Issues in Informing Science + Information Technology

An Official Publication of the Informing Science Institute InformingScience.org

IISIT.org

Volume 21, 2024

HOW DIFFERENT ARE JOHNSON AND WANG? DOCUMENTING DISCREPANCIES IN THE RECORDS OF ETHNIC SCHOLARS IN SCOPUS

Trang Pham | Singapore University of Social Science, Singapore | phamtrang4652@gmail.com

ABSTRACT

Aim/Purpose	This study captures and describes the discrepancies in the performance matrices of comparable Chinese and American scholars as recorded by Scopus.
Background	The contributions of Chinese scholars to the global knowledge enterprise are increasing, whereas indexing bibliometric databases (e.g., Scopus) are not optimally designed to track their names and record their work precisely.
Methodology	Coarsened exact matching was employed to construct two samples of comparable Chinese and American scholars in terms of gender, fields of work, educational backgrounds, experience, and workplace. Under 200 scholars, around a third being Chinese and the rest American scholars, were selected through this data construction method. Statistical tests, including logit regressions, Poisson regression, and fractional response models, were applied to both samples to measure and verify the discrepancies stored within their Scopus accounts.
Contribution	This study complicates the theory of academic identity development, especially on the intellectual strand, as it shows ethnic scholars may face more errors in how their track records are stored and presented. This study also provides inputs for the discussion of algorithmic discrimination from the academic context and to the scientific community.
Findings	This paper finds that Chinese scholars are more prone to imprecise records in Scopus (i.e., more duplicate accounts, a higher gap between the best-statistic accounts, and the total numbers of publications and citations) than their American counterparts. These findings are consistent across two samples and with different statistical tests.

Accepting Editor: Eli Cohen │ Received: January 5, 2024 │ Revised: May 6, May 9, 2024 │ Accepted: May 27, 2024

Cite as: Pham, T. (2024). How different are Johnson and Wang? Documenting discrepancies in the records of ethnic scholars in Scopus. *Issues in Informing Science and Information Technology, 21,* Article 4. https://doi.org/10.28945/5325

Recommendations for Practitioners	This paper suggests practitioners and administrators at research institutions treat scholars' metrics presented in Scopus or other bibliometric databases with caution while evaluating ethnic scholars' contributions.
Recommendations for Researchers	Scholars and researchers are suggested to dedicate efforts to monitoring their accounts on indexing bibliometric platforms.
Impact on Society	This paper raises awareness of the barriers that ethnic scholars face in participating in the scientific community and being recognized for their contributions.
Future Research	Future research can be built on this paper by expanding the size of the analytical samples and extending similar analyses on comparable data harvested from other bibliometric platforms.
Keywords	ethnic scholars, bibliometric data, research performance, Chinese vs. American scholars

INTRODUCTION

A search for a researcher named "Peng Wang" on Scopus resulted in the same author identification with 11 ways to record their names. This author published almost 3,000 journal articles in 24 subject areas ranging from molecular biology to finance. Such a genius scholar may exist, but chances are this is a glitch in the bibliometric system. The bibliometrics systems historically designed to keep track of scholarly works on the Westernized protocol are inefficient in tracking Romanized Chinese publishing names. Meanwhile, China was the country that produced the most scientific publications worldwide in 2022 (Curcic, 2023). Therefore, there is a need to scrutinize the noisy performance metrics applied to Chinese scholars in the age of globalized education.

Measuring research performance has been a long-time quest for both scholars and administrators. While scholars need metrics to demonstrate their research achievements and productivity, administrators in higher education or research institutes need a tool to ensure highly performed scholars receive fair credits (e.g., promotion, tenure, wage raise, etc.). In the global knowledge enterprise, a track record summarizing agendas, achievements, and contributions would provide impressions of researchers to employers, collaborators, and the public. Bibliometric databases such as Scopus, Google Scholar, and Web of Science are trusted sources for institutions, administrators, and the public to reference researchers and their works.

However, ethnic authors, many of whom use their names as publishing names, face difficulty making themselves unique or recognizable in Western-designed bibliometric databases. It is known knowledge that records under Chinese and Korean names in Scopus were more likely to be deemed imprecise (Ioannidis et al., 2018). Informally, scholars with ethnic or Romanized family names (e.g., the Chinese, Korean, Chinese and Korean descendants, etc.) occasionally take it as given when they are indexed with a wrong publication, missing publications, and citations, and have their profiles split or merged as different individuals.[1] In other words, their works are imprecisely recorded in bibliometric databases. Currently, there is not much effort to describe how noisy the current system is and what to do to address this issue. Despite some discussion about algorithmic discrimination (The White House, n.d.) and technology bias (Garcia, 2016), limited empirical studies have captured the magnitude of the potential technological discrimination against ethnic scholars.

[1] See some discussion threads on Academia Stack Exchange (https://academia.stackexchange.com/questions/68941/what-are-potential-hassles-of-publishing-papers-with-family-name-preceding-given); Reddit (https://www.reddit.com/r/AskAcademia/comments/o9fwwh/how_do_i_cite_chinese_names_written_as_1_english/); and ResearchGate (https://www.researchgate.net/post/How-to-cite-chinese-names).

