 1,124 other universities (hereafter referred to as *Tier 3* universities).

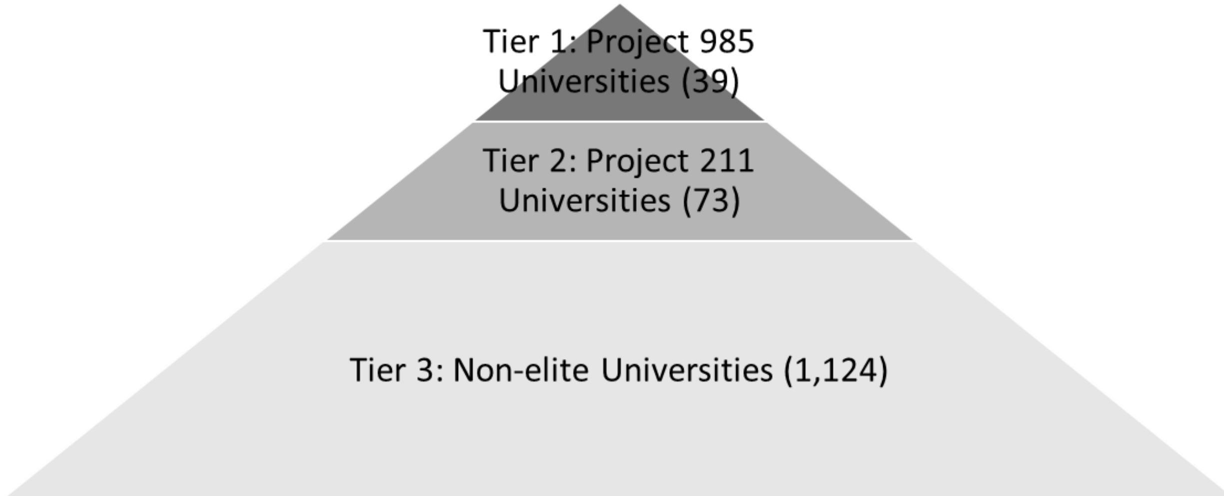

Figure 19. The Pyramid Hierarchy of Chinese Universities

To construct world-class universities, the Chinese government differentiates universities and allocates most funding to a few elite universities, which lead to the "Matthew Effect" among China's higher education system in which the gap between elite universities and non-elite universities has been increasing. Figure 20 presents a huge gap between the elite (Tier 1 and Tier 2) and non-elite (Tier 3) universities in terms of their average annual budgets. From 2002 to 2015, the average annual budget of Tier 1 and Tier 2 universities increased from 23.86 million USD to 113.05 million USD while the mean budget of Tier 3 universities increased from 1.89 million USD to 9.27 million USD. Tier 1 and Tier 2 universities' budget is, on average, 12 times more than Tier 3 universities budget at all times (Ministry of Education of China, 2003-2016). Although a new national research program, "Double World-Class", was announced by the Chinese government in 2016, it was not in effect during the period of our investigation.

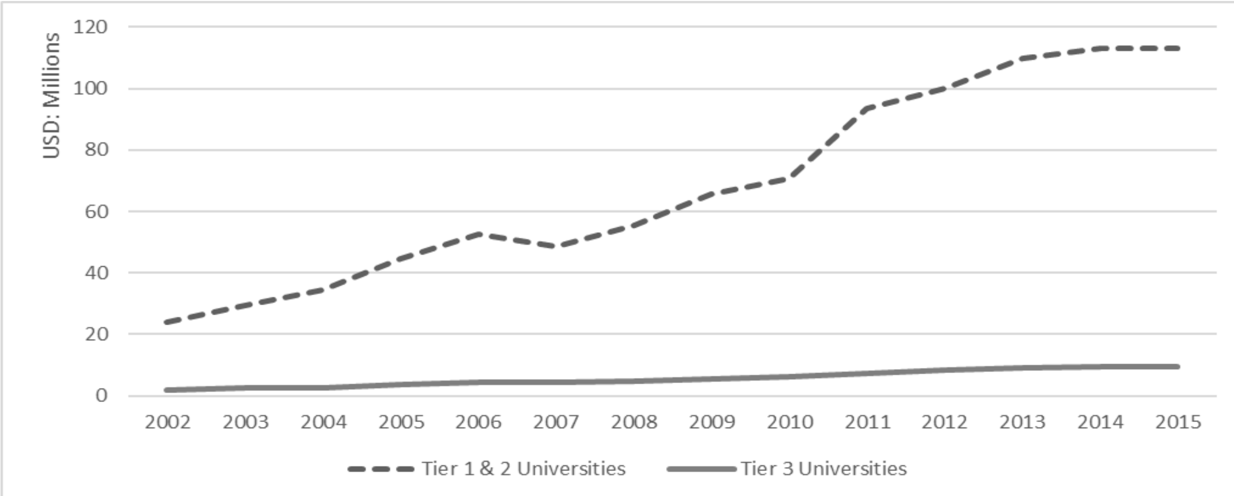

Figure 20. Comparison of Average University Budget between Elite Universities and Non-Elite Universities (2002-2014)

4.2.2. The Cash-per-publication Reward Policy in China

Since the 1980s, to increase the international visibility of Chinese research, the number of WoS papers has been used to evaluate the research output in China of both institutions and individuals (Gong & Qu, 2010). Chinese scholars are required to publish WoS papers to attain promotion, while their affiliated institutions need the number of WoS papers for ranking and funding application (YJ Wang & Li, 2015). Chinese universities and research institutions also offer preferential policies and monetary rewards to encourage their scholars to publish in journals indexed by WoS (Peng, 2011).

The first cash-per-publication reward policy (hereafter referred to as the cash reward policy) was launched by the Department of Physics at Nanjing University around 1990. Initially, scholars received 25 USD for each WoS paper, and the amount increased to between 60 and 120 USD in the mid-1990s (Swinbanks, Nathan, & Triendl, 1997). As the first to apply the WoS to research evaluation, Nanjing University topped the list of Chinese universities publishing most WoS papers seven years in a row in the 1990s (Gong & Qu, 2010); its research evaluation policy and cash reward policy were then copied by other universities and research institutions. Today, every university and research institution in China has established their own cash reward policies.

4.2.3. Regional Difference in China

China consists of 31 provincial-level divisions that were traditionally grouped into seven geographical regions: North, Northeast, Northwest, Center, East, Southwest, and South. Economic development in different regions differs significantly. The GDP per capita in the developed regions (i.e. the North, Northeast, East and South) are much higher than those in the developing regions (i.e. the Northwest, Center, and Southwest) as shown in Table 18.

Since Chinese universities are financially supported by not only the central government but also the local government, regional difference in economic development may lead to differences in the financial capacities of Chinese universities from different regions. Universities in developed regions may have adequate budgets, offering greater monetary reward compared to universities in developing regions. As Table 18 indicates, the average university budgets in the developed regions are much higher than those in the developing regions.

Table 18. GDP per capita and Average University Budget in China by Region in USD (2014)

	North	**Northeast**	**Northwest**	**Center**	**East**	**Southwest**	**South**
GDP per capita	$8,457.28	$7,853.83	$5,881.87	$6,095.42	$9,544.21	$5,017.62	$7,965.03
Avg. University Budget (in millions)	$29.63	$17.67	$11.58	$11.16	$18.05	$9.69	$15.47

Source: National Bureau of Statistics of China (2015). *China Statistical Yearbook*. Beijing, China Statistics Press. Ministry of Education of China (2015). *Scientific Statistics in Higher Education Institutions - 2015*. Beijing, Higher Education Press.

4.3. Literature Review

4.3.1. Reward System of Science

Merton (1973) presents the *sociology of science* with a rewards system and recognition model. He states that science could be regarded as a social institution, with values, norms, and organization (Merton, 1957, 1973). This institution can reward its members (scientists) for their performance. Members also would like to present their achievements in order to get the rewards (Merton, 1957). Merton (1973) also points out that rewards for scientific achievement can be given only if others recognize it. As a result, scientists are eager to publish their works; peers read the publications and recognize their achievements by citing or acknowledging them in their own works. Based on Merton's recognition model (Merton, 1973), the reward system of science is also described as a *reward triangle* consisting of authorship, citations, and acknowledgements (Cronin & Weaver-Wozniak, 1993).

Previous studies suggest that other forms of recognition in addition to the "reward triangle" should be added to the reward system of science. Blume and Sinclair (1973) point out that academic prizes, honorary fellowships, and service on academic committees should be recognized as academic achievement. Sugimoto, Russell, Meho, and Marchionini (2008) compare citation counts and academic mentoring impact, and indicate that academic mentorship should also be granted recognition for its contribution to the spread of knowledge. As social media and other forms of dissemination are being incorporated into scientific practices, a multifaceted reward system has been identified that includes social media mentions, readership counts, and so on (Desrochers et al., 2015).

4.3.2. Academic Monetary Rewards in History

In 1719, the first academic prize was initiated by *Académie des Sciences* in France to award scientists who contribute to the advancement of knowledge in Astronomy. Thereafter, some academic prizes, with or without cash incentives, were introduced by *Académie des Sciences* and *Royal Society of London* to reward new scientific findings or past accomplishments. The establishment of The Nobel Prize, the largest monetary prize in the academic world, turned the academic prize into a metaphor for accomplishment and prestige (Zuckerman, 1977).There are now various academic prizes with big monetary rewards recognizing academic achievement locally and internationally. These large awards reshape the reward system of science's upper reaches (Zuckerman, 1992).

However, since these rich academic prizes are awarded only to very few outstanding scholars, they are valued on the basis of their representations of honors rather than their cash values (Zuckerman, 1992). In addition, Merton (1957) even claims that winning the monetary reward should not be the main purpose of any scientific activity because it may break the norm of *disinterestedness*, referring to rewards for action unaffected by self-interest. He states that:

> Like other institutions also, science has its system of allocating rewards for performance of roles. These rewards are largely honorific, since even today, when science is largely professionalized; the pursuit of science is culturally defined as being primarily a disinterested search for truth and only secondarily a means of earning a livelihood. (Merton, 1957, p. 659)

In 1986, the UK adopted the Research Assessment Exercise (RAE), which allocates national funds to departments based on past performance and peer review. The monetary incentives to publish, regarded as a reform, spread over the world afterwards; some countries even introduced a system of cash bonuses to individuals rather than institutions for each article published in top international scientific journals (Franzoni et al., 2011). Indeed, the economic incentives affect the university research at both the institution (Thursby, Jensen, & Thursby, 2001; Thursby & Thursby, 2002) and the individual levels (Frey et al., 2013; Osterloh & Frey, 2014).

4.3.3. The Monetary Reward in China

As described above, the cash reward policy was launched in China to promote scientific productivity and publication. Through case studies, some Chinese scholars found that monetary rewards could increase scholars' motivation and improve productivity in publishing WoS papers (Z Li & Zhang, 2008; ZW Li & Zhong, 2013; Shan, Han, & Zhao, 2013; Zeng, An, & Wang, 2012). However, no study compares the cash reward policies in different universities nor presents the landscape of the monetary reward for publications in China.

Cash reward policies also create some negative effects. Chinese researchers may favour fast research that leads to quick, cashable publications as opposed to long-term research; in essence, publishing in WoS journals can become the only research goal (Jin & Rousseau, 2004). Some Chinese scientists may resort to plagiarized or fabricated research, purchase ghostwritten papers, or sell authorship (Hvistendahl, 2013; J. Qiu, 2010). Monetary reward also amplifies the existing "Matthew Effect" among Chinese universities (Zhong & Chen, 2008). Compared with Tier 3 universities, Tier 1 and Tier 2 universities dominate scientific resources, and thus could offer more cash rewards for publications, motivating their scholars to produce more (JX Li, 2013; Qi, 2009).

Since the monetary reward is an internal award, it is only announced by Chinese universities via internal documents. Some universities even keep it confidential to avoid competition from other universities. Although the cash reward policy has been applied for 20 years, we still know little about: 1) the range of amounts paid to individuals for publications; 2) if the cash award varies with the quality of journals; 3) if the cash reward policy differs significantly from one university to another. The purpose of this chapter is to present the landscape of the cash-per-publication reward policy in China and address these questions.

