Study of Motivation Effect from Peer Review Accolade: Observation based on DID analysis

Houqiang Yu [*], Yian Liang [**] and Yinghua Xie[**]

[*] *yuhq8@mail.sysu.edu.cn*
0000-0002-9241-6630
School of Information Management, Sun Yat-sen University, China

[**] *liang.yian@qq.com; xieyh53@mail2.sysu.edu.cn*
0000-0002-1234-1131; 0000-0002-6710-4395
School of Information Management, Sun Yat-sen University, China

Abstract
The utilization of accolade awards to incentivize reviewers is highly common, but its effectiveness remains uncertain. This study seeks to investigate how receiving an accolade award affects reviewers' subsequent motivation. We perform a large-scale, global, and all-disciplinary analysis based on quasi-natural experiments. 6604 awarded reviewers and 179,737 reviewers are matched as experimental and control groups respectively. The difference-in-differences (DID) method is conducted to examine the accolade awarding effects. After receiving an accolade award, reviewers, on average, reviews about four fewer manuscripts, and this reduction follows a V-shaped pattern. Additional analyses are conducted to examine how individual differences and economic-cultural factors influence the accolade awarding effects. The sharp decrease in marginal utility of accolades, the ethical nature of peer review, and the unexpected properties of accolade awards are the mechanisms that generate the negative effect. The academic community should reassess the existing incentive strategies on reviewers.

Keywords
Peer reviewers; Accolade awards; Motivation; Quasi-natural experiments; Difference-in-differences; Propensity score matching

1. Introduction

In 2018, a survey in *Nature News* (Vesper, 2018) reports that journal editors consider identifying suitable reviewers as the most challenging task. The term "peer-review crisis" is coined to characterize the scarcity of reviewers and their lack of motivation (Petrescu & Krishen, 2022). This issue has persisted for an extended period and continues to be an ongoing challenge (Bianchi & Squazzoni, 2015; DeLisi, 2022).

Nowadays, the primary approach to incentivize reviewers is by establishing accolade awards (Zaharie & Seeber, 2018), like "2022 Outstanding Reviewer". However, limited empirical research has been conducted to thoroughly examine how peer review

accolade awards influence the subsequent motivation of reviewers. The issue of motivating peer reviewers is of complexity. Various studies based on artificial and small-scale experiments reached conflicting conclusions (Robinson et al., 2021) under varying context. While individual differences and the economic-cultural environment are two significant influencing factors (Gagné & Deci, 2005; Ryan & Deci, 2020). Therefore, it is highly necessary to have a large and representative sample, and consider the influence of individual and economic-cultural factors simultaneously.

When theorizing the peer review incentives, there are certain points to note. According to the peer review motivation frameworks (Zaharie & Osoian, 2016), when analyzing incentive mechanisms of peer review accolades, taking account its impact on reviewers' self-achievement and prosocial motivation is necessary. For self-achievement motivation, the diminishing marginal utility of academic influence from peer review accolades is particularly pronounced. For the prosocial motivation, review activities are ethical in nature and are often seen as an academic responsibility. And it is important to emphasize the unexpected nature of accolade awards, as expected and unexpected awards have different effects (Neckermann & Yang, 2017).

Publons, a global peer review platform, grants the annual "Global Peer Review Award" to outstanding reviewers from 2017 to 2019. This presents an opportunity to examine the influence of the peer review accolade awards on reviewers' motivation in a representative way.

This study aims to address the following purposes. (1). Investigate the impact of peer review accolade awards on the subsequent motivation of reviewers, using a highly representative sample and quasi-natural experiment method. (2). Explore how individual differences and economic-cultural factors influence the accolade awarding effect. (3). Based on the empirical results, this study will try to provide a thorough explanation of the mechanisms behind the accolade awarding effect, and propose strategies for motivating reviewers.

2. Data and Methodology

Unlike artificial experiments, quasi-experiments allow for large-scale inferences and occur naturally, resembling real-life scenarios. The peer review activities are easily influenced by various factors, like economic-cultural backgrounds and individual differences. To achieve highly representative causal inferences, we aim to utilize quasi-experimental methods based on a large and representative sample.

