International Association for Management of Technology
IAMOT 2013 Proceedings

THE RELEVANCE OF CUSTOMER REVIEWS IN THE MOBILE APPLICATION MARKETPLACES

SAMI HYRYNSALMI

Turku Centre for Computer Science (TUCS) &
Department of Management, University of Turku
University of Turku, FI-20014 Turun yliopisto, Finland
sthyry@utu.fi

ARHO SUOMINEN

VTT Technical Research Centre of Finland,
Itäinen Pitkäkatu 4, Turku, P.O. Box 106, 20521 Turku, Finland
arho.suominen@vtt.fi

MARKO SEPPÄNEN

Center for Innovation and Technology Research
Department of Industrial Management
Tampere University of Technology
PL 541, FI-33101 Tampere, Finland
marko.seppanen@tut.fi

TUOMAS MÄKILÄ

Business and Innovation Development, University of Turku
University of Turku, FI-20014 Turun yliopisto, Finland
tuomas.makila@utu.fi

Abstract

The effect of user generated product ratings has been studied by academics in different contexts from books to movies and its importance has been noted by the industry. In this study, the relevance of customer ratings in a mobile application ecosystem with two different datasets, over 174,000 applications, gathered from Google Play in December 2011 and in November 2012 is analyzed. The results show that a high average rating correlates positively with sales improvement as well as a high variance correlates positively with the number of installations, both statistically significant. The correlations are, however, small. On contrary to the suggestions of previous studies, the price of the application does not seem to affect the importance of product ratings. Further research would investigate how application providers' brands and visibility in social media may influence on the popularity and its growth in mobile application ecosystems.

Keywords: Product ratings, eWOM, Mobile application ecosystems, application marketplaces, Google Play

Introduction

Product ratings by customers at the online marketplaces have been investigated in both conceptual studies (Awad and Etzion, 2006; Chen and Xie, 2005, 2008; Mayzlin, 2006) and empirical evaluations (Godes and Mayzlin, 2004; Gao, Gu, and Lin, 2006; Zhu and Zhang, 2010). Previous studies have analyzed the effects of the product ratings mainly for tangible products. For example, Sun (2012) presented recently a conceptual model, in which a high average of consumer ratings indicates a high quality product whereas a high variance in consumer ratings indicates a 'niche' product. The proposed model was verified with books sold through the Amazon web store.

In contrast to the majority of traditional online stores, the offering in the newly emerged mobile application marketplaces differs in two dimensions. Firstly, most of the applications offered in the mobile application marketplaces are cheap, only a few dollars or less — or even free with e.g. an advertisement based revenue stream. Secondly, several application developers offer customers an option to try the product before the purchase decision. These aspects combined to intangible nature of application business might challenge the traditional assumptions of consumer ratings since there is little or no financial interest to protect.

In this paper, we investigate the effects of the consumer ratings in the mobile application marketplace. Our assumption is that due to micro-prices and trial versions of the applications, consumer ratings are not as important as in other marketplaces. We are focusing on the following research questions:

RQ1 Does a high average of ratings correlate with the sales of an application in the mobile application ecosystem?

RQ2 Do the consumer ratings matter more at the sale of more expensive applications?

RQ3 Does a high variance of ratings combined with a high average indicate a niche product in the mobile application ecosystem?

In order to answer the research questions, we selected Google Play, the application marketplace of Android ecosystem, for a case study subject due to its popularity, high adaption rate, and from the practical point of view, the information provided in the marketplace. The study was conducted by crawling data on applications published in Google Play. Three datasets were gathered: Two complete datasets in December 2011 and November 2012, and one partial dataset in December 2012. The datasets include over 400,000 individual applications. The first set is used as a reference point for consumer ratings evaluation and the second one in analyzing the effects to sales. The third set contains variance data and it is used to answer to the RQ3.

The results indicate that there is a correlation between a high average of ratings and the sales improvements, and the niche products can possibly be identified with a high variance. The found correlations are rather small. However, we did not find evidence on that the ratings in more expensive applications would matter more than in the low-cost ones. As a managerial implication, the findings suggest the use of ratings in a mobile ecosystem as a measure of quality is difficult and require further studies to improve the existing models. It should be noted, that this study is limited to Google's mobile ecosystem only. Further analysis of other mobile ecosystems and application marketplaces (e.g. Apple's iOS ecosystem, Microsoft' Windows Phone ecosystem) is needed to further validate the results.