Please note that the scope of this study is limited to research in natural science, including engineering and medical science. Since social science and humanities have more localized interests in research and varied methods of disseminating knowledge, beyond journal publications, most Chinese universities have different systems for evaluating and awarding research output in social science and the humanities. For Chinese scholars in social science and humanities, there are additional approaches to winning cash rewards, such as publishing articles in domestic journals or publishing monographs. On the other hand, the cash rewards are only applied to publications in WoS journals in natural science.

4.4. Methodology

4.4.1. Data Collection

In order to present the landscape of the cash-per-publication reward policy in China, we sampled 100 Chinese universities and investigated their cash reward policies since the 1990s. Both stratified sampling and convenience sampling were used.

First, considering the 3-tier pyramid hierarchy of Chinese universities and regional differences, we classified all 1,236 Chinese universities into 21 categories by tiers and regions. Second, we tried to retrieve the cash reward policies from universities in each category to ensure that the sample is representative. Since most cash reward policies are recorded in internal documents that may not be externally accessible, we had to select universities from each category based on data availability. We used the Chinese search engine *Baidu* to locate such information and retrieved it from the official websites of each selected university[26]. Finally, a manual validation was conducted to ensure that the retrieved documents were official and valid.

As Table 19 shows, 100 Chinese universities were selected for the investigation: 25 universities in Tier 1, 33 universities in Tier 2, and 42 universities in Tier 3. The samples also represent Chinese universities from all seven regions in China, as discussed above. Since some Chinese universities had multiple cash reward policies (e.g., modified or new ones), two or more cash reward policies were found in some universities during the period of the investigation. Eventually, 168 cash reward policies were retrieved from these 100 universities. 45 universities contributed one policy each, while 45 universities contributed two; Zhejiang University and Guizhou Normal University issued five and four cash reward policies, respectively, while 8 universities contributed three each. The first cash reward policy that we found was issued in 1999; the number of cash reward policies increased afterwards and reached its peak of 21 in 2015. Eight policies were even issued in 2016 as we started this investigation.

[26] Some internal documents or information regarding the cash reward policies were provided privately by university staff.

Table 19. Distribution of the Sample Universities by Tier and Region

	North	Northeast	Northwest	Center	East	Southwest	South	Total
Tier 1	3(10)	3(4)	3(4)	4(5)	7(11)	3(3)	2(2)	25(39)
Tier 2	4(19)	3(7)	4(9)	4(7)	12(20)	3(7)	3(4)	33(73)
Tier 3	5(178)	3(130)	5(94)	4(162)	14(334)	7(127)	4(99)	42(1,124)
Total	12(207)	9(141)	12(107)	12(174)	33(365)	13(137)	9(105)	100(1,236)

Note: Numbers in brackets represent the total number of Chinese universities in each category.

Due to limited data availability, we did not use random sampling for our data collection, which is a limitation for this study. When comparing the science and technology personnel (S&T personnel), number of international publications, research funding received, and the number of graduate students between the sample and the population (see Table 20), we found that the means of these indicators from the sample Tier 1 universities were very close to those means from all Tier 1 universities, while the means from the sample Tier 2 universities were only a little higher than the means from all Tier 2 universities. The Tier 3 sample seemed to include many top Tier 3 universities so that the sample means were much higher than the average of all Tier 3 universities. We also did the one-sample T-test ($\alpha=0.05$) comparing the sample means with the population means to test whether the samples are representative. As Table 20 shows, we did not find any significant difference between sample and population in all four indicators in Tier 1 and Tier 2 and one indicator (S&T personnel) in Tier 3; significant difference was found between the Tier 3 sample and population in terms of the number of international publications, the research funding received, and the number of graduate students. The T-test indicated that the Tier 1 and Tier 2 samples represented the population well while the Tier 3 sample was a little weak in this study.

Table 20. Comparison of Stats between the Sample Universities and All Universities in Average (2014)

	S&T personnel	International Publications	Research funding (USD: in millions)	Number of graduate students*
Tier 1 Sample	5,182	3,071	205.97	16,176
All Tier 1	4,830	2,896	210.68	15,700
Diff (Tier 1)	0%	0%	0%	0%
Tier 2 Sample	2,228	807	62.16	7,937
All Tier 2	1,822	684	56.43	7,071
Diff (Tier 2)	0%	0%	0%	0%
Tier 3 Sample	1,045	290	19.05	3,348
All Tier 3	831	136	9.16	1,209
Diff (Tier 3)	0%	30.9%	32.8%	105.1%

Source: Ministry of Education of China (2014)
* Data regarding the numbers of graduate students were provided by Research centre for China Science Evaluation (RCCSE) at Wuhan University
Note: Diff in this table refers to the relative measure of hypothesized mean difference in the one-sample T-test (α=0.05). For example, 0% means that no significant difference between the sample mean and population mean while 30.9% means that the mean difference is equal to 30.9% of the population mean.

4.4.2. Data Analysis

Each cash reward policy contains various specifications about its criteria for the eligibility, amount, formula for calculation, and method of payment. It was difficult to compare different cash reward policies with different specifications. In order to compare the cash reward policies issued by different universities in different years, we selected some journals as examples and calculated the amounts of cash reward for a single research paper published in these journals according to different cash reward policies. The selected journals represent journals with different Journal Impact Factors (JIFs) and in different Journal Citation Report (JCR) Quartiles. For a good understanding of the comparison, we selected a list of nine popular journals that could be recognized by our readers, including four multidisciplinary science journals (the first four) and five library and information science journals (the last five) as shown in Table 21.

Table 21. List of Journals Selected for the Comparison

	Journal Impact Factor (5-year)	JCR Quartile (modified)
Nature	41.458	Q1
Science	34.921	Q1
Proceedings of National Academy of Sciences (PNAS)	10.285	Q1
PLOS One	3.535	Q3
MIS Quarterly	9.510	Q1
Journal of the Association for Information Science and Technology (JASIST)	2.762	Q1
Journal of Documentation	1.480	Q2
Library Hi Tech	0.741	Q3
International Journal of Library and Information Science (LIBRI)	0.469	Q4

Source: Journal Citation Report 2015

Both the Journal Impact Factor (JIF) of selected journals and the Journal Citation Report (JCR) Quartile in which these journals are located were used to calculate the amount of cash reward in most cash reward policies. Please note that we chose 5-year JIF instead of 2-year JIF because the former was used frequently by Chinese universities. Also, the JCR Quartile applied to the cash reward policies is not the original one with four equal quarters, but a modified one made by the Chinese Academy of Sciences. Compared to the original JCR Quartile grouping journals in each discipline into four equal quarters, the modified JCR Quartile use a pyramid hierarchy instead: only the top 5% of journals in each discipline are grouped into the Q1 while journals ranked in 5-20%, 20-50% and the bottom 50% are grouped into Q2, Q3 and Q4 in the modified JCR Quartile, respectively.

4.5. Results

A landscape of the cash-per-publication reward policy in China emerged as all 168 cash reward policies were analyzed. Chinese universities offer cash rewards that range from 30 to 165,000 USD for a single paper published in journals indexed by WoS, and the average reward amount has been increasing for the past 10 years. The results show us the overview of the cash-per-publication reward policies in terms of eligibility, amount, and their diversity and trends.

4.5.1. Eligibility and Method

WoS, which includes the Science Citation Index Expanded (SCIE), the Social Science Citation Index (SSCI), the Arts and Humanities Citation Index (AHCI) and the Conference Proceedings Citation Index (CPCI), plays a crucial role in China's cash reward policies. WoS data and the Journal Citation Report (JCR) are used as the eligibility criteria and the grade of the cash reward. Among all cash reward policies, only WoS papers are eligible for the cash reward, except that some universities offer small cash awards to papers indexed by *Engineering Index* (EI)[27]; WoS papers published in different journals may be awarded different amounts according to the journal's JIF and JCR quartile. Based on the analysis of these 168 cash reward policies, we grouped them into the following four major categories:

1. One-price reward (31): Universities pay the same amount to all WoS papers regardless of where these papers are published.

---

[27] Engineering Index is an engineering bibliometric database published by Elsevier. It indexes scientific literature pertaining to engineering materials.

2. Original JIF-based reward (49): Universities award eligible papers different amounts based on the JIF of the journals in which these papers are published. Some universities assigned different grades to eligible journals on the basis of their JIF and pay more for papers published in journals with a high grade; some universities use the JIF as the multiplier to differentiate the cash reward (e.g. the amount of cash reward is equal to a basic amount times the JIF).
3. JCR Quartiles-based reward (99)[28]: Universities award eligible papers different amounts based on the modified JCR Quartile of the journals provided by Chinese Academy of Science that these papers published.
4. Citation-based reward (15): Universities award papers on the basis of the number of citations they received in a given citation window[29]. Some universities set up a threshold of the number of citations and award papers over the threshold; some universities use the Essential Science Indicators (ESI) (e.g., hot paper and highly cited paper) as the threshold for cash award.

Among these 168 cash reward policies, we found 31, 49, 99, and 15 policies that fell into these four categories, respectively. These numbers sum to over 168 because some cash reward policies were grouped into more than one category when universities apply multiple methods to awarding their international publications. We also found trends in the cash reward policies in effect in the following three stages from the late 1990s onwards, as shown in Figure 21. Both the

---

[28] Although the JCR Quartile is also based on the JIF, JCR Quartile provides a structural hierarchy that groups all journals into four categories, which the original JIF-based method does not.
[29] The citation window varies in different universities; some universities use a 5-year citation window while other universities use a 3-year citation window; some universities even count the citations at any time.

one-price reward policies and the Original JIF-based reward policies were popular in the late 1990s and early 2000s, but their shares decreased when the JCR Quartile based policy was introduced. Since 2005, more and more Chinese universities have adopted the JCR Quartile based policy, which became the dominant policy from 2013 onwards.

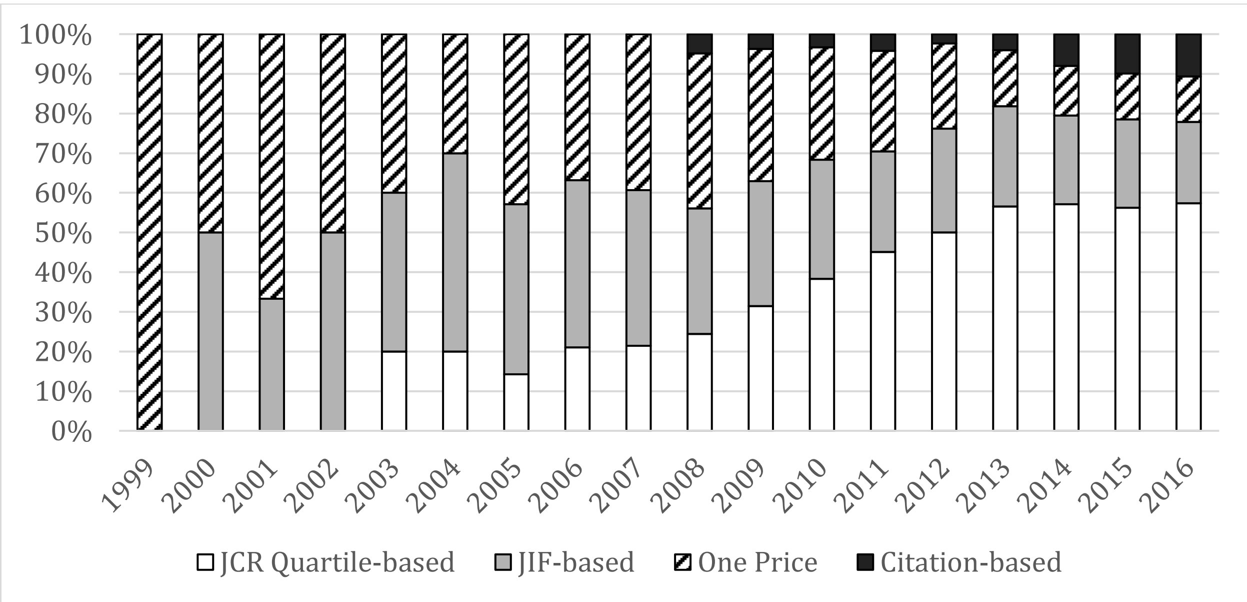

Figure 21. Share of the Cash Reward Policies in Effect by Category (1999-2016)

4.5.2. Authorship

The amount of individual cash rewards per WoS paper varies from 30 USD to 165,000 USD. Not all authors of a paper can claim cash rewards. In 118 out of 168 cash reward policies, universities only award cash to the first author; some universities even require that the awarded author must be both the first author and the corresponding author in 22 out of these 118 policies. Among 25 exceptional policies, universities award cash to non-first authors whose papers were published in particular prestigious journals (e.g., *Nature*, *Science*). Only 13 policies indicate that non-first authors may be awarded for all eligible publications, as they could get a discounted amount (e.g., half for the second author, a quarter for the third, etc.). In addition, there is no specific requirement for authorship in 12 out of 168 policies.