2.1 Data

The data in this study is acquired by integrating the Publons, ORCID, and OpenAlex database.

Publons (https://publons.com) is recognized as one of the largest global peer review platforms. Nature has also acknowledged and reported on Publons (Van Noorden, 2014).

This study aims to conduct empirical research based on the Publons data. Publons implements annually a selection process to identify the top 1% of reviewers in their respective fields from 2017 to 2019, and awards these individuals with the "Publons Global Peer Review Award". Due to various reasons, the award has ceased since 2019. To prevent biases in heterogeneous multi-period DID analysis (De Chaisemartin & d'Haultfoeuille, 2020), we focused on 6604 peer review winners in 2018, and make sure that the control group excludes winners in 2017 and 2019. Therefore, the pre-treatment duration of the experimental group extends from 2012 to 2017 (a total of six years), and the period of after-treatment spans from 2019 to 2023 (a total of five years).

We initially select the authors with more than 10 publications in OpenAlex database, a total of 2,678,929 authors. In ORCID database, 357,948 authors have established the link between ORCID and ResearcherID. The review data of scholars is acquired via ResearcherID.

Name disambiguation is also performed using three golden author databases, namely OpenAIRE (Manghi et al., 2019), EMAKG (Pollacci, 2022), and LAGOS-AND (Zhang et al., 2023).

In summary, a panel data of reviewers spanning from 2012 to 2023 is established, totally 179,737 reviewers. Data collection is conducted from August 10 to August 27, 2023.

2.2 Inferring method

The validity of the two-way fixed effects model (TWFE) commonly used in DID analysis, but has been questioned in recent years (Callaway & Sant'Anna, 2021; Goodman-Bacon, 2021). Liu et al. (Liu et al., 2022) introduced a framework for counterfactual estimation, named as "ifect estimation", which has been proved to be effective at relaxing the assumptions of TWFE model. Hence, we adopt this method for DID, using the *fect* package in R language developed by Liu et al.

The assumption of parallel trends presents a main challenge for the DID method. To meet this assumption, propensity score matching (PSM) is used to create a control group closely resembling the experimental group with a similar pre-treatment change pattern. In panel data matching, two commonly used methods are single-period matching and period-by-period matching. Multiple studies have highlighted concerns about the irrationality of period-by-period matching (Stuart, 2010). In this study, the former method is adopted, which typically focuses on the first period (2012) as the matching period. Hence, reviewers with one or more reviews are initially extracted from the dataset, as well as scholars' academic age, gender, discipline, cumulative number of publications, and cumulative citations as matched variables. Finally, 762 reviewers are identified from 179,737 samples as the control group by utilizing the MatchIt (Stuart et al., 2011) package in the R language. The research specifically focusses on the 2018 winners, totaling 6604 samples. We exclude the 2017 and 2019 winners, totaling 3251 individuals. The chosen period for comparison was 2012,

namely the first year. 711 reviewers are identified as the experimental groups.

2.3 Variables

The dependent variable in our study is the annual number of reviews. The independent variable pertains to the DID, whether is an accolade award winner and the year is over 2018 (after the 2018 Publons Global Peer Review Award). Prior to 2018 (specifically, from 2012 to 2017), the winners are considered as before treatment. The accolade award winners from 2018 onwards (2018-2023) will be regarded as having been treated. In the control group that does not receive the award, the time frame spanning from 2012 to 2023 are all considered as the untreated period.

It remains necessary to control for as many of these variables as possible to derive more precise and reliable conclusions. Based on prior evidence, number of publications, cumulative number of citations, academic age, number of collaborators, number of (second-level) disciplines, the intersection of a scholar's academic age and cumulative citations, as well as the intersection of academic age and number of publications will be controlled in this study.

The variables intended for use are presented in Table 1.

Table 1. Variables in this study.

