Journal of Informetrics 16 (2022) 101264

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Understanding the peer review endeavor in scientific publishing

Guangyao Zhang[a], Shenmeng Xu[b], Yao Sun[a], Chunlin Jiang[a], Xianwen Wang[a,*]

[a] WISE Lab, Institute of Science of Science and S&T Management, Dalian University of Technology, Dalian, China
[b] Vanderbilt University, Nashville, TN, United States

ARTICLE INFO

Keywords:
Peer review
Publons
Review length
Peer review endeavor

ABSTRACT

Peer review plays an essential role in the scholarly publishing life cycle. Using the verified peer review records of reviewers who use the Publons, we employed review length as a potential indicator of the effort researchers spend on peer review. We then examined the associations between various factors and review length. Special focus was placed on estimating the relationships between non-academic (economic and sociological aspects) factors and review length. Our results show that gender, country-level cultural backgrounds, and country-level economic backgrounds were significantly associated with review length. In addition, there are significant associations of disciplines (humanities & social sciences or hard sciences), English proficiency, publications, and verified reviews with review length.

1. Introduction

Peer review plays a momentous role in academic communication (Mulligan, Hall & Raphael, 2013) and represents the mechanism for quality control of science publications (Borgman & Furner, 2002; Rowland, 2002). Since the initial publish of Le Journal des Sçavans in 1665, peer review has become a widely accepted practice (Crawford & Stucki, 1990). The definite types (including single-blind, double-blind, triple-blind, open reports, open identities & open reports, open final-version commenting) and processes of peer review differ on a case-by-case basis, depending on disciplines, publishing venues and other factors. Peer review is valued by the academic community as an effective way for self-regulation in the selection of science articles in the development and evolution of scientific knowledge (Abelson, 1980; Goodman, Berlin, Fletcher & Fletcher, 1994; Shatz, 2004). The theory behind this mechanism is that a larger and more diverse group of people will usually find more flaws and will be able to make a more unbiased assessment of it than just the person or group who conducted this research (Sollenberger, 2002). However, academia has demonstrated complicated attitudes toward peer review (Mulligan et al., 2013). Peer review is considered as a unreliable system (Ernst, Saradeth & Resch, 1993; Eysenck & Eysenck, 1992). The effectiveness of the conclusions given by reviewers is also usually suspected (Armstrong, Shereene, Alexandra & Jeffrey, 2008; Chubin & Hackett, 1990). Although peer review can add value, sometimes they can lead to excellent papers being rejected (Siler, Lee & Bero, 2015).

Despite peer review is widely accepted, surprisingly little is known about its effects on the quality and utility of published information (Jefferson, Wager & Davidoff, 2002). One of the biggest problems with traditional peer review is that the content of peer review reports generally remains known only to the authors and to the handling editors. This implies that both excellent peer reviews, as well as useless or unprofessional ones, often remain a black box. Additionally, numerous publications on peer review are often based small-scale questionnaires (Godlee, Gale & Martyn, 1998; Jefferson et al., 2002; Paltridge, 2015; van Rooyen, Godlee, Evans, Black & Smith, 1999) or the ruminations of editors (Dalton, 1995) due to data limitations (Bruce, Chauvin, Trinquart, Ravaud & Boutron, 2016).

* Corresponding author.
E-mail address: xianwenwang@dlut.edu.cn (X. Wang).

https://doi.org/10.1016/j.joi.2022.101264
Received 8 March 2021; Received in revised form 9 January 2022; Accepted 16 February 2022
Available online 3 March 2022

However, with the development of open science (Zhang et al., 2021a, 2021b), this situation changed after the appearance of Publons. Publons was set up in 2012 as a platform for researchers to document and display their peer review activities, aiming to allow users to register their peer review efforts. Publons partners with academic journals so that the peer review exercises that researchers choose to display can be verified (Xu, Zhang, Sun & Wang, 2019). In 2018, Publons issued a report - *The Global State of Peer Review* (Publons, 2018). This report has been developed to investigate the state of scholarly journal peer review. Publons has combined one of the largest peer review surveys of the global research community with data derived from Clarivate Analytics' powerful cross-publisher sources and focused on four big questions: Who is doing the review? How efficient is the peer review process? What do we know about peer review quality? What does the future hold? (Publons, 2018).

Although peer review has attracted the attention of a large number of researchers, most of the studies are data-driven (Bravo, Grimaldo, López-Iñesta, Mehmani & Squazzoni, 2019; Buljan, Garcia-Costa, Grimaldo, Squazzoni & Marušić, 2020; Zhang et al., 2021a, 2021b), descriptive (Garcıa-Costa, Squazzoni, Mehmani & Grimaldo, 2021), and focusing on academic publishing (Godlee et al., 1998; Ho et al., 2013; Wicherts, 2016). This paper aims to further understand the peer review endeavor. Based on the dataset of Publons, we explored the review data and tried to explore the influencing factors for the first time. Compared with descriptive statistics based on a small data set (199 reviewers) (Xu et al., 2019), we expand our data set (24,151 reviewers), add more variables, and use regression analysis to advance our findings. Our research questions were, what factors and how do they relate to peer review endeavor? Specifically, using the review length as the agent of review endeavor, we want to examine the relationship between review length and factors from various perspectives for the first time, including economic, sociological and academic aspects.

2. Related works

2.1. Review length in bibliometrics

Peer review as a social process has attracted research interests in the scientometrics and bibliometrics community (Squazzoni, Brezis & Marusic, 2017). Numerous studies have revealed patterns and phenomena in peer review. Take Publons for example, Ortega explored the relationship between the peer review activities of scholars and their research performance (Ortega, 2017), and the relationship between Publons metrics and other impact measurements (bibliometric and altmetrics indicators) (Ortega, 2019).

Especially, review length has become one of the important entry points for observing peer review. Researchers have noted the differences in review length under different dimensions, such as gender (Bravo et al., 2019; Buljan et al., 2020), disciplines (Buljan et al., 2020; Xu et al., 2019). However, they have only noticed statistical differences and paid little attention to the significance of review length. Actually, in 1990, Laband proposed that examine referee reliability by measuring the length of the review report, assuming that the longer the text, the higher the quality of the referee comments (Laband, 1990; Squazzoni & Gandelli, 2013). Nowadays, Several frontier researches have followed Laband's views and considered the review length as a proxy of the amount of information provided by the reviewer to editors and authors (Casnici, Grimaldo, Gilbert, Dondio, et al., 2017a; Casnici, Grimaldo, Gilbert, & Squazzoni, 2017b; Bianchi, Grimaldo & Squazzoni, 2019; Bravo, Farjam, Grimaldo Moreno, Birukou & Squazzoni, 2018). They hold that a detailed report is not only expected to include more valuable information that helps editors to judge the quality, rigor and validity of a manuscript; it can also provide insights for authors to improve their work. For instance, Casnici used review length to measure the quality of the referee reports (Casnici, Grimaldo, Gilbert, & Squazzoni, 2017). Bianchi used review length as one of the dimensions for evaluating reviewers (Bianchi et al., 2019). Publons also took review length as a proxy for the quality of the review (Publons, 2018).

In addition, in terms of the functions of peer review, quality control is one of the functions of peer review. Besides, another important function of peer review is to help authors improve the quality of their manuscripts (Bornmann, 2011; Matsui, Chen, Wang & Ferrara, 2021). Therefore, a detailed review report is more helpful to authors. It is worth noting that whether review length reflects review quality is actually controversial (Xu et al., 2019). Therefore, we would like to take a step back and adopt a weaker hypothesis which assumes that at least the review length reflects the time and efforts that reviewers invest in the review process. Thus, we think it is reasonable to assume that the longer the review is, the more time and efforts reviewers will have put into it.

2.2. Non-academic factors related to peer review

Regarding the sociology of science, Cole argued that cognitive and social factors interact with each other, eliminating any one of these two types of factors will lead to a one-sided and unrealistic understanding of how science works (Cole, 1994).

2.2.1. Gender

"Review" indicates the expression of opinions. There are gender differences in expression of opinions. Gender subculture theory (Maltz & Borker, 1983; Tannen, 1992) argues that gender differences are cultural differences. Various cultural traditions have different expectations for male and female language behavior; men are often more direct and have more competitive speech styles, while women are often more cooperative, focusing on the establishment and maintenance of good relationships. The theory of male dominance (Thorne & Henley, 1975) posits that men have a more powerful style in discourse, and the dominant social roles of women are more passive and humbler.

