Journal of Database Management
Volume 35 • Issue 1

Examining the Usefulness of Customer Reviews for Mobile Applications:
The Role of Developer Responsiveness

Zhiying Jiang, Singapore University of Social Sciences, Singapore
https://orcid.org/0000-0001-8014-6963

Vanessa Liu, Singapore University of Social Sciences, Singapore*

Miriam Erne, Erasmus University, Rotterdam, The Netherlands

ABSTRACT

In the context of mobile applications (apps), the role of customers has been transformed from mere passive adopters to active co-creators through contribution of user reviews. However, customers might not always possess the required technical expertise to make commercially feasible suggestions. The value of customer reviews also varied due to their unmanageable volume and content irrelevance. In our study, over 189,000 user reviews with over 50 apps would be analyzed using review analysis and multivariate regression analysis to examine the impacts of innovation and improvement led by customers on app performance in terms of app revenues. The developers' lead time in responding to user reviews would be included as a moderator to investigate whether app performance would be enhanced if developers respond faster. This study should represent one of the first few attempts in offering empirical confirmation of the value of co-creation of apps with customers. The authors also present methodological contributions by establishing operationalization and analyses of user reviews.

KEYWORDS

Customer Led Improvement, Customer Led Innovation, Mobile Apps, User Involvement, User Reviews

INTRODUCTION

Nowadays application distribution platforms such as Apple App Store and Google Play provide millions of different mobile applications (apps) to users. As of the second quarter of 2022, there were around 3.50 million apps for android users and 2.18 million apps for App Store users available (Statista, 2022a). Survival in such a "hyper-competitive" mobile market was challenging to apps developers (Comino et al., 2019). Unwanted or unpopular apps could be phased out very shortly after launch, resulting in a waste of development cost and effort. To sustain competitiveness, it is therefore becoming increasingly important for app developers to pursue continuous improvement and launch novel features that meet customer needs (e.g., see Chen et al., 2014; Maalej and Hadeer, 2015; Maalej

DOI: 10.4018/JDM.343543
*Corresponding Author

et al., 2016). As customers are equivalent to users of mobile apps, the terms "customers" and "users" are used interchangeably in this paper.

Mobile apps often serve to provide users with functions in a specific domain, such as for productivity, gaming, lifestyle and entertainment, and etc. From this perspective, app development and maintenance can by and large be regarded as a form of service innovation and quality control. Management scholars have reached a consensus that understanding customer needs constitutes an essential foundation for innovative product or services and hence sustained competitiveness. Customer involvement is important as it reduces uncertainty that usually underlies the innovation process. Thomke and Von Hippel (2002) elaborated this point from an information asymmetry perspective that the 'need' information resides with the customers, and the 'solution' information lies with the producers. Hence, customers' perception of strengths and weaknesses of existing features as well as desires for new functions is critical to service providers at both strategic (e.g., recourse allocation) and operational (e.g., quality control) levels.

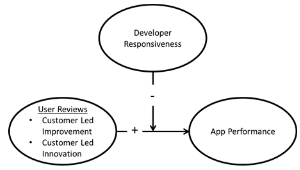
The need for customer-based information has prompted various information collection approaches such as satisfaction surveys, unsolicited customer complaints, interviews, focus groups or even personal observation. These methods have evolved into what has become known as customer involvement management. For app developers, the natural way to obtain customer-based information is through user reviews. User reviews, as a continuous flow of information, enables developers for quicker identification and action for problems (Finch, 1998; Nambisan, 2002; Parthasarathy and Daneva, 2021) highlighted how a virtual customer environment resolves two challenges often associated with customer involvement management in the offline context. Firstly, it enables customer involvement directed towards a diverse set of customers. Secondly, it becomes possible to get connected with the customers at a relatively lower cost.

Indeed, app developers are motivated to actively elicit customer comment due to many reasons. Good reviews visible to future adopters are positive signals about app quality and form potential user's quality perception. Positive reviews serves as word of mouth with impact on growth and revenue. One important reason for obtaining reviews is that they are enlightening to the app developers in terms of novel features. As customer needs vary significantly and the usage of the apps could differ across contexts, customers may be a good source of creative ideas for development of innovative functionalities. With actual usage experience of the apps, customers are able to spot a non-working feature. For example, the IOS based game Sky: Children of the lights perform fairly smoothly on IOS devices but not on Android devices due to the vast number of different brands and models of devices supported by the Android platform. User reviews could therefore help detecting bugs and enable continuous improvement of the apps.

Despite its potential usefulness for performance enhancement, screening through user reviews could be challenging. For instance, online gurus like Facebook could generate as high as at least 2,000 user reviews per day (Chen et al., 2014). The aspects covered in the reviews could be highly diverse, ranging from complaints about the price of the apps to the frequency of advertisements. Tools have accordingly been developed to enable automated categorization and mining of customer reviews (Maalej and Hadeer, 2015; Maalej et al., 2016). However, following up on these reviews remains highly time- and money-consuming. Little empirical evidence is available to prove it worth the resources to act upon user reviews. Considering not all customers are technically knowledgeable about app development, it is also not clear whether their involvement really offer constructive and commercially feasible suggestions for app improvement.

User involvement is only appropriate if certain involvement roles and development conditions are fulfilled (Ives and Olson, 1984). These conditions include, who should be involved, which type of software with which the users should be involved, and in which stage (i.e., when) of the software development the users should be involved. User involvement could be totally undesirable when technical expertise is required. While the potential value of customer feedback is not deniable, it

Figure 1. The conceptual model



may not always be economically justified for developers to translate their feedback into actual app features (Ives and Olson, 1984).

This study therefore aims to empirically investigate the impact of user reviews on app performance. Most prior researchers focused on the development of analytical tools for categorization of user reviews (e.g., Maalej and Hadeer, 2015; Maalej et al., 2016), presuming that customers could always provide useful feedback. In this study, over 189,000 user reviews associated with over 50 apps were categorized and analyzed to verify the impact of user reviews on app performance. Specifically, user reviews with innovative suggestions are conceptualized as "customer led innovation" and those with bug-fixing suggestions as "customer led improvement". In order to quantify the impact of addressing user reviews, app performance was measured in terms of revenues generated from the app and the number of downloads (Liang et al., 2015; Lee and Raghu, 2014). The time taken for app developers to respond to the user reviews was also taken into consideration. The value of the innovativeness of user inputs may depreciate over time as other competitors might have already launched similar features onto the market. Similarly, customers might get disappointed if the developers did not promptly address the errors they pointed out. Developer responsiveness was included as a moderator on the relationship between customer led innovation/ improvement and app performance.

The rest of this paper is organized as follows: first, the conceptual framework and the related past studies will be introduced. The research methodology and the data analysis procedure will then be presented. Finally, the findings will be discussed and the theoretical and managerial implications will be drawn.

CONCEPTUAL FRAMEWORK

The conceptual framework is presented in Figure 1. It is drawn on the notions of user involvement and service quality to propose a direct effect of customer led innovation/ improvement on app performance. Developer responsiveness is included as a moderator on this direct relationship.