Notations	**Comments**
$ResearcherID$	Identifiers of reviewers in Publons.
$Year_t$	Years
$Gender$	Gender is determined by the SciSciNet database using a machine learn algorithm, with a threshold of inference probability set at over 80%.
$Discipline$	Reviewer's discipline inferred by OpenAlex.
$IsWinner$	Whether to win the peer review accolade award.
$Treat_t$	Whether the treatment (be awarded) is implemented. If the year is greater than 2018, it is equal to 1, otherwise it is 0.
$DID_i(IsWinner_i \times Treat_i)$	Independent variable, equal to 1 if it is an accolade award winner and the year is over 2018.
R_t	Number of reviews in a given year
P_t	Number of publications in a given year.
CC_t	Cumulative citations in a given year.
CP_t	Cumulative number of publications in a given year.
Age_t	Academic age in a given year. The given year minus the year of the scholar's first publication.

Nod_t	Number of (second-level) disciplines, as proxy of interdisciplinary.
Noa_t	Average number of co-authors in the past three years.
Nof_t	Number of funding.
$Age_t \times CP_t$	Intersection of academic age and cumulative number of publications. Agency of scholars' seniority.
$Age_t \times CC_t$	Intersection of academic age and cumulative number of citations. Agency of scholars' seniority.

Hence, the modeling equation can be expressed as:

$$R_{i,t} = \beta_1 DID_{i,t} + \beta_2 P_{i,t-1} + \beta_3 CC_{i,t} + \beta_4 Age_{i,t} + \beta_6 Noa_{i,t} + \beta_7 Nod_{i,t} + \beta_8 Age_i \times CP_{i,t} + \beta_9 Age_i \times CC_{i,t} + year_t + u_i + \delta_{i,t}, \quad t = 2012, \ldots, 2023$$

3. Results

3.1 PSM results

In this study, the earliest period (2012) is selected for matching. Reviewers' career age, first-level discipline, seniority, gender, and bibliometrics data in 2012 are taken as matching variables. MatchIt utilizes logistic regression to estimate the propensity score, following the suggestion made by Ho et al. (Ho et al., 2007), and the non-parametric nearest distance method is employed in this study. After completing the matching, we assess the matching effect by comparing the SMD before and after matching. After matching, all SMDs are below 0.1, with only a few variables having SMDs exceeding 0.05. The matching effect is satisfactory.

3.2 PSM-DID results

3.2.1 Parameter estimations

The DID results are presented in Table 2. When including control variables and fixed effects, the coefficient for DID is -4.06. This suggests a decline in reviewer enthusiasm after receiving the Publons Global Peer Review Award.

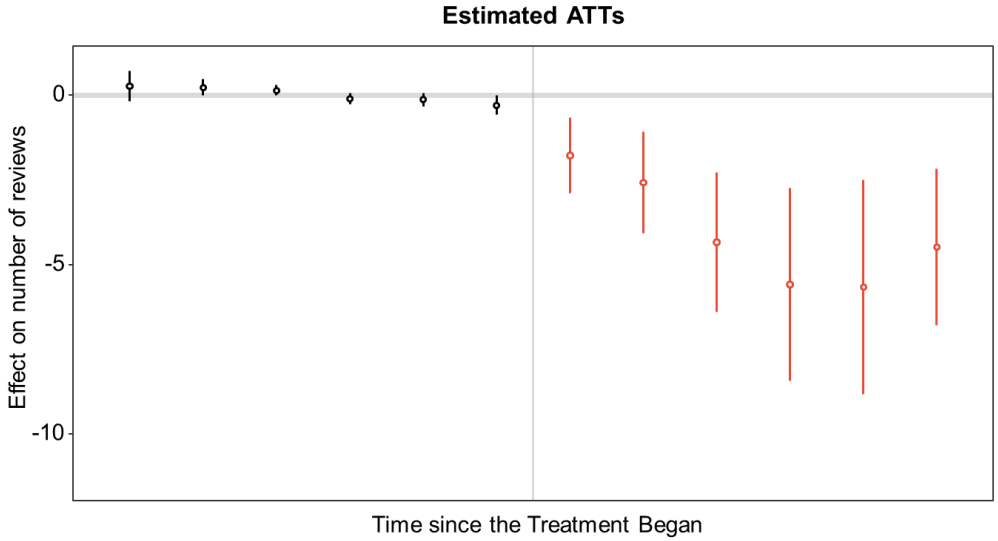
Table 2: Difference-in-differences analysis.

