Uncovering CHI Reviewers Needs and Barriers

Wanhae Lee
University of Seoul
Seoul, Republic of Korea
train96@uos.ac.kr

Minji Kwon
Kyunghee University
Yongin, Republic of Korea
minmin0916@khu.ac.kr

Yewon Hyun
POSTECH
Pohang, Republic of Korea
yewon@postech.ac.kr

Jihyun Lee
Kyung Hee University
Yongin, Republic of Korea
2018102130@khu.ac.kr

Joonho Gwon
University of Seoul
Seoul, Republic of Korea
unkn0wnpgr@uos.ac.kr

Hyunggu Jung
University of Seoul
Seoul, Republic of Korea
hjung@uos.ac.kr

ABSTRACT

A peer review process is an essential part of not only in the CHI conference but also in many scientific research areas to determine which paper submitted should be accepted. In order to improve the peer review process, tools have been developed to support reviewers by reducing the paper they should review or by highlighting key sentences to read faster. However, only a few studies investigated reviewers' perspectives regarding their tasks during the peer review process. In this paper, we conducted semi-structured interviews with CHI reviewers who have experienced in reviewing paper submitted to the CHI conference. As a result, we better understand how paper-reviewing tasks are performed, which tasks reviewers felt most challenging, and reviewers' needs for a better peer review experience.

Author Keywords

Needs assessment; Peer review; Reviewers; Qualitative; Interview study

CCS Concepts

•**Human-centered computing → Computer supported cooperative work;**

INTRODUCTION

The number of papers submitted to the CHI conference has increased each year [1]. To determine whether the submitted paper should be accepted, a peer review process is used. During the review process, reviewers are asked to perform multiple tasks, such as rating papers, giving feedback, and discussion. Even though most authors from CHI satisfy with the review they received [6], it is debated whether an existing peer review system is appropriate for a current situation where finding

experienced reviewer is harder and harder [14], where real-time communication everywhere on earth is possible [13], and where every decision totally depends on reviewers' subjective point of view [4]. Prior studies demonstrated a variety of tools that assist with editors' tasks [12], that automatically filter out papers that do not need to be reviewed [5, 8], that assign a paper to an appropriate reviewer [7], and that support reviewers by highlighting salient sentences [11]. In addition, one study revealed that reviewers felt burdened by conducting a survey with reviewers in Ecological Research [9]. Nevertheless, no prior studies conducted in-depth interviews with reviewers for uncovering their needs and barriers. In this paper, we focus on CHI reviewers who contribute to the biggest conference in human-computer interaction fields. We aim to answer the following questions to identify the needs and barriers of CHI reviewers during the peer review process.

- How do reviewers review CHI papers?
- What are reviewers' barriers when reviewing CHI papers?
- What are reviewers' needs for a better peer review process?

To our knowledge, this is the first study that aims to answer those questions qualitatively.

METHOD

Recruitment Process

To recruit study participants, We sent an email with an interview request and the purpose of our study to the potential study participants who has a name on papers published at CHI. The inclusion criteria are individuals who: (1) are between 18 and 65 years old, (2) are Korean (3) have participated as a full-paper reviewer at the CHI conference in recent three years, (4) have one or more publications at CHI as an author or a co-author and (5) have worked in the field of human-computer interaction (HCI) for five or more years.

Study Procedure

Prior to participating in the interview session, each participant was asked to fill out an online survey form. The questionnaire was created based on the following criteria:

- Basics for participant demographics, *e.g., age, area of expertise in HCI, what CHI conference did you participate in as a reviewer?*

	Mean	SD
# of reviewing papers at a recent CHI conference	5.0	4.76
Average # of reviewing papers at each CHI conference	2.8	1.21
Total time spent reviewing papers at a recent CHI conference (hours)	15.7	13.89
Average time spent reviewing papers at each CHI conference (hours)	10.1	5.30
Average time spent reviewing one paper (minutes)	265.7	68.03

Table 1. The number of papers reviewers reviewed, and time reviewers spent on reviewing process during the conference.