The paper is structured as follows. The next section will briefly present some of the previous studies done on the topic. A description of research procedure used follows. The fourth section presents the results of statistical analysis and it is followed by a discussion section. The final section concludes the study with some suggestions for further research on the topic.

Background

Theoretical background

De Maeyer (2012) reviews the theoretical background of online consumers' reviews and their impact on products sales and more precisely on price strategies. He identifies a significant increase in literature on the subject resulting in the increased use of online reviews. The challenge with the online review is to grasp the multidimensional nature of the system and understand the implications of different actors and actions in the system.

The theoretical background of customer online reviews stems from literature on electronics word-of-mouth (eWOM). One of the first works on the subject are based on Balasubramanian and Mahajan (2001) and Hennig-Thurau et al. (2004) that focused on eWOM on designated consumer-opinion platforms. Since then the increased adoption rate of eWOM has made the original research by the authors even more influential, but as noted by Hennig-Thurau et al. (2004), future research is needed to validate empirical results and create deeper understanding on the concepts and phenomena (De Maeyer, 2012).

One challenging issue is the impact of positive and negative feedback. Although the conventional logic of thinking would argue that positive eWOM, or positive feedback on a product, would result on higher product sales this might not be the case. In the work of Berger et al. (2010), the multidimensionality of reviews is clearly shown, as their results suggest that negative word-of-mouth can increase product sales through increasing awareness. As Hyrynsalmi et al. (2012a) have shown, the mobile application market is highly dynamic and this might suggest that overall product awareness is even more important than positive reviews. This leads us to the first research question:

> *RQ1: Does the high average of ratings correlate with the improvements in the sales in the mobile application ecosystem?*

Focusing clearly on a product centered approach, previous studies have found that the less popular the product the more it will gain or lose through the impact of online reviews (Zhu and Zhang, 2010; Vermeulen and Seegers, 2009). The online reviews clearly have a supporting role in decision-making. This support system might be of significance in the mobile application ecosystem only in cases that the applications are of high cost. The mobile ecosystems are to a significant extent based on free or low-cost applications – the consumer can test the majority of applications with no or little cost. This limits the need for an additional support system for the customer purchase decision. However, with an increasing cost, the customer might be more interested in the review as a decision support system. This leads us to the second research question:

> *RQ2: Do the consumer ratings matter more at the sale of more expensive products?*

In the recent study, Sun (2012) shows that in addition to the average of the ratings, the variance of the reviews might be useful for customers to identify interesting products. In her model, a high average indicates a product of overall good quality whereas the high variance of the ratings might indicate a niche product. She found empirical evidence to support the theorem by analyzing books sold in Amazon.com and in Barnesandnoble.com. The new theorem leads us to the third research question of this study:

> *RQ3: Does a high variance of ratings combined with a high average indicate a niche product in the mobile application ecosystem?*

Previous Studies on Mobile Application Marketplaces

Previous studies on the user ratings in online application marketplaces argue that the user ratings establish a fundament element in the applications marketplace. For example, Hao et al. (2011a) argue that "*– the app[lication] market where ratings play a central role in determining the consumer's* ex ante *perceived net utility as well as their willingness to pay*" while Apple states in their iOS Developer guidelines that "*[c]ustomer ratings and reviews on the App Store can have a big effect on the success of your app –*"[1].

Although user ratings have been studied to the great extent, in the context of application or software marketplaces, to the best of authors' knowledge, the existing literature is significantly thinner. Hao, Li, Tan, and Xu (2011a, 2011b) developed a theoretical framework to assess the importance of ratings to the application markets. Carare (2012) showed that the bestselling rank has an impact on determinant of demand. The result might indicate that the user ratings are not as crucial as the visibility in the top list. Ha and Wagner (2013) analyzed the content of review comments in Google Play, focusing specifically on the privacy and security issues. They found that most of the comments are about general quality of the application and only a few were concerning security issues.

