CEB Working Paper

Centre Emile Bernheim
Research Institute in Management Sciences

Solvay Brussels School
Economics & Management

Demographic factors as determinants of social scientists' productivity: SDC-based Scopus data, 2008-2017

Thu-Trang Vuong, Hong Kong T. Nguyen, Tung Manh Ho and Quan-Hoang Vuong

As collaboration became widespread in academia and educational systems, and the number of authors per article increased, the count of publication is no longer an accurate indicator of scientific output in many cases. In order to overcome this limitation, the study defined and computed a relative count of publications called 'CP' (credit-based contribution point), based on the sequence-determines-credit (SDC) method, as a proxy for the output of authors taking into account their level of contribution of each author. Analyses were done on a sample of 410 Vietnamese social scientists whose publications have been indexed in the Scopus database during 2008-2017. Results showed that the average CP of Vietnamese researchers in the field of social sciences and humanities is very low: more than 88% of authors have CP<5 over a span 10 years. Researchers with higher CP were mostly 40-50 years old; however, even for this sub-group, mean CP only amounted to 3.07. Multiple factors – including knowledge, research skills, critical thinking and comprehensiveness – could boost CP by a ratio of 1:1.06. In addition, there is no evidence of gender differences in average CP values, however a regional difference. These findings offer significant insights into the education system in what concerns science and technology, namely policy implications for science funding and management strategies for research funds.

Keywords: scientific productivity; social sciences; Vietnam; science funds management; administration.

JEL Classifications: D85, D91, Q01.

CEB Working Paper N° 17/030
November 2017

Université Libre de Bruxelles - Solvay Brussels School of Economics and Management
Centre Emile Bernheim
ULB CP114/03 50, avenue F.D. Roosevelt 1050 Brussels BELGIUM
e-mail: ceb@admin.ulb.ac.be Tel.: +32 (0)2/650.48.64 Fax: +32 (0)2/650.41.88

Demographic factors as determinants of social scientists' productivity: SDC-based Scopus data, 2008-2017

Thu-Trang Vuong [1], Hong Kong T. Nguyen [2], Tung Manh Ho [3,4] and Quan-Hoang Vuong [4,5,*]

[1] Sciences Po Paris, Campus de Dijon; thutrang.vuong@sciencespo.fr;
[2] Vietnam Panorama Media Monitoring, Hanoi, Vietnam; hongkongt.nguyen@gmail.com;
[3] Institute of Philosophy, Vietnam Academy of Social Sciences, Hanoi, Vietnam; tung.ho@wu.edu.vn;
[4] Centre for Interdisciplinary Social Research, Western University Hanoi, Vietnam;
[5] Centre Emile Bernheim, Université Libre de Bruxelles, Bruxelles 1050, Belgium;

* Correspondence: qvuong@ulb.ac.be;

Abstract: As collaboration became widespread in academia and educational systems, and the number of authors per article increased, the count of publication is no longer an accurate indicator of scientific output in many cases. In order to overcome this limitation, the study defined and computed a relative count of publications called 'CP' (credit-based contribution point), based on the sequence-determines-credit (SDC) method, as a proxy for the output of authors taking into account their level of contribution of each author. Analyses were done on a sample of 410 Vietnamese social scientists whose publications have been indexed in the Scopus database during 2008-2017. Results showed that the average CP of Vietnamese researchers in the field of social sciences and humanities is very low: more than 88% of authors have CP<5 over a span 10 years. Researchers with higher CP were mostly 40-50 years old; however, even for this sub-group, mean CP only amounted to 3.07. Multiple factors – including knowledge, research skills, critical thinking and comprehensiveness – could boost CP by a ratio of 1:1.06. In addition, there is no evidence of gender differences in average CP values, however a regional difference. These findings offer significant insights into the education system in what concerns science and technology, namely policy implications for science funding and management strategies for research funds.

Keywords: scientific productivity; social sciences; Vietnam; science funds management; administration.

JEL Classification: D85, D91, Q01,

This is an unpublished working manuscript

November 6, 2017

Demographic factors as determinants of social scientists' productivity: SDC-based Scopus data, 2008-2017

1. Introduction

1.1. Overview

Scientific research is the driving force in the development of a country and the pillar of higher education systems. According to a 2001-2005 study, Vietnam only ranked 4th among 10 ASEAN countries in terms of number of scientific research articles [1]. Hien (2010) also analyzed data from ISI journals to conclude that Vietnamese scientific output is low compared to East Asian countries and depends heavily on foreign authors [2], despite the fact that Vietnam was considered to be in a developmental, high-output stage [3-5]. Given the scant literature on this topic, especially in the context of Vietnam, a study on the performance of Vietnamese scientists would prove to be useful in assessing the true state of development of Vietnamese academia.

1.2. Literature review

Countless factors come into play when it comes to determining the sources of influence on productivity, especially in science [5-6].

Age, one such factor, is indeed of great interest in the study of scientific productivity; research conducted on this subject has reached numerous conclusions. Scientific productivity tends to be highest when scientists are in their thirties and forties then declines as they age further [7-10]. Explanations to this issue vary from author to author: cumulative advantage theory, benefit maximization theory, and intellectual impairment. As a strategic choice, scientists will reduce research productivity when they perceive a lack of expert recognition from the community [11]. In addition, the theory of maximizing benefits can also explain the decrease in productivity when scientists find that other jobs may be more beneficial for their promotion or income [12]. Lehman (1953) meanwhile concluded that the decline in scientific productivity with age may be due to intellectual weakness [7]; in a similar vein, Cooney et al (1988) argued that older scientists become obsolete compared to their younger colleagues [13]. The contradicting results between studies show that the effect of age on scientific productivity is unclear, especially when older authors can play a role in guiding young researchers and can thus contribute without directly writing articles [14].

Geographical factors, in some studies, also made a difference in research output among scientists. Most studies have shown that authors hailing from regions with underdeveloped economic conditions and/or limited communication networks and Internet tend to have a more modest body of published articles [15-17]. Prpić (2002) also concluded that female scientists are more concentrated in urban than in rural areas, demonstrating that highly productive scientists are often concentrated in economically developed regions [18]. However, the causes and effects of regional factors and its relationship to scientific productivity still receive little attention in the extant literature.

The issue of whether or not gender is a factor affecting scientific productivity is perhaps the most controversial. The majority of previous studies concluded that male researchers are more productive than females [19-21]. In the curriculum from primary school to university, key scientists are predominantly males, and only a very small number are women [22]. To explain this difference, many studies have suggested that childbirth and child care are the most important determinants of the productivity of female scientists [12, 23-25]. A study by Kyvik (1996) found that: Women with children under 11 years old published 41% less articles than men; the gap dropped to 8% with women and men with children over 10 years of age [26]. Women with children under 11 years old work 5.5 hours per week less than male counterparts in the same situation, and there is no difference in work hour between men and women with children over 10 years old. However, there are also results that deny the difference in scientific productivity between male and female authors [27-29]. Lemoine (1992) reported the same number of women as men in national journals publications in Venezuela [30]. In the fields of science, technology or mathematics, men do not surpass women in the quantity and quality of scientific research, although these are not traditionally considered to be areas of expertise of women [31]. In a study by Van Arensbergen et al. (2012), it has been observed in their samples of scientists that in the previous generation, women only took 22%; today, the figure has increased to 45%. There are also fields more dominated by women than men, such as psycho-social studies [32].