4.5.3. Amount of the Cash Reward

After analyzing 75 policies from 40 Chinese universities[30] that had the cash reward policy in effect between 2008 and 2016, we inferred a cash reward policy by calculating the average cash award for papers published in nine selected journals as described above. We also found that Chinese universities increased the amount of cash reward on average between 2008 and 2016 as shown in Table 22.

1. *Nature*, *Science*: Among most cash reward policies, publishing a paper in these two prestigious journals[31] would receive special treatment. Chinese universities offer the highest cash reward to *Nature* or *Science* papers. The author(s) may receive a prize up to 165,000 USD; some universities even announced that the amount of cash rewarded for a *Nature* or *Science* paper was negotiable. Indeed, the average amount of cash award for a *Nature* or *Science* paper increased 67% from 26,212 USD in 2008 to 43,783 USD in 2016.
2. *PNAS:* Although *Proceedings of National Academy of Sciences* is also a prestigious journal, it is not recognized by Chinese universities for special treatment. However, based

[30] In order to keep the analysis consistent, 60 universities were excluded because their first cash reward policies were issued after 2008.
[31] Some Chinese universities add *Cell* to this list of prestige journals; but most universities only recognize *Nature* and *Science*.

on its JIF and JCR Quartile ranking, the average cash award for a *PNAS* paper is more than 3,000 USD, increasing slightly from 3,156 USD in 2008 to 3,513 USD in 2016.

3. *PLOS ONE:* Although *PLOS ONE* is ranked as a Q1 journal in the original JCR Quartile, it is categorized as a Q3 journal by the modified version of JCR provided by the Chinese Academy of Science. As a result, the amount awarded to a *PLOS ONE* paper is only around 1,000 USD, and this even declined from 1,096 USD in 2008 to 984 USD in 2016.
4. *MIS Quarterly, JASIST*: Both journals are ranked as Q1 journals in their category (Library and Information Science) by JCR. *MIS Quarterly*'s JIF is higher than *JASIST*'s, although the latter is recognized as the top journal in Library and Information Science, and so the average amount of cash awarded for a *MIS Quarterly* article is higher than that for a *JASIST* paper. From 2008 to 2016, the average amount of cash awarded for a *MIS Quarterly* paper slightly increased from 2,613 USD (2008) to 2,938 USD (2016), while the average amount of cash award to a *JASIST* paper increased 43% from 1,737 USD in 2008 to 2,488 USD in 2016.
5. *Journal of Documentation*: As a journal ranked Q2 by the JCR, the average amount cash award to a paper published in *Journal of Documentation* is over 1,000 USD. The average amount increased from 1,082 USD in 2008 to 1,482 in 2016.
6. *Library Hi Tech, LIBRI*: Although both journals are indexed by WoS, they are respectively ranked as Q3 and Q4 journals in JCR because of their low JIF. These rankings are reflected in the cash awards: the average amount of cash awards to papers published in these two journals is below 800 USD. The average amount to LIBRI paper even decreased from 650 USD in 2008 to 484 USD in 2016.

Table 22. Comparison of Average Amount of Cash Awards for a Paper Published in Selected Journals (2008-2016)*

	2008	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Nature, Science	$26,212	$26,006	$25,781	$25,365	$33,990	$36,658	$38,908	$43,783	$43,783
PNAS	$3,156	$3,025	$3,353	$3,443	$3,664	$3,619	$3,751	$3,513	$3,513
PLOS One	$1,096	$1,086	$1,035	$994	$991	$915	$941	$984	$984
MIS Quarterly	$2,613	$2,570	$2,553	$2,654	$2,876	$2,861	$2,992	$2,938	$2,938
JASIST	$1,737	$1,758	$1,741	$1,887	$2,066	$2,303	$2,435	$2,488	$2,488
Journal of Documentation	$1,082	$1,087	$1,042	$1,111	$1,167	$1,265	$1,329	$1,408	$1,408
Library Hi Tech	$781	$775	$726	$741	$740	$768	$795	$783	$783
LIBRI	$650	$644	$577	$560	$538	$509	$517	$484	$484

* All the amounts are full amount (in USD) awarded to the first author

In summary, Chinese universities differentiate the amount of cash reward based on the JIF and JCR Quartile of journals in which the awarded papers are published. The average amount of cash award has increased over the past 10 years, except that the amount awarded to papers published in journals with low JIF has decreased. Publications in *Nature* and *Science* are awarded the largest amount of cash reward. This trend is also reflected by the change of five cash reward policies from Zhejiang University as shown in Table 23. The amount of cash awarded for publications in prestigious journals (i.e. *Nature*, *Science*, *PNAS*) increased while the amount for publications in other journals declined. Only the first author could receive cash rewards, except that non-first authors could get a discounted amount when publishing in *Nature* or *Science.*

Table 23. Comparison of Cash Awards in 6 Cash Reward Policies from Zhejiang University*

	2002	2005	2008	2010	2015
Nature, Science	$6,000	$30,000	$30,000	$30,000	$45,000
PNAS	$900	$2,100	$2,250	$1,500	$1,500
PLOS One	$900	$600	$600	$600	$0
MIS Quarterly	$900	$750	$900	$900	$1,050
JASIST	$525	$450	$450	$600	$0
Journal of Documentation	$525	$450	$450	$0	$0
Library Hi Tech	$525	$300	$225	$0	$0
LIBRI	$525	$300	$225	$0	$0
Eligible authorship	1st only	1st only except *Nature, Science*	1st only except *Nature, Science*	1st only except *Nature, Science*	1st only except *Nature, Science*
Type of policies	JIF	JIF	JIF	JIF	JCR Quartile

* All the amounts are full amount (in USD) awarded to the first author

4.5.4. Difference by Tier and Region

We also found that universities in different tiers have different preferences when choosing their cash reward policies. 14 out of 15 citation-based reward policies were issued by Tier 1 and Tier 2 universities, while 60% of the one-price reward policies were issued by Tier 3 universities. Such preferences also differ in universities from different regions. About 90% of universities in developed regions preferred either the original JIF-based reward policies or the JCR Quartile-based reward policies, while 60% of the one-price reward policies were favoured by universities in developing regions.

It was unexpected that Tier 3 universities would like to pay more for publications than Tier 1 and Tier 2 universities, despite having smaller budgets. As Table 24 shows, in 2016, Tier

3 universities paid double or even triple what Tier 1 and 2 universities did for a paper published in some journals. The average amounts of cash reward at Tier 2 universities for papers published in these journals are between the amounts paid by Tier 3 and Tier 1 universities, respectively. However, we did not find any significant difference in the average amount of cash reward among universities from different regions.

Table 24. Average Amount of Cash Awards for a Paper Published in Selected Journals by Tier (2016)*

	Nature, Science	*PNAS*	*PLOS One*	*MIS Quarterly*	*JASIST*	*Journal of Documentation*	*Library Hi Tech*	*LIBRI*
Tier 1	$38,846	$2,704	$401	$1,924	$1,465	$817	$283	$216
Tier 2	$53,823	$4,113	$783	$3,251	$2,695	$1,377	$679	$434
Tier 3	$63,187	$5,488	$1,661	$5,150	$3,902	$2,102	$1,172	$642

* All the amounts are full amount (in USD) awarded to the first author.

4.6. Discussion

Traditionally, the monetary reward incentive is used in business to reward employees with money for excellent job performance (Aguinis, Joo, & Gottfredson, 2013). Chinese universities apply this to awarding their scholars for research performance, thus promoting publication productivity. Considering the low annual salaries of Chinese scholars - the average annual salary of university professors is around 8,600 USD, while the average basic salary of new hired professors is only about 3,100 USD (Altbach, 2012)[32] - the amount of cash-per-publication

[32] In his book, Altbach used the purchasing power parity index instead of the pure salary based on the exchange rate for the comparison of professors' salary among 28 countries. It means that the annual salaries of Chinese scholars here (8,600 USD and 3,100 USD) represent the amount having the same purchase power as 8,600 USD and 3,100 USD in US respectively.

reward is a huge incentive: the reward value for a *JASIST* paper is equal to a single year's salary for a newly hired professor while the cash award for a *Nature* or *Science* article is up to 20 times a university professors' average annual salary. The cash reward policy has been successful as China's international scientific publication has experienced a period of exponential increase in the past 20 years.

On the other hand, the monetary reward incentive also brings some negative effects. Chinese scholars may regard the monetary reward as an extrinsic rather than an intrinsic motivator (Aguinis et al., 2013; Kohn, 1993). In other words, they are driven to publish just for the monetary reward rather than disseminating knowledge and receiving the recognition defined by the reward system of science (Merton, 1973). For example, Professor Gao from Heilongjiang University published 279 papers in a single journal, *Acta Crystallographica Section E,* and received more than half of the total cash rewards given by Heilongjiang University between 2004 and 2009 (Lei & Lai, 2010). In this case, the monetary reward incentive creates a negative goal displacement effect (Frey et al., 2013; Osterloh & Frey, 2014). Prof. Gao's only research focus in these five years was to find new crystal structures in his lab and always report the results of this to the same journal, because he could accomplish the goal of winning the cash bonus in a short term as contrasted with receiving fewer awards by conducting long-term research projects (Lei & Lai, 2010). In addition, academic fraud in China, such as plagiarism, academic dishonesty, ghostwritten papers, fake peer review scandal, and so on also appeared in a growing number of publications (Hvistendahl, 2013). After searching the WoS, we found that the number of paper corrections authored by Chinese scholars increased from 2 in 1996 to 1,234 in 2016, a historic high.

This study also indicates the abusive use of bibliometric indicators in these cash reward policies. Although the Journal Impact Factor (JIF) is widely recognized to be a poor metric for evaluating the quality of individual papers (Archambault & Larivière, 2009; Lozano, Larivière, & Gingras, 2012; Seglen, 1997) , it is used in almost all cash reward policies as the golden rule[33] to assess the value of individual research. In addition, this study reveals that the WoS is the only data source accepted in these cash reward policies except that some universities offer small cash awards to papers indexed by *Engineering Index* (EI). Although Scopus could be an alternative to WoS in bibliometric studies (Norris & Oppenheim, 2007; Torres-Salinas et al., 2009) and indexes more Chinese journals (Mongeon & Paul-Hus, 2016), it is not recognized by Chinese universities. It also means that publications not indexed by WoS, including millions of papers published in Chinese journals, are almost ignored and excluded from the cash reward.

We also found a positive trend among these cash reward policies. The focus of the cash reward policy changed from the quantity to the quality of the international publications when Chinese universities (especially Tier 1 and Tier 2 universities) had published an adequate number of WoS papers. This was why many one-price reward policies have been replaced by JCR Quartile-based or citation-based policies since 2008. In order to promote impact instead of quantity, these universities increased the amount of cash reward for papers published in Q1 and Q2 journals and decreased or stopped payment for papers published in Q3 and Q4 journals. A few universities even abandoned using the JIF and instead used the citation counts as the criterion for evaluating the quality of individual papers in citation-based reward policies. The

[33] Although the JCR Quartile is frequently used in the cash reward policies, the ranking of JCR Quartile is also on the basis of JIF.

hierarchical difference among universities was indicated in this trend change. When Tier 1 and Tier 2 universities reduced or cancelled the cash reward for papers published in journals with low JIF (e.g. Q3 and Q4 journals), some Tier 3 universities increased the amount for the same category papers. Indeed, Tier 3 universities have higher demand than Tier 1 and Tier 2 universities for both the quantity and the quality of international publications, so Tier 3 universities pay more cash for each individual paper.