Hyrynsalmi et al. (2012a) studied the correlation between the paid applications download category and the average ratings in Google Play. They assumed that in free to install applications, user ratings are less relevant as it is easier to try the application than read the review comments. They found small negative, although statistically significant, correlation between the number of downloads and the applications' average rating. The data of 52,679 applications included several products with only a few reviews which might have hampered the result. Furthermore, their dataset presented a static situation, and thus it cannot be estimated if the rating had effect on the growth of popularity of an application.

In addition to abovementioned studies, several have suggested to replace the user generated rating system. Yan and Chen (2011) remark that the user ratings requires laborious hand work and thus, potentially lacking reviews will be sparse. Thus, they suggested an application recommendation system. Similar work to replace star rating systems have been done by several authors, e.g. Girardello and Michahelles (2010a, 2010b); Lim et al. (2011); Davidsson and Moritz (2011).

[1] "Developing for the App Store: Publishing an App in the App Store" by Apple Inc. `http://developer.apple.com/library/ios/#documentation/General/Conceptual/ApplicationDevelopmentOverview/DeliverYourAppontheAppStore/DeliverYourAppontheAppStore.html`. Accessed on January 21st, 2013.

Research Procedure

This study is based on the application data collected from the Google Play digital marketplace. The high-level steps of the research method are quite straightforward and are based on the statistical analysis on the data gathered from the marketplace. These steps are

(i) Form the list of the applications using the third party listings;

(ii) Gather the individual application data from the Google Play;

(iii) Import the data to Microsoft Excel 2010 and combine two applications via unique package names; and

(iv) Analyze the data using IBM SPSS statistical software.

The technical details of the method are more complex and they are described in details in Hyrynsalmi et al. (2012b).

We use three datasets collected in December 2011, November 2012 and December 2012. The data variables collected were held same between the datasets. The only notable difference in the context of this study is that only the last dataset contains the exact numbers of votes in different rating categories. Due to constraints set by the data published by Google, before December 2012 we stored only the average of all ratings together with the number of votes for each application as given by Google. In the course of this study we have been able to widen the search to include the actual values of different votes. However, at the time of writing of this study, crawling this data was not ready. Thus, we were forced to use a random subset of data, which consists of more than 100,000 applications.

Subsequently, we use the following variables:

Number of ratings implies how many people have rated the application. A user can rate the application by using star ratings from one to five stars. In addition to the number of stars given, the user can also write a review of the application. The variable is parsed from the marketplace.

Variance of star ratings presents distribution of ratings for a single application in the marketplace. The variable is calculated from the parsed data.

Average rating is a value calculated and published by Google. This represents, as described by Google, an arithmetic mean calculated from the star based ratings. It should be noted, that using an arithmetic mean as an average measure of a Likert type value is challenging. Thus, we use the value as a proxy for consumer rating simultaneously being aware of this limitation. The variable is parsed from the marketplace.

Change in average ratings denotes difference in two average rating values. It is calculated by subtracting the newer value from the older.

Installation category illustrates how many times an application has been downloaded and installed[2] to a single device divided in the rough categories published by the marketplace. The published categories are on the half-logarithmic scale, i.e. ‘1–5’, ‘5–10’, ‘10–50’, ‘50–

[2] When the data gathering started in December 2011, Google used the term ‘downloads’ instead of ‘installs’. However, it is unknown if this change in terminology have any actual effect to the published numbers.

100', etc. A missing value is interpreted as zero. The highest noted installation category at the time of the study was '100,000,000–500,000,000'.

Change in installation category denotes difference between two installation category values. The variable is ordinal, as we calculate how many steps an application's installation category has advanced during the study period.

We used Spearman's rank correlation coefficient (also known as Spearman's rho) to analyze the correlation between the variables as we did not want to make any presumption on the distributions of variables – and we know the distributions are skewed. Furthermore, the installation category and changes in installation category are ordinal variables. The Spearman's rho values are calculated with IBM SPSS version 20.

Results

The crawled dataset contains 339,861 applications in December 2011 and 407,920 applications in November 2012. From these two datasets, we identified 242,620 applications that existed in both crawled datasets.