This study captures and describes the discrepancies in the performance matrices of comparable Chinese and American scholars as recorded by Scopus. I used coarsened exact matching to construct samples of about 200 scholars with rich demographic and professional details. I conducted multiple statistical tests to quantify the imprecision embedded within each individual's accounts and compared the differences between accounts belonging to Chinese and American scholars. This work contributes to the recent discussion on technological discrimination, contextualizing higher education. It also extends the theory of academic identity development when suggesting that the imprecise paper trace (or track record) of ethnic scholars could generate an incomplete impression of ethnic scholars' intellectual contribution.

CONCEPTUAL FRAMEWORK

IMPRECISIONS IN BIBLIOMETRIC DATABASES

Over the past decades, the number of published scientific papers has increased by about 8 to 9 percent annually (Landhuis, 2016). In 2022, a rough estimate showed a grueling 5.14 million academic articles published in over 46,000 journals worldwide (Curcic, 2023). The mere task of indexing this amount of new knowledge into a searchable filing system is tremendous. It requires the collaboration of the bibliometric platform management, the journals and conference organizers, and the researchers themselves. Indexing the research to the right authors provides the scholars with the correct reflection of their track records to showcase their scientific contributions. Such track records are essential for the scholars' career advancement, future collaborations, and promotion. The track records also help other scientist fellows, higher education administrators, and the public follow a scholar's work and performance for evaluation or appreciation. However, currently, scholars' presentation on most bibliometric platforms do not correctly reflect their track records and scientific publications, even on trustworthy sources, such as Scopus, Web of Science, or Google Scholar (Franceschini et al., 2016; Ioannidis et al., 2018).

The inputs of indexing platforms, like Scopus, come from a list of journals, publishers, conferences, and publishers that meet certain standards to be included. Once a researcher publishes their work in an outlet in the indexing list, Scopus generates an author profile for that researcher on the platform (Baas et al., 2020). Authors' profiles are verified, maintained, curated, and updated by Scopus' algorithm and staff so that all works done by the authors are filed together (Baas et al., 2020). This automation ensures that only scientific works reviewed, published, or presented in quality sources are recognized. This approach poses several disadvantages.

First, new and international journals and conferences may not be listed after several years (Elsevier, n.d.). Many of these newcomers are innovative, fast-growing, and author friendly. Therefore, they can manage to attract high-quality contributions from established researchers as well as novice and minority scholars. Lagging in including such new journals and conferences may lead to a gap in how indexing platforms reflect authors' scientific contributions. Second, discrepancies between indexed platforms are documented as each may maintain different input pools of journals or conferences (Meho & Yang, 2007). This means an author may have different numbers of publications, conference papers, and H-indexes if they are being looked up on different platforms. Third, it is a hassle for authors to correct their profile pages if wrong data are indexed due to their limited rights to edit their own profiles. For example, researchers are advised to contact the Scopus team for assistance in merging or deleting their author pages in case of duplicate profiles (National University of Singapore Library, 2024). Workarounds have been integrated to give authors more flexibility, such as linking a Scopus profile to an ORCID profile where authors can edit their data (Baylor University, 2021). However, imprecision remains in widely used indexing databases.

Imprecise Records of Ethnic Scholars' Publishing Names

"Your name is your brand," as an anonymous saying goes in the scientific community. Unsurprisingly, maintaining precise scientific profiles on bibliometric platforms is key to promoting scholars' works and representing their scholarly image. Scholars aspire to build their personal "brand" of their names using various channels and methods: committing to quality research, maintaining a professional online image, managing social media to promote their careers, and so on (Gander, 2014). Publishing names are found to correlate with one's citations and media coverage of a scientist (Einav & Yariv, 2006; Peng et al., 2024), which inherently affects one's social impact and research influence. A well-maintained author profile on trusted bibliometric platforms signals that the scholar is serious about their work with continuing endeavors.

The problem is it costs time and effort to keep all profiles on various bibliometric databases well-maintained – the resources that scholars can invest into their research. Furthermore, the likelihood of having imprecise records is higher for scholars with non-Western names due to typographical errors or name mismatching. Chinese and Korean family names are among those with the highest rate of imprecision (Ioannidis et al., 2018). By extension, it is more costly for a Chinese or Korean scholar to keep their track records precise in indexing platforms than their colleagues with a Western last name. Given that China is currently the country with the highest number of publications worldwide (Curcic, 2023), imprecise records of Chinese scholars may lead to misrepresentation and wrong information reported in indexing bibliometric platforms.

Names have long been established as a signal of one's origin, which may expose one to discrimination. Audit studies have shown that job candidates with Black or Hispanic names have a lower likelihood of callbacks than those with Anglo names (Bertrand & Mullainathan, 2004; Darolia et al., 2015; Gaddis, 2017). A similar form of discrimination applies to publishing names. In particular, Chinese family names are often indexed with ambiguity (Ioannidis et al., 2018), partly due to their Romanised names not being correctly recognized by the system. As Romanization does not capture intonations in the Chinese language, people with different family names may be indexed together. One person whose name is written in different orders is indexed into different accounts. These mixes create ambiguity when one needs to look up scholars' body of work. This may happen to Western names as well (Franceschini et al., 2016). However, Chinese (and Korean) scholars require more active efforts to verify, update, or contact the platforms to correct wrongly recorded data.