4.7. Conclusion

In this study, after investigating 168 cash-per-publication reward policies from 100 Chinese universities, we described the landscape of the cash-per-publication reward policy in China and revealed its trends since the late 1990s. Chinese universities apply the monetary reward incentive used in business to promote scientific publication productivity, which lead to a radical increase in China's international scholarly publication. The cash-per-publication reward policy also produces the "Matthew Effect" (Merton, 1968) among university professors, as the amount of cash reward for publications is much higher than professors' annual salaries. Publications bring scholars not only cash rewards but also the possibility of future funding and promotion, which reveals the golden rule of academia in China: Publish or Impoverish.

This study revealed that monetary reward policies had been widely used to promote research productivity; these monetary reward policies might bring some negative effects when improving the research productivity, which was not systematically investigated by previous studies. We still know little about the potential impact of these monetary reward policies on research activities, which should be explored in the future. The landscape presented in this study could form a foundation for future studies that investigates the consequence and determinants of monetary reward policies.

Some limitations exist in this study. Due to limited data availability, we used convenience sampling, which may influence the sample representation as compared with sampling randomly, and the sampling fraction in Tier 3 universities is much lower than that of Tier 1 and Tier 2 universities. Social science and humanities is also not included in this study. We hope that future research could collect more data and overcome such limitations. Although this study presents a landscape of the cash-per-publication reward policies in China, we did not investigate if a correlation between the cash reward policy and the number of publications exists, which could be explored by future research.

5. Conclusion

The objective of this thesis is to compare an international bibliometric database (i.e., WoS) with a Chinese bibliometric database, in terms of authors and their output, to demonstrate the extent of the overlap between the two groups of Chinese scientific elites in both international and Chinese bibliometric databases, and to determine the effect of disciplines. This was achieved in Chapter 2. International bibliometric databases and Chinese bibliometrics databases adopt different classification systems of science, which may influence the results of comparison between WoS and VIP in terms of their authors. The difference between the journal classification of science and the paper classification of science was compared in Chapter 3. Finally, Chapter 4 attempted to interpret the results indicated in Chapter 2 in the context of science policy and explore the reward system of science in China. This conclusion summarises the results, limitations, and implications of the thesis.

5.1. Summary

Paper 1 (Chapter 2) compares Web of Science (WoS) with Chinese Science and Technology Periodical Citation Database (VIP) in terms of the group of scientific elites defined as the most highly productive authors in 115 selected disciplines. After analyzing more than 20 million articles published in national or international journals by Chinese scholars between 2008 and 2015, paper 1 finds that the overlaps between two groups of scientific elites are 10.52%, on average, ranging from 0% to 33.98% across the 115 disciplines; the overlaps in the Natural Sciences including Life Sciences & Biomedicine, Physical Sciences, and Technology are higher than those in the Social Sciences and Humanities. The results suggest that WoS does not accurately represent Chinese research activities except for some disciplines, in which, Chinese scholars publish few papers in national journals compared to the large number of WoS papers. Thus, WoS cannot be accurately used to evaluate Chinese research output in Social Sciences and Humanities. On the other hand, using a combination of WoS and national Chinese bibliometric databases could be more appropriate to evaluate research output in Natural Sciences.

Paper 2 (Chapter 3) investigates the difference between the journal classification system and the paper classification system for the same data set, using the same classification scheme. It reveals the extent of paper misclassification, which may have influenced the results returned by paper 1. Based on more than 1.8 million journal articles indexed by the Chinese Science Citation Database (CSCD), paper 2 shows that almost half of papers are misclassified in the selected dataset; the findings raise questions concerning the accuracy of the existing journal classification systems of science, which offers the foundation for evaluative bibliometrics and interdisciplinary research.

Paper 3 (Chapter 4) reveals and presents the landscape of the cash-per-publication reward policy in China, and its evolution since the later 1990s. After investigating 168 university documents regarding the cash-per-publication reward policy at 100 Chinese universities, findings indicate that Chinese universities apply monetary reward incentives to promote scientific publication productivity by offering cash rewards ranging from $30 to $165,000 USD for papers published in journals indexed by Web of Science (WoS). Paper 3 also finds that the cash-per-publication reward policy differs between elite universities and non-elite universities, which could be used to explain the different publication patterns between elite universities and non-elite universities, which are found in Chapter 2.

In summary, this thesis presents three original research papers that advance knowledge in bibliometrics, scholarly communication, and science policy. Paper 1 is the first study comparing an international bibliometric database (i.e. WoS) and a national Chinese bibliometric database (i.e., VIP) in terms of most productive authors across all disciplines; it shows that WoS cannot adequately describe China's science production in almost all disciplines. Paper 2 is the first study comparing the classification system of science between the journal level and the paper level from the same data set using the same classification scheme; it shows that journal level classification, such as the one used by WoS, misrepresents the topics of scientific literature. Paper 3 is the first study that describes the landscape of the cash-per-publication reward policy in China; it shows that the drastic increase in China's research production coincides with the increase in cash-per-publication.

5.2. Limitations

The work presented has limitations grouped in three categories: research design, data, and findings.

5.2.1. Research design

In paper 1, the Chinese Science and Technology Periodical Citation Database (VIP) is selected to represent the Chinese bibliometric database because of its broad coverage of scholarly disciplines. However, VIP also indexes some non-academic sources while WoS only indexes the top academic journals in each discipline. The difference in terms of coverage between WoS and VIP is a limitation to this paper. In addition, Chinese scientific elites are defined as the most productive authors in terms of number of publications, but the impact of their research (i.e. citations) is not considered.

5.2.2. Data

Paper 1 develops a method for name disambiguation using a combination of an author's full name and her/his primary affiliated institution. This approach cannot disambiguate scholars who are affiliated to multiple institutions or those who may have changed their primary institution between 2008 and 2015; this accounts for approximately 3% of Chinese scholars in the dataset, which is unlikely to have a significant effect on overall results. In addition, paper 2 assumes that the CLC code of a paper represents its major research area (discipline) since it is chosen by its author(s); however, scholars are not experts in classification or topical indexing, which may produce some indexing inaccuracies and inconsistencies in the selected CLC codes. Meanwhile, the CLC code of a journal is included in its Chinese Serial Number (CN), which is assigned by the State Administration of Press, Publication, Radio, Film and Television of The People's Republic of China (SAPPRFT) when the journal is launched and registered with SAPPRFT. This one-time initial classification of a journal is generally permanent and may not reflect new or updated journal scope through time. Finally, in paper 3, due to the limited data availability a convenience sampling was used, and this may not be as reliable as a random sample.

5.2.3. Findings

Paper 2 reveals that almost half of the papers are misclassified in the journal classification system using the CSCD data. This finding may not be generalized to other bibliometric databases (e.g., WoS, Scopus, etc.) but it certainly suggests there will be significant classification differences between paper versus journal level topical indexing of individual articles. Finally, paper 3 only investigates China's monetary reward policies in Natural Sciences and it therefore excludes the Social sciences and Humanities.

5.3. Future Research

This thesis offers three implications concerning research evaluation, classification of science, and science policies. The most notable implications for the future research is to investigate the impact of metrics-based science policies, which focus on outcomes such as number of publications and rankings, on research output, either in China or worldwide.

This thesis shows differences in publishing behaviour among Chinese scholars: those from elite universities prefer publishing in international journals indexed by WoS, while those from non-elite universities publish more papers in Chinese journals. This finding suggests new compelling questions: 1) Why do Chinese scholars from non-elite universities prefer to publish in Chinese journals even though international publications are promoted by China's metrics-driven science policies? 2) Are different publishing preferences encouraged by China's metrics-driven science policies? 3) Do China's metrics-driven science policies treat elite universities differently than the rest? Future research should use bibliometric techniques to analyse and describe the relationship between China's science policies and Chinese scholars' publishing behaviour. This is uncharted research grounds that investigate the relationship between metrics-based science policies and research output. This new research direction should answer the following research questions:

1. In the past 20 years, is there a significant difference between elite and non-elite Chinese universities in research funding received?
2. Is there a significant difference between elite and non-elite Chinese universities in research output in terms of the number of publications?
3. Is there a significant difference between elite and non-elite Chinese universities in return of investment (ROI) of research?

Answers to these questions will reveal to what extent China's metrics-driven science policies influence Chinese universities' research performance; in other words, it will demonstrate if China's metrics-driven science policies produce the "Matthew Effect" that starves the non-elite universities to feed the elite universities, and therefore increases inequality and disparity in China's higher education system.

In addition, the proposed study advances knowledge in science policy evaluation by exploring the impacts of metrics-driven science policies on the higher education system. This is notable given that previous studies are generally conducted in the European context: this study broadens knowledge to include China, which is now the world's second most productive country in terms of the number of scientific articles published, and provides a knowledge base to investigate the relationship between science policies and research output in higher education institutions across the world.

Bibliography

Aguinis, H., Joo, H., & Gottfredson, R. K. (2013). What monetary rewards can and cannot do: How to show employees the money. *Business Horizons, 56*(2), 241-249.

Altbach, P. G. (2012). *Paying the professoriate: A global comparison of compensation and contracts*: Routledge.

Archambault, É., Beauchesne, O. H., & Caruso, J. (2011). *Towards a multilingual, comprehensive and open scientific journal ontology.* Paper presented at the 13th international conference of the international society for scientometrics and informetrics, Durban, South Africa.

Archambault, É., & Larivière, V. (2009). History of the journal impact factor: Contingencies and consequences. *Scientometrics, 79*(3), 635-649.

Basu, A. (2010). Does a countrys scientific productivity depend critically on the number of country journals indexed? *Scientometrics, 82*(3), 507-516.

Blume, S. S., & Sinclair, R. (1973). Chemists in British universities: A study of the reward system in science. *American Sociological Review, 38*(1), 126-138.

Börner, K., Klavans, R., Patek, M., Zoss, A. M., Biberstine, J. R., Light, R. P., . . . Boyack, K. W. (2012). Design and Update of a Classification System: The UCSD Map of Science. *PLoS ONE, 7*(7), e39464. doi:10.1371/journal.pone.0039464

Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology, 61*(12), 2389-2404.

Boyack, K. W., Newman, D., Duhon, R. J., Klavans, R., Patek, M., Biberstine, J. R., . . . Börner, K. (2011). Clustering more than two million biomedical publications: comparing the accuracies of nine text-based similarity approaches. *PloS one, 6*(3), e18029.

Cao, C., Li, N., Li, X., & Liu, L. (2013). Reforming China's S&T System. *Science, 341*(6145), 460.

Cargill, M., Oconnor, P., & Li, Y. (2012). Educating Chinese scientists to write for international journals: Addressing the divide between science and technology education and English language teaching. *English for Specific Purposes, 31*(1), 60-69.

Chen, X. (2006). Li xiang dao xiang xing de zheng ce zhi ding '985 zheng ce' guo cheng fen xi (Ideal-oriented policy making: An analysis on the process of Project 985). *Peking University Education Review, 4*(1), 145-157.

Cheng, Y., & Liu, N. (2008). Tracking the development of China's top universities using scientometric indicators, 1997-2005. *Frontiers of Education in China, 3*(3), 415-428.

Chi, P.-S. (2016). Differing disciplinary citation concentration patterns of book and journal literature? *Journal of Informetrics, 10*(3), 814-829. doi:10.1016/j.joi.2016.05.005

Cole, S. (1975). The growth of scientific knowledge: Theories of deviance as a case study. In L. A. Coser (Ed.), *The idea of social structure: Papers in honor of Robert K. Merton* (pp. 175-220). New York: Harcourt Brace Jovanovich.

Cole, S., & Cole, J. R. (1967). Scientific output and recognition: A study in the operation of the reward system in science. *American Sociological Review, 32*(3), 377-390.

Cronin, B., & Weaver-Wozniak, S. (1993). *Online Access to Acknowledgments.* Paper presented at the National Online Meeting.