This study will provide insights into the effects of gender, age, research time and region on the relative scientific output of Vietnamese authors, continuing the line of recent studies such as [4, 33].

2. Results

2.1. Descriptive statistics

The majority (238 out of 410 scientists) came from northern Vietnam, accounting for more than 58%. The remainder came from the South and Central regions, and about 17% of the sample were Vietnamese scientists living abroad.

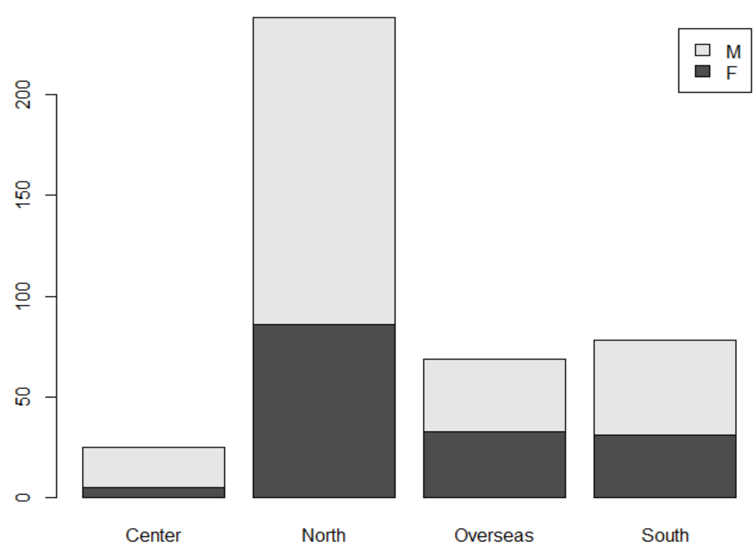


Figure 1. Regional and gender distribution of scientists

The number of male authors was higher, whether in general or by regions (Fig.1). In general, the number of male scientists accounted for 62.2%, corresponding to 255/410 people. Whether or not this translated to a male dominance in the Vietnamese scientific community would be concluded after the analyses.

Author contribution could be expressed in nominal and relative productivity. In numerical values, the highest publication count was 63 articles, in stark contrast to the modest average of only 3.6 articles (Table 1). To add to it, over 300 authors were below average in terms of productivity. Meanwhile, in relative values, average productivity was 2.405, maximum 61.16 points.

Table 1. Descriptive statistics for variables in the study

Variable	Min	Mean	Max	SD	95%CI	p-value
Adjusted productivity ("*cp*")	0.05	2.40	61.16	4.62	1.96–2.85	2.2×10^{-16}
Nominal productivity ("*ttlitems*")	1	3.60	63	5.89	3.03–4.17	2.2×10^{-16}
# articles in leading role ("*au.key*")	0	1.77	60	4.24	1.36–2.18	4.5×10^{-05}
Age ("*age*")	19	42.10	72	9.15	41.21–42.99	2.2×10^{-16}

Figure 1 showed the relationship between relative and nominal productivity by gender. It could be observed that most authors with a high number of published articles also had a high relative productivity, which led to believe that there existed a positive relationship between number of publications and author contribution in the publications. In fact, the correlation coefficient between the two factors was 0.926. However, given that each value represented a different aspect of author productivity, their relationship should be examined with care.

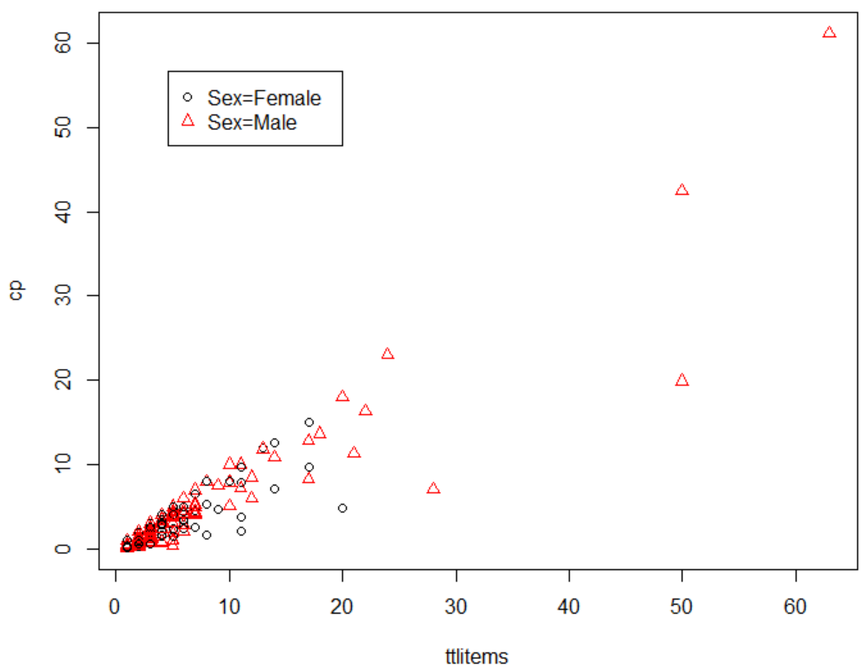


Figure 2. Relationship between nominal productivity and relative productivity by sex

Fig. 2 also showed that the two peripheral data points had a much higher nominal productivity ("*ttlitems*") value (≥50) than the rest (<30). Furthermore, both peripheral points turned out to be male scientists, thus creating a strong dispersion of nominal and relative productivity in terms of gender in this field. Concretely, men's average CP was 2.37 (p=6.39×10^{-11}), significantly higher than that of women, 1.79 (p=11×10^{-15}).

We then removed the aforementioned outlying data points from the dataset and obtained new means, reported in Table 2 below:

Table 2. Descriptive statistics for variables after the removal of peripheral data point

Variable	Min	Mean	Max	SD	95%CI	p-value
Adjusted productivity ("*cp*")	0.05	2.119	23	2.872	1.840 – 2.399	2.2×10^{-16}
Nominal productivity ("*ttlitems*")	1	3.226	28	3.934	2.843 – 3.609	2.2×10^{-16}
# articles in leading role ("*au.key*")	0	1.523	22	2.543	1.278 – 1.771	2.2×10^{-16}
Age ("*age*")	19	42.06	72	9.171	41.168 – 42.955	2.2×10^{-16}

Comparing the figures in Table 1 and Table 2, it could be seen that the means and standard deviations decreased after the removal of extreme outliers. Nominal productivity ("*ttlitems*") had the most significant drop, followed by relative productivity ("*cp*"): their respective means decreased by 0.374 and 0.286, their respective standard deviations by 1.953 and 1.752. This suggested that peripheral values might alter the general character of the dataset.