In an early stage of data inspection, we identified outliers and anomalous records. Applications' installation category was substantially decreased in 77 cases. We assume that these applications were removed and restored during the study period, and therefore the download counter was reset. Furthermore, we identified one particular developer who had several applications included in this anomaly list. In the case of 1,882 applications, the number of user ratings has decreased. These cases might be also a result of removing and restoring an application or the reviews might have been inappropriate and thus deleted, or even the accounts of the reviewing users might have been removed. The above records were not included into the analysis. We also excluded 65,706 (27.3 %) applications from the analysis as they did not have any ratings at the start date of analysis. As we are analyzing the relevance of customer reviews, we did not focus on applications that are so new that they have not been even reviewed yet. Finally, we selected a set of 174,967 (72.1 %) applications for the study.

A median growth in installation categories was a single step. The installation category of 80,719 (46.1 %) applications did not change during the measurement period. For 37.0 % of applications, the installation category advanced only one step and for 16.8 % of applications' installation category growth two or more steps. Only eight applications advanced eight steps during the one-year period of the study.

On average, the number of votes almost tripled (+282 %) in the set of studied applications in the given period. The variance on the increase of the votes was also significant. On one hand, 41,959 (24 %) applications did not get any new customer reviews. On the other hand, there was an application for which the number of votes grow from 14 votes to 246,698 at the same time when its installation category advanced from the category '1,000–5,000' to the category '5,000,000–10,000,000'. Interestingly, the application's average rating remained the same (4.4 stars in scale 1-5). The variance of the data can be explained by the fact that the applications were in different phases of their life cycles when the datasets were constructed.

Altogether, when analyzing the data for this study we should note that the large dataset contains a significant amount of outliers and anomalous records. The above procedure and rigorous validation process (e.g. calculating the data results in two computers individually and the validating the results) has significantly reduced the number of irregularities. In addition, the authors have no control over the data provided by Google – data being the result of a web crawling process – and we should note that the values as they are given are of commercial value to Google. In the following, when analyzing the results for the RQs, we should be aware of the aforementioned constraints.

RQ1. Does a high average of ratings correlate with the sales of an application? To answer the research question, we define the null hypothesis '*there is no correlation between changes in installation category and average ratings*' and the alternative hypothesis '*there is positive correlation between average ratings and changes in installation category*'.

We analyzed the correlation between the average ratings in 2011 and the changes in installation category during the measurement period of eleven months. The Spearman's rho revealed a statistically significant relationship between the average rating and the amount of change in download category, ϱ[174,967] = 0.18, $p < 0.001$. Thus, we can reject the null hypothesis. This indicates that there is a statistically significant correlation between high (and respectively low) average in star ratings and high (small) level of advancement in number of installations.

RQ2. Do the consumer ratings matter more at the sale of more expensive applications? In order to answer to the research question, we divided applications into five price groups: free, below 1 euro, 1 to 2 euros, 2 to 5 euros, and over 5 euros. The division is artificial and created by the authors in order to ensure that each price group has enough applications for the statistical analysis. The price groups and the correlations between average ratings and differences in installation categories are presented in Table 1.

We hypothesized that users' interest towards the reviews of the peers grow when the price of the product increases. From Table 1 we see that the Spearman's rho reveals a statistically significant relationship between average ratings and changes in installation categories in each price category (ϱ[174,967]= 0.13−0.19, $p < 0.001$). Although the correlations slightly differ in distinct categories, the values are still considerable similar. Furthermore, the numbers indicate that the correlation in free and cheap applications is slightly stronger than in the paid ones. Thus, our hypothesis is not supported, since the user reviews are not considerable more important in the expensive applications.

Table 1: Correlations between average ratings and changes in installation categories of applications in different price groups.

Price group	**N**	**%**	**ϱ**	**p**
0	132,547	75.6	0.19	<0.001
(0−1)	20,756	11.9	0.17	<0.001
[1−2)	11,094	6.3	0.18	<0.001
[2−5)	8,039	4.6	0.18	<0.001
[5−	2,531	1.4	0.13	<0.001
ALL	174,967	100	0.18	<0.001

RQ3. Does a high variance of ratings combined with a high average indicate a niche product? We used a random subset of data gathered in December 2012 to address this question. In total, the subset contains 105,069 applications and over 54 million reviews. Interestingly, the reviews are highly skewed towards the high end: 65.8 % of reviewers gave five stars, 17.3 % gave four stars, 7.0 % gave three stars, 2.5 % gave two stars, and 7.3 % gave one star.