User Involvement

The notion of user involvement was well documented in the literature, referring to the level of personal relevance and importance attached by users to the system (Barki and Hartwick, 1989). In broad terms, it is defined as "direct contact with users" (Kujala, 2008). Recently, it was observed that customers had become more and more involved in the product development ((Prahalad & Ramaswamy, 2004)). User involvement was essential and indispensable for system/ software developers as it helped to collect more accurate user requirements and enable quality improvement, resulting in better fulfillment of user needs and higher user satisfaction (Lederer, 1993; Kamel, 1995; Kaulio, 1998; Kujala, 2008). User involvement was therefore recognized by previous researchers as beneficial to the improvement of quality and performance (Berger et al., 2005). Terms such as co-creation or co-design had emerged to describe the collaboration between developers and users. Other terms included quality function deployment (QFD), user-oriented product development, concept testing, Beta testing, consumer idealized design, lead user method and participatory ergonomics (Kaulio, 1998). In the collaborative process, users may assume the roles of providers of information, commentators or objects for observations.

Customer involvement is critical to reduce uncertainty that often surrounds innovation process. Due to the information asymmetry, the 'need' information resides with the user and the 'solution' information lies with the provider (Thomke and Von Hippel, 2002). The need for customer based information has prompted a variety of collection methods that have developed into a domain known as customer involvement management. The alternative means for firms to listen to its customers include satisfaction survey, focus group, unsolicited customer complaints, personal observation, and etc. Internet discussions as a source for customer involvement was first discussed by Finch (1998). In his paper, Finch argues that internet discussion as a continuous flow of customer perception sheds insights on the strengths and weakness of the existing products or service features. Such customer-based information enables firms for quicker identification and action for problems, thus can be deemed as a form of customer involvement. Nambisan (2002) stated that a virtual customer environment resolves two challenges often associated with customer involvement management. Firstly, it enables customer involvement directed towards a diverse set of customers. Secondly, it becomes possible to get connected with the customers at a relatively lower cost. Hence, in recent literature, it is not uncommon to see user reviews are often regarded as a form of customer involvement, especially in the domain of software engineering (Dabrowski et al. 2022).

In the context of mobile apps, customers and apps developers may exchange ideas on shared platforms such as the App stores. Customers could submit their desirable new features or functionalities (Khalid et al., 2015; Panichella et al., 2015). Complaints from users on lack of certain features could shed light on potential new apps development (Barlow et al., 2016). Customers may highlight bugs such as incompatibility or poor functionality (Khalid et al., 2015; Panichella et al., 2015). However, the number of user reviews could be voluminous and hard to manage. More importantly, not all feedback is useful. Almost 65% of app reviews were found to be noisy and irrelevant (Chen et al., 2014). Some suggestions might be solely emotional and commercially infeasible for adoption.

Many tools were therefore developed to aid the search, screening, and extraction of useful information from user reviews. A review of the current literature showed that different tools were built with different mining objectives. Examples included MARK (Mining and Analyzing Reviews by Keywords) (Vu et al., 2015), MARA (Mobile App Review Analyzer) (Iacob and Harrison, 2013), ALERTme (Guzman et al., 2017), and AR-Miner (App Review Miner) (Chen et al., 2014). These tools made use of techniques like natural language processing, topic modeling, clustering and machine learning algorithms to search, classify, extract, group and rank user reviews based on pre-defined keywords or categories.

Mobile App Performance

The construct of performance is a nuanced and multifaceted aspect, encompassing various dimensions. In the specific context of mobile apps, performance can be conceptualized as the service quality (Kuo et al., 2016), sustained functionality or overall success of the application (Liang et al., 2015). This entails not only the operational aspects but also the app's ability to meet user expectations and achieve its intended goals over time.

Mobile app performance is a complex interplay of functional and non-functional characteristics (Hort et al., 2021). Functional characteristics are intricately tied to the specific nature of the app's services, making them highly app-specific. On the other hand, non-functional characteristics extend beyond the direct functionalities and can encompass broader service aspects such as responsiveness. In the scope of this study, our attention is directed towards exploring the non-functional performance characteristics of mobile apps, with a particular focus on aspects like responsiveness that contribute to user satisfaction and overall app success.

Examining non-functional performance in the mobile app landscape entails a consideration of both app-level and sell-level attributes. At the app level, attributes such as rankings and the number of downloads play a pivotal role in shaping the perceived success and reach of an application (Lee and Raghu, 2014). Additionally, some researchers have utilized sales revenues as a proxy for assessing an app's success, highlighting the financial dimension as an indicative measure of performance (Liang et al., 2015).

In this study, we adopt a non-functional approach, evaluating mobile apps based on their number of downloads and the revenues they generate.

Responsiveness and App Performance

Service quality is traditionally defined as a user's assessment of the "overall excellence or superiority" of a service (Parasuraman et al., 1988). A predominant method for measuring service quality is the application of the SERVQUAL scale, comprising five dimensions: tangibles, reliability, responsiveness, assurance, and empathy (Parasuraman et al., 1988).

The concept's definition has evolved with the rise of e-services (Zeithaml et al., 2000) and mobile services (Tan et al., 2008). Various models have been developed to gauge e-service quality, introducing distinct dimensions. For instance, Zeithaml et al. (2002) proposed a seven-dimensional e-service quality model, encompassing ease of use, privacy, graphic design, information availability, reliability, compensation, and contact. Alternative models have also surfaced, suggesting additional elements specific to the e-service context, such as fulfillment, efficiency, availability, and privacy (Parasuraman et al., 2005).

Considering mobile apps as a form of service, given the deployment of the mobile channel for delivering value to users (Balasubramanian et al., 2002; Kuo et al., 2016), Tan et al. (2008) posited that mobile service quality should include seven dimensions: perceived usefulness, perceived ease of use, content, variety, feedback, experimentation, and personalization. More recently, Huang et al. (2015) introduced the M-S-QUAL scale, distinguishing factors influencing virtual product shopping and physical product shopping experiences. Kuo et al. (2016) synthesized seven attributes for mobile service quality: efficiency, fulfillment, privacy, responsiveness, personalization, tangibility, and reliability.

In this study, we focus on the examination of the role of responsiveness in affecting mobile app performance.

Responsiveness manifests in various forms, with interpretations ranging from the technical functionality of the user interface (Mirzoev and Kane, 2017) to the accountability of service providers to users (Lodenstein et al., 2016). In the realm of mobile app development, responsiveness takes on a user-centric perspective, primarily concerned with enhancing the user experience. This involves the adept handling and incorporation of user feedback into the app's evolution (Khan et al., 2021). In essence, responsiveness encompasses both the actions undertaken to address user feedback and the timeliness with which these actions are executed. For instance, a user might submit a request for bug resolution or propose a new feature, and the subsequent addressing of these requests represents the responsive nature of the app developer. However, it is crucial to recognize that the time taken to fulfill these requests is equally pivotal. Prolonged delays in addressing issues may lead to user dissatisfaction, potentially resulting in app uninstallation.