Meanwhile, the gender divide in scientific communities is a topic of great concern (De Nicola & D'Agostino, 2021), and this inequality may also exist in peer review endeavors (Helmer, Schottdorf, Neef & Battaglia, 2017; Squazzoni et al., 2021). The gender differences in the review comments have attracted the interest of academia, although not unanimously. Bravo found that male

reviewers tended to write more constructive reviews (Bravo et al., 2019). Buljan revealed that man write longer review than woman (Buljan et al., 2020). However, several studies have yielded different results. According to the latest research, women tend to provide more positive recommendations than men in all fields except for the physical sciences (Squazzoni et al., 2021). Garcıa-Costa analyzed 1.3 million reports and found that review reports from women are more developmental than those from men (Garcıa-Costa et al., 2021). Zhang analyzed the active reviewers from Publons and showed that female reviewers write longer reviews (Zhang, Shang, Huang & Sivertsen, 2022). So, we expect that gender is related to review length, which leads to the first hypotheses 1a & 1b.

H1a. Review length of females is shorter than that of males.

H1b. Review length of males is shorter than that of females.

2.2.2. Economic backgrounds

The national economic backgrounds reflect, to some extent, the scientific system and academic environment. Firstly, established economies have a more comprehensive peer review mechanism and a more standardized peer review system. However, peer review is a relatively new experience to scientists in emerging economies (Xu et al., 2019). For example, China once implemented a system of "three-level review": an editor performs the first review, the head of the editorial department performs the second review and the editor-in-chief performs the final review (Fang, Lifang & Xiaochuan, 2008). So, researchers might be less familiar with peer review rules and have less understanding to the scientific peer review norms. Second, established economies are more inclined to accept the form of evaluation (i.e., peer review) used by Publons. The practice of Publons has not been generally accepted (Xu et al., 2019). Researchers can use their review records as evidence of their influence in western academic community (Bornmann, 2011). However, this approach is not accepted universally for emerging countries. Individual's evaluation in these countries mainly focus on researchers' scientific publications and fundings (King, 1987, 2004). Reviewers' work is an "unrequited contribution" and cannot play a role in current research evaluation. This may result in researchers spending less effort on reviewing manuscripts. Moreover, Publons show that review length varies drastically by region. Specifically, reviewers in established economies write significantly longer reviews than reviewers in emerging economies (Garcıa-Costa et al., 2021; Publons, 2018). So, we argue that national economic backgrounds may be related to review length and derived the second hypotheses.

H2. Review length of reviewers in established economies is longer than that of reviewers in emerging economies.

2.2.3. Cultural backgrounds

The relationship between culture and science is receiving increasing attention. Khosrowjerdi & Bornmann discussed the relationship between culture and national scientific performance in their latest research (Khosrowjerdi & Bornmann, 2021). They also believe that systems of thought of individuals are culturally shaped. Several studies also revealed that culture could influence research and innovation activities (Jones & Davis, 2000; Pedro Couto & Cabral Vieira, 2004; Yair, 2020) and productivity (Kedia, Keller & Jullan, 1992).

When people express themselves, they are constructing contexts with social attributes attached to expressions, such as economy, culture, and politics. They will inevitably form unique ways of thinking, with their own cognitive biases or stereotypes regarding specific concepts (Kottak, 2013). Researchers have noticed the relationships between culture and holistic mode of thinking. Nisbett found that the East Asians had relatively holistic mode of thinking while Westerns had analytic mode of thinking (Nisbett, Peng, Choi & Norenzayan, 2001). Furthermore, compared to the individualism espoused in the West, Easterners are more introverted (Hofstede & Hofstede, 2010). East Asian countries' high-context communication involves more indirect expression than direct expression (Hall, 1976; Kim, Pan & Park, 1998), which is believed to be related to the impact of Confucianism (Yum, 1988). Specially, providing negative opinions for authors involves issues relevant to modesty, politeness, and conflict control when reviewing manuscripts. The confrontation-avoiding concepts (Kim et al., 1998) in Confucian idea might influence reviewers' critical text. So, we expect that cultural backgrounds may be related to review length and propose third hypotheses.

H3. Compared with Confucianism influenced culture, there is a longer review length associated with Christianism influenced culture.

2.3. Academic factors related to peer review

The relationship between peer review and academic factors has been the conspicuous topic of much attention. We review these literatures on four aspects: disciplines, English proficiency, publications, and verified reviews.

2.3.1. Discipline

Xu and Buljan revealed that reports written by reviewers for HSS are significantly longer than those written by reviewers for HS (Buljan et al., 2020; Xu et al., 2019). It is easier to express opinions in the humanities and social sciences (HSS) than in the hard sciences (HS). In addition, this may be related to the review duration and review criteria. The review time in HSS is longer than in HS (Publons, 2018). As the reviewers in HSS have more time to complete their reviews, they may write longer reviews than reviewers in HS (Publons, 2018). Garcıa-Costa showed that HSS show higher standards compared to all areas (Garcıa-Costa et al., 2021). Thus, we expect that disciplines are related to review length, which leads to the hypothesis 4a.

H4a. Review length of reviewers in HSS is longer than that of reviewers in HS.

2.3.2. English proficiency

The extensive use of English throughout the world has a complicated history. English is the de facto official language in western countries, such as the United States and Australia. Besides, English is used as a de jure official language, just like in India. Thus, a number of reviewers may face the dilemma of linguistic imperialism phenomenon (Phillipson, 2018). Editors may have preferences toward regions where English is spoken as a first language (Publons, 2018). Moreover, English proficiency is directly related to ability to express opinions in English, and reviewers in emerging economies (most of which are Non-native English-speaking countries) may be less comfortable writing in English (Publons, 2018). So, reviewers in countries that seldom use English might tend to write less in English when reviewing manuscripts (Hinkel, 2011). In addition, empirical studies have also indicated that critical thinking is more difficult in a second language (Floyd, 2011). Considering that the majority of journals associated to Publons are published in English and reviews are usually written in English (Publons, 2018). So, we intend to examine that English proficiency is an important factor that could influence review length and propose the hypothesis 4b.

H4b. The more proficient a reviewer's English is, the longer the reviews.

2.3.3. Publications

Papers published each year are increasing rapidly, while reviewers are increasing only by a limited amount. In 2017, editors had to invite an average of 2.4 reviewers to get one review conducted and reviewers took an average of 16 days to complete a review. However, the figures were 1.9 and 16.8 in 2013 (Publons, 2018). As we all know, scientists are always busy with scientific research (Barnett, Mewburn & Schroter, 2019; Wang et al., 2012, 2013). They devote considerable time and effort for writing reviews (Barnett et al., 2019; Righi & Takács, 2017). However, they could have used these sacrificed time for their own research activities (Bernstein, 2013). Therefore, we argue that reviewing is a time-consuming activity that may relate to the scientific research output of researchers to some extent. Therefore, we propose the hypothesis 4c.

H4c. The more articles published, the shorter the reviews.

However, reviewers in different levels may write comments differently. Younger scholars were more willing to provide objective recommendations (Bravo et al., 2019) and more content than senior colleagues (Casnici, Grimaldo, Gilbert, & Squazzoni, 2017). Juniors may have more time for reviews (Merrill, 2014). In addition, they may have desires to build reputation by writing long reviews and show their diligence and reliability to journal editors (Garcıa-Costa et al., 2021).The researchers who have already published many papers may lose their desire to please academic journals, so their reviews may be shorter. Therefore, the relationship between publications and review length may be complex. We propose the hypotheses 4d & 4e.

H4d. For ordinary researchers, as publications increases, review length tends to increase.

H4e. For high-publications researchers, as publications increases, review length tends to decline.

2.3.4. Verified reviews

Reviewers may obtain several advantages in peer review (Copiello, 2018). Reviewers can build reputation and get in the good graces of editors. Besides, reviewers also obtain indirect advantages in peer review, such as "to keep up with current ideas and new results" (Engers & Gans, 1998). Several studies show that peer review is helpful to researchers, especially for those who are in early career (Rodríguez-Bravo et al., 2017; Wei, Bu, Kang & Li, 2021). In sum, peer review may be a process of accumulating experience. Adequate review experience helps reviewers to provide more reliable and substantial reviews. Researchers with more verified reviews may write longer reviews because they have more experience in peer review. So, the hypothesis 4f is stated as:

H4f. The more verified reviews, the longer the reviews.

Additionally, we hope to explore the influence of interactions between factors on the review length. Gender role theory suggests that society has different role expectations for men and women, with men being expected to pursue material success and women being expected to play the role of caregiver (Gutek, Nakamura & Nieva, 1981). These gendered role expectations may influence reviewers' peer review. Specifically, as the publications and verified reviews increase, female reviewers may behave more communicative. Therefore, we propose the following hypotheses 5a and 5b.

H5a. As publications increases, review length of females tends to decline less than that of males.

H5b. As verified reviews increases, review length of females tends to increase more than that of males.

In terms of countries, most of the emerging economies are non-English-speaking countries. The improvement of English proficiency may be of greater help to researchers in emerging economies. Finally, we expect that even under different economic backgrounds, cultural backgrounds may still be related to review length. So, we derived the following hypotheses 5c and 5d

H5c. Reviewers in emerging economies benefit more from English proficiency than those in established economies.