In order to study correlations, we assumed that the star ratings are similar kind of data as Likert scale data. Thus, we calculated the proportions of votes for each star class and for these proportions, we calculated variances and standard deviations. Moving on to study a correlation between the variance of user ratings and installation category, we found a significant medium positive correlation, $\varrho[105,069] = 0.36$, $p < 0.001$.

Our hypothesis is that the high variance correlates with the high installation category when the average rating is high. The null hypothesis is that there is no such correlation. To calculate this, we followed Sun (2012) and used the product of an average rating and a standard deviation as the variable. We found a medium positive correlation, which is statistically significant, between the new variable and the installation category, $\varrho[105,069] = 0.42$, $p < 0.001$. This suggests that, indeed, there is correlation with high variance and high installation category when the product average rating is high. However, it is worth to note that the variance alone correlates considerably well with the installation category, and in this analysis we were restricted to static data.

Discussion

There have been arguments presented by both academics and practitioners on the relevance of the consumer reviews to success of an application. In this study, we used data gathered from the mobile application marketplace, Google Play, in order to analyze the previously presented claims. We focused on three questions: does a high average of ratings correlate with sales of an application (RQ1), do the consumer ratings matter more at the sale of more expensive applications (RQ2), and does a high variance of ratings with a high average rating indicate a niche product (RQ3)?

Our data revealed a small positive and statistically significant correlation between average ratings and sales improvements during the study period of eleven months. This would suggest that, indeed, a higher average rating improve the sales (RQ1). However, it should be noted that the correlation is rather small. Also, we noticed 42,393 (24.2 % of all) applications, which average ratings was over four stars; however, their installation category did not change during the study period.) Although consumer reviews can be important to software vendors as a quick feedback channel, we would argue that focusing on improving the average ratings does not pay off. In addition, the users' rating conventions – almost two thirds of ratings were the highest one – skew the usefulness of the reviews. In further studies, we could compare this skewness and its potential effect with other studies of consumer/user based assessments e.g. in fields of psychology or marketing.

We hypothesized that due to micro-prices, the consumer reviews are not as important as with tangible products since the barrier to try an application product is low. On the other hand, with more expensive paid applications, the ratings would be more important because a user will consider the buying decision more closely. The data did not support this hypothesis: there seem to be no difference in correlations between average ratings and advanced installation levels between different price categories (RQ2). Thus, we suggest that users do not use the average rating presented in marketplace when making the buying decision.

Our third question was to study how the variance of ratings could be used in the marketplaces. Sun (2012) proposed that a high variance and a high average rating would suggest a niche product: some users like the application and praise it, the other do not find it useful and criticize it. We studied this theory by investigating the correlation between the product of the average rating and standard deviation of ratings, and advancements in installation category. There was a significant positive correlation (RQ3), thus supporting the hypothesis.

We studied more closely a dozen randomly selected applications with a high average and high variance. Based on the reviews written by the users, it seems that most of these applications really are niche products: some of the comments were negative either based on technical problems or uselessness of the product while the others praised its features. However, one of the studied applications had only negative verbal comments but still more than half of the users (n=97) had awarded five stars. We would assume that these reviews could have been done by, e.g., developers themselves, their close ones or by even an outsourced review service[3] – or by users with a strange sense of humor. However, this would again raise questions on the usefulness of the user ratings.

Our starting assumption was that the specific features of the mobile application ecosystems would cause consumer ratings to be less important than in other online marketplaces for tangible products. That is, when a consumer is buying a new computer, he invests a considerable amount of money to that purchase. Similarly, when he is buying a new book, he invests his free time to read through the book. In these cases, it is natural to browse through the review information in order to ensure good return value to the invested resource. Compared with a time-consuming book the decision process in purchasing, e.g., a DVD movie is much more straight-forward although the price-point between these two items is quite similar. The characteristics of the mobile applications favor the simple decision-making during the purchasing. Most of the applications are cheap and do not require much time and effort from the user. The existence of free trial versions lowers the barrier even further. In addition to these features, the delivery of a new application is instant whereas the shipment of a tangible product takes time.