This study takes an empirical approach to scrutinize the impact of responsiveness on mobile app performance, discerning between the actions initiated in response to user reviews and the timeframe within which these actions are executed. We categorize these actions as customer-led innovation and customer-led improvement, emphasizing the user-driven nature of the responsiveness concept. Simultaneously, we introduce the term "developer responsiveness" to encapsulate the temporal aspect, elucidating how the efficiency and timeliness of addressing user feedback contribute significantly to the overall responsiveness dynamics in the mobile app ecosystem.

Customer Led Innovation and Customer Led Improvement

User reviews, if carefully and properly screened and processed, could be vital to innovativeness of app development. For example, a customer might point out interesting and novel features that could be added for iPhone users. With many varieties of smartphones available and varied user profiles, it was difficult for app developers to consider all possible new features. User reviews could be a good source to identify creative solutions. Though some users may be tech-non-savvy, the imaginativeness may never be foreseen in the development process. Their feedback could help developers to visualize innovative features of the apps. Similarly, customers are in a better position to detect bugs based on their actual usage experience, such as the incompatibility of apps with certain phone models. User reviews with new feature requests are therefore conceptualized as customer led innovation. It denotes requests from users on new features to be added to the apps or new app development. Customer led innovation offer insights to developers to add novel features, resulting in greater efficiency of development and higher user satisfaction (Kujala, 2008). On the other hand, user reviews with suggestions on improvement are conceptualized as customer led improvement. It denotes reports from users about unwanted errors, bugs, annoying advertisements and other usability problems. If these user reviews are addressed properly, more customers will be attracted to purchase the apps and hence higher revenues could be generated (Kujala, 2008). Notably, it is not only the number of reviews on improvement or innovation matters. These reviews could only be effective in improving the app if they are being addressed.

Accordingly, it is hypothesized that:

Hypothesis One: Customer led improvement (i.e., customers' feedback on improvement being addressed) has a significant and positive impact on app performance.

Hypothesis Two: Customer led innovation (i.e., customers' feedback on new features being addressed) has a significant and positive impact on app performance.

Developer Responsiveness to User Reviews

The time taken by developers to respond to user reviews on app innovation and improvement may matter (Vaniea and Rashidi, 2016). After a customer submitted his/her feedback, he/she may tend to expect the developer to address the suggestion quickly. For example, if the developer response is slow, the current customers may continue to experience the bugs in the regular apps usage and may eventually rescind usage or even uninstall the apps. Apps are experience products, when it comes to experience goods, the impact of electronic Word of Mouth (eWOM) is particularly salient (Litvin et al., 2008). Given the critical influence of eWOM, delaying responding to the consumer requests will impede the app performance for new customer acquisition (Xie et. al., 2014, Kim & Kim, 2023). Conversely, customers may tend to be more positive about the apps if their concerns and problems were addressed promptly. Timely responses are even more crucial for suggestions of new features. The degree of novelty would be diminished and the risk of being copied by competitors would increase over time. In general, reasonable responsiveness should lead to better quality and performance of apps (Hort et al., 2021; Burgess et al., n.d.). In this paper, the developer responsiveness is measured as the time span between a suggestion raised by a customer and an action taken by the developer. Hence, the shorter the time taken to respond to user reviews, the greater the effect is the reviews on app performance.

Accordingly, it is hypothesized that:

Hypothesis Three: Developer responsiveness significantly moderates the relationship between customer led improvement and app performance. The longer the time taken to respond customers' requests (the more responsive a developer), the weaker the impact of customer lead improvement on the app performance.

Hypothesis Four: Developer responsiveness significantly moderates the relationship between customer led innovation and app performance. The longer the time taken to respond customers' requests (the more responsive a developer), the weaker the impact of customer lead innovation on the app performance.

RESEARCH METHODOLOGY

Data and Operationalization

Data Source

The data was obtained through App Annie, the largest business intelligence company in app industry. It collects key metrics of most apps on both IOS and Android platform. For this study, apps in the iOS Health and Fitness category in the United States were collected. Apps active between 1st March, 2016 and 28th Feb, 2017 were sampled. In this way, the sampled apps have survived at least for one year. The reason is that in a long-tail industry like mobile apps, many apps phase out before they actually take off. The "Top App" rank lists the best 1,000 apps in the category. In these 1,000 apps, 43 apps accounted for 75% of the total revenue in the Health and Fitness market. Nine of these apps had both, a free and a paid version. This results in 52 apps, of which two were excluded due to insufficient reviews. Hence, eventually 50 apps with 189,537 reviews were analyzed in this study. App updates, app reviews and ratings were publicly available data, which had been consolidated and could be retrieved from App Annie data base through its API. Private app performance data such as app revenue, app downloads and app ranking are obtained from App Annie.

Review Analysis

Review data is unstructured text data. To generate any insights based on quantitative analysis, the unstructured text data needed to be transformed to structured numeric data. Latest techniques in natural language processing (NLP) such as text categorization, topic modelling and thematic analysis are often applied to analyze review data. However, these frequently applied NLP techniques are not able to address the research questions. For supervised learning methods such as text categorization, a training data and a list of classifiers is supposed to be provided. Take sentiment analysis for example, emotions are categorized as positive, negative or neutral. A pre-designed list of phrases is fed to the algorithm to indicate the emotional valence. In this research, how new features proposed by customers could lead to app success was studied. The proposed new features were specific to each app and could not be pre-determined by the researchers without examining the review data. In other words, it is not possible to generate a list of features that have been proposed by customers to implement text categorization procedure. Topic modelling and thematic analysis belong to unsupervised learning NLP procedure. Hence, no pre-determined list of tags is needed to train the algorithm. However, given the massive volume of reviews, the topics summarized can be quite irrelevant to the research goals and screening these topics can be equally time-consuming. Similarly, thematic analysis is useful in recognizing the underlying patterns, yet it is still challenging to recognize specific key information. Hence, these automated NLP procedures cannot address the data requirements of this study.

This study examines the impact of customer led innovation and improvement on apps' performance. Hence, it is necessary identify reviews proposing new features and reviews reporting bugs and issues. Such reviews were then matched with app update data to see if the customer suggested innovations or improvements had been adopted by the developers. The workflow of this review analysis involves two steps (see illustration in the Appendix). Step 1: separating specific user reviews by filtering out generic reviews. Step 2: searching for matches between the feature updates and reviews, and simultaneously categorizing the reviews into (1) bugs, (2) feature requests, (3) user experiences, (4) ratings (5) pricing, (6) "too many advertisements".

Step 1: Extracting Useful Reviews

Similar to Chen et al, (2014), reviews that did not contain relevant information were eliminated. Reviews were categorized into two broad categories, generic and specific user reviews. Generic reviews are comments to feedback users' overall evaluation or experiences about the app. Examples of such reviews are ‘love it’, ‘by far the best app on meditation!!’, etc. On the other hand, specific reviews often indicate an actionable function that the app developer can fix, improve or create. Examples of specific reviews include ‘better to have a timer’, ‘the interface background should be personalized’, etc. In this research, generic reviews were deemed as not informative, and thus were filtered out.