H5d. Compared with reviewers from a country with a Confucianism influenced culture, reviewers from a country with a Christianism influenced culture have longer reviews irrespective of whether they are in a country with an established or emerging economy.

3. Data collection and cleaning

Reviewers' contributions to journals are verified and displayed on their Publons homepage. We collected reviewers' identity and affiliation information, as well as the number of verified reviews and the average review length. We initially collected these data (including names, research fields, verified reviews, institutions, countries/regions and average review length) on 174,560 reviewers in April 2018 from the overview page for all countries.[1]

To explore the relationship between review length and language proficiency, we collected the Education First English Proficiency Index (EF EPI) for each country. EF EPI is a global ranking of adult English proficiency in non-native English-speaking countries.[2] We gained data on 156,513 researchers after excluding countries without available EPI data.

We updated the data set in May 2019. We obtained the data on 124,412 researchers after removing the data that were "not found" (i.e., involving homepages that were no longer in operation). The data set which became 88,697 researchers after deleting those without verified reviews. We found duplicate items in the data set because some researchers were affiliated with multiple institutions. Fortunately, the first institution registered for each researcher is the institution at which the researcher is currently working. Therefore, 86,901 researchers were obtained after removing duplicate researchers (based on institution information).

We noticed some suspicious entries among inexperienced reviewers (or reviewers who have relatively incomplete and erroneous records) during the above-mentioned data cleaning. These reviewers may lack review experience or have insufficient information to record, and the validity of the data is questionable. So, we excluded reviewers with verified reviews of less than 3. We employed a gender inference system that uses three gender inference methods (facial recognition [FR]/Facebook-Generated Name List [FGNL]/Web-Based Gender Predictor [WebGP]) with a specific voting model to make the final prediction (Gu, Yang, Tang & Zhang, 2016; Tang et al., 2008). So, we can judge the reviewers' gender based on their names. We divided the reviewers manually into two large categories: HSS and HS based on the research fields of the reviewers on Publons. We adopted an approach similar to Publons that countries are categorized into two categories (Publons, 2018). We divided the countries' economic backgrounds where reviewers worked into two categories: established and emerging economies according to the Human Development Report (2010).

As for country-level cultural backgrounds, we divided the reviewers into three categories: Confucianism influenced culture, Christianism influenced culture, and Other. To designate each reviewer with a country-level cultural backgrounds based on their name. We first distinguished the Chinese names among the reviewers by regular expression.[3] Then we used a surname dictionary (Proper Names And Translation Service, Xinhua News Agency, 2007) to match the remaining surnames with countries, and then mapped each country to the corresponding cultural backgrounds.

Although the validity of the evaluation function of the h index is controversial (Bornmann, 2014; Ding, Liu & Kandonga, 2020; Hirsch, 2005; Viiu, 2016), traditionally, h index is one of the most widely used indicators in academia. Hirsh pointed out that "h index is an indicator of the impact of a researcher on the development of his or her scientific field" (Hirsch & Buela-Casal, 2014). Actually, in recent years, some researchers have used h index to quantify scientists' prestige or reputation in academia in peer review research (Matsui et al., 2021; Teplitskiy, Acuna, Elamrani-Raoult, Körding & Evans, 2018). So, we consider using h index to replace publications in this study to get more solid results. The h index and number of publications of each reviewer were collected from Scopus, the world's largest citation database of peer-reviewed literature. We aim to use more accurate data to improve the reliability of our results. However, the publications and citations provided by Publons are not accurate. Scopus[4] has become an important data source for peer reviewed research because of the comprehensiveness and availability of its data (Teplitskiy et al., 2018). We can obtain reviewers' publications and h index in Scopus by qualifying the full name and institutional affiliation for each author.

We excluded reviewers that did not fill in the research field (because we were unable to identify their disciplines), their name couldn't be matched to a specific cultural background, or did not have an h index or data on publications. Additionally, we believe that reviews that are very long or very short are invalid data points, so we truncated the data by deleting researchers with a mean review length < 54 words (the shortest 1%) or > 1534 words (the longest 1%). Finally, a data set containing 24,151 reviewers was obtained.

4. Basic descriptives

The factors that we explored comprised gender, economic and cultural backgrounds, disciplines, English proficiency, publications, and verified reviews. For the first four factors, we visualized the results using boxplots (Fig. 1).

By showing the distribution of review length, boxplots reveal the degree of its dispersion in Fig. 1. The horizontal line in middle is median value. The span between the upper quartile and the lower quartile and the span between the upper and the lower bounds indicate that the difference in various categories. The mean and median review length are greater for females than for males, but this is not intuitive in Fig. 1. a). Reviewers in established economies and Christianism influenced culture have greater mean and median

[1] https://publons.com/country/

[2] https://www.ef.edu/epi/about-epi/

[3] (a[*io*]?|ou?|e[*inr*]?|ang?|ng|[*bmp*](a[*io*]?|[*aei*]ng?|ei|ie?|ia[*no*]|o|u)|pou|me|m[*io*]u|[*fw*](a|[*ae*]ng?|ei|o|u)|fou|wai|[*dt*](a[*io*]?|an|e|[*aeio*]ng|ie?|ia[*no*]|ou|u[*ino*]?|uan)|dei|diu|[*nl*](a[*io*]?|ei?|[*eio*]ng|i[*eu*]?|i?ang?|iao|in|ou|u[*eo*]?|ve?|uan)|nen|lia|lun|[*ghk*](a[*io*]?|[*ae*]ng?|e|ong|ou|u[*aino*]?|uai|uang?)|[*gh*]ei|[*jqx*](i(ao?|ang?|e|ng?|ong|u)?|u[*en*]?|uan)|([*csz*]h?|r)([*ae*]ng?|ao|e|i|ou|u[*ino*]?|uan)|[*csz*](ai?|ong)|[*csz*]h(ai?|uai|uang)|zei|[*sz*]hua|([*cz*]h|r)ong|y(ao?|[*ai*]ng?|e|i|ong|ou|u[*en*]?|uan))

[4] https://www.elsevier.com/solutions/scopus/how-scopus-works/content

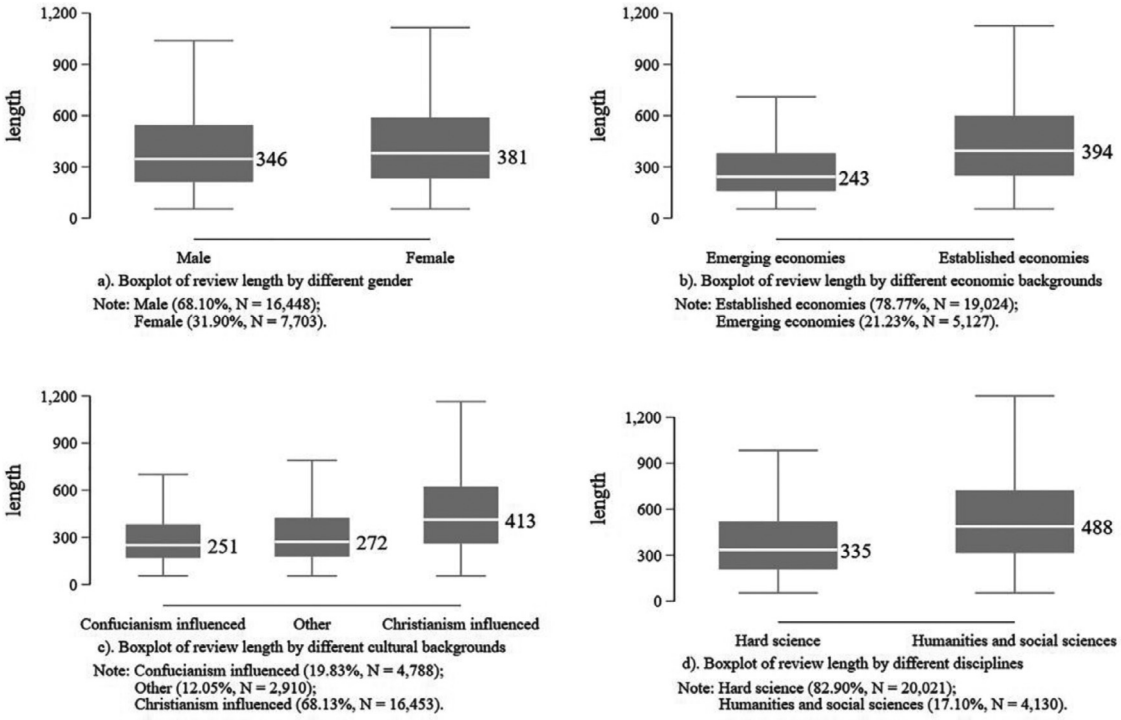

Fig. 1. Boxplots of review length a). Boxplot of review length by different gender b). Boxplot of review length by different economic backgrounds c). Boxplot of review length by different cultural backgrounds d). Boxplot of review length by different disciplines.