Variables	**Model (1)**	**Model (2)**
DID	**-3.81*****	**-4.06*****
Control variables	No	Yes
Individual fixed effect	Yes	Yes
Year fixed effect	Yes	Yes
N	15839	15839
R^2	.0003	.075

Note: *: p<0.1, **: p<0.01, ***: p<0.001. The same below.

Furthermore, we plot the average treatment effects on treated (ATTs), as shown in Figure 1. It is evident that the average ATTs is about -4 after treatment, and the trend demonstrates a V-shape pattern. After undergoing a consecutive decline for four years, this negative trend will slow down to some extent.

Figure 1: The average treatment effects on treated.



3.2.2 Parallel trend tests and placebo tests

The F test is commonly used for pre-trend testing. A higher p-value from the F-test suggests a better fit of the parallel trend. In the *fect* package, the TOST checks whether the 90% confidence intervals for the estimated ATTs in pretreatment periods exceed the equivalence range. A smaller p-value for the TOST indicates a better fit to the pre-trend. In this study, the p-value of F test and TOST are 0.218 and 0 respectively, passing both the F test and the TOST test at a significance level of 5% (F>0.05, T<0.05).

The placebo test in this study involves the p-value of the F test and the p-value of TOST. The p-value of the F-test in our analysis is 0.518, significantly greater than 0.05, and the TOST is 0, passing the placebo test.

3.3 Additional analysis

3.3.1 Individual differences and awarding effects

By analysing existing literature on peer review motivation, this study examines six variables: academic age, seniority, funding, social capital, interdisciplinarity, and gender of reviewers, to understand their influence on the awarding effect. Moderating effects are mainly employed. For gender effect analysis, subgroup modelling will be used.

Table 3 shows that academic age, seniority, social capital, and gender significantly influence the awarding effect. The remaining variables, including funding and interdisciplinarity, are not statistically significant. For reviewers with higher seniority and social capital, winning the accolade awards will make a more significant negative impact. There is a noticeable decrease in motivation among male reviewers after receiving an accolade. For women, this accolade award has no impact. Both male and female subgroup models passed the parallel trend test, with p-values being both 0, and TOST test values are 0.384, 0.582 respectively. The lack of significance within the female group is not attributed to the smaller sample size. When a sample of 131 men is randomly selected, the p-value of the DID coefficient obtained is also less than 0.01.

Table 3 Individuals factor and awarding effect.

Variables	Model (3) (Academic age)	Model (4) (Seniority)	Model (5) (Seniority)	Model (6) (Social capital)	Model (7) (Interdisciplinary)	Model (8) (Funding)	Model (9) (Male)	Model (10) (Female)
DID	**-2.29*****	**-2.49*****	**-2.46*****	**-2.52*****	**-2.57*****	1.637	**-3.07*****	-0.09
Age	**11.07*****							
$Age \times DID$	**-1.834***							
$Age \times P$		**-9.10e-5***						
$Age \times P \times DID$		**-9.45e-5****						
$Age \times CC$			-2.79e-7					
$Age \times CC \times DID$			**-2.54e-6*****					
Noa				**.002*****				
$Noa \times DID$				**-.003*****				
Nod					**.301*****			
$Nod \times DID$					.013			
$Fund$						**.039***		
$Fund \times DID$						5.22		
Other Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	1488	1488	1488	1488	1488	180	628	131
R^2	.070	.069	.070	.070	.069	.180	.005	.171

Note: Moderator variables are standardized. The fund data includes NSF and NIH, so we only selected data from American reviewers when building the moderating effect model.

3.3.2 Economic-cultural factors and awarding effects

This study will use country-based group modeling to analyze how scholars from different nations react to the peer review accolade. To maintain the sample size, we increased the number of matched control groups and created the control group in a 1:3 ratio. The construction methods' matching also passed the parallel test. Then, countries with sample sizes below 30 are excluded. Countries with p-values below 0.2 are retained.

Of the 28 countries studied, relevant indicators, such as income level (IMF standards), Confucian culture, official language (English usage), and scientific research performance (national-level citations, number of publications, and H-index) are collected. The variables of H-index, number of citations, and number of publications are the average values of countries from 2012 to 2023. A regression analysis is conducted to examine the impact of these factors on the awarding effects by analyzing the coefficients of the difference-in-differences (DID) estimator.