Arguably, the operation of academic databases is established with foundational algorithms that are, to some extent, embedded with Eurocentric mindsets. For example, the system is programmed to recognize an individual with a full name recorded in alphabetical characters. A name comprises a given name, a middle name, and a family name – in that sequence. The middle name is optional or can be initialized. An ethnic name, like a Chinese name, may not subscribe to this system. The original name is not alphabetical, and the family name goes before the given name. Chinese scholars have to adjust their names to conform with the Western identification system, such as flipping the family names and given names, removing or combining the middle name with the given name. These changes may not be consistent as scholars decide to adjust their names to be more conforming or more unique. This leads to duplicate bibliometric profiles that underreport scholars' track records. I hypothesize that Chinese scholars are more likely to have duplicated profiles than scholars with Western names.

Wrongly recorded bibliometric accounts have implications for scholars' careers. First and foremost, the scholar is negatively affected if their publications and citations are stored in multiple profiles (i.e., "leakages"). This means underreported accreditations of their scholarly works. Derivatively, the H-index is not calculated correctly and is usually underestimated due to the "leakages" in the reported numbers of publications and citations. Rarely would anyone verify the H-index generated by the database; therefore, colleagues, administrators, and the public may have an impression that ethnic scholars have a lower-than-reality H-index.

OTHER FACTORS OF IMPRECISE RECORDS

Besides names, the literature has also established other factors that may affect how scholars and their works are recorded in bibliometric databases. Scholars' field of work is one important factor. Research productivity and contributions vary by field, partly due to the disciplinary protocol of authorship recognition (Gervits & Orcutt, 2016). A scholar in sociology is considered productive in publishing from five to ten papers a year (Warren, 2019), whereas a physicist would need at least 20 papers a year, depending on subfields and team sizes (Battiston et al., 2019). For physicists, a publication or several citations missing from their profiles are neglectable. Professorial ranks are another determinant of imprecision. A young scholar or doctoral student at the beginning of their career usually has only a few conference proceedings or co-authored papers on hand. Several may not be consolidated under one bibliometric account. The prestige of one's doctoral degree is argued to affect the number of publications and citations of a study, as graduates from prestigious universities usually have more chances to be part of a large project (Battiston et al., 2019; Li & Koedel, 2017), which has wide coverage and hence more influence. Bibliometric data may advantage women (Thelwall et al., 2023), making gender a factor that contributes to any imprecision in the records.

METHODOLOGY

In this section, I present the data construction and analysis conducted on a sample of about 200 scholars extracted from Scopus to examine the miss-documentation of ethnic scholars' achievements. The choice of Scopus is because this database has a larger coverage than the Web of Science and more precise records than Google Scholar (Bar-Ilan, 2008; Norris & Oppenheim, 2007). Scopus also has credits for providing curated and high-quality data (Baas et al., 2020). Furthermore, it is the choice of convenience as Premium Scopus access is granted through my home institution.

DATA CONSTRUCTION

The analytical data in this paper is part of a dataset in use for a larger project investigating the cultural differences revealed in the behaviors of American and international professors in US higher education. An initial pool consists of under 3,000 instructors who worked in three universities from 2011 to 2017. These academics were selected from three flagship universities, and the dataset contains their names, gender, race, number of publications, number of citations, H-index, years of experience, professorial ranks, and so on. These are public research-intensive universities. Each housed about 23,000 to 33,000 undergraduate students during the time of the study, who were enrolled in approximately 260 to 300 courses. In one university, the faculty population was around 15% international, whereas in the other two, this portion was less than 10%. All three universities are prominently White. In the second step, the Chinese scholars were identified from the initial pool and matched with comparable American scholars in all observable dimensions. Around 4% (or 123 scholars) of the initial pool were Chinese, and over 2,000 were American.

The data construction steps are elaborated as follows:

Step 1: Build the initial pool of around 3,000 academics

To find the teaching pool, I collected data on all courses offered during all semesters between 2011-2017 from each focal university. Data include class-average grades, the full distribution of grades (percent As, Bs, etc.), class level, student enrollment, department, semester, and instructor's (or scholar's) full name. Within the teaching pool, I used a conservative power calculation on the class-grade estimation to find that a sample of around 3,000 instructors would be sufficient to detect a difference of at least 10%.

I randomly selected scholars, stratified by university (~1000 per university), from the grades data to create the targeted sample. Then, using their full names, I conducted a manual search through each instructor's CV and website to collect information on instructors' qualifications, employment history, demographics (e.g., gender), and information to identify their country of origin (e.g., the country where bachelor's degree was obtained). In the initial pool, I omitted graduate student instructors (N=766) and instructors who could not be found during the manual search (N=211). Ultimately, the initial pool is made up of 2,023 scholars working in three universities during the 21 semesters of the study period.