Table 1
Summary statistics of review length for non-academic factors.

	Obs	Mean	Median	Std. Dev.
Gender				
Female	7703	443.22	381	276.85
Male	16,448	414.64	346	269.74
Economic backgrounds				
Established	19,024	456.65	394	278.12
Emerging	5127	301.73	243	208.58
Cultural backgrounds				
Confucianism	4788	306.39	251	200.32
Other	2910	332.81	272	226.27
Christianism	16,453	474.00	413	283.12
Disciplines				
HS	20,021	399.14	335	258.84
HSS	4130	543.12	488	303.01

review length than those in emerging economies, Confucianism-influenced culture, as shown in Fig. 1. b) and c). We also provided a summary statistics table with median, means and standard deviation in Table 1.

To test whether there was a significant difference, statistical tests were performed. The review length data for reviews by different gender, economic backgrounds, cultural backgrounds did not meet the normal distribution assumptions based on the Kolmogorov-Smirnov test. So, we conducted a Wilcoxon test and found that there were significant differences in review length between different genders ($p = 0.0006$), economic backgrounds ($p = 0.0001$). As for cultural backgrounds, due to a failed Bartlett's test ($p<0.05$), we could not conduct analysis of variance (ANOVA). So, the nonparametric Kruskal–Wallis test was conducted. The results show that there were significant differences in review length with different cultural backgrounds ($p = 0.0001$). We conducted further Kruskal-Wallis tests between reviewers with different cultural backgrounds, and there were differences in the review length between any two cultural backgrounds (adjusted p-value for significance was 0.008). Similarly, HSS has greater mean and median review length than that in HS, as shown in Fig. 1. d). We found that there are differences between HSS and HS through a Wilcoxon test ($p = 0.0001$) Tables 2. and 3 summarize the variables explored in this study. A pairwise correlation matrix of the variables is shown in Table 3.

Table 2
Variable description.

Variable	Attribute		Description
Review length	RL		Average review length of each reviewer
Gender	GE		1= "Male"; 0= "Female"
Economic backgrounds	EB		1= "Emerging"; 0= "Established"
Cultural backgrounds	CB	CBO	1= "Non-Confucianism"; 0= "Confucianism"
		CBC	1= "Christianism"; 0= "Other"
Disciplines	DI		1= "HSS"; 0= "HS"
English proficiency	EP		EPI for each country
Publications	PU		Value for reviewers' publications
H index	HI		Value for reviewers' h index
Verified reviews	VR		Value for reviewers' verified reviews

Table 3
Summary statistics and correlation matrix of variables (N = 24,151).

Variable	Mean	Median	Min	Max	Variance	1	2	3	4	5
RL	423.76	357.00	54.00	1534.00	74,174.90	1.00				
EP	69.00	80.00	45.71	80.00	155.85	0.26	1.00			
PU	80.43	40.00	1.00	4570.00	31,104.50	−0.11	−0.11	1.00		
HI	16.91	13.00	0.00	172.00	216.08	−0.08	−0.06	0.68	1.00	
VR	36.10	19.00	3.00	3891.00	4034.93	0.03	−0.02	0.05	0.16	1.00

Notes: Dummy variables (gender, economic backgrounds, cultural backgrounds and disciplines) were not included in the correlation analysis.

5. Results

Review length is a discrete variable involving non-negative integers. Therefore, we adopted "*ln(RL)*" as the dependent variable instead of "*RL*". We standardized *EP, PU, HI*, and *VR* so that we can compare the relative magnitudes of parameters, since the different variables have different units of measurement.

To evaluate the associations of the different explanatory variables with review length, the following benchmark model was employed:

$$ln(RL) = f(GE, EB, CB, DI, EP, PU, VR) \quad (1)$$

Table 4 provides the results of regression model, i.e., of Eq. (1). Column 1 (m1) shows the results when the non-academic and academic variables was included in the model. Gender, economic backgrounds, cultural backgrounds, disciplines, English proficiency, publications and verified reviews were significantly associated with the review length (p<0.05). These results initially validate our hypothesis (H1b, H2, H3, H4a, 4b, 4c, and 4f). The magnitude and significance of these coefficients did not change significantly when we replaced publications with h index in column 2 (m2).

We divided our data set into three equal parts according to the number of publications to distinguish the differences at different publications levels. We defined researchers with low publications as "Ordinary" researchers and researchers with high publications as "High-pub" researchers. Column 3 (m3) considered only Ordinary researchers, and column 4 (m4) considered only High-pub researchers. Publications remained significantly associated with review length. However, the directions of the coefficients changed. Specifically, as the publications increases 1 standard deviation, review length increases by 1.45% in m3; when the publications increases 1 standard deviation, review length decreases by 0.015% in m4. Our results support H4d & 4e to some extent. We dropped the reviewers who are from other culture in m5. The magnitude and significance of coefficients in column 5 are similar to m1. The PU at the 99th (p<676) and 97.5th percentile (p<325) were truncated in appendix-m1 and appendix-m2 to avoid the interference of outliers. We have added these results to the appendix.

We specified a poisson regression (m6) and a negative binomial regression (m7) to cater for the count data nature of review length. Besides, we also show the average marginal effects (dy/dx) of variables to indicate the effects of variables intuitively. Taking the poisson model as an example, as English proficiency and verified reviews increases 1 standard deviation, the review length increases by 39 and 3 words, respectively. Moreover, publications reduces review length, an increase in publications by 1 standard deviation decreases 29 words. Additionally, review length for HSS was greater than that for HS. As to cultural backgrounds, review length associated with Christianism influenced culture was the longest, while the review length associated with Confucianism influenced culture was the shortest. However, compared to other variables, the coefficients of gender and verified reviews are relatively smaller. In addition, the coefficient of gender is significant only at the 0.1 level in m7 (Table 5).

We further analyzed the interaction effects of variables to test the remaining hypotheses and showed the results in Table 6. M8, m9, m10, and m11 show results of our tests of Hypotheses 5a, 5b, 5c, and 5d, respectively. M8 tested the interaction between gender and publications. Hypothesis 5a was not supported. Although the interaction term was not significant, the direction of its coefficient was negative. M9 tested the interaction between gender and verified reviews. The coefficient of the interaction term was negative

Table 4
Regression results.

	m1 Ols	m2 H-index	m3 Ordinary	m4 High-Pub	m5 Excluding CBO
GE	0.019**	0.018**	0.064***	−0.005	0.020**
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
EB	−0.146***	−0.153***	−0.282***	−0.066***	−0.128***
	(0.01)	(0.01)	(0.03)	(0.02)	(0.01)
CBO	0.029**	0.036**	−0.026	0.040	
	(0.01)	(0.01)	(0.02)	(0.02)	
CBC	0.291***	0.300***	0.272***	0.275***	0.296***
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
DI	0.230***	0.227***	0.244***	0.206***	0.231***
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
EP	0.104***	0.104***	0.043***	0.124***	0.103***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
PU	−0.036***		1.451***	−0.015***	−0.035***
	(0.00)		(0.19)	(0.00)	(0.00)
VR	0.017***	0.018***	0.060***	0.014***	0.015***
	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)
HI		−0.019***			
		(0.00)			
Cons	5.622***	5.617***	6.252***	5.561***	5.613***
	(0.01)	(0.01)	(0.08)	(0.02)	(0.01)
Obs	24,151	24,151	8300	7942	21,241
Adj R^2	0.156	0.154	0.135	0.155	0.142

*$p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

Table 5
Regression results with alternative model specifications.

	m6 Poisson				m7 Nbr			
	Coef	Std.	dydx	Std.	Coef	Std.	dydx	Std.
GE	0.024***	(0.01)	10.181***	(3.53)	0.015*	(0.01)	6.147*	(3.50)
EB	−0.146***	(0.01)	−58.775***	(5.10)	−0.139***	(0.01)	−56.431***	(4.99)
CBO	0.051***	(0.01)	17.715***	(5.21)	0.048***	(0.01)	16.536***	(5.09)
CBC	0.308***	(0.01)	121.325***	(3.85)	0.310***	(0.01)	122.150***	(3.83)
DI	0.209***	(0.01)	94.516***	(4.79)	0.225***	(0.01)	102.047***	(4.84)
EP	0.093***	(0.01)	39.235***	(2.15)	0.099***	(0.01)	42.061***	(2.16)
PU	−0.069***	(0.01)	−29.179***	(2.96)	−0.039***	(0.00)	−16.452***	(1.69)
VR	0.007**	(0.00)	2.962**	(1.34)	0.007*	(0.00)	3.069*	(1.57)
Cons	5.779***	(0.01)			5.781***	(0.01)		
lnalpha					−1.125***	(0.01)		
Obs	24,151		24,151		24,151		24,151	
Pse R^2	0.152		–		0.013		–	
ll	−1,721,987.90		–		−163,254.80		–	
aic	3,443,993.79		–		326,529.60		–	
bic	3,444,066.62		–		326,610.52		–	

* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

and significant (p<0.01). This result show that, as verified reviews increases, review length of females tends to increase more than that of males. Hypothesis 5b was supported. M10 tested the interaction between EPI and economic backgrounds. The coefficient of the interaction term was positive and significant (p<0.01). This result shows that, reviewers in emerging economies benefit more from English proficiency than those in established economies. Hypothesis 5c was supported. M11 tested the interaction between economic backgrounds and cultural backgrounds. The coefficient of the interaction term was positive and significant (p<0.01). This result shows that, compared with reviewers from a country with a Confucianism influenced culture, reviewers from a country with a Christianism influenced culture have longer reviews irrespective of whether they are in a country with an established or emerging economy. Hypothesis 5d was supported.