On the other hand, details on the contributions of scientists in publications based on age groups were reported in Table 3.

Table 3. Distribution of authors by age group and relative productivity

Age group	<30		[30,40)		[40,50)		≥50		**Total**	
CP	F	CS	F	CS	F	CS	F	CS	# people	Proportion
<1	6	6	61	61	49	49	34	34	**150**	**36.59%**
[1,2)	4	10	45	106	43	92	26	60	**118**	**28.78%**
[2,3)	1	11	19	125	18	110	12	72	**50**	**12.20%**
[3,4)	1	12	9	134	14	124	0	72	**24**	**5.85%**
[4,5)	0	12	6	140	10	134	3	75	**19**	**4.63%**
[5,6)	0	12	5	145	8	142	0	75	**13**	**3.17%**
[6,7)	0	12	0	145	1	143	1	76	**2**	**0.49%**
[7,∞)	0	12	5	155	22	187	7	90	**34**	**8.29%**

Notes: F: Frequency; CS: Cumulative sum

Table 3 showed quantities of authors by age group and relative productivity. The majority of authors (36.59%) scored less than 1 in relative productivity. Within this majority, no author had published more than 5 articles.

In all age groups, the proportion of scientists with a relative productivity under 1 was the largest. Furthermore, as relative scientific output ("*cp*") gradually increased from 1-2 points to 5-6 points, the number of authors decreased. It is worth noting that, from 7 points of relative productivity and higher, the 40-to-50 age group was the largest (21 authors).

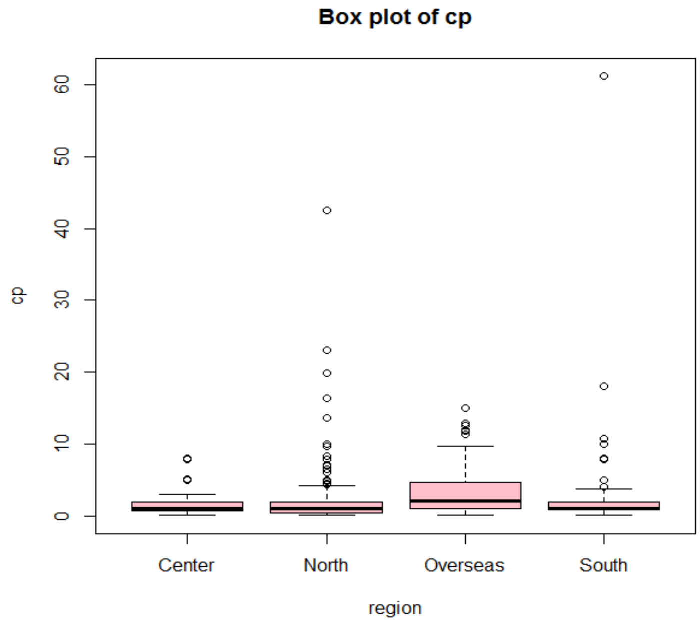


Figure 3. Relative productivity distribution by region

Fig. 3 depicted relative productivity by region, which indicated the dispersion of "*cp*" values in different categories of "*region*". Overseas authors had the highest relative productivity – 3.58 points, which is above average – and are also the most dispersed in terms of relative productivity.

These basic descriptive statistics provided background evaluations regarding relative productivity and influential factors. Regression models shall help bring about more specific assessments.

2.2. Estimation results

2.2.1. The influence of age and first-authorship on relative scientific productivity

To test the first hypothesis, relative scientific productivity ("*cp*") was examined against age groups ("*age_gr*") and first-authorship ("*au.key*"). While "*au.key*" was a continuous variable, "*age_gr*" was a categorical variable and was divided into four groups: 30 years and under ("less30"), from 30 to less than 40 years old ("b3040"), from 40 to less than 50 years old ("b4050"), and 50 years old and above ("g50").

After data were processed in R, the following results were obtained:

Table 4. Estimation results of "*cp*" against "*age_gr*" and "*au.key*"

	Intercept	"*au.key*"	"*Age_gr*"		
			"b4050"	"g50"	"less30"
	β_o	β_1	β_2	β_3	β_4
"*cp*"	0.403^{***} [4.687] (3.8×10^{-6})	1.060^{***} [86.876] (2×10^{-16})	0.243^{*} [2.064] (0.0396)	0.146 [1.025] (0.3058)	-0.136 [-0.437] (0.6626)

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1; *z*-value in square brackets; *p*-value in round brackets; baseline category for: "*Age_gr*"="b3040". Residual *s.e.*=1.038 on 405 *df*. Multiple R^2=0.9501, Adj. R^2= 0.9496 . *F*-stat= 1929 on 4 and 405 *df*, $p<2.2\times10^{-16}$.

Table 4 showed that first-authorship ("*au.key*") had a statistically significant coefficient. For the age variable ("*age_gr*"), only the 40-to-50 age group had a statistically significant coefficient. Since all independent variables had at least one significant coefficient, the regression model could be accepted. The first hypothesis, maintaining first-authorship and age had an impact on relative productivity was thus confirmed.

Using the regression results, we constructed an equation expressing the relationship between variables as follows:

$$cp = 0.403 + 1.060 \times au.key + 0.243 \times b4050$$

Note that only the 40-to-50 age group could be included in the equation, the other two admitting statistically insignificant coefficients. The coefficient of greatest significance (β_1) was that of the variable "*au.key*" and was positive, indicating that, one additional publication in which the scholar was the lead author would increase relative productivity by 1,060 point (other factors being kept constant). Using this equation, if a 45-year-old author played the leading role in two articles, the estimated relative productivity could be calculated as follows:

$$0.403 + 1.060 \times 2 + 0.243 \times 1 = 2.766$$

On another note, the insignificance of the other two age groups in the regression model, "less30" and "g50", was telling in itself, and shall be discussed towards the end of the article.

2.2.2. The influence of gender and region on relative scientific productivity

In this model, factors considered to have an impact on relative productivity included regional ("*region*") and gender ("*sex*"). Regression models similar to above were used, yielding the following estimation results:

Table 5. Estimation results of "*cp*" against "*Region*" and "*Sex*"

	Intercept	"*Sex*" "M"	"*Region*" "North"	"Overseas"	"South"
	β_0	β_1	β_2	β_3	β_4
"*cp*"	1.550 [1.560] (0.119)	0.724 [1.534] (0.126)	-0.032 [-0.033] (0.974)	1.655. [1.532] (0.126)	0.764 [0.721] (0.471)

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1; *z*-value in square brackets; *p*-value in round brackets; baseline category for: "*Sex*"="F", "*Region*"="Center". Residual *s.e.*=4.594 on 405 *df*. Multiple R^2=0.0228, Adj. R^2=0.0132. *F*-stat=2.363 on 4 and 405 *df*, *p*-value: 0.0525.