De Bellis, N. (2009). *Bibliometrics and citation analysis: from the Science citation index to cybermetrics*. Lanham, Md.: Scarecrow Press.

Delanty, G. (2005). *Social science*: McGraw-Hill Education (UK).

Desrochers, N., Bowman, T. D., Haustein, S., Mongeon, P., Quan-Haase, A., Paul-Hus, A., . . . Tsou, A. (2015). *Authorship, patents, citations, acknowledgments, tweets, reader counts and the multifaceted reward system of science.* Paper presented at the Proceedings of the Association for Information Science and Technology, St. Louis, MO.

Ding, Z., Zheng, X., & Wu, X. (2012). Strategies for Expanding the International Influences of Academic Journals: An Example from Chinese Pharmaceutical Journals. *Serials Review, 38*(2), 80-85.

Franzoni, C., Scellato, G., & Stephan, P. (2011). Science policy. Changing incentives to publish. *Science, 333*(6043), 702-703.

Frey, B. S., Osterloh, M., & Homberg, F. (2013). Organizational Control Systems and Pay-for-Performance in the Public Service. *Organization Studies, 34*(7), 949-972.

Fu, H., Chuang, K., Wang, M., & Ho, Y. (2011). Characteristics of Research in China Assessed with Essential Science Indicators. *Scientometrics, 88*(3), 841-862.

Fu, H., & Ho, Y. (2013). Comparison of independent research of China's top universities using bibliometric indicators. *Scientometrics, 96*(1), 259-276. doi:10.1007/s11192-012-0912-5

Gao, X., & Guan, J. (2009). A scale-independent analysis of the performance of the Chinese innovation system. *Journal of Informetrics, 3*(4), 321-331.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science, 178*(4060), 471-479.

Garfield, E. (1979). *Citation indexing - its theory and application in science, technology, and humanities*. New York: Wiley.

Garfield, E., Malin, M. V., & Small, H. (1978). Citation data as science indicators. In Y. Elkana, J. Lederberg, R. K. Merton, A. Thackray, & H. Zuckerman (Eds.), *Toward a Metric of Science: The Advent of Science Indicators*. New York: John Wiley & Sons.

Ge, M. (2015). English Writing for International Publication in the Age of Globalization: Practices and Perceptions of Mainland Chinese Academics in the Humanities and Social Sciences. *Publications, 3*(2), 43-64. doi:10.3390/publications3020043

General Administration of Press and Publication of China. (1989). *Guanyu Pizhun Zhongguobiaozhunkanhao Guojia Biaozhunhan de Tongzhi (Notice of the Approval of Chinese Serial Number)*. Beijing: General Administration of Press and Publication of China Retrieved from http://www.people.com.cn/electric/flfg/d2/890111.html.

Gilbert, N. (1977). Referencing as Persuasion. *Social Studies of Science, 7*(1), 113-122.

Glänzel, W. (2003). *Bibliometrics as a research field: a course on theory and application of bibliometric indicators*. Universidade Federal de Pernambuco. Retrieved from http://yunus.hacettepe.edu.tr/~tonta/courses/spring2011/bby704/Bib_Module_KUL.pdf

Glänzel, W., & Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics, 56*(3), 357-367. doi:10.1023/a:1022378804087

Gong, F., & Qu, M. (2010). Nan jing da xue ge an: SCI ying ru ping jia ti xi dui zhong guo da lu ji chu yan jiu de ying xiang (A case study of Nanjing University: The influence of introducing SCI into assessment system on the quality of basic research in Mainland Chinese universities). *Higher Education of Sciences, 2010*(3), 4-17.

Griffith, B. C., Small, H. G., Stonehill, J. A., & Dey, S. (1974). The structure of scientific literatures II: Toward a macro-and microstructure for science. *Science studies, 4*(4), 339-365.

Guan, J., & He, Y. (2005). Comparison and evaluation of domestic and international outputs in Information Science & Technology research of China. *Scientometrics, 65*(2), 215-244.

He, T. (2009). International scientific collaboration of China with the G7 countries. *Scientometrics, 80*(3), 571-582.

Hennemann, S., Wang, T., & Liefner, I. (2011). Measuring regional science networks in China: a comparison of international and domestic bibliographic data sources. *Scientometrics, 88*(2), 535-554.

Hicks, D., & Wang, J. (2009). *Towards a bibliometric database for the social sciences and humanities*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.211.1429&rep=rep1&type=pdf

Hu, Z. G., Guo, F. Q., & Hou, H. Y. (2017). Mapping research spotlights for different regions in China. *Scientometrics, 110*(2), 779-790. doi:10.1007/s11192-016-2175-z

Huang, M.-C., Fang, S.-C., & Chang, S.-C. (2011). Tracking R&D behavior: bibliometric analysis of drug patents in the Orange Book. *Scientometrics, 88*(3), 805. doi:10.1007/s11192-011-0400-3

Hvistendahl, M. (2013). China's publication bazaar. *Science, 342*(6162), 1035-1039.

ISTIC. (2010). *Statistical Data of Chinese S&T Papers 2010*. Beijing: ISTIC.

ISTIC. (2014). *Statistical Data of Chinese S&T Papers 2014*. Beijing: ISTIC.

ISTIC. (2017). *Statistical Data of Chinese S&T Papers 2017*. Beijing: ISTIC.

Janssens, F., Zhang, L., De Moor, B., & Glänzel, W. (2009). Hybrid clustering for validation and improvement of subject-classification schemes. *Information Processing & Management, 45*(6), 683-702.

Javitz, H., Grimes, T., Hill, D., Rapoport, A., Bell, R., Fecso, R., & Lehming, R. (2010). *U.S. Academic Scientific Publishing*. Arlington, VA: National Science Foundation, Division of Science Resources Statistics.

Jin, B., & Rousseau, R. (2004). Evaluation of research performance and scientometric indicators in China. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research : the use of publication and patent statistics in studies of S&T systems* (pp. 497-514). Dordrecht; London: Kluwer Academic Publishers.

Jin, B., & Rousseau, R. (2005). *China's quantitative expansion phase: exponential growth but low impact.* Paper presented at the the 10th International Conference on Scientometrics and Informetrics, Stockholm.

Jin, B., & Wang, B. (1999). Chinese Science Citation Database: its construction and application. *Scientometrics, 45*(2), 325-332.

Jin, B., Wang, S., Wang, B., Rousseau, R., Wu, Z., Liu, X., & Zhu, X. (1999). A unified method of counting international and domestic articles. *Journal of Management Sciences in China, 2*(3), 59-65.

Jin, B., Zhang, J., Chen, D., & Zhu, X. (2002). Development of the Chinese Scientometric Indicators (CSI). *Scientometrics, 54*(1), 145-154.

Keathley-Herring, H., Van Aken, E., Gonzalez-Aleu, F., Deschamps, F., Letens, G., & Orlandini, P. C. (2016). Assessing the maturity of a research area: bibliometric review and proposed framework. *Scientometrics, 109*(2), 927-951. doi:10.1007/s11192-016-2096-x

Klavans, R., & Boyack, K. W. (2009). Toward a Consensus Map of Science. *Journal of the American Society for Information Science and Technology, 60*(3), 455-476.

Klavans, R., & Boyack, K. W. (2010). Toward an objective, reliable and accurate method for measuring research leadership. *Scientometrics, 82*(3), 539-553. doi:10.1007/s11192-010-0188-6

Klavans, R., & Boyack, K. W. (2017). Which Type of Citation Analysis Generates the Most Accurate Taxonomy of Scientific and Technical Knowledge? *Journal of the Association for Information Science and Technology, 68*(4), 984-998.

Kohn, A. (1993). Why incentive plans cannot work. *Harvard business review, 71*(5), 54.

Kostoff, R. N., Briggs, M. B., Rushenberg, R. L., Bowles, C. A., Icenhour, A. S., Nikodym, K. F., . . . Pecht, M. (2007). Chinese science and technology - Structure and infrastructure. *Technological Forecasting and Social Change, 74*(9), 1539-1573. doi:10.1016/j.techfore.2007.02.008

Kostoff, R. N., Briggs, M. E., Rushenberg, R. L., Eowles, C. A., Bhattacharya, S., Johnson, D., . . . Pecht, M. (2007). Assessment of science and technology literature of China and India as reflected in the SCI/SSCI. *Current science., 93*(8), 1088-1092.

Kuhn, T. S. (2012). *The structure of scientific revolutions*: University of Chicago press.

Larivière, V., Archambault, É., Gingras, Y., & Vignola-Gagnè, É. (2006). The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities. *Journal of the American Society for Information Science and Technology, 57*(8), 997-1004.

Larivière, V., Macaluso, B., Archambault, É., & Gingras, Y. (2010). Which scientific elites? On the concentration of research funds, publications and citations. *RESEARCH EVALUATION, 19*(1), 45-53.

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology, 64*(1), 2-17.

Lei, Y., & Lai, Y. (2010, Jan 4, 2010). Hei-long-jiang da-xue yi jiao shou nian jian Acta Crystallographica fa lun wen 279 pian (A professor of Heilongjiang University published 279 papers in Acta Crystallographica). *China Youth Daily*. Retrieved from http://tech.sina.com.cn/d/2010-01-04/07293733450.shtml

Leydesdorff, L. (2012). World Shares of Publications of the USA, EU-27, and China Compared and Predicted Using the New Web of Science Interface Versus Scopus. *Profesional de la Informacion, 21*(1), 43-49.

Leydesdorff, L., & Bornmann, L. (2016). The operationalization of "fields" as WoS subject categories (WCs) in evaluative bibliometrics: The cases of "library and information science" and "science & technology studies". *Journal of the Association for Information Science and Technology, 67*(3), 707-714. doi:10.1002/asi.23408

Leydesdorff, L., & Jin, B. (2005). Mapping the Chinese Science Citation Database in terms of aggregated journal–journal citation relations. *Journal of the American Society for Information Science and Technology, 56*(14), 1469-1479.

Leydesdorff, L., & Rafols, I. (2009). A Global Map of Science Based on the ISI Subject Categories. *Journal of the American Society for Information Science and Technology, 60*(2), 348-362.

Leydesdorff, L., & Zhou, P. (2005). Are the contributions of China and Korea upsetting the world system of science? *Scientometrics, 63*(3), 617-630.

Li, J. (2013). Ma tai xiao ying xia de gao xiao jiao shi xin chou ji li ce lue chong gou (Restructure the incentive policies for university professors in the context of Matthew Effect). *Journal of Henan Normal University (Philosophy and Social Science Edition), 2013*(5), 163-166.

Li, J., & Li, Y. T. (2015). Patterns and evolution of coauthorship in China's humanities and social sciences. *Scientometrics, 102*(3), 1997-2010. doi:10.1007/s11192-014-1471-8

Li, Z., & Zhang, H. (2008). Cong jin liu nian ke yan lun wen chan chu tan tao bei jing da xue lin cahung zhong liu xue yuan lun wen jiang li zhi du (The discussion on the cash-per-publication policy of Beijing University Cancer Hospital based on the publication productivity in the past six years). *Chinese Journal of Medical Science Research Management, 21*(1), 38-39.

Li, Z., & Zhong, R. (2013). Jiang jin ji li yu xue shu cheng guo, lai zi Zhe-Jiang da xue SCI lun wen jiang li cheng guo de zheng ju (Incentive reward and research performance: A proof from the cash-per-publication policy of Zhejiang University). *Journal of Industrial Engineering/Engineering Management, 27*(2), 220-226.

Liang, L. (2003). Evaluating China's research performance: how do SCI and Chinese indexes compare? *Interdisciplinary Science Reviews, 28*(1), 38-43.

Liang, L., Havemann, F., Heinz, M., & Wagner-Döbler, R. (2006). Structural similarities between science growth dynamics in China and in western countries. *Scientometrics, 66*(2), 311-325.

Liang, L., Wu, Y., & Li, J. (2001). Selection of databases, indicators and models for evaluating research performance of Chinese universities. *RESEARCH EVALUATION, 10*, 105-114.

Liang, L., & Zhu, L. (2002). Major factors affecting Chinas inter-regional research collaboration: Regional scientific productivity and geographical proximity. *Scientometrics, 55*(2), 287-316.