As these data are publicly available, there is no requirement for IRB approval for my data collection process. However, to maintain confidentiality, the data were de-identified before analyzing and aggregated in relevant reports and papers. Details of the data construction process are provided in the Appendix.

Step 2: Identify the Chinese scholars and their comparable American counterparts

Within this pool, I continue to identify the subset of Chinese scholars and match them with comparable American scholars based on the rich observables the initial dataset offers. First, I identified all Chinese scholars based on their country of origin. I excluded 73 scholars of Chinese origin but with Western family names, and the sample resulted in 53 Chinese scholars with Chinese last names. Next, I used coarsened exact matching (CEM) to match them with comparable American peers based on gender, doctoral degree prestige, field of work, professorial rank, and years of experience.

I constructed two comparison groups of American scholars: *the best-match* (91 American individuals with exact values in terms of the five interested criteria to 50 Chinese) and *the good-match* (137 Americans with exact values in the first four interested criteria and their years of experience is within +/1-year difference from 53 Chinese). Twenty Chinese scholars are without an American match and excluded in the next step. *The best-match* sample (n=141) is a sub-group of *the good-match* sample (n=190).

Second, I consolidated the bibliometric data of Chinese and American scholars selected into *the good-match* and *best-match* samples from Scopus. Of both samples, 90% can be found with at least one account. The data from Scopus include the numbers of accounts, publications, and citations in all accounts, as well as the H-index in the account with the highest statistics. I manually searched for each scholar by their name. When several accounts showed up, I verified that the works in the different accounts belonged to the same person through institution affiliation, names, and web search. A research assistant also developed a program to automate the search and verify the manual searches.

Descriptive statistics of the final samples are provided in Table 1. Ninety-six percent of the Chinese in the *good-match* sample have an account, and 46% of them have more than one account, compared with 88% and 8% for Americans, respectively. In both samples, successively, 7.5% and 9.1% of American scholars have duplicate accounts, whereas this rate of Chinese scholars is over 40%.

MODEL

The use of coarsened exact matching (CEM) ensures that those who were included in the analytical sample are comparable on all observable dimensions. CEM is the preferred method because, in this context, it is essential to establish comparable groups. As stated in the literature, the records of scholars working in different fields can be widely different (Battiston et al., 2019; Li & Koedel, 2017). A regression that controls for the field of work, professorial rank, or years of experience could only compare the average performance between the two groups of American and Chinese scholars, but it would not be able to compare the similar individuals from the two groups. Furthermore, due to the small sample size, adding control variables (many are categorical) would reduce the model's degree of freedom and lead to imprecise estimates. Similarly, these analytical samples are underpowered for fixed-effect models.

With the two samples constructed by CEM, I model the relationship between having a Chinese name (aka being Chinese) and the precision of their Scopus records using the following equation:

$$record\ precision_i = \beta_0 + \beta_1 * Chinese_i + \varepsilon_i \quad (1)$$

In which, $record\ precision$ of scholar *i* is measured by several indicators, including (i) whether scholar *i* has a duplicated profile, (ii) numbers of duplicate accounts of scholar *I*, (iii) the gap between the best statistic account and the total publications and citations of scholar *i*. $Chinese_i$ is an indicator variable of the Chinese name, taking a value of 1 if scholar *i* has a Chinese family name and 0 if scholar *i* is American. The coefficient of interest is β_1, which shows the discrepancy between a Chinese scholar and a comparable American scholar. I apply three different statistical tests, namely a logit regression, a Poisson regression, and a fractional response model, to check if the difference is robust.

RESULTS

LIKELIHOOD OF ACCOUNT DUPLICATES

Table 1 shows the breakdown of the *good-match* (Panel A) and *best-match* (Panel B) samples by American and Chinese scholars. In both good-match and best-match samples, over 90% can be found on Scopus (i.e., having at least one Scopus account). Chinese scholars are suggestively more likely to be found on Scopus ($p \cong 0.1$) and more likely to have multiple Scopus accounts ($p<0.001$) than American scholars. Both American and Chinese scholars encounter issues of duplicate accounts; however, the numbers of duplicates among Chinese are significantly higher than those of American scholars (1.745 versus 1.140 per individual for *good matches* and 1.792 versus 1.125 per individual for *best matches*, $p<0.001$). This breakdown suggests more noise in the metrics of Chinese scholars' accounts than American scholars in Scopus.