Although in this study we find a positive, statistically significant, correlation between an average rating and sales improvement, the correlation is remarkably small. This suggests our starting assumption might be justified. Furthermore, the swift nature of the mobile applications in contrast to a tedious work of reviewing the application in the marketplace might cause that the user reviews are sparse and uninformative. We saw this phenomenon when reviewing the verbal comments of users: most of reviews were only a few words long and focused on either praise the goodness or bash the problems of the application. We did not found any long review on the pros and cons of a product although these kinds of comments are quite common in other online marketplaces such as Amazon.com.

As an implication, we suggest that use of ratings, in this context, as a measure of quality or an indicator of future sales is more complex than earlier considered and thus requires further study. Although there is a correlation between high ratings and sales improvements, there are many different factors, most of them outside of this study and the dataset used, affecting to the outcome. For instance, in the most installed applications, i.e. the superstars of the ecosystem, there seem to be many negative and positive reviews. Therefore, the average rating as an indicator does not seem to be justified. This does not hamper the use of the user reviews as a way to gather fast feedback to improve the application.

One clear factor affecting to the results and overall usefulness of the user ratings is that the marketplaces rarely ask users to review an application. The marketplace, however, ensure that

[3] See “PR firm settles with FTC over alleged App Store Astroturfing” by Chris Foresman, arstechnica `http://arstechnica.com/tech-policy/2010/08/pr-firm-settles-with-ftc-over-alleged-app-store-astroturfing/` Accessed on January 20th, 2013.

the reviewing users had installed the application before the review can be placed. Therefore, the users seem to evaluate applications only when they are either extremely disappointed or satisfied with the application. Some applications have a built-in feature, which reminds every now and then a user to review the application in the marketplaces[4]. Some of these applications filter users so that those with positive feedback are directed to the marketplace's rating service and those with negative are directed to the application vendor's feedback page[5]. Again, these kinds of behavior distort data and decrease the usefulness of reviews as an indicator of quality.

Generalization of the results is limited by focusing on the analysis only on one mobile ecosystem. Further analysis of different mobile ecosystems and their marketplaces is needed to further validate the results. For discussion about the limitations of the data gathering process and completeness of data, see Hyrynsalmi et al. (2012b). It should also be noted that data contain lots of 'noise', e.g. applications that are launched as a hobby, and applications meant only for a small group of users. These applications most likely will not fit the presented theorems that suppose each application is meant for commercial use.

Conclusion

This study analyzed the relevance of user ratings in Android operating system's application marketplace. We used large datasets of applications gathered from the Google Play. We found that a high average rating correlates positively with improvements in sales, and a high variance of user ratings together with a high average rating correlate positively with a high number of installations. However, we did not find evidence on that consumer ratings would be more important to costly applications than to the low-cost ones. The study is limited by focusing only on one marketplace and the results should be validated with other mobile ecosystems. In future studies, also the effect of strong brands should be noted instead of only analyzing user ratings. Furthermore, also other factors, e.g. visibility in social media, influencing to the sales improvement should be studied carefully.

Acknowledgements

Authors wish to thank B.Sc. (Tech.) Miika Oja-Nisula for excellent technical support. In addition, Hyrynsalmi is grateful for Nokia Foundation for funding his dissertation work on mobile ecosystems.

[4] See e.g. *Apprirater* utility library for iPhone `https://github.com/arashpayan/appirater/`. Accessed on January 18th, 2013.

[5] See e.g. "Improving Your App Store Ratings" by Cory, Mobile Orchard `http://mobileorchard.com/improving-your-app-store-rating/`. Accessed on January 18th, 2013

References

Awad, N. and H. Etzion (2006). Stay out of my forum! evaluating firm involvement in online ratings communities. In Workshop on Information Systems and Economics (WISC), Evanston, IL. Northwestern University.

Balasubramanian, S., and V. Mahajan (2001).The Economic Leverage of the Virtual Community. International Journal of Electronic Commerce 5 (3), 103–138.

Berger, J., A. T. Sorensen, and S. J. Rasmussen (2010). “Positive Effects of Negative Publicity: When Negative Reviews Increase Sales. Marketing Science 29 (5), 815–827.

Carare, O. (2012). The impact of bestseller rank on demand: Evidence from the app market. International Economic Review 53(3), 717–742.

Chen, Y. and J. Xie (2005). Third-party product review and firm marketing strategy. Marketing Science 24(2), 218–240.

Chen, Y. and J. Xie (2008). Online consumer review: A new element of marketing communication mix. Management Science 54(2), 477–491.