Liu, W. S., Tang, L., Gu, M. D., & Hu, G. Y. (2015). Feature report on China: a bibliometric analysis of China-related articles. *Scientometrics, 102*(1), 503-517. doi:10.1007/s11192-014-1371-y

Lozano, G. A., Larivière, V., & Gingras, Y. (2012). The weakening relationship between the impact factor and papers' citations in the digital age. *Journal of the American Society for Information Science and Technology, 63*(11), 2140-2145.

Martin, B., Tang, P., Morgan, M., Glanzel, W., Hornbostel, S., Lauer, G., . . . Zic-Fuchs, M. (2010). *Towards a bibliometric database for the Social Sciences and Humanities: A European scoping project*. Retrieved from https://globalhighered.files.wordpress.com/2010/07/esf_report_final_100309.pdf

Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google scholar. *Journal of the American Society for Information Science and Technology, 58*(13), 2105-2125.

Melkers, J. (1993). Bibliometrics as a tool for analysis of R&D impacts. In B. Bozeman & J. Melkers (Eds.), *Evaluating R&D impacts : methods and practice*. New York: Springer Science+ Business Media.

Mely, B., El Kader, M. A., Dudognon, G., & Okubo, Y. (1998). Scientific publications of China in 1994: evolution or revolution? *Scientometrics, 42*(1), 3-16.

Meng, W., Hu, Z., & Liu, W. (2006). Efficiency evaluation of basic research in China. *Scientometrics, 69*(1), 85-101.

Merton, R. K. (1957). Priorities in scientific discovery: A chapter in the sociology of science. *American Sociological Review, 22*(6), 635-659.

Merton, R. K. (1968). The Matthew Effect in Science: The reward and communication systems of science are considered. *Science, 159*(3810), 56-63.

Merton, R. K. (1973). *The sociology of science: theoretical and empirical investigations*. Chicago: University of Chicago Press.

Ministry of Education of China. (2000). *Introduction to the Project 211*. Beijing Retrieved from http://www.moe.edu.cn/publicfiles/business/htmlfiles/moe/moe_846/200804/33122.html.

Ministry of Education of China. (2003-2016). *Scientific Statistics in Higher Education Institutions*. Beijing: Higher Education Press.

Ministry of Education of China. (2014). *Scientific Statistics in Higher Education Institutions - 2014*. Beijing: Higher Education Press.

Ministry of Education of China. (2015). *Scientific Statistics in Higher Education Institutions - 2015*. Beijing: Higher Education Press.

Ministry of Education of China. (2016). *List of higher education institutions*. Beijing Retrieved from http://www.moe.gov.cn/srcsite/A03/moe_634/201606/t20160603_248263.html.

Moed, H. (1996). Differences in the construction of SCI based bibliometric indicators among various producers: A first over view. *Scientometrics, 35*(2), 177-191.

Moed, H. (2002a). The impact-factors debate: the ISI's uses and limits. *Nature, 415*(6873), 731-732.

Moed, H. (2002b). Measuring China's research performance using the Science Citation Index. *Scientometrics, 53*(3), 281-296.

Moed, H. (2005). *Citation analysis in research evaluation*. Dordrecht, UK: Springer.

Mohrman, K. (2005). World-class universities and Chinese higher education reform. *International Higher Education, 2005*(39), 22-23.

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics, 106*(1), 213-228.

Montgomery, S. L. (2013). *Does science need a global language? English and the future of research*. Chicago, IL: The University of Chicago Press.

National Bureau of Statistics of China. (1996-2016). *China Statistical Yearbook on Sciencce and Technology*. Beijing: China Statistics Press.

National Bureau of Statistics of China. (2015). *China Statistical Yearbook*. Beijing: China Statistics Press.

National Science Board. (2018). *Science and Engineering Indicators 2018*. (NSB-2018-1). Alexandria, VA: National Science Foundation Retrieved from https://www.nsf.gov/statistics/indicators/.

National Science Library of Chinese Academy of Science. (2015). *CSCD JCR Annual Report*. Beijing: Intellectual Property Publishing House Co.. Ltd.

National Science Library of Chinese Academy of Science. (2017). *Zhongguokexueyingwenshujuku (CSCD) Laiyuan Qikan Linxuan Baogao: 2017-2018 (Report of CSCD Journal Selection: 2017-2018)*. Retrieved from http://sciencechina.cn/style/report17_18.pdf

Nicholas, D., & Ritchie, M. (1978). *Literature and bibliometrics*. London; Hamden, Conn.: C. Bingley ; Linnet Books.

Niu, F., & Qiu, J. (2014). Network structure, distribution and the growth of Chinese international research collaboration. *Scientometrics, 98*(2), 1221-1233.

Norris, M., & Oppenheim, C. (2007). Comparing alternatives to the Web of Science for coverage of the social sciences' literature. *Journal of Informetrics, 1*(2), 161-169.

OECD. (2002). *The measurement of scientific and technological activities proposed standard practice for surveys on research and experimental development: Frascati manual 2002*. Paris: Organisation for Economic Co-operation and Development.

Okubo, Y. (1997). *Bibliometric Indicators and Analysis of Research Systems Methods and Examples*. Paris: OECD Publishing.

Osterloh, M., & Frey, B. S. (2014). Ranking Games. *Evaluation Review, 39*(1), 102-129. doi:10.1177/0193841X14524957

Otlet, P. (1934). *Traité de documentation: le livre sur le livre, théorie et pratique*. Bruxelles, Belgium: Editiones mundaneum.

Park, H. W., Yoon, J., & Leydesdorff, L. (2016). The normalization of co-authorship networks in the bibliometric evaluation: the government stimulation programs of China and Korea. *Scientometrics, 109*(2), 1017-1036. doi:10.1007/s11192-016-1978-2

Peng, C. (2011). Focus on quality, not just quantity. *Nature, 475*(7356), 267.

Porter, A. L., & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics, 81*(3), 719. doi:10.1007/s11192-008-2197-2

Porter, A. L., Roessner, D. J., & Heberger, A. E. (2008). How interdisciplinary is a given body of research? *RESEARCH EVALUATION, 17*(4), 273-282.

Price, D. J. d. S. (1963). *Little science, big science*. New York: Columbia Univ. Press.

Price, D. J. d. S. (1965). The science of science. *Bulletin of the Atomic Scientists, 21*(8), 2-8.

Pudovkin, A. I., & Garfield, E. (2002). Algorithmic procedure for finding semantically related journals. *Journal of the American Society for Information Science and Technology, 53*(13), 1113-1119. doi:10.1002/asi.10153

Qi, Y. (2009). Ma tai xiao ying dui ke ji jiang li de ying xiang (The impact of Matthew Effect on scientific incentive). *Journal of Lanzhou Jiaotong University, 28*(5), 167-169.

Qiu, J. (2010). Publish or perish in China. *Nature, 463*(7278), 142-143.

Qiu, J. P., Yang, R., & Zhao, R. (2010). Competition and Excellence: Ranking of World-class Universities 2009 and Advance of Chinese Universities. *Journal of Library and Information Studies, 8*(2), 11-27.

Quan, W., Chen, B., & Shu, F. (2017). Publish or impoverish: An investigation of the monetary reward system of science in China (1999-2016). *Aslib Journal of Information Management, 69*(5), 1-18.

Ren, S., & Rousseau, R. (2002). International visibility of Chinese scientific journals. *Scientometrics, 53*(3), 389-405.

Rousseau, R. (2015). The tip of the Chinese publication iceberg. *ISSI newsletter, 11*(4), 100-102.

Rousseau, R., Jin, B., & Yang, N. (2001). Observations concerning the two- and three-year synchronous impact factor, based on the Chinese science citation database. *Journal of Documentation, 57*(3), 349-357.

Royle, J., Coles, L., Williams, D., & Evans, P. (2007). Publishing in international journals: An examination of trends in chinese co-authorship. *Scientometrics, 71*(1), 59-86.

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ: British Medical Journal, 314*(7079), 498.

Shan, Z., Han, Z., & Zhao, J. (2013). Xin-jiang yi-ke da-xue di yi fu shu yi yuan ke yan lun wen jiang li zhi du gai ge qian hou lun wen chan chu de dui bi fen xi (A comparison of the number of publication before and after the reform of cash-per-publication policy in Xinjiang Medical University first affliated hospital). *Beijing Medical Journal, 35*(2), 147-149.

Shelton, R. D., Foland, P., & Gorelskyy, R. (2009). Do New SCI Journals Have a Different National Bias? *Scientometrics, 79*(2), 351-363.

Shen, S. X. (2016). Negotiating Authorship in Chinese Universities : How Organizations Shape Cycles of Credit in Science. *Science, Technology, & Human Values, 41*(4), 660-685.

Shu, F., Dinneen, J. D., Asadi, B., & Julien, C.-A. (2017). Mapping science using Library of Congress Subject Headings. *Journal of Informetrics, 11*(4), 1080-1094.

Shu, F., & Larivière, V. (2015). Chinese-language articles are biased in citations. *Journal of Informetrics, 9*(3), 526-528. doi:10.1016/j.joi.2015.05.005

Shu, F., Larivière, V., & Julien, C. (2016). *National and international scientific elites: an analysis of Chinese scholars.* Paper presented at the 21st International Conference on Science and Technology Indicators-STI 2016. Book of Proceedings.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science, 24*(4), 265-269.

Small, H., & Griffith, B. C. (1974). The structure of scientific literatures I: Identifying and graphing specialties. *Science studies, 4*(1), 17-40.

Song, Y. H., Ma, F., & Yang, S. L. (2015). Comparative study on the obsolescence of humanities and social sciences in China: under the new situation of web. *Scientometrics, 102*(1), 365-388. doi:10.1007/s11192-014-1410-8

Strotmann, A., & Zhao, D. (2012). Author Name Disambiguation: What Difference Does it Make in Author-Based Citation Analysis? *Journal of the American Society for Information Science and Technology, 63*(9), 1820-1833.

Sugimoto, C. R., Russell, T. G., Meho, L. I., & Marchionini, G. (2008). MPACT and citation impact: Two sides of the same scholarly coin? *Library & Information Science Research, 30*(4), 273-281.

Sun, X., & Zhang, J. (2010). Qian yi gao xiao xue shu lun wen jiang li xian zhuang (A discussion on the cash-per-publication policy in universities). *Science and Technology Innovation Herald, 2010*(35), 227.

Sun, Y., & Cao, C. (2014). Demystifying central government R&D spending in China. *Science, 345*(6200), 1006-1008.

Swinbanks, D., Nathan, R., & Triendl, R. (1997). Western research assessment meets Asian cultures. *Nature, 389*(6647), 113-117.

Tang, J., & Yang, C. (2008). Over 10 billion yuan to be invested in "211 Project" [Press release]. Retrieved from http://en.people.cn/90001/6381319.html

Thursby, J. G., Jensen, R., & Thursby, M. C. (2001). Objectives, characteristics and outcomes of university licensing: A survey of major US universities. *The journal of Technology transfer, 26*(1-2), 59-72.

Thursby, J. G., & Thursby, M. C. (2002). Who is selling the ivory tower? Sources of growth in university licensing. *Management Science, 48*(1), 90-104.

Torres-Salinas, D., Lopez-Cózar, E. D., & Jiménez-Contreras, E. (2009). Ranking of departments and researchers within a university using two different databases: Web of Science versus Scopus. *Scientometrics, 80*(3), 761-774.

Tseng, Y.-H., & Tsay, M.-Y. (2013). Journal clustering of library and information science for subfield delineation using the bibliometric analysis toolkit: CATAR. *Scientometrics, 95*(2), 503-528. doi:10.1007/s11192-013-0964-1

Waltman, L., & Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology, 63*(12), 2378-2392.