Table 1. Scopus accounts by American and Chinese scholars

	American	**Chinese**	**Total**	**p-value**
Panel A: Good match	mean (%; SD)			
N	137 (72.1%)	53 (27.9%)	190 (100.0%)	
have Scopus account(s)	0.883 (0.322)	0.962 (0.192)	0.905 (0.294)	0.096
have more than one Scopus account	0.091 (0.289)	0.431 (0.500)	0.192 (0.395)	<0.001
number of Scopus accounts found	1.140 (0.567)	1.745 (1.036)	1.320 (0.785)	<0.001
Panel B: Best match				
N	91 (64.5%)	50 (35.5%)	141 (100.0%)	
have Scopus account(s)	0.879 (0.328)	0.960 (0.198)	0.908 (0.290)	0.114
have more than one Scopus account	0.075 (0.265)	0.458 (0.504)	0.219 (0.415)	<0.001
number of Scopus accounts found	1.125 (0.603)	1.792 (1.051)	1.375 (0.860)	<0.001

To further evaluate the errors contained in Scopus metrics of Chinese and American scholars, the following analyses only consider the samples of scholars who have at least one Scopus account. This results in the analytical sample size of 172 for the *good matches* and 128 for the *best matches*.

COMPARISONS OF ACCOUNT DUPLICATES

To quantify the likelihood of duplicate accounts, Table 2 presents the estimates of record precision measured by the chance of having more than one account and the number of accounts. In columns 1 and 3, logit models were applied to the *good-match* and *best-match* samples to show that a Chinese scholar is 26.5 to 32 percentage points higher than an American scholar to have more than one Scopus account, respectively ($p<0.01$). Columns 2 and 4 are the results of the Poisson regression model on the two samples, where the dependent variable of record precision is measured by the counts of Scopus accounts associated with each individual. In the *good-match* sample, the number of Scopus accounts associated with one Chinese scholar is 75.3% higher than that of an American scholar ($p<0.01$). In the *best-match* sample, this gap is 89.6% ($p<0.01$).

Table 2. Regression results on duplicate accounts and the number of duplicate accounts

VARIABLES	Good match		Best match	
	have more than one Scopus account	number of accounts	have more than one Scopus account	number of accounts
	(logit regression, margins)	(Poisson regression, margins)	(logit regression, margins)	(Poisson regression, margins)
	(1)	(2)	(3)	(4)
Chinese	0.265***	1.753***	0.320***	1.896***
	(0.0442)	(0.236)	(0.0498)	(0.294)
Observations	172	172	128	128

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

This result is interpreted as a Chinese scholar is more likely to incur duplicate accounts in Scopus, and the number of duplicates among Chinese is significantly higher than that of American scholars. In relative terms, the average rate of duplicate Scopus accounts among Chinese is almost double the duplicate rate of Americans. These estimates are consistent across two samples regardless of the measurements of record precision (i.e., a binary variable of duplicate accounts or the raw counts of Scopus accounts).

PERFORMANCE RECORDED IN THE HIGHEST STATISTIC ACCOUNT

Although one scholar may be associated with more than one Scopus account, there is usually one account that consolidates the majority of their records. When looking up a scholar, users tend to rely on the highest statistic account to obtain an overview of the individual's track record. Following, I examine the performance statistics recorded in the best-statistic account to compare the potential discrepancies this account may record for Chinese versus American scholars.

Among the 172 scholars of the *good-match* sample and 128 scholars of the *best-match* sample with Scopus accounts, I report the numbers of publications and citations as shown in the highest statistic Scopus accounts and the total numbers of publications and citations in all the associated accounts. I also report the H-index that Scopus calculated in the highest statistic account. To measure potential noise in this account, I take the complement of the best statistics (numbers of publications and citations in the highest statistic account) over the total statistics (total numbers of publications and citations in all accounts) for each individual. If an individual has only one Scopus account, this value is 0. The higher this value is, the more discrepancies there are in the highest statistic account.

Table 3 summarizes key performance metrics that were recorded in the highest and the total statistics of all Scopus accounts. Expectedly, the differences between the numbers of publications, citations, and h-index in the best account and the total sum are not statistically significant ($p>0.1$) for both *best-match* and *good-match* samples. This is due to the sample construction process, which ensures comparability. The only two significantly different metrics are the complements of the publication ratio and citation ratio in both samples. These values among Chinese scholars are consistently larger than those among American scholars. In the *good-match* sample, the discrepancy in the publication number in the highest statistic account among Americans is 0.011, whereas that among Chinese is 0.065 – or six times higher ($p<0.01$). This rate of the citation ratio is eight times higher ($p<0.01$). In the *best-match* sample, these rates are even higher, reaching nine and 17 times higher ($p<0.01$).

Table 4 models this correlation between the potential discrepancy and being Chinese using a fractional response model with the complements of publication and citation ratios as dependent variables. Overall, being Chinese increases the complement fractions by over 3% in the *good-match* sample and over 5% in the *best-match* sample. These estimates are statistically significant at 5% for the publication ratios and 10% for the citation ratios in both samples. I conclude that in the best statistical account, American scholars show more precise track records than Chinese scholars.

