Weiyi Shang

EV 3.129, 1515 Ste. Catherine Street West

Department of Computer Science and Software Engineering, Concordia University

Montreal, QC, Canada H3G 1M8

1 514-848-2424 ext. 7801

July 21, 2016

Dear Editor and Reviewers:

Thank you for your insightful feedback and comments, both positive and constructive, and for allowing us the opportunity to improve our manuscript. We have taken each of your comments into consideration and made the appropriate changes and extensions to our manuscript.

In the end, we added new analysis results as suggested by the reviewers. We studied topics in user reviews in Google Play Store. We also made several clarifications throughout the paper, improved paper presentation, and improved the discussion of related work. We feel that the paper is now much stronger.

Below, we include a description of the changes that we made to our manuscript with respect to each of the reviewers comments. We denote the reviewers comments in italic typeface, and our responses follow below each reviewer comment.

# Reviewer 1 Comment

**Comment R1.1.** *This article studies the user reviews for top mobile Apps in the Google Play Store and compares its findings with those on IOS apps. The study is quite impressive in the scale as they collected user reviews of 10,000+ top apps on a daily basis over a period of two months. Their findings are quite interesting to learn, which have some commonalities as well as differences from previous results. It would be even more interesting if the analysis can have something on the topics and content summaries of the user reviews, e.g., ratings, popular words in the reviews, and so on. Currently it is mostly on the statistics, e.g., number of reviews.*

**Response.** Thanks for the suggestion! We studied the topics in the top 100 most reviewed apps and investigated the global and local topics in app reviews. The study results are briefly discussed in the paper as an infobox due to the imposed word count limits on the actual article. We hope this infobox satisfies your request while still ensuring that we do not exceed the imposed word count limits.

# Reviewer 2 Comment

**Comment R2.2.** *In this paper, the author conducted data analysis over the reviews on Mobile Apps in Google playstore. They found some interesting facts, such as a very small number of Apps received much more reviews than the others. Though the authors argued strongly that their findings are different from Pagano and Maalej's and Hoon et al.'s observations, in fact, there is no New finding in this study. The real difference is that the previous study was conducted over Appstore and this one was conducted on Google Playstore. Thus, I do not think this paper should be published without a new finding.*

**Response.** App stores (and in particular user and developer interactions through app reviews) are a whole new domain of study in computing today. Our study is important in that it raises a warning flag about the importance of looking across such stores and that the computing community cannot just assume that

the observed patterns are universal. Instead according to our study, there are differences between stores. For example, while prior studies highlight that many iOS developers are overwhelmed with reviews, we note that this is not the case for most Android apps. Furthermore, we are the first to dig deeper into the relation between app reviews and other meta data (as pointed by reviewer 3). Table 2 details the key differences between our study and prior studies

# Reviewer 3 Comments

**Comment R3.3.**  *On the other hand, the study remains mostly at the descriptive level, leaving out implications, interpretation, and discussions. I would really appreciate more context in a discussion section on why the study is important and which implications it has, e.g. for automated extraction of topics, triaging etc. as the beginning of the manuscript might "promise".*

**Response.**  The greatest implication and take-home message of this paper is that practitioners and researchers on mobile app should be mindful of the differences among different app stores. Assumptions of having the same observations/patterns across different apps stores may not hold. With the awareness of the differences, techniques that are designed to assist mobile app developers should be optimized for different app stores. Similarly, researchers should examine whether other empirical findings do hold across app stores. We added the discussion of the implication of our paper to the new revision of the paper.

**Comment R3.4.**  *Related work is complete, but could be set more into the context of this manuscript.*

**Response.** The related work is set into the context of this paper in the follow 3 aspects

1. The importance of mobile app user reviews motivates our study.
2. The findings of this paper ascertains the context of prior techniques that analyze mobile app user reviews.
3. Prior empirical studies on mobile app user reviews focus on iOS apps, while our findings on Google Play Store are different.

These points are highlighted in bold in Section "Mobile App Analytics" in the new draft.

**Comment R3.5.**  *Due to the app selection and sampling process, the presented results are only valid for free, established, popular apps - a fact that should be carefully noted throughout the manuscript.*

**Response.** We noted such limitation in the new revision of the paper.

**Comment R3.6.**  *Language and typos should be checked.*

**Response.** Fixed.

**Comment R3.7.**  *Introduction - Yet little is known about the reviewing dynamics of the Google Play Store =>This seems a bit exaggerated.*

**Response.** The sentence is rephrased in the new draft.

**Comment R3.8.**  *The number of downloads and releases impact the number of received reviews =>Should*

be "correlates with" instead of "impacts". In the case of downloads and releases - other than categories - the authors do not know which influences which, so a causal relation is not given per se.