Results from Table 5 showed that no regression coefficients were statistically significant (in all cases). On the other hand, the *F*-test performed on the null hypothesis H_0 (coefficients not simultaneously equal to zero) suggests that the null hypothesis could not be rejected decisively at 10% conventional level. In other words, it was impossible to verify that the coefficients of "*Sex*" and "*Region*" were not simultaneously equal to zero. Hypothesis 2 was thus rejected: the gender of the scholar and the geographic region where the scholar was based had no empirically verifiable effect on their scientific output in terms of relative productivity.

2.2.3. Verifying the impact of peripheral observations on the overall character

Results from the two above models have shown that only first-authorship had a general impact on productivity, while age only affected scientific contribution for one single age group, and gender and region did not influence author output, in terms of relative productivity. However, could these results be due to peripheral data points? To answer this question, we eliminated outliers (3 observations with a nominal productivity of ≥50 works) and reran the models (Appendix B) on the modified dataset. Results were then compared with those of the two original models.

It was remarked that the number of significant variables obtained in the new regression models was the same. More specifically, in the model without outliers, results also showed that the effect of first-authorship ("*au.key*") and age ("*age_gr*") among the 40-to-50-year-old age group (at "b4050") on relative productivity ("*cp*") were statistically significant with . The magnitude of the coefficients corresponding to these two variables, "*au.key*" and "*age_gr*" were also not significantly different from that of the original models. It could be affirmed that the outlying observations did not affect the general characteristics of the dataset. Therefore, the above results were reliable.

In addition, it could be seen that the values of R^2 and adjusted R^2 in the original regression models were higher than those of the new models (i.e., outliers removed). This showed that the outlying values contributed significantly to the increased meaningful interpretation of the model.

3. Discussion

By analyzing images and statistics, we hope this article would reach valuable conclusions on the factors affecting scientific output in terms of relative productivity of Vietnamese authors and the current status of academia in Vietnam.

First, results obtained from the first regression model indicated that age only had an effect on relative productivity for scientists of 40 to fewer than 50 years old. This was also the largest category, grouping 165 out of 410 scholars (over 40%). Furthermore, results also showed that authors aged between 40 and under 50 had a higher mean relative productivity (3.07, with p=1.96×10^{-14}), had the largest average nominal productivity (4.41 articles) and largest average number of articles in first-authorship (2.29 articles). Meanwhile, all other age groups admitted insignificant coefficients, suggesting that first-authorship and age had no effect on scientists outside of the 40 to 50 age range. This curious occurrence could be explained by various reasons. It should first be remarked that, as said above, the majority of authors in our sample belonged to this age category; thus, it is possible that the other groups were not large enough to show the impact of these factors. Technical aspects aside, authors in the 40-50 age range were also the most productive, which contrasted with the general inactivity of authors in other age categories. Perhaps the period from 40 years of age to fewer than 50 was a sort of "critical period" in the career of a scholar, and it was consequently the only age range in which author productivity could be defined with its base characteristics, including the impact of older age and first-authorship.

Secondly, gender and region did not affect scientific output in terms of relative productivity. While results of previous studies suggested that men were more productive than women in research [19-21], our finding showed no difference in CP-based productivity between male and female. As for regions, it seemed that scientists were not affected by differences in culture, infrastructure and policy that were specific to each region of Vietnam. This could be explained by the fact that scientists today have more access to academic resources which have become more and more readily available on the Internet. The task of researching and reviewing literature has thus become much less demanding. With these results, it could be said that when the contribution into each publication was taken into account, and not merely the count of articles, the scientific output of authors in social sciences in Vietnam did not depend on objective factors such as gender and geographical affiliation [34]. The main determinant of scientific output was thus the aptitude of the researcher, but also the motivation to do scientific research, especially in the case of social sciences which were still more or less underappreciated in Vietnam. In this respect, university administrations could play a role in supporting Vietnamese social scientists, for example by simplifying procedures of updating scientist curriculum vitae in official records or of application for project funding.

Finally, relative productivity turned out low for most Vietnamese authors – in other words, the level of contribution in their published works was not high. Although maximum relative productivity has reached 61.16 points, more than 88% of the authors scored under 5 points of relative productivity. A direct explanation for this phenomenon was that their Nominal productivity – which was the total number of publications – was also low: about 79% of scientists published less than 5 articles in the span of 10 years of research. This indicates not only the nascent state of Vietnamese social sciences, but perhaps also limitations in current science and technology funding policies and management strategies of scientific research funds in Vietnam. In fact, there exist few fund sources for science and technology in Vietnam, and even fewer for social sciences specifically. They are mostly public and inefficiently managed, with an inadequate coverage in terms of researchers taken into consideration. A suggested solution would be for these funds to take a concrete direction by putting focus on one or a few particular disciplines in order to devise specific funding policies.

Another factor directly linked to the low relative productivity was first-authorship. The average Vietnamese social scientist rarely assumed the role of lead author, while first-authorship was found to be an excellent way to boost productivity, with its coefficient being $\beta_{au.key}$=1.060 (p=2×10^{-16}). The experience of being a lead author could enrich a scientist in terms of knowledge, insights and research skills – all of which are critical qualities that would boost productivity. A lack or scarcity of this experience, as part of productivity cultures in the academia [35], would set authors back on this productivity boost, which would explain the lower relative output.

The international trend of collaboration in scientific research also came into play. Nowadays, the number of authors participating in the same project was increasing more and more. It could be observed in our dataset that Vietnamese social scientists were not exempt from this tendency. In this context, the contribution of the author into their scientific projects becomes even more important; and the highest level of contribution per article would be achieved by being either solo or lead author. Yet, about 60% authors in our sample did not play a leading role in more than half of their articles. The average ratio of solo publications to nominal productivity is only 0.16. Put into perspective with the fact that relative productivity, which represents not only the quantity of works but also the general contribution of the author to their published articles, was low in our sample, this raised a number of questions: What is the value of collaboration in scientific research for authors? Does the experience of collaborating truly boost the scientists' research productivity as the conclusions of previous studies? [36-38] Or do they only boost the absolute number of articles they have worked on, but not the amount of contribution they bring to academia, the significance of their role in a project? And ultimately, would working in collaboration hone the research skills of co-authors as scientists – as much as the experience of single-authoring would do? These questions give rise to the need of understanding the upcoming acculturation process both within the academia (Vuong 2017) and beyond [34].

4. Materials and Methods

4.1. Research hypothesis and analytical framework

4.1.1. Research hypothesis

Both nominal and relative productivity can measure the academic contribution of a scientist. Nominal productivity is defined as the quantity of publications of a scientist, whether written alone or with peers and regardless of position in the author list, if any. Relative productivity (or adjusted productivity) is measured using the placement of a scholar on the authors list of an article and suggests an indicator for author productivity based on their role and contribution in their final products.

Our analyses were conducted on the basis that the measure of relative productivity assures fairness among key authors and authors who play a more minor role in publications. Moreover, given that Vietnamese scientists have a tendency to co-author, basic descriptive statistics around relative productivity ("*cp*") will help provide a more in-depth overview of the current state of scientific contributions in Vietnam.