- A review process of reviewers, *e.g., devices used for reviewing. # of reviews at last CHI Conference, # of papers typically reviewed at each CHI conference, how many hours used on reviewing papers.*

- How burdensome to a reviewer at each task, *e.g., reading, understanding, rating, commenting, discussion*

Less than ten minutes was expected to complete the survey. We copied the questions of expertise from Precision Conference Solutions (PCS) [3] to better understand areas of their expertise. Prior to identifying tasks that generate a burdensome to reviewers, we divided the review process into five steps: *reading, rating, commenting, reflecting other reviewers reviews, and reflecting rebuttals* based on information gathered from a pilot interview. The pilot interview was conducted with a researcher who has served as a CHI Associate Chair (AC) responsible for recruiting expert reviewers during the peer review process. We also defined six evaluation criteria *(contribution to the CHI community, significance, originality, validity, written presentation, and relevant previous work)* based on information posted on the CHI webpage [2]. To determine the level of burdensome of each step and criteria, we used a seven-point Likert scale where one is not burdensome at all, and seven is very burdensome. Next, we prepared interviews by devising a template that covers following topics:

- Peer review process from reviewers' perspectives, *e.g., what device do you use for reviews? how long does it take to review?*
- Challenges for reviewers, *e.g., what is the most complicated task in the peer review process? which article is most challenging when you serve as a reviewer?*
- Tasks that the reviewer wants some improvement, *e.g., which process do you think that should be improved when reading, rating and commenting?*

All interviews were conducted remotely over Zoom[1] or phone call. Verbal consent was given to each participant at the beginning of each interview for recording and analyzing the data after anonymization. Each interview lasted 30 to 45 minutes. All interviews were audio-recorded and transcribed. We assigned a unique identifier to each participant for analyzing collected interview data. All participants received a $10 worth gift card after participation.

[1] https://zoom.us/

Tasks and rating criteria	Mean	SD
Commenting	4.57	1.988
Reading	3.57	1.618
Rating	3.14	1.573
Reflect other reviewers reviews	2.85	1.471
Reflect Rebuttals	2.85	1.573
Originality	4.28	1.603
Validity	3.85	1.463
Relevant previous work	3.71	1.799
Significance	3.57	1.618
Written presentation	3.42	1.397
Contribution to the CHI community	2.85	1.345

Table 2. Average perceived burdensomeness of each reviewing task and rating criteria to reviewers. (1 to 7 Likert-scale)

Data Analysis

To analyze qualitative data from the interviews, we collected statements from transcripts that described the experiences, thoughts, needs, and barriers of reviewers. An affinity diagramming session was conducted to identify key insights and themes that occurred in the interview data repeatedly by a team of researchers: five junior researchers who have less than a year experience of HCI fields and one senior researcher who has more than ten years experience of HCI fields.

RESULTS

Study Participants

Seven reviewers participated in the study. Their average age is 33.1 years old, with a minimum of 28 and a maximum of 36 years old. In questionnaire about participants expertise that allows duplicate answers, six responded that they have expertise in *mobile devices: phones / tablets* and three responded that they have expertise in *ambient devices / internet of things* in a devices field, four responded *user experience design* and three responded *design methods* is their domain that they have expertise in, and lastly, four participants each responded *prototyping / implementation* and *interaction design* are methods that they have expertise in. The analysis of the questionnaire about participants reviewing tasks revealed the following stats (see Table 1): 1) reviewers, on average, spend 10.1 hours on reviewing process at a conference with standard derivation is 5.30; 2) on average, they spend 265.7 minutes per full-paper to review with standard derivation is 68.03; 3) On average, they reviewed 2.8 papers with a minimum of 2 and a maximum of 5 papers. In the case of the questionnaire that asks how burdensome of each task is (see Table 2), we found that there is a tendency that commenting task is the most burdensome task among five. To be specific, five out of seven participants said commenting is the most burdensome task. About the questionnaire that asks about how burdensome of each evaluation criteria is, Originality is the one that five among seven participants said the most burdensome criteria. We will discuss these results in the discussion section. Lastly, we found that all the participants used a desktop or laptop when reviewing. This is because PCS, web-based service used at CHI for the reviewing process, is designed for desktop user experience. Four out of seven participants said that they used a tablet with a stylus (e.g., an iPad Pro) or papers with a pen to leave a memo while reading, and the other three said that they conducted the
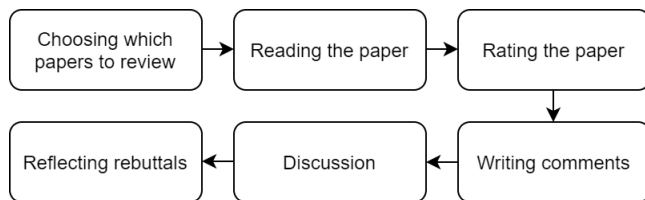
Figure 1. Six review stages identified on the affinity diagramming session. Each stage points to the next stage.

whole review process by desktop or laptop. Also, the three participants who did not use devices that support writing tools said that they leave a brief memo while reading with basic writing software such as Notepad and word processor.