Step 2: Review Categorization

With the generic reviews filtered out, the specific reviews were further categorized into six different categories. In a prior study, Maalej and Hadeer (2015) used keywords to indicate the categorization. They proposed keywords for four categories of reviews, namely, “bug”, “feature request”, “rating”, and “user experience”. Keywords for the categories of “pricing” and “too many advertisements” were also added for this study. The definition of these six categories is elaborated below with examples.

1. Bug

Keywords: bug, fix, problem, issue, defect, crash, solve

A bug report is in general a not wanted error in a program or system, they arise mainly because of programming failures by developers. A bug is any kind of problem with the app, a crash, an erroneous behavior, or a performance issue (Maalej and Hadeer, 2015). Therefore, a bug is literally anything a user is complaining that is not working right, but it is not a bug if a user is wishing for something new.

Examples of bug reviews:

“It’s not letting me sign up and I deleted the app and re downloaded it but it’s not working”

“It isn’t letting me make an account and says error try again later”

“If you open the app in the watch it tries to connect for a minute (literally a minute) then crashes”

2. Feature Request

Keywords: add, please, could, would, hope, improve, miss, need, prefer, request, should, suggest, want, wish

In general, a request for a new feature is when the user thinks something should have been developed, that does not exist yet, therefore, it requires new code (Cheung, 2013; Wiggins, 2015). A feature request is the wish for new and missing functionality, if users speak about new ideas to make an application better in the future, or if they compare the app with missing features that similar apps offer (Maalej and Hadeer, 2015). If somebody is wishing for new content in the application, this is treated as a feature request too.

Examples of feature request reviews:

“Needs to have a value for calories burned for strength training too”

“Missing Apple Watch compatibility”

3. User Experience

Keywords: help, support, assist, when, situation

User experiences derive from the individual, not social, and is this individual's response and perception that emerges from the use of the product or service as in the ISO 9241-210. In this research, user experience is described as the reflection of users experience with the app and app features (Maalej and Hadeer, 2015). User experience reviews are mostly positive with high star ratings and users are talking about why they like the app and how it changed their life.

Examples of user experience reviews:

"This app is fantastic in every way. If used correctly, it can be life changing. It integrates with MapMyWalk, transferring calorie expenditure to your daily calorie requirements. The community is warm and supportive and you'll have all the tools you need to lose weight and get healthy. Good luck!"

"The WW app keeps me focused on the choices of food I eat and how each selection effects my well-being. It's a great app!"

"I love this app! Tracking is so easy. I can find foods with the search bar quickly. The Connect feature gives support from members all over the world."

4. Rating

Keywords: Great, good, nice, very, cool, love, hate, bad, worst

Ratings are often text reflections of the numeric star rating (Maalej and Hadeer, 2015). They are very generic and are less relevant for the app development process as they do not hold useful information.

Examples of rating reviews:

"Love it!"

"It's a pretty good app, I like it."

5. Pricing

Keywords: expensive, price, pricing, rip off, $, cost, overpriced, fee, pay, payment, paid, cheaper

Pricing refers to the user evaluation of the price of the app. Examples could be whether the price is over-priced, too expensive, unfair or underpriced etc.

Examples of pricing reviews:

"Overpriced...over rated...only tracks calories...$10 a month to track macros...ridiculous"

"Your initial fee only gives you a handful of exercises. To get additional workouts it costs more $$."

6. Too Many Advertisements

Keywords: ads, advertisements, popping up, annoying, banner

This keyword refers to user comments on the frequency and quantity of advertisements inherent in the usage of the app. For example, there may be too many pop up advertisements or banner. Alternatively, the advertisements may have taken up too much of the user interface of the app.

Examples of reviews for too many ads:

"The avalanche of ads makes it unusable unless you pay $3 each and every month."

"Paid for the ap. Still get ads pushed to me. Don't advertise to me if I paid the money for the non-ad version."

Descriptive Statistics for Review Analysis. Based on categorization between generic and specific reviews, 82% of the reviews were generic reviews while 18% were specific reviews. This first step of categorization was to filter out major irrelevant information and facilitate further categorization. In terms of ratings associated with reviews, they were quite positive as majority of the reviews scored a 5-star rating (76.8%), followed by 4-star (, 13.2%), 1-star (5.1%), 3-star (2.9%) and 2-star (2%) respectively. The data shows that the user satisfaction (90%) with the selected 50 apps was higher than the average user satisfaction with mobile apps (78%) (pagano's and maalej, 2013). The reason could be that the selected apps were the top 50 apps in the category.

One interesting finding is that generic reviews were much more positive than the specific review. Almost 98% of the generic reviews had at least 4 stars while it is only 54% for specific review. Though the ratings varied across different apps, the result held even at the app level. Table 1 presents apps with a high number of generic reviews versus apps with a high number of specific reviews. It shows that apps with more generic reviews in general received higher ratings than apps with more specific reviews. This is consistent with the conjecture that specific reviews carried more information on app quality improvement. As shown in Table 1, if app feature improvement and innovation were mostly proposed in specific reviews and apps with more specific reviews were likely to receive low ratings, then how would customer led innovation and improvement help app performance? It echoes back to the research question that the key might lie in whether the proposed improvement or innovation had been adopted by the developer.

Further Categorization of Specific Reviews. The specific reviews were then further classified. The classification was accomplished by workforce employed from Amazon MT. Various validation procedures had been implemented to ensure the workforce understand the task requirements and their task quality. Out of these reviews, 25.75% were classified as "user experience", 24.13% as 'rating", 24.03% as "bug", 16.01% as "feature request", 5.29% as "pricing", 2.79% as "too many advertisements", and 2.00% as "others".

Table 1. Rating differences across app type

High Generic				High Specific			
App Name	Total Number of Reviews	% Generic	Ø Rating	App Name	Total Number of Reviews	% Specific	Ø Rating
Calm	14964	85,6%	4.8	Sweat with Kayla	3455	62.8%	2.7
Fitness Buddy	10251	85.7%	4.6	Beachbody	1902	71.1%	2.9
Instant Heart Rate+	9898	88.5%	4.7	Fitplan	194	57.1%	3.3
Life Period Tracker	16713	90.6%	4.9	Lifesum	2856	49.6%	3.7
Relax Melodies	17902	92.2%	4.8	Weight Watchers	14393	50.4%	3.4

Variable Operationalization

Customer Led Innovation. If customers are part of the innovation process. He or she expresses the need for a new feature in a review and app developers implement that feature in a future update. This is a match between a review and a feature from the update. Examples of this type of reviews from the data set are "the app works great, but has become very out of date with all the new blends available. For the price of the app, i would expect regular updates as new oils and blends become available." The corresponding app launched the new feature "added new oils and blends" in their next update amongst others things. Hence, there was a match. The match between a user review and feature update counts only if the review date was prior to the update date, otherwise the feature was not innovated or led by customers. The level of measurement was interval scaled, where one unit represented one match between a review and feature update. If three different reviews matched all with the same feature update, the variable had a value of three.