Thus we aim to assess factors of impact on relative productivity by testing the following two hypotheses:

Hypothesis 1: *First-authorship in articles is one of the key determinants of the author's contribution to research output; in addition, scientists of different age groups will have different levels of contribution.*

Hypothesis 2: *Gender and regional differences contribute to scientific output as a factor of influence.*

4.1.2. Analytical framework

The dependent variable in this paper is a continuous variable. Therefore, we use the OLS multivariable linear regression model with the general model as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

The condition is that k independent variables X_i must have the same sample size n as the dependent Y. Y must be numerical, while X_i can be numerical or categorical. Regression results provide the values of β_i, which represent the linear impact of X_i on Y. The statistical significance of predictor variables in the model are determined based on z-value and p-value; with p<0.05 being the conventional level of statistical significance required for a positive result.

To ensure the validity of the model, the F-test was performed with the hypothesis pair of H_0: $\beta_1=\beta_2=\ldots=\beta_i=\ldots= 0$, and H_1: at least one coefficient is not zero. The test result would determine the value of F and the coefficient p: in effect, if p<0.05, hypothesis H_0 would be rejected, which would confirm that regression coefficients in the model are not simultaneously null.

4.2. Dataset and variables

4.2.1. Data

The subjects of this study are Vietnamese scientists who have published in Scopus-indexed scientific journals over the past 10 years, from 2008 to 2017. Through gathering author information from personal and institutional websites, along with databases such as Google Scholar and Scopus, the research team has obtained a dataset of 410 Vietnamese authors. For each scientist we created a separate profile which includes all relevant factors such as age, sex, region, years of study since master graduation, the number of times is the key author, affiliated units, field of study, academic title of Professor, Associate Professor, and the like.

Data were then entered into a Microsoft Excel spreadsheet, cleaned and saved in .csv form. Data processing and analyses were done in R (3.3.1).

4.2.2. Dependent variable

This study will focus on the contribution of Vietnamese authors to scientific research in the world and the impact of several objective factors on their productivity. We measure scientific output not by number of publications, but rather by relative productivity. This indicator can also be called credit-based contribution point, abbreviated as 'CP', and is coded in our dataset as "*cp*". CP represents the output of a scholar not only based on the amount of papers they authored, but also taking into account their contribution in each article they produced.

The relative productivity of an author was calculated by converting the nominal productivity using the method of Sequence-Determines-Credit (SDC) [39]. Accordingly, the 1st author will be marked of 1/1, i.e. 100%, the 2nd author has the value of ½ (corresponding to 50%), 3rd author 1/3 (i.e. 33%), 4th author is ¼ (i.e. 25%),... In this manner, the further an author is placed from first position (implying a more modest contribution compared to the lead author), the smaller the count added to their relative productivity.

The relative productivity of the author is then calculated as the sum of all adjusted values from his or her publications.

4.2.3. Other variables

Factors considered to be of impact on relative productivity include gender, age and region b. They were used as predictor variables in our models, and are coded in our dataset as follows:

- "*sex*", a categorical variable representing author gender, with two categories: "male" and "female";
- "*age*", a numerical variable recording author age. It was not directly employed in our models, but was used to calculate the next variable, which would in turn be used in our regressions;
- "*age_gr*", a categorical variable representing age groups, calculated using "age", with 4 categories: "less30" (30 years old or younger), "b3040" (older than 30 to 40), "b4050" (older than 40 to 50), "g50" (over 50 years old);
- "*region*", a categorical variable with 4 options: "north", "center", "south" and "overseas".
- In our analysis, we also drew parallels between relative productivity ("*cp*") on one hand, and total number of publications (Nominal productivity) in relation to first-authorship on the other. The two latter factors would be examined as the following variables:
- "*ttlitems*", indicating the number of articles that an author has published;
- "*au.key*", recorded as the number of articles in which an author held first position.

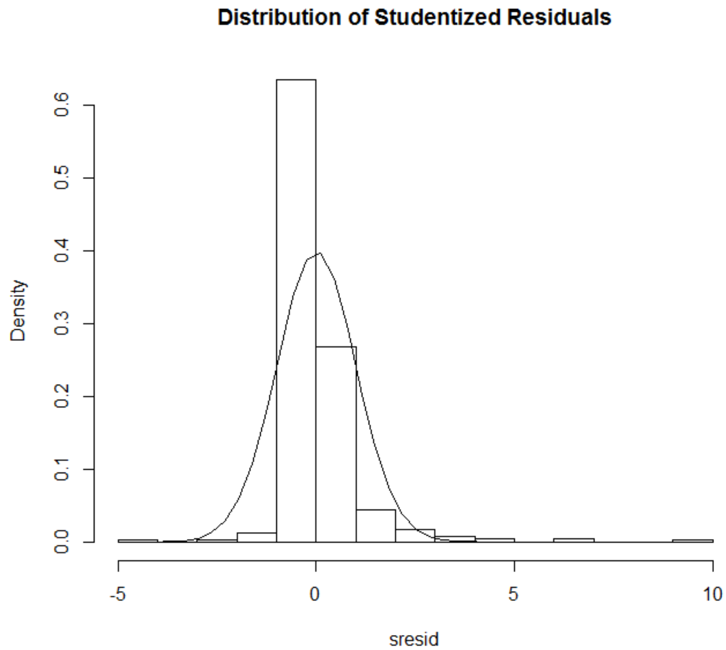
5. Conclusions

To put it briefly, in the disciplines of social science in Vietnam, the majority of authors belong to the 40-to-50 age range; this age category displayed basic characteristics of their academic career which included the impact of first-authorship and age on their relative productivity (measured as CP using the SDC method) and which did not manifest in other age groups. On the contrary, age and geographical region did not have any impact on scientific output. Attention should be paid to the effect of first-authorship and its boost to scientific productivity, particularly in the administration of academic institutions and policy management in science and technology funds. Namely, simplification of administrative procedures for scientists as well as more efficient financial support for research activities – especially for single authors – could prove to be good incentives for scholars and increase their contribution to the national education system.

Acknowledgments: We would like to thank Vuong & Associates for their research initiative *The Network of Vietnamese Social Scientists* ("NVSS"), which enabled the research process and provided the raw data for the study. We particularly thank Dam Thu Ha and Nghiem Phu Kien Cuong for their excellent research assistance. We specially thank Nguyen To Viet Ha, Pham Hung Hiep, Nancy K. Napier and Dam Quang Minh for their valuable comments. We are also grateful for useful comments from participants—who came from different academic disciplines—in the seminar "Beauty and the Light of Hope" co-organized by Western University Hanoi and Vuong & Associates in Hanoi, on October 3, 2017, particularly Nguyen Tu Cuong (mathematics), Nguyen Ngoc Chau (biology), Pham Duc Chinh (mechanics), Khuat Thu Hong (sociology), Le Van Canh (linguistics), Nguyen Viet Cuong and Tran Quang Tuyen (economics), Hoang Anh Tuan Kiet (physics), Tran Kien (law), Bui Quang Khiem (fine arts), Nguyen Pham Muoi (journalism/media), Bach Ngoc Chien (public governance), to name just a few.