Interview Results from CHI Reviewers
Through an affinity diagramming session [10], we grouped qualitative data into 55 categories and assigned each category into six stages as follows: 1) choosing which papers to review, 2) reading the paper, 3) rating the paper, 4) writing comments, 5) discussion, and 6) reflecting rebuttals (see Figure 1).

The analysis of the grouped qualitative data revealed three key findings: discussions are not preferred to reviewers, discontinuity in space and time during the review, and needs and hopes of reviewers.

Discussions are not preferred to reviewers
What caught our attention was that the reviewers do not prefer to discuss each other:

"For reviewers, In the event of disagreement, it is a very arduous task to unite opinions. To me, reviewing task is just a voluntary task, and I don't think I have to go through my arguments until after the debate." (P6)

Similarly, P2 stated that

"People really do not discuss each other. Because it's a very touchy subject to rebut what others say." (P2)

In addition, The "lack of discussion even though there is a disagreement" situation stated by all participants except P4:

"I have an experience that there were five reviewers, four including me gave 4 points, and one gave 2 points because his point of view was a little different. However, there was no further discussion." (P5)

(was the discussion lively when there is a disagreement?)
"I don't think so. The discussion didn't take place any more. Rather than discussion, it solved by simply attaching another AC." (P6)

However, P5 and P7 mentioned that sometimes the discussion during the peer review process updates the score of reviewed papers.

"Sometimes, some people lead the discussion. Most of them are negative reviewers. they gave low scores with piles of specific reasons. In that case, Those who gave 3.5 points score down to three points." (P7)

"In case of persuasiveness, other reviewers sometimes adjusted their scores. I saw one case that everyone gave a

bad score, but one person gave incredibly good and made a convincing argument about it. So the other reviewers all raised their score one or half points more, and the paper initially rejected was accepted." (P5)

Discontinuity in space and time during the review
Although this is not a dominant opinion, there were difficulties related to spatial and temporal separation. For example, P3 said that he felt uncomfortable about comparing text in the related work section with text in the reference section when reviewing the paper.

"The reference section is at the end of the paper while the related work section is located almost at the front. So I have to check it by clicking on it. It's hard to know which corresponds to what because the signs are all numeric." (P3)

Similarly, discontinuity in time can be seen in the rebuttal process:

"In the case of Rebuttal, I have to review what I have reviewed again. It's been a while since I read it, so it's a little hard to remember what I thought." (P5)

Needs and hopes of reviewers
We found participants' needs and hopes that drawing our attention. First, we noticed that reviewers need some support from determining whether the study described in the paper offers really an original idea or not.

"I thought that this research is really potential and promising, but when I read other reviewer's reviews, I found out that this research is already done by someone else. I was embarrassed then." (P3)

"Yes, I sometimes get confused about whether this paper is a new study or not. Is this a really new one? I wish I could find out whether the paper is a new study more easily and convincing way." (P1)

One of the interesting findings is that they hope some additional features on PCS. In particular, P5 wanted to figure out the characteristics of his review activities from the stored history data for fun, and P4 wanted review deadline countdown features.

"It would be fun to record all my review history and check my avg review score, reviewing tendency, reviewing timeline, etc." (P5)

"I hope PCS to add 'review deadline countdown' feature, which shows me how many days, hours, minutes, and seconds left until the deadline. I always finished my review two days before the deadline. I'm not someone who does the work ahead of time." (P4)

Lastly, one participant stated that an additional procedure that evaluates the quality of the paper in accordance with the length of the paper needed.

"If you don't consider the number of pages, make sure you appreciate the short paper. Obviously, the CHI homepage stated that the page volume is evaluated depending

on the contribution of the article. That's not happening at all... I think there's need for a resurgence of a short paper (to evaluate shorter paper properly)." (P7)

DISCUSSION

From the survey results, we found that CHI reviewers had difficulty when commenting and evaluating the originality of the paper. We believe it was because writing a good comment requires corresponding efforts and that reviewers should check out the latest related work to examine the originality of some studies, which is time-consuming tasks. However, we only interviewed seven reviewers, and our questions are not focused on this; we can not be sure about that. In case of the analysis of the qualitative responses from interviews, there are three key findings: discussions are not preferred to reviewers, discontinuity in space and time during the review, and needs and hopes of reviewers.