Table 3. Performance in the best-statistic accounts

Good match (n = 172)		American	Chinese	p-value
		(only those with Scopus account)		
		mean (%; SD)		
N		121 (70.349%)	51 (29.651%)	
publication number by author (max)	[1]	65.686 (145.927)	63.510 (63.455)	0.919
publication number by author (total)	[2]	66.777 (147.578)	68.157 (67.736)	0.949
citation number by author (max)	[3]	2,259.339 (7,014.095)	1,655.725 (2,193.543)	0.548
citation number by author (total)	[4]	2,268.372 (7,030.952)	1,705.608 (2,203.346)	0.577
compliment of publication number in highest stat. acct./total	1 - [1]/[2]	0.011 (0.056)	0.065 (0.130)	<0.001
compliment of citation number in highest stat. acct./total	1 - [3]/[4]	0.006 (0.036)	0.049 (0.133)	0.001
h-index (max)	[5]	16.521 (15.843)	16.647 (12.142)	0.959
Best match (n = 128)				
N		80 (62.5%)	48 (37.5%)	
publication number by author (max)	[1]	72.737 (177.399)	64.854 (64.949)	0.768
publication number by author (total)	[2]	73.662 (179.363)	69.792 (69.288)	0.886
citation number by author (max)	[3]	2,724.438 (8,545.914)	1,690.812 (2,237.404)	0.414
citation number by author (total)	[4]	2,730.625 (8,567.413)	1,743.812 (2,246.744)	0.436
compliment of publication number in highest stat. acct./total	1 - [1]/[2]	0.008 (0.053)	0.070 (0.133)	<0.001
compliment of citation number in highest stat. acct./total	1 - [3]/[4]	0.003 (0.021)	0.052 (0.136)	0.002
h-index (max)	[5]	16.462 (17.992)	16.708 (12.265)	0.933

Table 4. Marginal differences in max. records of Chinese vs. American scholars

VARIABLES	Good match		Best match	
	compliment of publication number in highest stat. acct. / total	compliment of citation number in highest stat. acct. / total	compliment of publication number in highest stat. acct. / total	compliment of citation number in highest stat. acct. / total
	(1)	(2)	(3)	(4)
	(fractional response model, margins)			
Chinese	0.033**	0.030*	0.054**	0.056*
	(0.016)	(0.018)	(0.027)	(0.031)
Observations	172	172	128	128

Standard errors in parentheses

**** p<0.01, ** p<0.05, * p<0.1*

DISCUSSION

BIBLIOMETRIC ACCOUNTS OF CHINESE SCHOLARS CONTAIN MORE ERRORS THAN COMPARABLE AMERICAN PEERS

This paper compares the performance metrics recorded in Scopus between two comparable groups of Chinese and American scholars working at three research-intensive universities in the United States. Coarsened exact matching was used to create two samples of comparable scholars (the *good-match* and the *best-match*). Different statistical tests were applied to verify and measure the gap between the records of Chinese and Americans, as well as the discrepancies between the performance records and their contributions. Overall, I find that Chinese scholars have a higher chance of being present on Scopus and a higher chance of having duplicate accounts on Scopus than American scholars. The performance recorded in the best-statistic accounts of Chinese scholars is noisier than that of American scholars, giving a less precise image of the former group when being looked up on Scopus. This result is robust across the examinations between two samples and different statistical tests.

CAUTION AND AWARENESS WHEN EVALUATING ETHNIC SCHOLARS

A scholar's name signals one's academic identification, which is important and relevant for scholars to develop an impression of themselves to educational administrators, the science community, and the public. However, one's name also signals one's ethnic origin, which has recently been acknowledged to expose one to algorithmic discrimination (The White House, n.d.). One way this systemic problem manifested itself in the scientific community is the higher probability of errors in ethnic scholars' accounts on bibliometric platforms like Scopus. One may argue that authors can log in, correct the information, and request to change or merge accounts. However, these administrative steps pose barriers for those who are unfamiliar with the platform (especially non-US, non-European scholars). Moreover, due to the time lags for incorrect records to be fixed, imprecise impressions might endure for a long time around the internet or inside archival data. This study, therefore, puts forth implications for ethnic authors to check their accounts on important academic databases and make sure their virtual appearances correctly reflect their scholarly contributions. For university leaders, given the scope of possible technological errors in any sources, it is suggested that standard metrics in use to evaluate faculty, e.g., one's track records such as numbers of publications, citations and H-index, should be more mindfully taken in cases of ethnic scholars.

This paper echoes recent studies on ethnic scholars to depict behaviors among ethnic scholars that deviate from what the theory of academic identity development may predict. A scholar's track record, theoretically, has been recognized as a key strand of academic identity. An academic identity is formed as an intimate connection between a scholar, their work, their work environment, and peers through three strands, namely intellectual, networking, and institutional strands (McAlpine, 2012). Among these three, the intellectual strand represents past and continuing contributions to one's specialism, which is conventionally demonstrated by papers published, grants awarded, and staying abreast of the latest developments in the fields (McAlpine, 2012). However, Peng et al. (2024) find that a scholar with an ethnic name is less likely to be accredited in science news. Pham (2022) proposes that international early-career scholars do not strictly take their track records to define themselves as academics. Lee and Haupt (2020) suggest that new developments in the US-China relationship generate some positive externalities for all Chinese research teams in their publications, which may lead to more records of Chinese authors being added to bibliometric databases. It is suggested that more work be done to examine how these scholars navigate and integrate into academia. Given this finding, this paper also calls for more caution when individuals, research institutions, or the public want to formulate an evaluation of ethnic scholars' scientific contributions.