Fig. 2.a–.d present these plots Fig. 2..a shows that the slop is much steeper when the gender is male Fig. 2..b shows that the slop is much steeper when the gender is female. Both graphs reflect femininity to some extent. In other words, female reviewers may behave

Table 6
Regression results with interactions.

	m8 GE*PU		m9 GE*VR		m10 EB*EP		m11 EB*CB	
GE	0.018**	(0.01)	0.014	(0.01)	0.019**	(0.01)	0.019**	(0.01)
EB	−0.146***	(0.01)	−0.146***	(0.01)	−0.067**	(0.03)	−0.115***	(0.02)
CBO	0.029**	(0.01)	0.029**	(0.01)	0.030**	(0.01)	0.076***	(0.02)
CBC	0.291***	(0.01)	0.291***	(0.01)	0.287***	(0.01)	0.302***	(0.01)
DI	0.230***	(0.01)	0.230***	(0.01)	0.230***	(0.01)	0.230***	(0.01)
EP	0.104***	(0.01)	0.104***	(0.01)	0.101***	(0.01)	0.102***	(0.01)
PU	−0.027***	(0.01)	−0.036***	(0.00)	−0.036***	(0.00)	−0.037***	(0.00)
VR	0.017***	(0.00)	0.057***	(0.01)	0.017***	(0.00)	0.018***	(0.00)
GE*PU	−0.011	(0.01)						
GE*VR			−0.045***	(0.01)				
EB*EP					0.071***	(0.03)		
EB*CBO							−0.111***	(0.03)
EB*CBC							−0.023	(0.02)
Cons	5.623***	(0.01)	5.627***	(0.01)	5.626***	(0.01)	5.610***	(0.01)
Obs	24,151		24,151		24,151		24,151	
Adj R^2	0.156		0.157		0.156		0.156	

* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

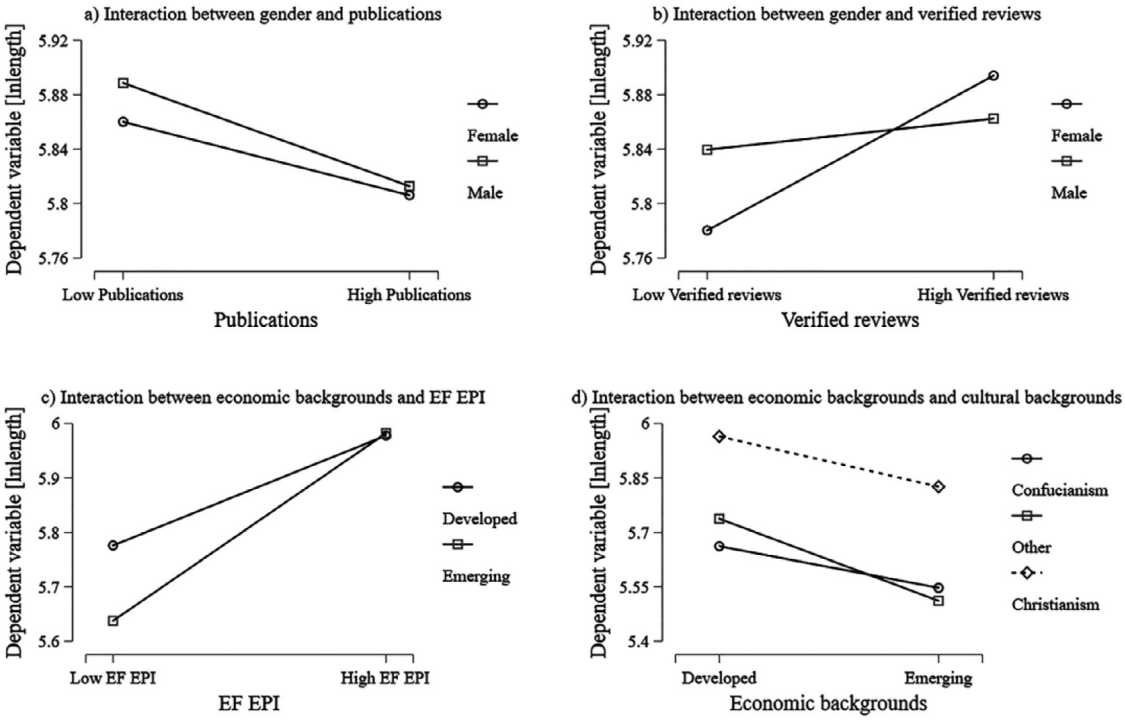

Fig. 2. Interactions between variables a) gender and publications b) gender and verified reviews c) economic backgrounds and EF EPI d) economic backgrounds and cultural backgrounds.

more communicative in peer review Fig. 2..c displays a steeper slope when the economic backgrounds are emerging Fig. 2..d shows that even working in developed countries, reviewers from Confucianism influenced culture write shorter reviews.

6. Discussion

The above analysis discussed several factors that might be related to review length, including non-academic factors (gender, economic backgrounds and cultural backgrounds) and academic factors (disciplines, English proficiency, publications, h index and verified reviews). In terms of their effects on review length, each of these factors may have small effects that compound when considered together.

Part of our findings are consistent with previous studies that some factors may reduce female scientists' time investment in work (Cole & Zuckerman, 1984; Xie & Kimberlee, 2003). We found that male reviewers tended to write longer reviews, which fit with Buljan's results (Buljan et al., 2020). However, the gender effect is relatively small compared to the other variables. Consistent with Publons (Publons, 2018), we also found that reviewers in HSS tended to writer longer reviews. Furthermore, we have also made some new discoveries. Previous studies have discussed the reports from Western Europe are generally more developmental (Garcıa-Costa et al., 2021), and our study further illustrates that reviewers from developed economies tend to write longer reviews. We also discuss the role of English proficiency, which is significantly associated with the expression of reviewers' comments. Specifically, the reviewers working in emerging economies are more likely to benefit from English proficiency.

Ortega noticed the weak relationship between reviewers' peer review activities and their research performance (Ortega, 2017, 2019). Our results reveal that publications and verified reviews are related to review length. Peer review experience may help reviewers write more informative reviews. However, the relationship between publications and review length varies for reviewers with different levels of publications. For reviewers with low publications, the relationship between publications and review length is positive. This result further suggests that reviewers who have not achieved high achievements may have a willingness to build nice relationships with editors. We considered gender effect in our analysis of publications and verified reviews, and revealed the role of femininity to some extent. Existing research has focused on the women's status or "penalization" in peer review (Card, DellaVigna, Funk & Iriberri, 2020; Pinho-Gomes et al., 2021; Squazzoni et al., 2021). The unique strengths of women have been confirmed in medical research, including the fact that female doctors are more patient-centered (Roter & Hall, 2004) and are less overconfidence (Tsugawa et al., 2017). However, the unique role and contribution of women in peer review has not been noted. Our study argues that women may play a more communicative role in peer review.

More importantly, our study reveals an association between culture and peer review to some extent. Reviewers influenced by Confucian culture tend to write shorter reviews, even if they are working in developed economies. This finding contributes to further understanding of reviewers' behavior. Edward T. Hall suggested that a high-context communication is one in which little has to be said or written because much of the information is in the objective environment, while very little is in explicit part of the message. A low-context communication is one in which the mass of information is vested in the explicit code, which is typical for individualist cultures. For example, American business contracts are much longer than Japanese business contracts (Hall, 1976; Hofstede & Hofstede, 2010). In terms to the uncertainty avoidance (Cyert & March 1963). Japanese and Chinese have higher tolerance for uncertainty than Westerners. It is socially acceptable to express one's opinions and show one's emotions in Western expressive cultures (Hofstede & Hofstede, 2010). This may give western reviewers more incentive to question and clarify doubts (Hofstede, 2001).