Customer Led Improvement. The variable customer led improvement is similar to customer led innovation. When the user helped to improve the app, it was likely to be a report about bugs, annoying advertisements, or usability problems. The level of measurement was interval scaled, and one unit represented a match with a review. The match counts only if the review date was prior the update date.

Developer Responsiveness. Developer responsiveness to customer led innovation was differentiated with that to customer led improvement.

Developer responsiveness examines whether the time taken for app developers to respond has a moderating effect on the impact of user reviews on app performance. The level of measurement was scaled and one unit represented one week. Developer responsiveness to customer led innovation measures time span between the date an update is announced and the date the first review that requires this new feature. The variable was constructed as the average of all features that matched the update.

Similarly, Developer responsiveness to customer led improvement refers to the time interval from the first user review until the bug was fixed / advertisement was removed in days. Again, the variable was constructed as the average time of all fixed bugs in the study period.

App performance. App performance could be operationalized in a number of ways like apps ratings etc. In this study, app performance was measured using both the revenue generated from the apps (Liang et al., 2015) and the number of total downloads (Lee and Raghu, 2014) during the research time frame. Revenue allows the examination of the financial impact on the app developers more directly. Revenue as a performance indicator measures how well the product and service is sold on the market. Revenues could include purchases of apps, micro-transactions within an app or in-app advertisement (iadv) (ghose and han, 2014). The revenues for each app were computed by a summation of the daily revenues for the research time frame. It is a key measure for any for-profit organization to sustain and develop their business. However, given most of the apps under study adopt a freemium monetization model, it is not uncommon to see the top 10% to 20% of the users subsidize the rest of the users in the app industry. The large pool of the free users might not bring immediate monetary value to the app, but they help accelerate the app adoption and they also constitute as the main source of potential customers. Hence, using revenue as the only performance indicator is myopic and ignores the growth potential of the apps under study. The number of downloads is therefore also used as an additional proxy of app performance. The number of total downloads enables the measurement of app adoption and diffusion in the research period. The number of total downloads was obtained by a summation of the daily downloads.

Table 2. Variable descriptive statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Revenue	50	233	1965418	292013.80	428985.40
Downloads	50	133	898599	176441.84	188303.53
Customer_led_Improvement	50	0	176	11.92	31.11
Customer_led_Innovation	50	0	41	4.74	8.61
Responsiveness_Improvement	50	0	51	12.13	16.54
Responsiveness_Innovation	50	0	50	13.74	17.51

Variable Descriptive Statistics

The descriptive statistics of variables used in the conceptual model are presented in table 2. In total 50 apps were studied, thus the valid sample for each variable is 50 with no missing values. Revenue is measured as the sum of daily revenue over the study period, which ranges from $233 to $1.96 million. The number of downloads is also measured as the sum of the daily downloads during the research period. Multiple downloads by the same user are counted as a unique download. For apps with a free and a premium version, downloads of both versions are recorded separately. The number of downloads across different apps ranges from 133 to 898,599 with a mean of 176,441.84. The mean of customer-led improvement (11.92) is almost three times of the mean of the customer-led innovation (4.74). As expected, a longer responsive time for app developers to take action to the innovative ideas proposed by users than the responsive time taken to act upon the proposed app improvement. To the surprise of the authors, the mean responsiveness to customer-led innovation (13.74) and customer-led improvement (12.13) do not differ much.

To examine the interrelationships among the constructs, we computed Pearson correlations for the variables in our study as shown in Table 3. As anticipated, a robust correlation emerged between the performance indicators: App Revenue and App Downloads. There is a moderate correlation between Customer-led Improvement and both App Revenue and App Downloads. However, Customer-led Innovation shows only a weak correlation with App Downloads and no significant correlation with App Revenue. Consistent with our expectations, Customer-led Innovation and Customer-led Improvement are strongly correlated. Regarding the moderating variables, neither Responsiveness Improvement nor

Table 3. Pearson Correlation Table

	Revenue	Downloads	Customer_led_ Improvement	Customer_led_ Innovation	Responsiveness_ Improvement	Responsiveness_ Innovation
Revenue	1	**.743****	**.420****	.193	.140	.026
Downloads		1	**.466****	**.370****	.148	.106
Customer_led_ Improvement			1	**.665****	**.540****	**.424****
Customer_led_ Innovation				1	**.399****	**.424****
Responsiveness_ Improvement					1	**.350***
Responsiveness_ Innovation						1

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Responsiveness Innovation shows a correlation with the dependent variables. However, both exhibit a weak or moderate positive correlation with the independent variables."

RESULTS AND DISCUSSION

Multiple regression analysis was deployed to test the relationships between the key constructs. Namely, how customer led innovation and improvement impacted the app performance (revenue and number of downloads). These linear models were estimated with SPSS OLS procedure. The linear models (1) are used to test H1 and H2 (see table 4).

$$\text{App Performance} = \beta_0 + \beta_1 Customer\ Led\ Improvement + \beta_2 Customer\ Led\ Innovation + \varepsilon \quad (1)$$

Model (2) and (3) were used to test the moderation effect indicated in H3 and H4.

$$\text{App Performance} = \beta_0 + \beta_1 Customer\ Led\ Improvement$$

$$+\beta_3 Customer\ Led\ Improvement \times Responsiveness\ Cusomter\ Led\ Improvement + \varepsilon \quad (2)$$

$$\text{App Performance} = \beta_0 + \beta_2 Customer\ Led\ Innovation$$

$$+\beta_3 Customer\ Led\ Innovation \times Responsiveness\ Cusomter\ Led\ Innovation + \varepsilon \quad (3)$$

Direct Effect of Customer Led Improvement and Customer Led Innovation on App Performance

As shown in table 4 that customer led improvement has a positive effect on both app revenue and total number of downloads. However, such impact is not supported for customer led Innovation in both revenue and download model. The overall model is significant, with an F-value of 5.517 for the revenue model and an F-value of 6.777 for the downloads model, hence hypothesis 1 is supported ($p= 0.005$). In terms of fitness of good, an R2 of 0.190 is satisfactory with cross-sectional data, where values of 0.10 are typical (Sarstedt & Mooi, 2019). Holding other factors constant, increasing customer led improvement by one unit leads to revenue increase by $7,292.185 and 2390.485 more downloads. Note that the increase in revenue refers to the time frame of 12 weeks. It makes sense that reported bugs by users can increase the app performance as it is challenging for developers to locate every bug in the jungle of different smartphones, different platforms, and different system updates. In addition, if a user takes effort to report a bug or propose a feature improvement, it implies their user experience might have been quite impaired. Addressing such issues enhances user satisfaction, hence better app performance.