Wang, L. (2016). *Gao xiao zai guo jia ke ji jiang li zhi du zhong huo jiang xian xiang ji qi ying xiang yin su de yan jiu: yi jiao yu bu zhi shu gao xiao wei shu ju cai ji dui xiang.* (Master of Arts), East China Normal University, Shanghai. Retrieved from https://mall.cnki.net/lunwen-1016146006.nh.html

Wang, L. L. (2016). The structure and comparative advantages of China's scientific research: quantitative and qualitative perspectives. *Scientometrics, 106*(1), 435-452. doi:10.1007/s11192-015-1650-2

Wang, L. L., & Wang, X. W. (2017). Who sets up the bridge? Tracking scientific collaborations between China and the European Union. *RESEARCH EVALUATION, 26*(2), 124-131. doi:10.1093/reseval/rvx009

Wang, S., Wang, H., & Weldon, P. (2007). Bibliometric analysis of English-language academic journals of China and their internationalization. *Scientometrics, 73*(3), 331-343.

Wang, S., & Weldon, P. R. (2006). Chinese academic journals: quality, issues and solutions. *Learned publishing : journal of the Association of Learned and Professional Society Publishers., 19*(2), 97-106.

Wang, W., Wu, Y., & Pan, Y. (2014). An investigation of collaborations between top Chinese universities: a new quantitative approach. *Scientometrics, 98*(2), 1535-1545.

Wang, X., Xu, S., Wang, Z., Peng, L., & Wang, C. (2013). International scientific collaboration of China: collaborating countries, institutions and individuals. *Scientometrics, 95*(3), 885-894.

Wang, Y., & Li, L. (2015). Gao xiao ke yan Ji yao yang wang xing kong geng yao jiao ta shi di (University research needs both the long term goal and the short term objective). *Chinese University Technology Transfer, 2015*(7), 4-7.

Wang, Y., Wu, Y., Pan, Y., Ma, Z., & Rousseau, R. (2005). Scientific collaboration in China as reflected in co-authorship. *Scientometrics, 62*(2), 183-198.

Wang, Z., Li, G., Li, C., & Li, A. (2012). Research on the semantic-based co-word analysis. *Scientometrics, 90*(3), 855-875.

White, B. (2006). Examining the claims of Google Scholar as a serious information source. *New Zealand Library & Information Management Journal, 50*(1), 11-24.

Wu, Y., Pan, Y., Zhang, Y., Ma, Z., Pang, J., Guo, H., . . . Yang, Z. (2004). China Scientific and Technical Papers and Citations (CSTPC): history, impact and outlook. *Scientometrics, 60*(3), 385-397.

Wu, Y., Zhang, Y., Cheng, Y., Ou, B., Su, Z., & Li, Z. (1991). Bibliometric studies in China and ISTIC's approach. *Journal of the Hong Kong Library Association*(15), 161-171.

Yan, E., Ding, Y., & Zhu, Q. (2010). Mapping library and information science in China: a coauthorship network analysis. *Scientometrics, 83*(1), 115-131.

Yang, S., Ma, F., Song, Y., & Qiu, J. (2010). A longitudinal analysis of citation distribution breadth for Chinese scholars. *Scientometrics, 85*(3), 755-765.

Young, H., & Belanger, T. (Eds.). (1983). *The ALA glossary of library and information science*. Chicago: American Library Association.

Yuan, H., Xu, W., & Hu, H. (2013). Young Chinese doctors and the pressure of publication. *Lancet, 381*(9864), e4. doi:http://dx.doi.org/10.1016/S0140-6736(13)60174-9

Zeng, L., An, Z., & Wang, L. (2012). SCI lun wen jiang li zhu du dui gao xiao ke ji chuang xin de cu jin zuo yong, yi dian zi ke ji da xue wei li (The influence of cash-per-publication policy on scientific innovation: A case study of University of Electronic Science and Technology of China). *Journal of UESTC (Social Sciences Edition), 14*(5), 110-112.

Zhang, H., & Guo, H. (1997). Scientific Research Collaboration in China. *Scientometrics., 38*(2), 309.

Zhang, L., Janssens, F., Liang, L., & Glänzel, W. (2010). Journal cross-citation analysis for validation and improvement of journal-based subject classification in bibliometric research. *Scientometrics, 82*(3), 687-706.

Zhang, L., Rousseau, R., & Glänzel, W. (2016). Diversity of references as an indicator of the interdisciplinarity of journals: Taking similarity between subject fields into account. *Journal of the Association for Information Science and Technology, 67*(5), 1257-1265.

Zhao, R.-Y., Lei, j., Ma, R., & Qiu, J. (2008). A comparison among five bibliometric database in terms of citation index. *Information Studies: Theory and Application, 2008*(4), 589-605.

Zheng, J., Zhao, Z., Zhang, X., Chen, D., Huang, M., Lei, X., . . . Zhao, Y. (2012). International scientific and technological collaboration of China from 2004 to 2008: a perspective from paper and patent analysis. *Scientometrics, 91*(1), 65-80.

Zhi, Q., & Meng, T. G. (2016). Funding allocation, inequality, and scientific research output: an empirical study based on the life science sector of Natural Science Foundation of China. *Scientometrics, 106*(2), 603-628. doi:10.1007/s11192-015-1773-5

Zhong, Z., & Chen, X. (2008). Ke ji jiang li zhong de "ma tai xiao ying" ji qi dui ce (Matthew Effect in the scientific incentive). *Journal of Guangdong Institute of Socialism, 2008*(3), 73-77.

Zhongguo Tushuguan Fenleifa [Chinese Library Classification]. (2010). (5 ed.). Beijing: National Library of China Publishing House.

Zhou, P., & Glänzel, W. (2010). In-depth analysis on Chinas international cooperation in science. *Scientometrics, 82*(3), 597-612.

Zhou, P., & Leydesdorff, L. (2006). The emergence of China as a leading nation in science. *Research Policy, 35*(1), 83-104.

Zhou, P., & Leydesdorff, L. (2007). A comparison between the China Scientific and Technical Papers and Citations Database and the Science Citation Index in terms of journal hierarchies and interjournal citation relations. *Journal of the American Society for Information Science and Technology, 58*(2), 223-236.

Zhou, P., Thijs, B., & Glänzel, W. (2009a). Is China also becoming a giant in social sciences? *Scientometrics, 79*(3), 593-621.

Zhou, P., Thijs, B., & Glänzel, W. (2009b). Regional Analysis on Chinese Scientific Output. *Scientometrics, 81*(3), 839-857.

Zhu, J., Hassan, S.-U., Mirza, H. T., & Xie, Q. (2014). Measuring recent research performance for Chinese universities using bibliometric methods (English). *Scientometrics, 101*(1), 429-443.

Zuckerman, H. (1977). *Scientific elite : Nobel laureates in the United States*. New York: Free Press.

Zuckerman, H. (1992). The proliferation of prizes: Nobel complements and Nobel surrogates in the reward system of science. *Theoretical Medicine, 13*(2), 217-231. doi:10.1007/bf02163629

Appendices

Appendix I. List of 115 Disciplines (Named as WoS Categories) Investigated in This Study

Discipline	Domain	Number of VIP Paper	Number of Chinese WoS Paper	Number of WoS Paper	Overlap Rate
Archaeology	Arts and Humanities	60,841	296	17,510	14.96%
Architecture	Arts and Humanities	1,282,473	198	14,886	1.47%
Arts	Arts and Humanities	707,497	233	18,683	0.57%
Film, Radio & Television	Arts and Humanities	103,061	6	7,912	0.00%
Folklore	Arts and Humanities	24,024	2	2,386	0.00%
Language & Linguistics	Arts and Humanities	371,720	1,115	30,471	2.99%
Literary Theory & Criticism	Arts and Humanities	49,982	5	4,684	1.61%
Literature	Arts and Humanities	1,083,364	1,203	26,095	2.73%
Music	Arts and Humanities	178,012	27	13,056	1.53%
Religion	Arts and Humanities	53,125	352	24,704	2.91%
Theater	Arts and Humanities	56,306	12	5,356	0.00%
History*	Social Science, Arts and Humanities	359,834	342	51,483	1.87%
Anthropology	Social Science	9,368	527	25,498	12.93%
Business, Finance	Social Science	587,650	1,858	30,592	1.91%
Communication	Social Science	68,526	677	21,351	4.38%
Demography	Social Science	17,098	159	6,957	11.02%
Education & Educational Research	Social Science	4,223,457	2,141	71,150	2.19%
Education, Special	Social Science	9,807	193	10,415	9.92%
Ethics	Social Science	59,975	302	15,695	0.00%
Ethnic Studies	Social Science	12,469	21	4,840	0.00%
Health Policy & Services	Social Science	315,090	700	37,186	14.44%
Hospitality, Leisure, Sport & Tourism	Social Science	131,945	1,197	14,261	8.93%
International Relations	Social Science	24,756	704	24,353	2.28%
Law	Social Science	481,443	747	35,197	2.27%
Linguistics	Social Science	70,251	1,448	33,511	1.66%
Management	Social Science	70,363	4,143	60,533	10.23%
Nursing	Social Science	446,419	1,258	52,821	18.80%
Political Science	Social Science	1,090,508	352	45,981	0.00%
Psychology	Social Science	208,608	1,270	47,447	11.26%
Psychology, Applied	Social Science	7,043	948	24,938	5.28%
Psychology, Clinical	Social Science	39,476	726	51,799	6.56%
Sociology	Social Science	513,485	794	38,776	2.17%
Acoustics	Natural Science	14,305	4,681	33,853	10.53%
Agricultural Engineering	Natural Science	114,559	5,143	26,330	0.00%
Agriculture, Dairy & Animal Science	Natural Science	232,323	3,917	52,085	26.05%
Agronomy	Natural Science	39,822	7,548	63,396	7.34%
Anatomy & Morphology	Natural Science	34,583	1,027	15,768	8.70%
Anesthesiology	Natural Science	38,024	1,100	30,347	18.75%
Astronomy & Astrophysics	Natural Science	30,644	12,022	139,717	0.95%
Biology	Natural Science	332,325	9,899	112,652	2.88%

Biophysics	Natural Science	5,694	15,673	99,344	1.84%
Cardiac & Cardiovascular Systems	Natural Science	253,165	7,390	142,009	33.98%
Cell Biology	Natural Science	14,288	26,601	198,327	4.76%
Chemistry, Analytical	Natural Science	65,589	39,511	162,528	11.94%
Chemistry, Medicinal	Natural Science	5,161	18,069	103,474	13.51%
Chemistry, Organic	Natural Science	27,613	30,992	164,894	2.79%
Chemistry, Physical	Natural Science	43,667	92,342	385,141	15.00%
Construction & Building Technology	Natural Science	357,718	6,970	43,424	13.95%
Crystallography	Natural Science	6,570	21,102	68,079	2.93%
Dentistry, Oral Surgery & Medicine	Natural Science	80,447	4,534	67,970	19.91%
Dermatology	Natural Science	58,222	2,157	52,110	23.24%
Ecology	Natural Science	7,717	8,051	125,916	11.21%
Electrochemistry	Natural Science	9,206	26,621	94,616	1.00%
Emergency Medicine	Natural Science	23,076	817	24,226	8.37%
Endocrinology & Metabolism	Natural Science	147,325	9,111	126,476	26.42%
Engineering, Aerospace	Natural Science	142,500	3,698	22,053	21.78%
Engineering, Biomedical	Natural Science	36,667	9,921	79,322	11.43%
Engineering, Chemical	Natural Science	603,071	43,147	201,554	4.98%
Engineering, Civil	Natural Science	83,742	17,560	104,356	0.97%
Engineering, Electrical & Electronic	Natural Science	491,942	70,318	359,127	4.88%
Engineering, Environmental	Natural Science	317,078	17,441	81,213	17.39%
Engineering, Geological	Natural Science	17,619	3,514	19,807	10.67%
Engineering, Mechanical	Natural Science	586,926	24,712	119,071	5.74%
Engineering, Petroleum	Natural Science	187,725	3,019	15,296	22.64%
Entomology	Natural Science	13,912	4,671	46,880	26.13%
Environmental Sciences	Natural Science	104,259	44,478	273,654	15.53%
Fisheries	Natural Science	106,487	3,975	37,800	15.61%
Food Science & Technology	Natural Science	286,279	20,159	154,335	23.41%
Forestry	Natural Science	179,587	2,361	35,370	9.52%
Genetics & Heredity	Natural Science	11,777	19,456	150,383	9.26%
Geography, Physical	Natural Science	17,219	4,784	36,195	9.05%
Geology	Natural Science	178,065	3,993	19,561	14.88%
Geriatrics & Gerontology	Natural Science	3,381	1,826	35,573	6.82%
Hematology	Natural Science	21,523	4,976	85,876	16.59%
Horticulture	Natural Science	296,112	3,545	26,362	8.62%
Immunology	Natural Science	28,681	15,274	168,214	11.16%
Instruments & Instrumentation	Natural Science	60,216	18,377	103,010	0.93%
Mathematics	Natural Science	220,520	34,142	189,692	3.79%
Mathematics, Applied	Natural Science	9,311	47,499	188,452	0.00%
Mechanics	Natural Science	45,435	24,187	131,955	11.37%
Meteorology & Atmospheric Sciences	Natural Science	74,917	12,552	87,958	22.97%
Microbiology	Natural Science	18,955	13,802	149,520	13.33%
Mineralogy	Natural Science	5,084	2,653	19,115	13.40%
Nuclear Science & Technology	Natural Science	21,896	6,869	72,051	23.96%
Obstetrics & Gynecology	Natural Science	206,861	4,732	88,736	22.64%
Oceanography	Natural Science	35,647	5,470	46,960	22.22%
Oncology	Natural Science	439,499	42,484	264,782	14.63%