It is remarkable that western societies are not homogeneous and reviewers from different cultures work at universities. However, national culture is the shared, collective values that differentiates one group of individuals from another (Hofstede, 2011; Khosrowjerdi & Bornmann, 2021). As Hofsted emphasized in his classic monograph, *Cultures and Organizations, Software of the Mind*, "The values scores (culture) do not imply that all Africans are short-term thinkers, nor that all East Asians are long-term thinkers. They do mean that these ways of thinking are sufficiently general to affect common behavior patterns and the structure and functioning or malfunctioning of national institutions. Through these processes, thinking affects economic development." (Hofstede & Hofstede, 2010).

The majority of reviewers are still undertaking the task of reviewing manuscripts due to their sense of responsibility and academic ideals. However, the efforts of reviewers have not been fully recognized. Given that the increase in papers is much larger than the increase in reviewers, the phenomenon of "peer review fatigue" worries the academic community (Publons, 2018). Academics should improve the recognition of peer review in their academic system design, improve the design of the peer review system, and guide reviewers to put forward their review opinions.

Journals should fully recognize the contributions made by reviewers based on peer review quality, select the best peer reviewers of the year, and increase the chances that the best peer reviewers will become editorial board members. Regarding the academic evaluation system, especially in emerging economies, the academic contribution of reviewers should be recognized. Peer review information should be accepted gradually and incorporated into the evaluation system. Peer review should be converted into visible and measurable research results, and academic resources such as Publons should be fully utilized. Of course, due to the Goodhart's law (when an indicator becomes the basis for policy formulation, it is no longer valid), incorporating the academic contributions of reviewers precisely into evaluation systems is very important.

Peer review is a process of finding truth by extensively drawing on the wisdom of expert researchers in the relevant fields, providing a broader platform for academic exchanges and integrating different information. Academic progress would be hindered without scholarly communication or skepticism. Active expression of opinions is also important for gaining peer recognition and improving academic reputations. So, it is very important for researchers, especially those in emerging economies, to improve their English proficiency in order to fully express their opinions and strengthen peer review training. Providing suggestions actively and helping authors to improve the quality of manuscripts both are functions of peer review. Thus, researchers who are modest and conservative should improve their ability to express their opinions, even in the face of more established authors. It's essential to combat gender bias in peer review. Furthermore, making female reviewers' work more visible (i.e. Publons) is equally important. Moreover, promoting more gender diversity in editorial teams is also a necessary initiative to help the female reviewers who are underrepresented.

6.1. Limitation

We acknowledge that our research has several limitations. Firstly, Publons provides an approach to understanding peer review endeavor. However, the self-reported review records from Publons may be incomplete because of inactive users. The review length

on Publons was calculated based on the review text submitted by reviewers themselves (Xu et al., 2019). Reviewers may submit their review text selectively because the submission process was not mandatory. In addition, we can't distinguish the round (the first review or the final review) that submitted review texts belong to. This might lead to reviewers submitting different content to Publons. Moreover, review length may depend on how familiar the reviewer is to the content of the submission. Similarly, due to the availability of data, we could only obtain verified reviews displayed on Publons.

Second, the division of cultures could be improved by more literature investigation. The reason we chose to classify the cultures into Confucianism influenced culture, Christianism influenced culture, and Other was because the eastern/western classification of countries only allows dichotomous classification (Xu et al., 2019). It is worth mentioning that under the circumstance of extensive mobility of scientists, most of them may be cross-culture reviewer. It is also not enough to simply classify them into, for example, cultures that are directly or indirectly critical, or cultures that rely on high or low-context communication (Hall, 1976; Kim et al., 1998). In addition, it can be rough dividing reviewers into different countries and then into different cultures based on their names. However, this approach enables us to discuss differences in peer review across cultures without a better algorithm for dividing cultures. Perhaps a better culture recognition algorithm can help us carry out our research in the future.

Third, we drew on Publons (Publons, 2018) and Buljan's (Buljan et al., 2020) division of disciplines (HSS and HS) in our manuscript. Indeed, it's important to discuss interdisciplinary research in peer review. However, our data only support our division of disciplines into humanities and social sciences and hard sciences.

Finally, our research is only exploratory research. Although we can find relationships between review length and potential factors based on cross-sectional data, it is important to note that our findings are still confined to discussing the correlation between these factors. It needs the support of diachronic data for further causal inference.

7. Conclusion

This study contributes to exploring peer review endeavor in scientific publishing. In particular, we explored the associations between potential factors and review length. We found that reviewers with different genders, economic backgrounds, and cultural backgrounds wrote reviews of different lengths. Specifically, males, when controlling for other factors, wrote longer reviews than females. Second, reviewers in established economies wrote longer reviews compared to those in emerging economies. Third, reviewers from Western countries, which are countries with Christianism influenced culture, wrote longer reviews than reviewers from countries with Confucianism influenced culture and Other culture.

In addition, we found that disciplines, English Proficiency, publications, and verified reviews were all associated with review length. First, reviewers in humanities and social sciences wrote longer reviews than reviewers in hard sciences. Second, reviewers in countries/regions with higher English proficiency and more verified reviews wrote longer reviews. Third, reviewers with more publications wrote longer reviews. However, regarding reviewers with different publications levels, publications played a complex role: for high-publications reviewers, as publications increases, review length tends to decline, while for ordinary reviewers, as publications increases, review length tends to increase.

We also explored the interactions between factors. The interactions between gender and verified reviews, economic backgrounds and English proficiency, and economic backgrounds and cultural backgrounds were all associated with review length. Specifically, as verified reviews increases, review length of females tends to increase more than that of males. Second, English proficiency may increase review length for both countries with established and emerging economies, but it was more helpful to emerging economies. Finally, even among countries with different economic backgrounds, culture still associated with review length. Christianism influenced culture was associated with the longest review length, while Confucianism influenced culture was associated with the shortest review length.

Overall, this work lays the groundwork for future studies using Publons data. We take an initial step toward understanding review length as a potential indicator of the time and effort researchers devote to the process of "scientific gatekeeping" based on the one of the largest data sets on peer review so far. Investigating potential factors associated with review length increased additional evidence into how the endeavor of peer review is being performed.

CRediT authorship contribution statement

Guangyao Zhang: Methodology, Software, Formal analysis, Writing – original draft, Visualization, Investigation. **Shenmeng Xu:** Writing – original draft. **Yao Sun:** Investigation. **Chunlin Jiang:** Supervision. **Xianwen Wang:** Conceptualization, Methodology, Funding acquisition, Supervision.

Acknowledgments

We thank four anonymous reviewers for their insightful comments and suggestions.We would like to thank Dr. Zhigang Hu, Chao Song, Chunbo Zhang, Yuqi Wang, Jiao Ma, Xiaolan Zhang and Licheng Wang for their advice and support in data analysis and model design. This work is supported by National Natural Science Foundation of China (71974029).

An earlier version of this study was presented at the ASIS&T Annual Meeting 2019 proceedings held at Melbourne, Australia.

Appendix

See Table A1

Table A1
Regression results with truncated PU.

	appendix-m1		appendix-m2	
	PU_99th		PU_97.5th	
GE	0.033***	(0.01)	0.035***	(0.01)
EB	−0.162***	(0.01)	−0.165***	(0.01)
CBO	0.029**	(0.01)	0.029**	(0.01)
CBC	0.301***	(0.01)	0.303***	(0.01)
DI	0.208***	(0.01)	0.206***	(0.01)
EP	0.095***	(0.01)	0.095***	(0.01)
PU	−0.175***	(0.01)	−0.195***	(0.01)
VR	0.028***	(0.01)	0.030***	(0.01)
Cons	5.597***	(0.01)	5.592***	(0.01)
Obs	23,669		23,429	
Adj R^2	0.160		0.160	

References

Abelson, P. (1980). Scientific communication. *Science (New York, N.Y.), 209*(4452), 60–62. 10.1126/science.7280664.

Armstrong, A., Shereene, Z. I., Alexandra, B. K., & Jeffrey, D. B. (2008). Fate of manuscripts declined by the journal of the American academy of dermatology. *Journal of the American Academy of Dermatology, 58*(4), 632–635. 10.1016/j.jaad.2007.12.025.

Barnett, A., Mewburn, I., & Schroter, S. (2019). Working 9 to 5, not the way to make an academic living–Observational analysis of manuscript and peer review submissions over time. *British Medical Journal, 367*, Article l6460. 10.1136/bmj.l6460.

Bernstein, J. (2013). Free for service–The inadequate incentives for quality peer review. *Clinical Orthopaedics and Related Research, 471*(10), 3093–3097. 10.1007/s11999-013-3216-z.

Bianchi, F., Grimaldo, F., & Squazzoni, F. (2019). The F-3-index. Valuing reviewers for scholarly journals. *Journal of Informetrics, 13*(1), 78–86. 10.1016/j.joi.2018.11.007.

Borgman, C. L., & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology, 36*(1), 2–72. 10.1002/aris.1440360102.

Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology, 45*(1), 197–245. 10.1002/aris.2011.1440450112.