In Table 4, the impact of income level and national scientific research influence (H-index) on awarding suggests that high levels of income and advanced scientific research can mitigate the negative effects of accolades. All other factors are not significant.

Table 4. Economic-cultural factors and award effect.

	Model (11)	**Model (12)**	**Model (13)**	**Model (14)**	**Model (15)**	**Model (16)**
Income level	**3.642*****					
Is English country		2.616				
Is Confucianism			7.039			
H-index				**.005***		
Number of citations					4.78e-7	
Number of publications						5.31e-6

Note: Low-income countries=0, Middle-income countries=1, middle-high-income countries=2, high income countries=3.

4. Discussions and conclusions

4.1 Mechanisms of awarding reviewers

After receiving a peer review accolade award, reviewers' sequent motivation decreased. This seems counterintuitive.

The theoretical framework of peer review motivation emphasizes the academic impact obtained from peer review as a primary motivator. As reiterated numerous times, peer review is a voluntary undertaking. For reviewers motivated by self-achievement, peer review awards only can primarily offer them academic influence. However, the reduction in the marginal utility of academic influence offered by peer review accolade is significant. Thus, reviewers who are motivated by self-achievement may experience a natural decline in motivation of reviewing after receiving an accolade award.

The ethical nature of peer review contributes to the negative incentive effect. Most reviewers are typically scholars. For a scientific system to operate in a sustainable way, it requires enough reviewers. Numerous studies emphasize that publishing is a privilege for scholars, while peer

review is seen as an obligation (Nash, 2023). Accolade awards winners may perceive awards as a sign that their review performance exceeds the academia's expectations and obligation (Gallus & Frey, 2017), then they may decline number of reviews (Monin & Miller, 2001; Mullen & Monin, 2016).

The unexpected nature of peer review accolades will increase this negative impact (Robinson et al., 2021). Unexpected accolade awards significantly reinforce the winner's higher awareness of surpassing social norms upon receiving the award. Because the winners are unaware that they have surpassed social norms before receiving the award (Neckermann & Yang, 2017).

These mechanisms are supported by various findings outlined in the *Additional Analysis* section. Male reviewers are often seen as more driven by self-achievement motivations, especially in terms of the academic influence which benefits career advancement (Meece et al., 2009). Reviewers with higher seniority, higher career years, and more social capital experience a significant drop in motivation after receiving peer review accolades. This phenomenon may be due to the fact that they have likely already achieved a higher level of academic influence, resulting in a more significant diminishing marginal effect. In regions with less developed economies and research strength, the negative effects on scholars' incentives will be more significant. This phenomenon can be explained at two different levels. (1) Conducting reviews is more expensive in developing regions. In developing regions, scholars may face more economic limitations, leading them to prioritize scientific and teaching activities with higher economic benefits, thus more potentially neglect the review activities. (2) The accolade award in this study is an international honor. In developing regions, the international award may be more surprising, making the award more unexpected. Individuals are more likely to perceive themselves as surpassing international standards after receiving accolades, leading to a reduction in their motivation.

4.2 Conclusions

Upon a representative investigation, we first find that the peer review accolade awards will reduce the sequent number of reviews. We examine the impact of relevant factors on this awarding effect, explain the mechanism of this awarding effect. The research questions outlined in *Introduction* are well addressed. The answers to these questions and main research findings can be summarized as follows:

(1) Peer review accolades will create a negative awarding effect. After receiving the Publons Global Peer Review Award, there is typically an average reduction of about four reviews.

(2) The awarding effect is influenced by individual and economic-cultural factors. Higher academic seniority, more social capital, higher academic age, female scholars, scholars from lower income level or less advanced scientific research regions are likely to reduce a greater number of reviews after receiving the accolade award.

(3) Certain mechanisms are proposed to explain negative awarding effect. The negative incentive may stem from the diminishing marginal utility nature of academic influence, the ethical nature of peer review, and the unexpected properties of accolade awards. According to these proposed mechanisms, we put forward potential strategies for advancement.