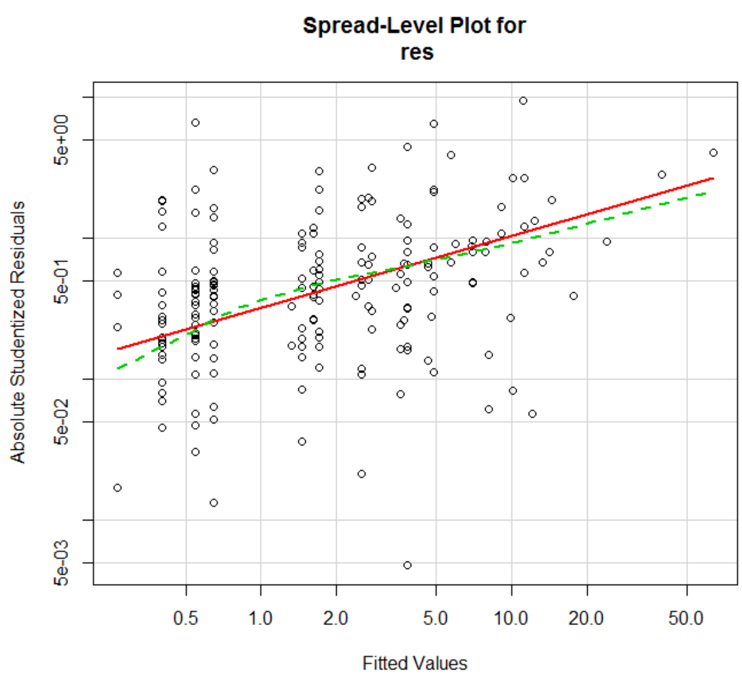
Appendix A. Examining the OLS model

A1. Standard distribution of residuals



A2. Result of the test for constant error variance

χ^2=182.18 with *df*=1, *p*=1.619×10^{-41}



Appendix B. The estimation result of the influence of the leading role, gender, age group, and region on adjusted productivity with full data (a) and data have eliminated the 'outliers' (b).

(N=410)									
	Intercept	"au.key"	"*Age_gr*" "b4050"	"g50"	"less30"	"*Sex*" "M"	"*Region*" "North"	"Overseas"	"South"
	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
"*cp*"	0.144 [0.621] (0.535)	1.061*** [85.841] (2×10^{-16})	0.238* [2.009] (0.045)	0.142 [0.966] (0.335)	-0.069 [-0.218] (0.827)	-0.007 [-0.061] (0.951)	0.318 [1.415] (0.148)	0.317 [1.290] (0.198)	0.123 [0.514] (0.608)
Residual *s.e.*=1.038 on 401 *df*. Multiple R^2=0.951, Adj. R^2=0.9496. *F*-stat=964.5 on 8 and 401 *df*, *p*-value: 2.2×10^{-16}.									
(n=407)									
"*cp*"	0.183 [0.877] (0.381)	1.061*** [56.096] (2×10^{-16})	0.226* [2.120] (0.035)	0.086 [0.653] (0.514)	-0.096 [-0.344] (0.731)	-0.029 [-0.300] (0.764)	0.245 [1.351] (0.177)	0.301 [1.372] (0.171)	0.150 [0.697] (0.486)
Residual *s.e.*=0.925 on 398 *df*. Multiple R^2=0.898, Adj. R^2=0.896. *F*-stat= 439.1 on 8 and 398 *df*, *p*-value: 2.2×10^{-16}									

Notes: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1; *z*-value in square brackets; *p*-value in round brackets; baseline category for: "*Age_gr*"="b3040", "*Sex*"="F", "*Region*"="Center".

Appendix C. Tables

Table 1: Some descriptive statistics for the variables used in the article

Variable	Min	Mean	Max	SD	95%CI	p-value
Adjusted productivity ("*cp*")	0.05	2.41	61.16	4.62	1.96 – 2.85	2.2×10^{-16}
Nominal productivity ("*ttlitems*")	1	3.60	63	5.89	3.03 – 4.17	2.2×10^{-16}
Number of articles in leading role ("*au.key*")	0	1.77	60	4.24	1.36 – 2.18	4.48×10^{-5}
Age ("*age*")	19	42.10	72	9.15	41.21 – 42.99	2.2×10^{-16}

Table 2: Descriptive statistics for variables after the removal of peripheral data points

Variable	Min	Mean	Max	SD	95%CI	*p*-value
Adjusted productivity ("*cp*")	0.05	2.12	23	2.87	1.84–2.40	2.2×10^{-16}
Nominal productivity ("*ttlitems*")	1	3.23	28	3.93	2.84–3.61	2.2×10^{-16}
Number of articles in leading role ("*au.key*")	0	1.52	22	2.54	1.28–1.77	2.2×10^{-16}
Age ("*age*")	19	42.06	72	9.17	41.17–42.96	2.2×10^{-16}

Table 3: Distribution of authors by age group and relative productivity ("cp")

Age group	Age<30		30≤Age<40		40≤Age<50		≥50		**SUM**	
Adj. prod. ("*cp*" point)	Freq.	Cum.	Freq.	Cum.	Freq.	Cum.	Freq.	Cum.	# people	Proportion
cp<1	6	6	61	61	49	49	34	34	150	36.59%
1≤*cp*<2	4	10	45	106	43	92	26	60	118	28.78%
2≤*cp*<3	1	11	19	125	18	110	12	72	50	12.20%
3≤*cp*<4	1	12	9	134	14	124	0	72	24	5.85%
4≤*cp*<5	0	12	6	140	10	134	3	75	19	4.63%
5≤*cp*<6	0	12	5	145	8	142	0	75	13	3.17%
6≤*cp*<7	0	12	0	145	1	143	1	76	2	0.49%

cp≥7	0	12	5	155	22	187	7	90	34	8.29%

Table 4. Estimation results of "*cp*" against "*age_gr*" and "*au.key*"

	Intercept	"*au.key*"	"*Age_gr*"		
			"b4050"	"g50"	"less30"
	β_0	β_1	β_2	β_3	β_4
"*cp*"	0.403*** [4.687] (3.8×10^{-6})	1.060*** [86.876] (2×10^{-16})	0.243* [2.064] (0.0396)	0.146 [1.025] (0.3058)	-0.136 [-0.437] (0.6626)

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1; *z*-value in square brackets; *p*-value in round brackets; baseline category for: "*Age_gr*"="b3040". Residual *s.e.*= 1.038 on 405 *df*. Multiple R^2=0.9501, Adj. R^2= 0.9496 . *F*-stat=1929 on 4 and 405 *df*, *p*-value = 2.2×10^{-16}.

Table 5. Estimation results of "*cp*" against "*Region*" and "*Sex*"

	Intercept	"Sex"	"Region"		
		"M"	"North"	"Overseas"	"South"
	β_0	β_1	β_2	β_3	β_4
"*cp*"	1.550 [1.560] (0.119)	0.724 [1.534] (0.126)	-0.032 [-0.033] (0.974)	1.655. [1.532] (0.126)	0.764 [0.721] (0.471)

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1; *z*-value in square brackets; *p*-value in round brackets; baseline category for: "*Sex*"="F", "*Region*"="Center". Residual *s.e.*=4.594 on 405 *df*. Multiple R^2=0.02281, Adj. R^2=0.0132. *F*-stat=2.363 on 4 and 405 *df*, *p*-value=0.0525.