We expect the findings of this study could be used to generate a guideline to improve the reliability and satisfaction of an existing peer review system in CHI. However, our study has three limitations. First, as our interview participants have five or more years of experience in the field of HCI, we do not know that findings in this study can be applied to junior reviewers. Second, we conducted research with only a small set of seven reviewers. We do not know what we found can be generally applied. Lastly, we do not know that assisted tools that have features revealed at our qualitative data are genuinely beneficial to reviewers. Future work still remains to develop tools based on the findings of this study (i.e., reviewers' needs) and evaluate the usability and feasibility through a deployment study. Such additional studies may confirm that it is the potential to conduct a large-scale deployment study with reviewers in the real-world peer review process.

CONCLUSIONS

The goal of this study was to understand CHI reviewers' needs and barriers when they perform tasks during a peer review process. The contribution of this study is to define the needs and barriers of reviewers' tasks during the reviewing process. By identifying them, we now know which task we should improve to increase the productivity of the reviewers' reviewing process. Through the survey, we found that reviewers think commenting and rating originality of papers are the most challenging tasks, while reflects other reviewers comment and author's rebuttals are not; it takes about 4 to 5 hours to complete review, and on average, they reviewed 2.8 papers at each conference. Through the interview, interview data revealed three key findings: discussions are not preferred to reviewers, discontinuity in space and time during the review, and needs and hopes of reviewers. Based on our findings, our next study will build a platform that could assist reviewers by showing metrics that could help reviewers to evaluate the originality of papers, a new way of discussion platform to make more discussion-able mood, and a new way of the reading environment to reduce the gap caused by time and space.

ACKNOWLEDGMENTS

REFERENCES

[1] 2019. Conference history of CHI. `https://sigchi.org/conferences/conference-history/CHI/`. (2019). [Online; accessed 2-Jan-2020].

[2] 2019. Guide to a Successful Submission. `https://chi2020.acm.org/authors/papers/guide-to-a-successful-submission/`. (2019). [Online; accessed 2-Jan-2020].

[3] 2019. Precision Conference. `https://new.precisionconference.com/pcs`. (2019). [Online; accessed 2-Jan-2020].

[4] Michael Bernstein, Dan Cosley, Carl DiSalvo, Sanjay Kairam, David Karger, Travis Kriplean, Cliff Lampe, Wendy Mackay, Loren Terveen, Jacob Wobbrock, and et al. 2012. Reject Me: Peer Review and SIGCHI. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12)*. Association for Computing Machinery, New York, NY, USA, 1197–1200.

[5] Jia-Bin Huang. 2018. Deep Paper Gestalt. *CoRR* abs/1812.08775 (2018).

[6] Yvonne Jansen, Kasper Hornbæk, and Pierre Dragicevic. 2016. What did authors value in the CHI'16 reviews they received?. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 596–608.

[7] Jr. Kasenchak, R. T. 2017. Scholarly triage: Advances in manuscript submission using text analytics techniques. *Information Services and Use* 37, 3 (2017), 313–318.

[8] K. Koyamada, Y. Onoue, M. Kioka, T. Uetsuji, and K. Baba. 2018. Visualization of JOV abstracts. *Journal of Visualization* 21, 2 (2018), 309–319.

[9] Y. Okuzaki, S. Nakamura, and M. Nakaoka. 2019. Toward more rapid and efficient peer review: A case study on publication in Ecological Research. *Ecological Research* 34, 5 (2019), 563–574.

[10] Michael Quinn Patton. 2002. Qualitative research and evaluation methods . Thousand Oakes. (2002).

[11] Ágnes Sándor and Angela Vorndran. 2009. Detecting key sentences for automatic assistance in peer reviewing research articles in educational sciences. In *Proceedings of the 2009 workshop on text and citation analysis for scholarly digital libraries*. Association for Computational Linguistics, 36–44.

[12] S. Sikdar, M. Marsili, N. Ganguly, and A. Mukherjee. 2017. Influence of Reviewer Interaction Network on Long-Term Citations: A Case Study of the Scientific Peer-Review System of the Journal of High Energy Physics. In *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. 1–10.

[13] Jeannette Wing and Ed Chi. 2011. Reviewing Peer Review. *Commun. ACM* 54 (07 2011), 10–11.

[14] A. Zoubir. 2014. Peer reviews: What would be the best model? [From the Editor]. *IEEE Signal Processing Magazine* 31, 2 (March 2014), 4–4.