Customer led product innovation (H2) is not significant in either model ($p=0.388$ and $p=0.536$). There are several reasons for this. From the users' perspective, most of the users were passive users in an app's user pool. Users who proposed new feature were often experienced users with creativity. Unlike requests for feature improvement, most of the users were less compelled to propose new features than reporting bugs. After all, bugs or feature failures interrupted their usage and were easier to detect. Hence, much fewer requests on new features (16%) were found than requests on feature improvement (24%). From developers' perspective, developing a new feature was costly and risky. It also involved changes to a user's habit, such as adapting to new interfaces, finding functions in new places, etc. Hence, if the new features did not enhance user experiences significantly for majority of

Table 4. Moderation Regression Estimation

	Revenue					Downloads				
Model 1										
	Estimators	**Std. Error**	**Sig**	**Part**	**F**	**Estimators**	**Std. Error**	**Sig**	**Part**	**F**
β_0	241998.174	63816.770	**.000**		**5.517**	136822.467	27414.767	**.000**		**6.777**
β_1	7292.185	2456.607	**.005**			2390.485	1041.910	**.026**		
β_2	-7761,506	8906.398	.388			2347.002	3763.517	.536		
Model 2										
	Estimators	**Std. Error**	**Sig**	**Part**	**F**	**Estimators**	**Std. Error**	**Sig**	**Part**	**F**
β_0	201049.292	59989.023	**.002**		**6.745**	128125.850	24530.433	**.000**		**11.351**
β_1	22054.394	9860.088	**.030**	.288		13719.751	4031.941	**.001**	.408	
β_3	-397.484	237.114	.100	-.216		-266.394	96.960	**.008**	-.329	
Model 3										
	Estimators	**Std. Error**	**Sig**	**Part**	**F**	**Estimators**	**Std. Error**	**Sig**	**Part**	**F**
β_0	285009.429	75781.101	.000		1.644	157563.278	31266.829	.000		4.959
β_2	-11415.171	19003.868	.551	-.085		-2531.637	7840.882	.748	-.043	
β_3	26.411	22.183	.24	.168		13.345	9.152	1.458	.193	

the users, the developer would not initiate such update. Therefore, even fewer requests for new features were observed to be addressed by the developer. During the study period of the dataset, each app on average had five matched new feature requests, but 12 requests for feature improvement.

Moderating Effect of Developer Responsiveness on the Relationship Between Customer Led Improvement/Innovation and App Performance

Table 4 reveals significant main and moderating effects in Model 2. Specifically, customer-led improvement significantly positively influences both app revenue and downloads ($p = 0.02$ and $p = 0.001$, respectively). The moderating variable, developer responsiveness to customer-led improvement, shows a weak significance ($p = 0.100$) in the revenue model and strong significance ($p = 0.008$) in the download model. This suggests that slower responsiveness (i.e., longer response times) diminishes the positive impact of customer-led improvement on app revenue. In line with Hypothesis H3, quicker resolution of issues leads to enhanced app performance.

In Model 3, however, both the main and moderating effects lack significance. Echoing the findings of Model 1, the impact of customer-led innovation on app performance (Hypothesis H4) is

Table 5. Summary of Findings from Data Analysis

Hypotheses	Results
Hypothesis One (H1): Customer led improvement has a significant and positive impact on app performance.	Accepted
Hypothesis Two (H2): Customer led innovation has a significant and positive impact on app performance.	Rejected
Hypothesis Three (H3): Developer responsiveness significantly moderates the relationship between customer led improvement and app performance. The shorter the time taken to respond customers' requests (the more responsive a developer), the larger the impact of customer lead improvement on the app performance.	Accepted
Hypothesis Four (H4): Developer responsiveness significantly moderates the relationship between customer led innovation and app performance. The shorter the time taken to respond customers' requests (the more responsive a developer), the larger the impact of customer lead innovation on the app performance.	Rejected

not significant ($p = 0.551$ and $p = 0.748$) in either the revenue or download models. Surprisingly, the moderating effect of responsiveness on app innovation also shows no significance ($p = 0.24$ and $p = 1.458$) in both models. This pattern may be attributed to two factors. First, developing innovative features typically takes longer than improving existing ones. Unlike delayed improvements (e.g., bug fixes), the absence of new features does not disrupt current app usage, leading to greater user tolerance for prolonged development times. Second, introducing new features often requires redesigning user interfaces, which can overwhelm users. While users expect immediate attention and frequent updates for improvements that affect their immediate use of the app, they are more accepting of lower frequency responses for innovative features that add utility but might alter the user interface. Therefore, we do not observe a significant moderating effect of developer responsiveness on Customer_led innovation.

The findings from the data analysis were summarized in table 5.

CONCLUSION

In this new era of digital transformation, customers have been empowered to take a much more active role in value co-creation with product developers. In particular, in the context of mobile apps, the role of customers has emerged from merely adopters to co-creators through voicing out their ideas in app reviews. The extant literature documented that such user involvement should lead to enhanced app performance. However, this presumption might not hold in the context of mobile apps, where hundreds or even thousands of user reviews may be easily generated online. The volume of user reviews might be hardly manageable and the quality and relevance of reviews might also vary significantly. This study therefore attempted to provide empirical evidence on the effect of user reviews on app performance.

The findings of this study confirmed that addressing customer led improvement reviews could significantly lead to improvement in app revenues. Such positive effect is even more remarkable if follow-up actions on the user suggestions are taken promptly by the app developers. Conversely, customer led innovation was not found to have a significant impact on app revenues. Responsiveness to these suggestions, however, has a significant yet weak moderating effect on such link between reviews on innovation and app revenues.

THEORETICAL AND PRACTICAL CONTRIBUTIONS AND IMPLICATIONS

The results present important theoretical and managerial implications. To the information systems (IS) literature, this study offers empirical evidence on the value of user reviews on app performance. Specifically, the financial impact of addressing user reviews on innovation and improvement respectively on revenues generated from the apps was measured and quantified. While the extant literature documents a positive effect of user involvement on app performance, the findings found that only reviews on feature improvement (i.e., customer led improvement) could significantly increase app revenues. On one hand, this establishes the role of customers as value co-creators and their impacts in the mobile app development process. They are no longer mere app adopters. The added value of customers is salient to improvement of app features. The insignificant effect of customer led innovation implies that user involvement may not be applicable to all contexts of app development. When creativity and innovativeness is required as in the case of new feature suggestions, customers may not be able to provide substantial insights.

This research also provides methodological contributions to the IS literature. It demonstrated how user reviews of customer led improvement and customer led innovation could be categorized and analyzed using review analysis. It also illustrated how developer responsiveness could be operationalized by matching app updates with their corresponding user reviews.