Open science practices

Bibliometric and geographical location data are accessible in the Openalex (https://docs.openalex.org/) and ORCID (https://orcid.org/) databases. The SciSciNet database

provides free access to sex determination data and NIH and NSF grant data (https://www.nature.com/articles/s41597-023-02198-9). Peer-review data can be accessed through the Web of Science Researcher API (https://developer.clarivate.com/apis/wos-researcher).

Funding information
This article is supported by National Natural Science Foundation of China (No. 72274227) and Humanity and Social Science Foundation of Ministry of Education of China (NO. 22YJA870016).

Author contributions
Houqiang Yu (Corresponding Author): Conceptualization; Funding acquisition; Methodology; Writing - review & editing
Yian Liang: Conceptualization; Data curation; Formal analysis; Investigation; Roles/Writing - original draft
Yinghua Xie: Investigation; Writing - review & editing

Competing interests
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References
Bianchi, F., & Squazzoni, F. (2015). Is three better than one? Simulating the effect of reviewer selection and behavior on the quality and efficiency of peer review. 2015 Winter simulation conference (WSC),

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200-230.

De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, *110*(9), 2964-2996.

DeLisi, L. E. (2022). Where have all the reviewers gone?: Is the peer review concept in crisis? In (Vol. 310, pp. 114454): Elsevier.

Gagné, M., & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational behavior*, *26*(4), 331-362.

Gallus, J., & Frey, B. S. (2017). Awards as strategic signals. *Journal of Management Inquiry*, *26*(1), 76-85.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254-277.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis*, *15*(3), 199-236.

Liu, L., Wang, Y., & Xu, Y. (2022). A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data. *American Journal of Political Science*.

Manghi, P., Bardi, A., Atzori, C., Baglioni, M., Manola, N., Schirrwagen, J., Principe, P., Artini, M., Becker, A., & De Bonis, M. (2019). The OpenAIRE research graph data model. *Zenodo*.

Meece, J. L., Glienke, B. B., & Askew, K. (2009). Gender and motivation. *Handbook of motivation at school*, 425-446.

Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of personality and social psychology*, *81*(1), 33.

Mullen, E., & Monin, B. (2016). Consistency versus licensing effects of past moral behavior. *Annual review of psychology*, *67*, 363-385.

Nash, C. (2023). Roles and Responsibilities for Peer Reviewers of International Journals. *Publications*, *11*(2), 32.

Neckermann, S., & Yang, X. (2017). Understanding the (unexpected) consequences of unexpected recognition. *Journal of Economic Behavior & Organization*, *135*, 131-142.

Petrescu, M., & Krishen, A. S. (2022). The evolving crisis of the peer-review process. *Journal of Marketing Analytics*, *10*(3), 185-186.

Pollacci, L. (2022). Emakg: An enhanced version of the microsoft academic knowledge graph. *arXiv preprint arXiv:2203.09159*.

Robinson, C. D., Gallus, J., Lee, M. G., & Rogers, T. (2021). The demotivating effect (and unintended message) of awards. *Organizational Behavior and Human Decision Processes*, *163*, 51-64.

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary educational psychology*, *61*, 101860.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *25*(1), 1.

Stuart, E. A., King, G., Imai, K., & Ho, D. (2011). MatchIt: nonparametric preprocessing for parametric causal inference. *Journal of statistical software*.

Van Noorden, R. (2014). The scientists who get credit for peer review. *Nature*.

Vesper, I. (2018). Peer reviewers unmasked: largest global survey reveals trends. *Nature*, 7-8.

Zaharie, M. A., & Osoian, C. L. (2016). Peer review motivation frames: A qualitative approach. *European Management Journal*, *34*(1), 69-79.

Zaharie, M. A., & Seeber, M. (2018). Are non-monetary rewards effective in attracting peer reviewers? A natural experiment. *Scientometrics*, *117*(3), 1587-1609.

Zhang, L., Lu, W., & Yang, J. (2023). LAGOS‐AND: A large gold standard dataset for scholarly author name disambiguation. *Journal of the Association for Information Science and Technology*, *74*(2), 168-185.