Appendix D. Figures

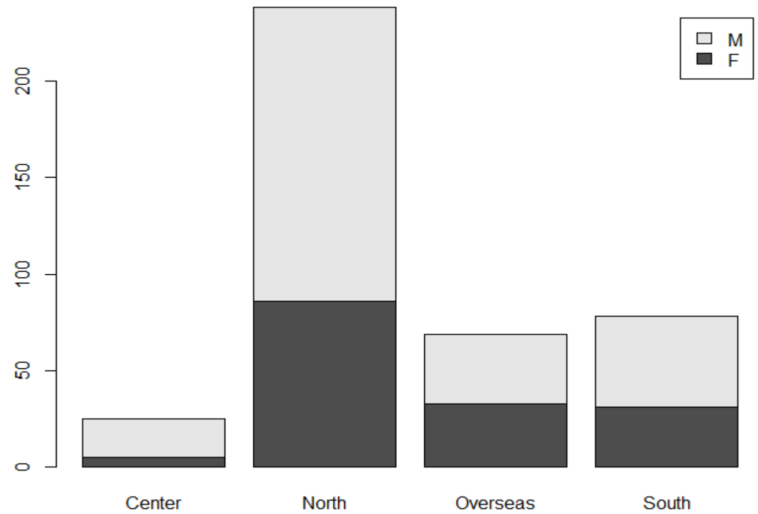


Figure 1. Regional and gender distribution of scientists

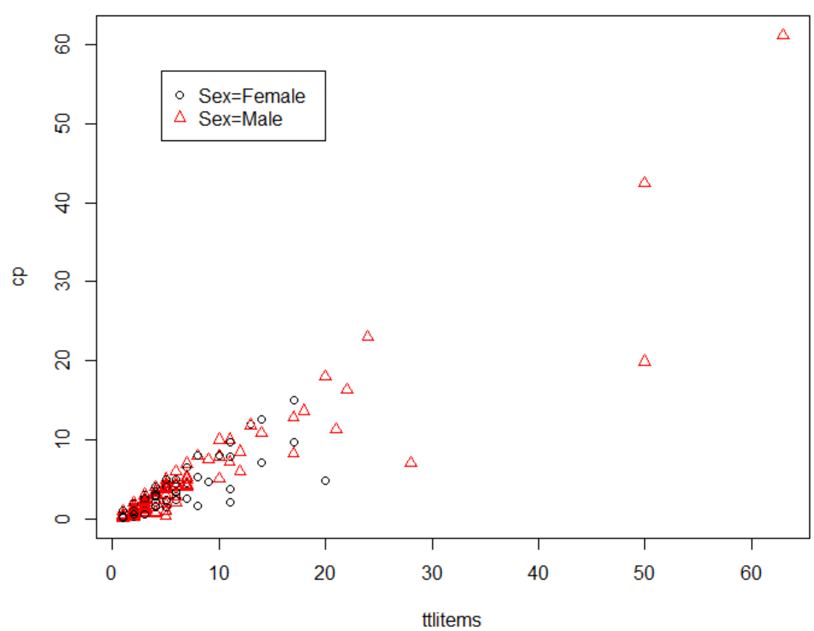


Figure 2. Relationship between nominal productivity and relative productivity by sex

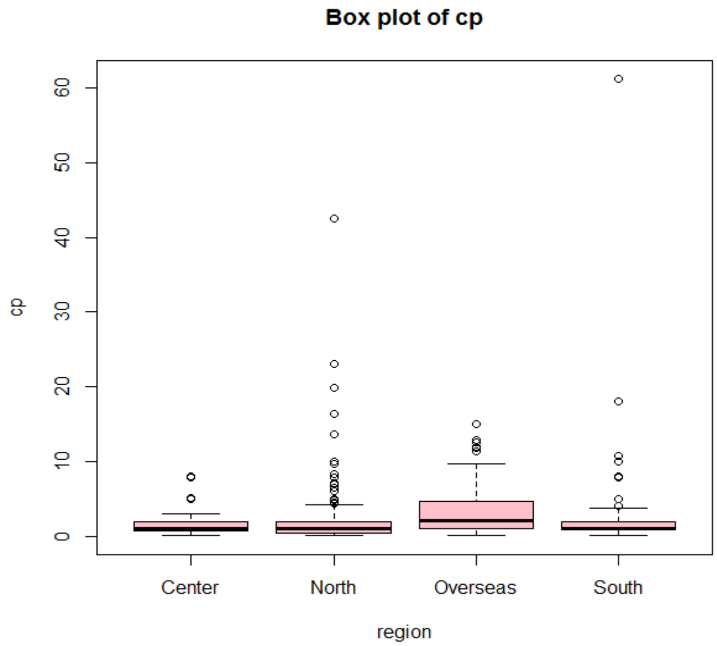


Figure 3. Relative productivity distribution by region

References

1. Nguyen, T.V.; Pham, L.T. Scientific output and its relationship to knowledge economy: an analysis of ASEAN countries. *Scientometrics* **2011**; *89* (1): 107-117. DOI: 10.1007/s11192-011-0446-2.
2. Hien, P.D. A comparative study of research capabilities of East Asian countries and implications for Vietnam. *Higher Education* **2010**; *60* (6): 615-625. DOI: 10.1007/s10734-010-9319-5.
3. Nguyen, T.V.; Ho-Le, T.P.; Le, U.V. International collaboration in scientific research in Vietnam: an analysis of patterns and impact. *Scientometrics* **2017**; *110* (2): 1035–1051, DOI: 10.1007/s11192-016-2201-1.
4. Vuong, Q.H.; Ho, T.M.; Vuong, T.T.; Nguyen V.H.; Napier, N.K.; Pham, H.H. Nemo solus satis sapit: trends of research collaborations in the Vietnamese social sciences, observing 2008-2017 Scopus data. *Publications* **2017**; *5* (4): 24, doi:10.3390/publications5040024.
5. Vuong, Q.H.; Ho, T.M.; Vuong, T.T.; Napier, N.K.; Pham, H.H. Gender, age, research experience, leading role and academic productivity of Vietnamese researchers in the social sciences and humanities: exploring a 2008-2017 Scopus dataset. *European Science Editing* **2017**; *43* (3): 51-55, DOI: 10.20316/ESE.2017.43.006.
6. Vuong, Q.H.; Napier, N.K. Academic research: the difficulty of being simple and beautiful. *European Science Editing* **2017**; *43* (2): 32-33, DOI: 10.20316/ESE.2017.43.002.
7. Lehman, H.C. *Age and Achievement*. Princeton University Press: Princeton, NJ, USA, 1953 [2017]; pp. 3-21, ISBN: 9870691627120.
8. Dennis, W. Creative productivity between the ages of 20 and 80 years. *Journal of Gerontology* **1966**; *21* (1): 1-8, DOI: 10.1093/geronj/21.1.1.
9. Cole, S. Age and scientific performance. *American Journal of Sociology* **1979**; *84* (4): 958-977, DOI: 10.1086/226868.
10. Over, R. Does research productivity decline with age? *Higher Education* **1982**; *11* (5): 511-520, DOI: 10.1007/BF00194416.