This research should offer fundamental value for both research and software engineers. Extensive extant research in both management and information system confirms software engineers or app developers derive valuable user information from online review data to guide software feature refinement, to bridge the gap between the developers and the users, to increase market transparency and improve release management (AISubaihin et al. 2021, Martin et al. 2017, Zhang et al. 2019 and Dabrowski et al. 2022). For example, for feature requirement, analysing app reviews can help software engineers to elicit new features desired by the users (Johann et al. 2017). For testing, app reviews can help in identifying bugs (Iacob et al. 2016; Shams et al. 2020). For release management, app reviews may help prioritize requested changes (Villarroel et al. 2016; Gao et al. 2018; Gao et al. 2019). For the richness of the app review data, app review analysis becomes an important source that software engineers seek information on app development. In a recent publication, Dabrowski and his co-authors (Dabrowski, 2022) presented a comprehensive survey research, covering 182 papers on app review analysis published from 2012 to 2020. This stream of literature posits that app review contains critically valuable information from the customers' side and responding to such customer requests enhances app performance. However, no empirical research ever established such relationship between user review analysis and the app performance. The purpose of this study is to link review analysis with app performance and test empirically whether such presumption establishes or not. It is believed that this research is the first to test such relationship in the literature, thus the research has fundamental implications to the domain of review analysis.

App developers may benefit from the findings in several ways. First, this research empirically examined and proved that launching update in response to a customer led improvement review could lead to an increase in app revenues by $7,292.19 within 12 weeks. This should be good reference for the developers to assess the costs and benefits in responding to customers' suggestions on feature improvement. As the moderating effect of developer responsiveness was found to be significant, app developers should therefore take prompt and timely actions to address requests from customers on bug fixing and feature improvement. However, app developers should exercise discernment in addressing customer led innovation reviews. The results indicate that responding to these reviews may not necessarily lead to any significant revenue growth, even if timely actions are taken.

LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH

Remarkable effort in both research and industry has been devoted to review analysis under the assumption that customer related information extracted from user reviewers can guide app development, thus better app performance. This research aims to test empirically if such effort establish in the first place. The proposed model links customer involvement with app performance, a relationship moderated by developer's responsiveness. As an explorative study and given limited information, the current model is a reduced form model, which does not have many structures built in. The only structure that the authors are able to build into the model is the way the independent variables are constructed, which meaningfully represents the behavioral process a developer incorporates the customer's request into the app development. However, it is believed such reduced model is sufficient to address the research goal. As shown in both the revenue and download model, the main effect of customer involvement is consistent with the same sign across the models. However, the model will surely be improved and provide more insights, should the constructs such as leadership quality, marketing campaign data etc. are accessible.

As this study was conducted in the context of health-related mobile apps only, future researchers may examine the effect of user reviews on app performance of other nature. This should offer insights on whether the nature of apps would have an impact on user involvement. This research also involves mainly apps in the United States. The generalizability of the findings may also be affected by factors specific to the US context, such as languages supported by the apps, phone models prevalent in the US or even network infrastructural issues. Future researchers may examine apps developed in other countries as a comparative study. Moreover, only 50 apps were analyzed in the current study. With over 2.65 million apps on Google play store (Statista.com, 2022a), the scope of sampling could be expanded in future research to enhance the representativeness of the app data.

To offer more insights for mobile app developers, future researchers could also look into the construct of developer responsiveness. This study focuses mainly on the time taken to respond to user reviews. Subsequent researchers could adopt an experimental design to ascertain the optimal responding and feedback cycle for improvement of the app performance. They may also compare and identify the optimal number of reviews required to lead to development of novel features.

The measurement of app performance is also not without limitations. App performance could be affected by factors other than software updates, such as promotional deals or other marketing efforts. Future researchers may attempt to isolate the effects of these contaminating factors in explaining app performance. In addition, app performance was only operationalized as app revenues and user downloads in this research. In the future, this construct could be operationalized in other ways, such as customer satisfaction and customer ratings, to provide a qualitative perspective of app performance. Other moderators could also be examined. In addition to developer responsiveness, the comprehensiveness of the customer review content or the expertise level of customers may also affect the usefulness and relevance of their reviews. Finally, future researchers should extend the study to a greater software development context outside mobile apps. For example, DevOps and Agile Software Development are prevalent methodologies that require rapid testing by customers and feedback is also constant offered by customers. Expanding the current study using these software development methodologies should further enhance data representativeness.

COMPETING INTERESTS

All authors of this article declare there are no competing interest.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. Funding for this research was covered by the author(s) of the article.

PROCESS DATES

Received: February 23, 2022, Revision: January 31, 2024, Accepted: March 19, 2024

CORRESPONDING AUTHOR

Correspondence should be addressed to Vanessa Liu; vanessaliusw@suss.edu.sg

REFERENCES

Al-Subaihin, A. A., Sarro, F., Black, S., Capra, L., & Harman, M. (2021). App Store Effects on Software Engineering Practices. *IEEE Transactions on Software Engineering*, *47*(2), 300–319. doi:10.1109/TSE.2019.2891715

Balasubramanian, S., Peterson, R. A., & Jarvenpaa, S. L. (2002). Exploring the implications of m-commerce for markets and marketing. *Journal of the Academy of Marketing Science*, *30*(4), 348–361. doi:10.1177/009207002236910

Barki, H., & Hartwick, J. (1989). Rethinking the Concept of User Involvement. *Management Information Systems Quarterly*, *13*(1), 53. doi:10.2307/248700

Barlow, J., Koncevich, S., Moller, C., & Olimp-business. (2016). *A Complaint Is a Gift [Russian Edition]: Using Customer Feedback as a Strategic Tool*. Olimp-business.

Berger, C., Möslein, K., Piller, F., & Reichwald, R. (2005). Co-designing modes of cooperation at the customer interface: Learning from exploratory research. *European Management Review*, *2*(1), 70–87. doi:10.1057/palgrave.emr.1500030

Burgess, C., Hill, M., Ferrill, T., Constantin, L., Hughes, C., Venkat, A., ... Pratt, M. K. (n.d.). CSO. *CSO Online*. Retrieved November 2, 2022, from https://www.csoonline.com/article/2132061/security-awareness/References+48+why-users-don-t-often-upgrade-software-when-they-should.html

Chen, N., Lin, J., Hoi, S., Xiao, X., & Zhang, B. (2014). AR-miner: Mining Informative Reviews for Developers from Mobile App Marketplace. In *Proceedings of the 36th International Conference on Software Engineering* (pp. 767–778). doi:10.1145/2568225.2568263

Cheung, K. (2013). A Feature Request Is A Bug. . .Is a Dumb Detail. *Getdonedone*. https://www.getdonedone.com/a-feature-request-is-a-bug-is-a-task-is-a-to -do-is-a-dumb-detail/

Comino, S., Manenti, F. M., & Mariuzzo, F. (2015). Updates Management in Mobile Applications. iTunes vs Google Play. SSRN *Electronic Journal*. 10.2139/ssrn.2664463

Dąbrowski, J., Letier, E., Perini, A., & Susi, A. (2022). Correction to: Analysing app reviews for software engineering: A systematic literature review. *Empirical Software Engineering*, *27*(2), 58. Advance online publication. doi:10.1007/s10664-022-10135-4

Finch, B. J. (1999). Internet discussions as a source for consumer product customer involvement and quality information: An exploratory study. *Journal of Operations Management*, *17*(5), 535–556. doi:10.1016/S0272-6963(99)00005-4