Bornmann, L. (2014). h-Index research in scientometrics–A summary. *Journal of Informetrics, 8*(3), 749–750. 10.1016/j.joi.2014.07.004.

Bravo, G., Farjam, M., Grimaldo Moreno, F., Birukou, A., & Squazzoni, F. (2018). Hidden connections–Network effects on editorial decisions in four computer science journals. *Journal of Informetrics, 12*(1), 101–112. 10.1016/j.joi.2017.12.002.

Bravo, G., Grimaldo, F., López-Iñesta, E., Mehmani, B., & Squazzoni, F. (2019). The effect of publishing peer review reports on referee behavior in five scholarly journals. *Nature Communications, 10*(1), Article 322. 10.1038/s41467-018-08250-2.

Bruce, R., Chauvin, A., Trinquart, L., Ravaud, P., & Boutron, I. (2016). Impact of interventions to improve the quality of peer review of biomedical journals–A systematic review and meta-analysis. *BMC Medicine, 14*(1), Article 85. 10.1186/s12916-016-0631-5.

Buljan, I., Garcia-Costa, D., Grimaldo, F., Squazzoni, F., & Marušić, A. (2020). Large-scale language analysis of peer review reports. *eLife, 9*, Article e53249. 10.7554/eLife.53249.

Card, D., DellaVigna, S., Funk, P., & Iriberri, N. (2020). Are referees and editors in economics gender neutral? *The Quarterly Journal of Economics, 135*(1), 269–327. 10.1093/qje/qjz035.

Casnici, N., Grimaldo, F., Gilbert, N., Dondio, P., & Squazzoni, F. (2017a). Assessing peer review by gauging the fate of rejected manuscripts–The case of the Journal of Artificial Societies and Social Simulation. *Scientometrics, 113*(1), 533–546. 10.1007/s11192-017-2241-1.

Casnici, N., Grimaldo, F., Gilbert, N., & Squazzoni, F. (2017b). Attitudes of referees in a multidisciplinary journal–An empirical analysis. *Journal of the Association for Information Science and Technology, 68*(7), 1763–1771. 10.1002/asi.23665.

Chubin, D. E., & Hackett, E. J. (1990). *Peerless science–Peer review and US science policy*. Albany: State University of New York Press.

Cole, J., & Zuckerman, H. (1984). The productivity puzzle–Persistence and change in patterns of publication of men and women scientists. *Advances in motivation and achievement*. Greenwich: JAI Press Inc.

Cole, S. (1994). *Making science–Between nature and society*. Cambridge: Harvard University Press.

Copiello, S. (2018). On the money value of peer review. *Scientometrics, 115*(1), 613–620. 10.1007/s11192-018-2664-3.

Crawford, S., & Stucki, L. (1990). Peer review and the changing research record. *Journal of the American Society for Information Science, 41*(3), 223–228 3<223::AID-ASI14>3.0.CO;2-3. 10.1002/(SICI)1097-4571(199004)41.

Cyert, R. M., & March, J. G. (1963). *A behavioral theory of the firm*. Englewood Cliffs: Prentice-Hall.

Dalton, M. S. (1995). Refereeing of scholarly works for primary publishing. *Annual Review of Information Science and Technology, 30*, 213–250.

De Nicola, A., & D'Agostino, G. (2021). Assessment of gender divide in scientific communities. *Scientometrics, 126*(5), 3807–3840. 10.1007/s11192-021-03885-3.

Ding, J., Liu, C., & Kandonga, G. A. (2020). Exploring the limitations of the h-index and h-type indexes in measuring the research performance of authors. *Scientometrics, 122*(3), 1303–1322. 10.1007/s11192-020-03364-1.

Engers, M., & Gans, J. S. (1998). Why referees are not paid (enough). *The American Economic Review, 88*(5), 1341–1349.

Ernst, E., Saradeth, T., & Resch, K. L. (1993). Drawbacks of peer review. *Nature, 363*(6427), 296 296–296. 10.1038/363296a0.

Eysenck, H., & Eysenck, S. (1992). Peer-Review–Advice to referees and contributors. *Personality and Individual Differences, 13*(4), 393–399. 10.1016/0191-8869(92)90066-X.

Fang, Q., Lifang, X., & Xiaochuan, L. (2008). Peer-Review practice and research for academic journals in China. *Journal of Scholarly Publishing, 39*(4), 417–427. 10.3138/jsp.39.4.417.

Floyd, C. B. (2011). Critical thinking in a second language. *Higher Education Research & Development, 30*(3), 289–302. 10.1080/07294360.2010.501076.

Garcıa-Costa, D., Squazzoni, F., Mehmani, B., & Grimaldo, F. (2021). Measuring the developmental function of peer review–A multi-dimensional, cross-disciplinary analysis of peer review reports from 740 academic journals. *SSRN.*. 10.2139/ssrn.3912607.

Godlee, F., Gale, C. R., & Martyn, C. N. (1998). Effect on the quality of peer review of blinding reviewers and asking them to sign their reports–A randomized controlled trial. *JAMA: The Journal of the American Medical Association, 280*(3), 237–240. 10.1001/jama.280.3.237.

Goodman, S. N., Berlin, J., Fletcher, S. W., & Fletcher, R. H. (1994). Manuscript quality before and after peer review and editing at annals of internal medicine. *Annals of Internal Medicine, 121*(1), 11–21. 10.7326/0003-4819-121-1-199407010-00003.

Gu, X., Yang, H., Tang, J., & Zhang, J. (2016). Web user profiling using data redundancy. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 358–365).

Gutek, B. A., Nakamura, C. Y., & Nieva, V. F. (1981). The interdependence of work and family roles. *Journal of Organizational Behavior, 2*(1), 1–16. 10.1002/job.4030020102.