LIMITATIONS AND AVENUES FOR FUTURE RESEARCH

The CEM method of building two strictly comparable samples proved its efficiency, as demonstrated by the insignificant differences between the two groups of scholars regarding publications, citations, and H-index in the best-statistic accounts. However, standard errors among the American group are higher than those of the Chinese group, which is due to the larger number of Americans and variations in their performance. Similar analyses conducted among larger scholar samples may produce more nuances in the performance and scholars' bibliometric records. Data from other databases, such as Google Scholar or Web of Science, can be extracted and analyzed for further insights.

Nevertheless, the key message of this study remains relevant in the current context of the global knowledge enterprise. Publishing with an ethnic name exposes a scholar to a higher scope of errors in their bibliometric accounts, which may give an imprecise impression of a scholar to the scientific audience. It is also suggested that administrators treat Scopus and data from other inputs cautiously.

REFERENCES

Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, *1*(1), 377–386. https://doi.org/10.1162/qss_a_00019

Bar-Ilan, J. (2008). Which h-index? — A comparison of WoS, Scopus and Google Scholar. *Scientometrics, 74*(2), 257–271. https://doi.org/10.1007/s11192-008-0216-y

Battiston, F., Musciotto, F., Wang, D., Barabasi, A.-L., Szell, M., & Sinatra, R. (2019). Taking census of physics. *Nature Reviews Physics*, *1*, 89–97. https://doi.org/10.1038/s42254-018-0005-3

Baylor University. (2021). *Scopus author ID*. https://libguides.baylor.edu/orcid/scopus

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, *94*(4), 991–1013. https://doi.org/10.1257/0002828042002561

Curcic, D. (2023, June 1). Number of academic papers published per year. *Wordsrated*. https://wordsrated.com/number-of-academic-papers-published-per-year/

Darolia, R., Koedel, C., Martorell, P., Wilson, K., & Perez-Arce, F. (2015). Do employers prefer workers who attend for-profit colleges? Evidence from a field experiment. *Journal of Policy Analysis and Management*, *34*(4), 881–903. https://doi.org/10.1002/pam.21863

Einav, L., & Yariv, L. (2006). What's in a surname? The effects of surname initials on academic success. *Journal of Economic Perspectives*, *20*(1), 175–188. https://doi.org/10.1257/089533006776526085

Elsevier. (n.d.). *Scopus content policy and selection.* https://www.elsevier.com/products/scopus/content/content-policy-and-selection

Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2016). The museum of errors/horrors in Scopus. *Journal of Informetrics*, *10*(1), 174–182. https://doi.org/10.1016/j.joi.2015.11.006

Gaddis, S. M. (2017). Racial/ethnic perceptions from Hispanic names: Selecting names to test for discrimination. *Socius*, *3*. https://doi.org/10.1177/2378023117737193

Gander, M. (2014). Managing your personal brand. *Perspectives: Policy and practice in higher education*, *18*(3), 99–102. https://doi.org/10.1080/13603108.2014.913538

Garcia, M. (2016). Racist in the machine: The disturbing implications of algorithmic bias. *World Policy Journal*, *33*(4), 111–117. https://doi.org/10.1215/07402775-3813015

Gervits, M., & Orcutt, R. (2016). Citation analysis and tenure metrics in art, architecture, and design-related disciplines. *Art Documentation*, *35*(2), 218–229. https://doi.org/10.1086/688724

Ioannidis, J. P. A., Klavans, R., & Boyack, K. W. (2018). The scientists who publish a paper every five days. *Nature*, *561*, 167–169. https://media.nature.com/original/magazine-assets/d41586-018-06185-8/d41586-018-06185-8.pdf

Landhuis, E. (2016). Scientific literature: Information overload. *Nature*, *535*, 457–458. https://doi.org/10.1038/nj7612-457a

Lee, J. J., & Haupt, J. P. (2020). Winners and losers in US-China scientific research collaborations. *Higher Education*, *80*, 57–74. https://doi.org/10.1007/s10734-019-00464-7

Li, D., & Koedel, C. (2017). Representation and salary gaps by race-ethnicity and gender at selective public universities. *Educational Researcher*, *46*(7), 343–354. https://doi.org/10.3102/0013189X17726535

McAlpine, L. (2012). Doctoral journeys from past to present to future. *Australian Universities' Review, 54*(1), 38-46.