Davidsson, C. and S. Moritz (2011). Utilizing implicit feedback and context to recommend mobile applications from first use. In Proceedings of the 2011 Workshop on Context-awareness in Retrieval and Recommendation, CaRR ’11, New York, NY, USA, pp. 19–22. ACM.

De Mayer, P. (2012) Impact of online consumer reviews on sales and price strategies: a review and directions for future research. Journal of Product and Brand Management 21(2), 132–139.

Gao, G. G., B. Gu, and M. Lin (2006). The dynamics of online consumer reviews. In Workshop on Information Systems and Economics (WISE), Evanston, IL. Northwestern University.

Girardello, A. and F. Michahelles (2010a). Bootstrapping your mobile application on a social market. In Proceedings of UbiComp’10. ACM.

Girardello, A. and F. Michahelles (2010b). Explicit and implicit ratings for mobile applications. In K.-P. Fähnrich and B. Franczyk (Eds.), Informatik 2010: Service Science - Neue Perspektiven für die Informatik, Beiträge der 40. Jahrestagung der Gesellschaft für Informatik e.V. (GI), Band 1, 27.09. - 1.10.2010, Leipzig, Volume 175 of LNI, pp. 606–612. GI.

Godes, D. and D. Mayzlin (2004). Using online conversations to study word-of-mouth communication. Marketing Science 23(4), 545–560.

Ha, E. and D. Wagner (2013). Do android users write about electric sheep? examining consumer reviews in google play. In Proceedings of the 10th Annual IEEE Consumer Communication & Networking Conference, CCNC 2013. IEEE.

Hao, L., X. Li, Y. Tan, and J. Xu (2011a). The economic role of rating behavior in third-party application market. In Proceedings of Thirty Second International Conference on Information Systems, ICIS 2011, pp. 1–15. Paper 26.

Hao, L., X. Li, Y. Tan, and J. Xu (2011b). The economic value of ratings in app market (July 18, 2011). Available at SSRN: http://ssrn.com/abstract=1892584.

Hennig-Thurau, T., K. P. Gwinner, G. Walsh, and D. D. Gremler (2004). Electronic Word-of-mouth via Consumer-opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet? Journal of Interactive Marketing 18(1), 38–52.

Hyrynsalmi, S., A. Suominen, T. Mäkilä, and T. Knuutila (2012a). The emerging mobile ecosystems: An introductory analysis of Android Market. In The 21st International Conference on Management of Technology, pp. Hsinchu, Taiwan. International Association for Management of Technology.

Hyrynsalmi, S., T. Mäkilä, A. Järvi, A. Suominen, M. Seppänen, and T. Knuutila (2012b). App Store, Marketplace, Play! an analysis of multi-homing in mobile software ecosystems. In S. Jansen, J. Bosch, and C. Alves (Eds.), Proceedings of the International Workshop on Software Ecosystems, IWSECO'2012, MIT Sloan School of Management, Cambridge, MA, USA, pp. 55–68. CEUR-WS. Available at CEUR-WS: http://ceur-ws.org/Vol-879/.

Lim, Y., H.-M. Kim, S. Kang, and T.-h. Kim (2011). Recommendation algorithm of the app store by using semantic relations between apps. In T.-h. Kim, H. Adeli, R. Robles, and M. Balitanas (Eds.), Ubiquitous Computing and Multimedia Applications, Volume 151 of Communications in Computer and Information Science, pp. 139–144. Springer Berlin Heidelberg.

Mayzlin, D. (2006). Promotional chat on the internet. Marketing Science 25(2), 155–163.

Sun, M. (2012). How does the variance of product ratings matter? Management Science 58(4), 696–707.

Vermeulen, I. E. and D. Seegers (2009) Tried and Tested: The Impact of Online Hotel Reviews on Consumer Consideration. Tourism Management 30(1), 123–127.

Yan, B. and G. Chen (2011). Appjoy: personalized mobile application discovery. In Proceedings of the 9th international conference on Mobile systems, applications, and services, MobiSys '11, New York, NY, USA, pp. 113–126. ACM.

Zhu, F. and X. M. Zhang (2010). The impact of online consumer reviews on sales: the moderating role of product and consumer characteristics. Journal of Marketing 74(2), 133–148.