Human development report 2010–The real wealth of nations–Pathways to human development (2010). New York, UNDP.
Hall, E.T. (.1976). Beyond Culture. Doubleday.
Helmer, M., Schottdorf, M., Neef, A., & Battaglia, D. (2017). Gender bias in scholarly peer review. ELife, 6, e21718. 10.7554/eLife.21718
Hinkel, E. (2011). What research on second language writing tells us and what it doesn't. *Handbook of Research in Second Language Teaching and Learning, 2*. 10.4324/9780203836507.
Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. In *Proceedings of the national academy of sciences of the United States of America: 102* (pp. 16569–16572). Proceedings of the national academy of sciences of the United States of America. 10.1073/pnas.0507655102.
Hirsch, J. E., & Buela-Casal, G. (2014). The meaning of the h-index. *International Journal of Clinical and Health Psychology, 14*(2), 161–164. 10.1016/S1697-2600(14)70050-X.
Ho, R.C.-M., Mak, K.-K., Tao, R., Lu, Y., Day, J.R., .& Pan, F. (2013). Views on the peer review system of biomedical journals–An online survey of academics from high-ranking universities. *BMC Medical Research Methodology*, 13(1), 1–15. 10.1186/1471-2288-13-74
Hofstede, G. (2001). *Culture's consequences–Comparing values, behaviors, institutions and organizations across nations*. Thousand Oaks, CA: Sage.
Hofstede, G. (2011). Dimensionalizing cultures–The hofstede model in context. *Online Readings in Psychology and Culture, 2*(1). 10.9707/2307-0919.1014.
Hofstede, G., & Hofstede, G. J. (2010). *Cultures and organizations, software of the mind*. New York: McGraw Hill Professional.
Jefferson, T., Wager, E., & Davidoff, F. (2002). Measuring the quality of editorial peer review. *JAMA: The Journal of the American Medical Association, 287*(21), 2786–2790. 10.1001/jama.287.21.2786.
Jones, G. K., & Davis, H. J. (2000). National culture and innovation–Implications for locating global R&D operations. *MIR: Management International Review, 40*(1), 11–39.
Kedia, B. L., Keller, R. T., & Jullan, S. D. (1992). Dimensions of national culture and the productivity of R&D units. *The Journal of High Technology Management Research, 3*(1), 1–18. 10.1016/1047-8310(92)90002-J.
Khosrowjerdi, M., & Bornmann, L. (2021). Is culture related to strong science? An empirical investigation. *Journal of Informetrics, 15*(4), Article 101160. 10.1016/j.joi.2021.101160.
Kim, D., Pan, Y., & Park, H. S. (1998). High-versus low-context culture–A comparison of Chinese, Korean, and American cultures. *Psychology & Marketing, 15*(6), 507–521 6<507::AID–MAR2>3.0.CO;2-A. 10.1002/(SICI)1520-6793(199809)15.
King, D. A. (2004). The scientific impact of nations. *Nature, 430*(6997), 311–316. 10.1038/430311a.
King, J. (1987). A review of bibliometric and other science indicators and their role in research evaluation. *Journal of Information Science, 13*(5), 261–276. 10.1177/016555158701300501.
Kottak, C. (2013). *Anthropology–Appreciating human diversity*. New York: McGraw-Hill.
Laband, D. N. (1990). Is there value-added from the review process in economics?–Preliminary evidence from authors*. *The Quarterly Journal of Economics, 105*(2), 341–352. 10.2307/2937790.
Maltz, D. N., & Borker, R. A. (1983). A cultural approach to male–female miscommunication. In *Language and social identity* (pp. 196–216). Cambridge: Cambridge University Press. 10.1017/CBO9780511620836.013.
Matsui, A., Chen, E., Wang, Y., & Ferrara, E. (2021). The impact of peer review on the contribution potential of scientific papers. *PeerJ, 9*, Article e11999. 10.7717/peerj.11999.
Merrill, E. (2014). Reviewer overload and what can we do about it. *The Journal of Wildlife Management, 78*(6), 961–962. 10.1002/jwmg.763.
Mulligan, A., Hall, L., & Raphael, E. (2013). Peer review in a changing world–An international study measuring the attitudes of researchers. *Journal of the American Society for Information Science and Technology, 64*(1), 132–161. 10.1002/asi.22798.
Nisbett, R. E., Peng, K. P., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought–Holistic versus analytic cognition. *Psychological Review, 108*(2), 291–310. 10.1037//0033-295X.108.2.291.
Ortega, J. L. (2017). Are peer-review activities related to reviewer bibliometric performance? A scientometric analysis of Publons. *Scientometrics, 112*(2), 947–962. 10.1007/s11192-017-2399-6.
Ortega, J. L. (2019). Exploratory analysis of Publons metrics and their relationship with bibliometric and altmetric impact. *Aslib Journal of Information Management, 71*(1), 124–136. 10.1108/AJIM-06-2018-0153.
Paltridge, B. (2015). Referees' comments on submissions to peer-reviewed journals–When is a suggestion not a suggestion? *Studies in Higher Education, 40*(1), 106–122. 10.1080/03075079.2013.818641.
Pedro Couto, J., & Cabral Vieira, J. (2004). National culture and research and development activities. *Multinational Business Review, 12*(1), 19–36. 10.1108/1525383X200400002.
Phillipson, R. (2018). Linguistic imperialism. In *The encyclopedia of applied linguistics* (pp. 1–7). New Jersey: Wiley-Blackwell. 10.1002/9781405198431.wbeal0718.pub2.
Pinho-Gomes, A.-. C., Vassallo, A., Thompson, K., Womersley, K., Norton, R., & Woodward, M. (2021). Representation of women among editors in chief of leading medical journals. *JAMA Network Open, 4*(9), Article e2123026. 10.1001/jamanetworkopen.2021.23026.
Proper Names And Translation Service, Xinhua News Agency. (2007). *Names of the world's peoples–A comprehensive dictionary of names in Roman-Chinese (revised edition)*. Beijing: ChinaTranslation & Publishing Corporation https://book.douban.com/subject/2149292/.
Publons. (2018). *Publons' global state of peer review 2018*. London: Publons. 10.14322/publons.GSPR2018.
Righi, S., & Takács, K. (2017). The miracle of peer review and development in science–An agent-based model. *Scientometrics, 113*(1), 587–607. 10.1007/s11192-017-2244-y.
Rodríguez-Bravo, B., Nicholas, D., Herman, E., Boukacem-Zeghmouri, C., Watkinson, A., Xu, J., et al. (2017). Peer review–The experience and views of early career researchers. *Learned Publishing, 30*(4), 269–277. 10.1002/leap.1111.
Roter, D. L., & Hall, J. A. (2004). Physician gender and patient-centered communication–A critical review of empirical research. *Annual Review of Public Health, 25*, 497–519. 10.1146/annurev.publhealth.25.101802.123134.
Rowland, F. (2002). The peer-review process. *Learned Publishing, 15*(4), 247–258. 10.1087/095315102760319206.
Shatz, D. (2004). *Peer review–A critical inquiry*. Lanham: Rowman & Littlefield Publishers.
Siler, K., Lee, K., & Bero, L. (2015). Measuring the effectiveness of scientific gatekeeping. *Proceedings of the national academy of sciences of the United States of America, 112*(2), 360–365. 10.1073/pnas.1418218112.
Sollenberger, J. F. (2002). Editorial peer review–Its strengths and weaknesses. *Journal of the Medical Library Association, 90*(1), 114–116.
Squazzoni, F., Bravo, G., Farjam, M., Marusic, A., Mehmani, B., Willis, M., et al. (2021). Peer review and gender bias–A study on 145 scholarly journals. *Science Advances, 7*(2), Article eabd0299. 10.1126/sciadv.abd0299.
Squazzoni, F., Brezis, E., & Marusic, A. (2017). Scientometrics of peer review. *Scientometrics, 113*(1), 501–502. 10.1007/s11192-017-2518-4.
Squazzoni, F., & Gandelli, C. (2013). Opening the black-box of peer review–An agent-based model of scientist behavior. *Journal of Artificial Societies and Social Simulation, 16*(2), 1–3.
Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., & Su, Z. (2008). ArnetMiner–Extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 990–998). 10.1145/1401890.1402008.
Tannen, D. (1992). *You just don't understand*. New York: William Morrow Paperbacks.
Teplitskiy, M., Acuna, D., Elamrani-Raoult, A., Körding, K., & Evans, J. (2018). The sociology of scientific validity–How professional networks shape judgement in peer review. *Research Policy, 47*(9), 1825–1841. 10.1016/j.respol.2018.06.014.
Thorne, B., & Henley, N. (1975). *Language and sex–Difference and dominance*. Rowley: Newbury House Publishers, Inc.
Tsugawa, Y., Jena, A. B., Figueroa, J. F., Orav, E. J., Blumenthal, D. M., & Jha, A. K. (2017). Comparison of hospital mortality and readmission rates for medicare patients treated by male vs female physicians. *JAMA Internal Medicine, 177*(2), 206–213. 10.1001/jamainternmed.2016.7875.

van Rooyen, S., Godlee, F., Evans, S., Black, N., & Smith, R. (1999). Effect of open peer review on quality of reviews and on reviewers' recommendations–A randomized trial. *British Medical Journal, 318*(7175), 23–27.

Viiu, G. A. (2016). A theoretical evaluation of Hirsch-type bibliometric indicators confronted with extreme self-citation. *Journal of Informetrics, 10*(2), 552–566. 10.1016/j.joi.2016.04.010.

Wang, X., Peng, L., Zhang, C., Xu, S., Wang, Z., Wang, C., et al. (2013). Exploring scientists' working timetable–A global survey. *Journal of Informetrics, 7*(3), 665–675. 10.1016/j.joi.2013.04.003.

Wang, X., Xu, S., Peng, L., Wang, Z., Wang, C., Zhang, C., et al. (2012). Exploring scientists' working timetable–Do scientists often work overtime? *Journal of Informetrics, 6*(4), 655–660. 10.1016/j.joi.2012.07.003.

Wei, C., Bu, Y., Kang, L., & Li, J. (2021). Directionality of paper reviewing and publishing of a scientist–A Granger causality inference. *Data Science and Informetrics, 1*(01), 68–80.

Wicherts, J. M. (2016). Peer review quality and transparency of the peer-review process in open access and subscription journals. *PloS one, 11*(1), Article e0147913. 10.1371/journal.pone.0147913.

Xie, Y., & Kimberlee, A. S. (2003). *Women in science–Career processes and outcomes*. Cambridge: Harvard University Press.

Xu, S., Zhang, G., Sun, Y., & Wang, X. (2019). Understanding the peer review endeavor. In *Proceedings of the Association for Information Science and Technology: 56* (pp. 316–325). 10.1002/pra2.26.

Yair, G. (2020). Hierarchy versus symmetry in German and Israeli science. *American Journal of Cultural Sociology, 8*(2), 214–245. 10.1057/s41290-019-00069-8.

Yum, J. (1988). The impact of confucianism on interpersonal relationships and communication patterns in East-Asia. *Communication Monographs, 55*(4), 374–388. 10.1080/03637758809376178.

Zhang, G., Wang, L., Xie, W., Shang, F., Xia, X., Jiang, C., et al. (2021a). This article is interesting, however"–Exploring the language use in peer review comment of articles published in the BMJ. *Aslib Journal of Information Management* ahead-of-print. 10.1108/AJIM-06-2021-0172.

Zhang, G., Wang, Y., Xie, W., Du, H., Jiang, C., & Wang, X. (2021b). The open access usage advantage–A temporal and spatial analysis. *Scientometrics, 126*(7), 6187–6199. 10.1007/s11192-020-03836-4.

Zhang, L., Shang, Y., Huang, Y., & Sivertsen, G. (2022). Gender differences among active reviewers: An investigation based on Publons. *Scientometrics, 127*, 145–179. https://link.springer.com/article/10.1007/s11192-021-04209-1.