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. https://doi.org/10.1002/asi.20677

National University of Singapore Library. (2024). *Author profiles*. https://libguides.nus.edu.sg/researchimpact/profiles

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. https://doi.org/10.1016/j.joi.2006.12.001

Peng, H., Teplitskiy, M., & Jurgens, D. (2024). Author mentions in science news reveal widespread disparities across name-inferred ethnicities. *Quantitative Science Studies*, *5*(2), 351–365. https://doi.org/10.1162/qss_a_00297

Pham, T. (2022). Academic identity development of Asian international doctoral students at a public university – A reflexive thematic analysis. *International Journal of Doctoral Studies*, *17*, 323–344. https://doi.org/10.28945/5013

Thelwall, M., Kousha, K., Stuart, E., Makita, M., Abdoli, M., Wilson, P., & Levitt, J. (2023). Do bibliometrics introduce gender, institutional or interdisciplinary biases into research evaluations? *Research Policy*, *52*(8), 104829. https://doi.org/10.1016/j.respol.2023.104829

Warren, J. (2019). How much do you have to publish to get a job in a top sociology department? Or to get tenure? Trends over a generation. *Sociological Science*, *6*(7), 172–196. https://doi.org/10.15195/v6.a7

The White House. (n.d.). *Algorithmic discrimination protections*. https://www.whitehouse.gov/ostp/ai-bill-of-rights/algorithmic-discrimination-protections-2/

APPENDIX

This section describes how the dataset of under 3,000 randomly selected university instructors is built to fit the analysis. The dataset is built from two sources. The first is from the course data from three focal universities spanning all semesters between 2011 and 2017 inclusively (hereafter: grade dataset). The data were publicly available and downloaded from the university registrars' websites. The grade dataset has information about class-average grades, class levels, class enrollment, numbers of each letter grades, departments, semesters, and most importantly, class instructors' full names. The second source is a unique dataset with information on instructors' qualifications and demographics (hereafter: instructor dataset), which were manually collected based on instructors' full names, universities, and departments. With these two data sources, I address the limitations of widely used survey databases in the existing literature. The grade dataset is based on administrative records of courses taught by faculty at the three universities, which ensures no concerns about non-random sample attrition. The instructor dataset contains instructors' nationality information, which enables me to attain specific country-level statistics.

The construction of our dataset proceeded in the following steps:

Step 1: Identify the sample size

First, the initial downloaded grade file yielded a universe of 18,000 instructors who taught over 166,000 classes during 21 semesters. Given that the instructor dataset requires manual data construction, my first objective was to identify a sample size that would be large enough to detect a meaningful difference in grading behavior and ensure the feasibility of a manual data search. A pre-analysis showed that in order to achieve *a minimum detectable effect size of 10% of a standard deviation of the average-grade distribution*, my target standard error size needed to be 0.025 for the main analysis, per the following:

$$Z = \frac{\sigma * t}{St.Dev(G)} \Rightarrow 0.1 = \frac{\sigma * 1.96}{0.5} \Rightarrow \sigma \approx 0.025$$

In this equation, Z is the effect size, σ is the target standard error of the parameter of interest (i.e., international status), t is the critical t-value, and St. Dev(G) is the standard deviation of the class-average grades of the population.

To perform an *ad hoc* power calculation, we collected information from 300 randomly selected instructors as a test sample (100 from each university) and calculated the standard errors of the main effect when I artificially duplicated the test sample repeatedly. My calculation showed that when the sample increased 10-fold to 3,000 faculty observations, the standard error was approximately 0.025. Although this *ad hoc* power calculation is not perfectly accurate because it replicates the same exact 300 observations, it gives a good estimate of the actual standard error as the real sample size grows. Based on this calculation, we collected data for 3,000 instructors to ensure a well-powered model with the ability to statistically detect meaningful grading differences between domestic and international faculty.

Step 2: Randomly select 3,000 instructors

Next, I utilized a stratified sampling strategy to create a drawing pool of instructors from departments and then randomly selected 3,000 instructors from within. To ensure that exceptionally large or small departments were not overrepresented in the drawing pool, I removed small departments of less than 10 instructors because too small departments will not satisfy department fixed-effect designs. For large departments of over 100 instructors, I randomly selected 100 to enter the pool. Departments with instructor populations from 11 to 99 enter the pool as they are. Then, I randomly selected 1,000 instructors from the stratified population at each university to make up the instructor dataset of 3,000 instructors. Given that the distribution of international instructors by field and department was unknown to me at the sampling stage, I applied this strategy to ensure that the sampled 3000 instructors were representative of departments in three universities.

Step 3: Manually collect instructors' data

With the list of 3,000 instructors in hand, I moved on to the manual data collection. This was the most labor-extensive step. I conducted a manual search to obtain the qualifications and demographic information of each instructor. I collected this information from instructors' curricula vitae and websites. In rare instances when these sources were unavailable, I substituted them with information from other sources, such as Scopus, LinkedIn, news articles, and university bulletins. This completed the instructor dataset.

Step 4: Assemble the final data

At this step, I merged the instructor dataset with the grade dataset and removed those whose profiles are not available online. Also, 766 instructors were excluded. This completes a dataset of 2,023 instructors working in three universities during 21 semesters. 384 instructors are international, accounting for under 20% of the initial pool. 17% of them (or 123 instructors) are Chinese. Over 2,000 instructors are American. Finally, all data were de-identified before analysis.

AUTHOR

Trang Pham, Ph.D., is a lecturer at the Singapore University of Social Sciences, Singapore. She studies issues of diversity in educational contexts and immigration policy. She teaches Social Policy and Social Sciences courses.