Operations Research & Management Science	Natural Science	20,382	11,303	61,132	5.50%
Ophthalmology	Natural Science	71,419	5,247	66,322	28.57%
Optics	Natural Science	26,339	49,038	187,128	6.90%
Orthopedics	Natural Science	15,362	4,423	81,182	0.00%
Otorhinolaryngology	Natural Science	55,025	1,911	43,136	33.49%
Paleontology	Natural Science	5,966	1,602	19,843	26.96%
Parasitology	Natural Science	8,605	2,629	41,367	22.97%
Pathology	Natural Science	18,972	7,104	63,964	2.67%
Pediatrics	Natural Science	111,694	3,390	118,406	17.72%
Pharmacology & Pharmacy	Natural Science	321,875	36,836	278,616	15.46%
Physics, Atomic, Molecular & Chemical	Natural Science	7,009	18,154	128,245	9.17%
Physics, Condensed Matter	Natural Science	1,667	43,319	214,623	0.00%
Physiology	Natural Science	16,931	6,109	82,223	3.01%
Plant Sciences	Natural Science	59,040	25,831	156,607	3.88%
Radiology, Nuclear Medicine & Medical Imaging	Natural Science	90,285	9,535	142,693	23.58%
Rehabilitation	Natural Science	14,862	1,220	53,165	10.43%
Remote Sensing	Natural Science	10,004	5,871	25,553	9.00%
Respiratory System	Natural Science	101,222	3,831	62,561	20.28%
Soil Science	Natural Science	52,090	5,013	30,875	18.18%
Sport Sciences	Natural Science	405,058	1,403	63,092	2.79%
Statistics & Probability	Natural Science	33,284	6,946	64,472	12.15%
Surgery	Natural Science	619,556	15,559	257,887	12.68%
Telecommunications	Natural Science	673,943	19,766	90,328	4.98%
Toxicology	Natural Science	8,960	9,497	79,098	7.48%
Urology & Nephrology	Natural Science	112,183	4,970	83,229	28.99%
Veterinary Sciences	Natural Science	241,600	5,581	115,576	14.88%
Water Resources	Natural Science	55,016	13,910	90,912	14.95%
Zoology	Natural Science	56,057	5,850	88,376	13.33%

* History is classified as a discipline under both Social Sciences and Arts and Humanities.

Appendix II. Number of CSCD papers and journals by level-1 discipline classified in the journal classification system

Level-1 Discipline	# of Papers	# of Journals
Engineering & Technology	664,629	260
Medical Science & Health	478,946	180
Physical Science and Chemistry	190,815	99
Agricultural Science	117,290	95
Astronomy & Earth Science	177,959	84
Biology	85,026	53
Environmental & Safety Science	77,037	45
General Social Science	66,554	28
Transportation	33,735	18
Aviation & Aerospace	21,019	15
General Natural Science	1,696	3
General Social Science	5,218	3

Appendix III. Number of CSCD papers and journals by level-2 discipline classified in the journal classification system

Level-2 Discipline	# of Papers	# of Journals
Automation & Computer Technology	125,023	33
Physics	92,013	30
Chemical Engineering	78,872	37
Multidiscipline	77,037	45
Chemistry	76,590	32
Preclinical Medicine	72,987	29
Electronic Technology & Telecommunications	72,010	33
Medical Science & Health – Other Topics	69,664	23
Environmental & Safety Science – Other Topics	66,554	28
Pharmaceutical Science	63,678	22
Machinery & Instruments	56,880	20
Surgery	53,323	21
Geology	51,778	45
Light Industry	50,609	11
Internal Medicine	50,530	14
Agricultural Science – Other Topics	45,403	24
Metal Science	45,167	12
Electrical Engineering	44,635	12
General Engineering & Technology	44,214	18
Agricultural Basic Science	42,227	18
Biology – Other Topics	41,025	20
Clinical Medicine	35,389	13
Aviation & Aerospace	33,735	18
Special Medicine	32,714	14
Architecture	32,155	12
Chinese Medicine	31,663	9
Metallurgical Engineering	31,592	13
Military Science	25,517	10
Petroleum Engineering	23,845	15
Agronomy	21,994	11
Engineering & Technology – Other Topics	21,833	10
Forestry	21,046	11
Transportation	21,019	15
Mining Engineering	20,719	8
Mathematics	20,084	24
Botany	19,815	12
Geophysics	19,423	11
Oceanology	18,157	13
Agricultural Engineering	17,701	4
Neurology & Psychiatry	17,098	7
Animal Science & Veterinary Medicine	16,558	8
Disease Prevention & Hygiology	15,691	7
Mechanics	14,848	14
Hydraulic Engineering	14,541	9
Physical Geography	14,234	11
Oncology	14,051	8
Meteorology	12,014	10

Topography	11,686	5
Energy & Power Engineering	11,573	7
Biology Principle & Theory	11,567	8
Nuclear Science & Technology	9,120	8
Obstetrics & Gynecology	8,847	4
Pediatrics	8,556	4
Fisheries	7,619	3
Microbiology	6,932	5
General Social Science	5,218	3
Plant Protection	5,125	4
Dentistry	4,784	3
Horticulture	4,590	2
Otorhinolaryngology	3,726	2
Ophthalmology	3,497	2
Zoology & Anthropology	2,589	3
Entomology	1,972	2
General Natural Science	1,696	3
Astronomy	1,446	4
Paleontology	1,126	3

Appendix IV. Number of CSCD papers and journals by discipline classified in the paper classification system

Level-1 Discipline	Level-2 Discipline	# of Papers
Agricultural Science	Agricultural Basic Science	26,066
	Agricultural Engineering	6,619
	Agricultural Science – Other Topics	1,259
	Agronomy	33,712
	Animal Science & Veterinary Medicine	22,521
	Fisheries	7,824
	Forestry	19,346
	Horticulture	18,508
	Plant Protection	13,690
	Total	**149,545**
Arts and Humanities	**Total**	**766**
Astronomy & Earth Science	Astronomy	1,926
	Geology	45,859
	Geophysics	12,967
	Meteorology	14,666
	Oceanology	9,039
	Physical Geography	5,147
	Topography	13,537
	Total	**103,141**
Aviation & Aerospace	**Total**	**29,111**
Biology	Biochemistry	8,535
	Biology Principle & Theory	38,341
	Botany	23,718
	Entomology	4,091
	Microbiology	9,411
	Paleontology	1,668
	Zoology & Anthropology	6,227

	Total	**91,991**
Engineering & Technology	Architecture	46,379
	Automation & Computer Technology	173,865
	Chemical Engineering	78,494
	Electrical Engineering	43,834
	Electronic Technology & Telecommunications	75,355
	Energy & Power Engineering	19,193
	Engineering & Technology – Other Topics	463
	General Engineering & Technology	29,719
	Hydraulic Engineering	12,789
	Light Industry	39,000
	Machinery & Instruments	28,315
	Metal Science	64,952
	Metallurgical Engineering	8,425
	Military Science	11,451
	Mining Engineering	8,562
	Nuclear Science & Technology	8,908
	Petroleum Engineering	21,713
	Total	**671,417**
Environmental & Safety Science	**Total**	**85,239**
General Natural Science	**Total**	**4,555**
General Social Science	**Total**	**34,998**
Medical Science & Health	Chinese Medicine	55,943
	Clinical Medicine	36,506
	Dentistry	6,152
	Disease Prevention & Hygiology	27,084
	Internal Medicine	63,468
	Medical Science & Health – Other Topics	18,948
	Neurology & Psychiatry	18,290
	Obstetrics & Gynecology	11,320
	Oncology	50,943
	Ophthalmology	4,293
	Otorhinolaryngology	5,099
	Pediatrics	10,455
	Pharmaceutical Science	49,312
	Preclinical Medicine	61,699
	Special Medicine	13,350
	Surgery	54,601
	Total	**487,463**
Physical Science and Chemistry	Chemistry	87,560
	Mathematics	32,510
	Mechanics	17,678
	Physics	56,160
	Total	**193,908**
Transportation	**Total**	**32,736**

Appendix V. Number of papers by WoS SU classified in the journal classification system

WoS SU	# of Papers
Engineering	428,690
General & Internal Medicine	281,482
Agriculture	143,963
Automation & Control Systems	123,611
Physics	86,369
Geology	82,589
Science & Technology - Other Topics	70,005
Chemistry	69,255
Environmental Sciences & Ecology	64,615
Pharmacology & Pharmacy	63,700
Metallurgy & Metallurgical Engineering	58,277
Surgery	53,333
Life Sciences & Biomedicine - Other Topics	39,994
Energy & Fuels	33,163
Architecture	32,527
Transportation	24,313
Forestry	21,273
Mathematics	20,402
Plant Sciences	18,170
Neurosciences & Neurology	17,103
Public, Environmental & Occupational Health	15,747
Oceanography	15,201
Mining & Mineral Processing	14,217
Mechanics	14,075
Oncology	14,055
Meteorology & Atmospheric Sciences	10,600
Obstetrics & Gynecology	8,849
Nuclear Science & Technology	8,743
Pediatrics	8,558
Biotechnology & Applied Microbiology	8,526
Microbiology	6,991
Fisheries	5,701
Arts & Humanities - Other Topics	5,221
Dentistry, Oral Surgery & Medicine	4,785
Otorhinolaryngology	3,726
Ophthalmology	3,497
Genetics & Heredity	2,456
Zoology	2,239
Entomology	1,976
Astronomy & Astrophysics	1,448
Paleontology	910
Anthropology	367

Appendix VI. Number of papers by WoS SU classified in the paper classification system

WoS SU	# of Papers
Engineering	349,105
General & Internal Medicine	246,825
Automation & Control Systems	179,797
Mathematics	147,682
Agriculture	126,726
Chemistry	86,660
Environmental Sciences & Ecology	85,383
Metallurgy & Metallurgical Engineering	73,318
Physics	62,318
Surgery	54,077
Pharmacology & Pharmacy	51,745
Oncology	50,632
Architecture	47,701
Geology	45,645
Energy & Fuels	41,859
Science & Technology - Other Topics	36,628
Transportation	32,609
Public, Environmental & Occupational Health	31,930
Social Sciences	28,907
Plant Sciences	23,478
Life Sciences & Biomedicine - Other Topics	20,580
Forestry	19,294
Neurosciences & Neurology	18,931
Mechanics	17,880
Genetics & Heredity	14,382
Meteorology & Atmospheric Sciences	14,317
Obstetrics & Gynecology	11,309
Mining & Mineral Processing	10,684
Pediatrics	10,445
Nuclear Science & Technology	9,228
Microbiology	9,189
Oceanography	9,026
Fisheries	7,805
Dentistry, Oral Surgery & Medicine	6,146
Biotechnology & Applied Microbiology	5,997
Zoology	5,578
Otorhinolaryngology	5,092
Ophthalmology	4,293
Entomology	4,068
Astronomy & Astrophysics	1,961
Paleontology	1,655
Anthropology	622
Arts & Humanities - Other Topics	581
No classification	51