**Response.** The sentence is rephrased as suggested.

**Comment R3.9.** *Mobile App Analytics - ll. 47ff: The related work (apart from the two studies that are also shown in the table) could be presented in more detail. What were concrete outcomes? How do they concretely relate to the authors' work?*

**Response.** Please refer to our response to comment 3.2.

The related work are set to present three aspects as our response to comment 3.2. We added more information about the outcomes of the related work. Related work does not consider that the assumptions of having the same observations/patterns across different apps stores may not hold. Our work makes the implication and take-home message that practitioners and researchers on mobile app should be mindful of the differences among different app stores. With the awareness of the differences, techniques that are designed to assist mobile app developers should be optimized for different app stores. Similarly, researchers should examine whether other empirical findings do hold across app stores.

**Comment R3.10.** *Studied Apps - The fact that only 500 reviews per day were sampled at max is a pity. This does not really allow to describe the full spectrum of how reviews distribute. However, the median _should_ still be stable, given that the authors only found few apps with more than 500 reviews.*

**Response.** Thanks for the suggestion. We find the median number of reviews per app per day to be 0. We elaborate such finding in table 2 in the new revision of the paper.

**Comment R3.11.** *- Only top apps are studied =>this reduces the subjects and generalizability to popular apps - Only "stable" (old) apps are studied =>this reduces the subjects and generalizability to popular, established apps - Only free apps are studied =>this reduces the subjects and generalizability to popular, established, free apps.*

**Response.** We agree that our findings are limited to top, stable and free apps in Google Play Store. We mention such limitations in the new revision of the paper.

**Comment R3.12.** *- Why would spam reviews be a problem? I don't see the argument that there are usually few reviews as an argument for not having spam. Spammers might choose relevant apps. Also I don't think it is relevant that you must login before reviewing, as this is no limitation for spammers. And: In the end there is spam, according to [3] (the study is for iOS, but there you have to login as well).*

**Response.** The paragraph is removed from the new revision.

**Comment R3.13.** *Our Findings - I don't really see much on the beanplots. Maybe a traditional boxplot with whiskers and outliers would give a better impression?*

**Response.** We tried to use boxplots as well. For figure 1(a) and figure 1(b): All other apps, box plots and bean plot would show similar information since the distribution is highly skewed (more than half of the data is 0). On the other hand, the figure 1 (b): 100 most reviewed apps, the bean plot can show the density of apps in different amount of reviews. For example, we can observe that most of the top 100 reviewed apps have around 10,000 reviews during the studied period. Such information can't be shown by boxplots.

However, if the reviewer feels that box plots are more appropriate, we can replace them in the next revision.

**Comment R3.14.** *- ... "4,275 reviews in a single day such large numbers would lead to increasing the overall reported average of the received reviews on a daily basis." => Yes, but not necessarily the median!*

**Response.** We do not mean that the the findings from prior study on Facebook represent the normal situation (like median) of other apps in iOS store. We try to explain that some apps, like Facebook, may have way over 500 reviews per day. Therefore, our limitation of seeing only 500 reviews may be the reason that we observe lower amount of reviews compared to prior studied. However, only 20 apps have more than 500 reviews per day. Therefore, we still calculate the average number of reviews in a single day in order to compare to prior study results.

To further address such limitation, we check the median number of reviews in a single day for each app. We find that the median number is 0. Although Hoon et al. in prior study reports that the median number of reviews per year is 50 and 30 for free and paid apps, respectively, we cannot directly compare such results (median 50 or 30 reviews per year) with our result (median 0 review per day). We report the median number of reviews for each app per day in table 2.

**Comment R3.15.** *- Figure 1b should read "All other apps". It is wrong as it is now.*

**Response.** Fixed.

**Comment R3.16.** *- Most apps receive few feedback and thus would not benefit from automated analyses. =>20 reviews per day is already an amount where you would need a person for reading, triaging, detecting duplicates etc. So I would not subscribe this actually.*

**Response.** We calculate the number of tokens from all the reviews of everyday for every app. The median number of words for app to receive for one day is 46. Such results confirm that most apps do not receive large amounts of feedback. We include such results in the new revision of the paper.

Again, we thank all of you for your valuable feedback, which has made this a stronger manuscript. We look forward to hearing your feedback on the updated manuscript.

Sincerely,
Stuart McIlroy, Weiyi Shang, Nasir Ali, Ahmed E. Hassan