11. Allison, P.D.; Stewart, J.A. Productivity differences among scientists: Evidence for accumulative advantage. *American Sociological Review* **1974**; *39* (4): 596-606, DOI: 10.2307/2094424.
12. Kyvik, S. Age and scientific productivity. Differences between fields of learning. *Higher Education* **1990**; *19* (1): 37-55, DOI: 10.1007/BF00142022.
13. Cooney, T.M.; Schaie, K.W.; Willis, S.L. The relationship between prior functioning on cognitive and personality dimensions and subject attrition in longitudinal research. *Journal of Gerontology* **1988**; *43* (1), P12-P17.
14. Bonaccorsi, A.; Daraio, C. Age effects in scientific productivity. *Scientometrics* **2003**; *58* (1): 49-90, DOI: 10.1023/A:1025427507552.
15. Pan, R.K.; Kaski, K.; Fortunato, S. World citation and collaboration networks: uncovering the role of geography in science. *Scientific Reports* **2012**; *2*: 902, DOI: 10.1038/srep00902.
16. Uthman, O.A.; Uthman, M.B. Geography of Africa biomedical publications: an analysis of 1996–2005 PubMed papers. *International Journal of Health Geographics* **2007**; *6* (1): 46, DOI: 10.1186/1476-072X-6-46.
17. Duque, R.B.; Ynalvez, M.; Sooryamoorthy, R.; Mbatia, P.D.; Dan-Bright S.; Shrum, W. Collaboration paradox: scientific productivity, the Internet, and problems of research in developing areas. *Social Studies of Science* **2005**; *35* (5): 755-785, DOI: 10.1177/0306312705053048.
18. Prpić, K. Gender and productivity differentials in science. *Scientometrics* **2002**; *55* (1): 27-58, DOI: 10.1023/A:1016046819457.
19. Fulton, O. *Rewards and Fairness: Academic Women in the United States*. Centre for Research in the Educational Sciences: University of Edinburg 1973; ISBN: 978-0904170146.
20. Cole, J.R.; Zuckerman, H. The productivity puzzle: persistence and change in patterns of publication of men and women scientists. In *Women in Science*; Steinkemp, M.W., Maehr, M.L., Eds.; JAI Press: Greenwich, CT, USA, 1984; Vol. 2, pp. 217-258, ISBN: 0892322888.
21. Bernard, J. *Academic Women*. Pennsylvania State University Press: Pennsylvania, PA, USA, 1964; ISBN: 9780452003897.
22. Alic, M. The history of women in science: A women's studies course. *Women's Studies International Forum* **1982**; *5* (1): 75-81, DOI: 10.1016/0277-5395(82)90066-8.
23. Kyvik, S. Motherhood and scientific productivity. *Social Studies of Science* **1990**; *20* (1): 149-160, DOI: 10.1177/030631290020001005.
24. Reskin, B.F. Scientific productivity, sex, and location in the institution of science. *American Journal of Sociology* **1978**; *83* (5): 1235-1243, DOI: 10.1086/226681.
25. Suitor, J.J.; Mecom, D.; Feld, I.S. Gender, household labor, and scholarly productivity among university professors. *Gender Issues* **2001**; *19* (4): 50-67, DOI: 10.1007/s12147-001-1007-4.
26. Kyvik, S.; Teigen, M. Child care, research collaboration, and gender differences in scientific productivity. *Science, Technology and Human Values* **1996**; *21* (1): 54-71. DOI: 10.1177/016224399602100103.
27. Leta, J.; Lewison, G. The contribution of women in Brazilian science: A case study in astronomy, immunology and oceanography. *Scientometrics* **2003**; *57* (3): 339-353, DOI: 10.1023/A:1025000600840.
28. Gupta, B.M.; Kumar, S.; Aggarwal, B. S. A comparision of productivity of male and female scientists of CSIR. *Scientometrics* **1999**; *45* (2): 269-289, DOI: 10.1007/BF02458437.
29. Breimer, L.H.; Nilsson, T.K. A longitudinal and cross-sectional study of Swedish biomedical PhD processes 1991–2009 with emphasis on international and gender aspects. *Scientometrics* **2010**; *85* (2): 401-414. DOI: 10.1007/s11192-010-0227-3.
30. Lemoine, W. Productivity patterns of men and women scientists in Venezuela. *Scientometrics* **1992**; *24* (2): 281-295, DOI: 10.1007/BF02017912.
31. Sotudeh, H.; Khoshian, N. Gender differences in science: the case of scientific productivity in Nano Science & Technology during 2005–2007. *Scientometrics* **2014**; *98* (1): 457-472, DOI: 10.1007/s11192-013-1031-7.
32. van Arensbergen, P.; van der Weijden, I.; van den Besselaar, P. Gender differences in scientific productivity: a persisting phenomenon? *Scientometrics* **2012**; *93* (3): 857-868, DOI: 10.1007/s11192-012-0712-y.

33. Ho, T.M.; Nguyen, V.H.; Vuong, T.T.; Dam, Q.M.; Pham, H.H; Vuong, Q.H. *Exploring Vietnamese co-authorship patterns in social sciences with basic network measures of 2008-2017 Scopus data*. F1000Research **2017**; *6*: 1559, DOI: 10.12688/f1000research.12404.1.
34. Vuong, Q.H. Global mindset as the integration of emerging socio-cultural values through mindsponge processes: a transition economy perspective. In *Global Mindsets: Exploration and Perspectives*; Kuada, J., Ed.; Routledge: London, UK, 2016; pp. 109-126, ISBN: 1138831778.
35. Vuong, Q.H. Learning to love the reviewer. *European Science Editing* **2017**; *43* (4): 83.
36. Bordons, Maria, Gómez, Isabel, Fernández, M. Teresa, Zulueta, M. Angeles, and Méndez, Aida. 1996. Local, domestic and international scientific collaboration in biomedical research. *Scientometrics* 37 (2): 279-295. DOI: 10.1007/BF02093625.
37. Hampton, S.E.; Parker, J.N. Collaboration and productivity in scientific synthesis. *BioScience* **2011**; *61* (11): 900-910, DOI: 10.1525/bio.2011.61.11.9.
38. Lee, S.; Bozeman, B. The impact of research collaboration on scientific productivity. *Social Studies of Science* **2005**; *35* (5): 673-702, DOI: 10.1177/0306312705052359.
39. Teja, T.; Hochberg, M.E.; Rand, T.A.; Rey, V.H.; Krauss, J. Author Sequence and Credit for Contributions in Multiauthored Publications. *PLoS Biol* **2007**; *5* (1): e18, DOI: 10.1371/journal.pbio.0050018.