Gao, C., Zeng, J., Lyu, M. R., & King, I. (2018). Online app review analysis for identifying emerging issues. In *Proceedings of the 40th International Conference on Software Engineering*, ICSE '18 (pp. 48–58). ACM. doi:10.1145/3180155.3180218

Gao, C., Zheng, W., Deng, Y., Lo, D., Zeng, J., Lyu, M. R., & King, I. (2019). Emerging app issue identification from user feedback: Experience on WeChat. In *Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Practice*, ICSE-SEIP '19 (pp. 279–288). IEEE Press. doi:10.1109/ICSE-SEIP.2019.00040

Ghose, A., & Han, S. P. (2014). Estimating Demand for Mobile Applications in the New Economy. *Management Science*, *60*(6), 1470–1488. doi:10.1287/mnsc.2014.1945

Gustafsson, A., Kristensson, P., & Witell, L. (2012). Customer co-creation in service innovation: A matter of communication? *Journal of Service Management*, *23*(3), 311–327. doi:10.1108/09564231211248426

Guzman, E., Ibrahim, M., & Glinz, M. (2017). A little bird told me: Mining tweets for requirements and software evolution. In *2017 IEEE 25th International Requirements Engineering Conference* (pp. 11–20).

Hort, M., Kechagia, M., Sarro, F., & Harman, M. (2021). A survey of performance optimization for mobile applications. *IEEE Transactions on Software Engineering*, *48*(8), 2879–2904. doi:10.1109/TSE.2021.3071193

Huang, E. Y., Lin, S. W., & Fan, Y. C. (2015). MS-QUAL: Mobile service quality measurement. *Electronic Commerce Research and Applications*, *14*(2), 126–142. doi:10.1016/j.elerap.2015.01.003

Iacob, C., Faily, S., & Harrison, R. (2016). Maram: Tool support for mobile app review management. In *Proceedings of the 8th EAI International Conference on Mobile Computing*, Applications and Services. doi:10.4108/eai.30-11-2016.2266941

Iacob, C., & Harrison, R. (2013). Retrieving and analyzing mobile apps feature requests from online reviews. In *2013 10th IEEE Working Conference on Mining Software Repositories*, MSR (pp. 41–44). doi:10.1109/MSR.2013.6624001

Ives, B., & Olson, M. H. (1984). User Involvement and MIS Success: A Review of Research. *Management Science*, *30*(5), 586–603. doi:10.1287/mnsc.30.5.586

Johann, T., Stanik, C., Bama, B. A., & Maalej, W. (2017). Safe: A simple approach for feature extraction from app descriptions and app reviews. In *2017 IEEE 25th International Requirements Engineering Conference (RE)* (pp. 21–30). doi:10.1109/RE.2017.71

Kamel, M. (1995). Identifying, classifying, and resolving semantic conflicts in distributed heterogeneous databases: A case study. *Journal of Database Management*, *6*(1), 20–32. doi:10.4018/jdm.1995010103

Kaulio, M. A. (1998). Customer, consumer and user involvement in product development: A framework and a review of selected methods. *Total Quality Management*, *9*(1), 141–149. doi:10.1080/0954412989333

Khalid, M., Asif, M., & Shehzaib, U. (2015). Towards Improving the Quality of Mobile App Reviews. *International Journal of Information Technology and Computer Science*, *7*(10), 35–41. doi:10.5815/ijitcs.2015.10.05

Khan, G., Kagwanja, N., Whyle, E., Gilson, L., Molyneux, S., Schaay, N., Tsofa, B., Barasa, E., & Olivier, J. (2021). Health system responsiveness: A systematic evidence mapping review of the global literature. *International Journal for Equity in Health*, *20*(112), 112. Advance online publication. doi:10.1186/s12939-021-01447-w PMID:33933078

Kim, D. Y., & Kim, S. Y. (2023). Investigating the effect of customer-generated content on performance in online platform-based experience goods market. *Journal of Retailing and Consumer Services*, *74*, 103409. Advance online publication. doi:10.1016/j.jretconser.2023.103409

Kujala, S. I. (2008). Effective user involvement in product development by improving the analysis of user needs. *Behaviour & Information Technology*, *27*(6), 457–473. doi:10.1080/01449290601111051

Kuo, T., Tsai, G. Y., Lu, I. Y., & Chang, J. S. (2016, December). Relationships among service quality, customer satisfaction and customer loyalty: A case study on mobile shopping APPs. In *Proceedings of the 17th Asia Pacific Industrial Engineering and Management System Conference* (pp. 7-10).

Lederer, A. L. (1993). INDUSTRY AND PRACTICE. *Journal of Database Management*, *4*(4), 40–44. doi:10.4018/jdm.1993100104

Lee, G., & Raghu, T. S. (2014). Determinants of Mobile Apps' Success: Evidence from the App Store Market. *Journal of Management Information Systems*, *31*(2), 133–170. doi:10.2753/MIS0742-1222310206

Liang, T. P., Li, X., Yang, C. T., & Wang, M. (2015). What in Consumer Reviews Affects the Sales of Mobile Apps: A Multifacet Sentiment Analysis Approach. *International Journal of Electronic Commerce*, *20*(2), 236–260. doi:10.1080/10864415.2016.1087823

Litvin, S., Goldsmith, R., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, *29*(3), 458–468. doi:10.1016/j.tourman.2007.05.011

Lodenstein, E., Dieleman, M., Gerretsen, B., & Broerse, J. E. W. (2016). Health provider responsiveness to social accountability initiatives in low- and middle-income countries: A realist review. *Health Policy and Planning*, *32*(1), 125–140. doi:10.1093/heapol/czw089 PMID:27375128

Maalej, W., & Hadeer, N. (2015). Bug Report, Feature Request, or Simply Praise? On Automatically Classifying App Reviews. In *2015 IEEE 23rd International Requirements Engineering Conference*.

Maalej, W., Kurtanović, Z., Nabil, H., & Stanik, C. (2016). On the automatic classification of app reviews. *Requirements Engineering*, *21*(3), 311–331. doi:10.1007/s00766-016-0251-9

Martin, W., Harman, M., Jia, Y., Sarro, F., & Zhang, Y. (2015). The app sampling problem for app store mining. In *Proceedings of the 12th Working Conference on Mining Software Repositories (MSR '15)* (pp. 123–133). IEEE Press. doi:10.1109/MSR.2015.19

Mirzoev, T., & Kane, S. (2017). What is health systems responsiveness? Review of existing knowledge and proposed conceptual framework. *BMJ Global Health*, *2*(4), e000486. doi:10.1136/bmjgh-2017-000486 PMID:29225953

Nambisan, S. (2002). Designing Virtual Customer Environments for New Product Development: Toward a Theory. *Academy of Management Review*, *27*(3), 392–413. doi:10.2307/4134386

Pagano, D., & Maalej, W. (2013). User feedback in the Appstore: An empirical study. In *2013 21st IEEE International Requirements Engineering Conference* (pp. 125-134). doi:10.1109/RE.2013.6636712

Panichella, S., Di Sorbo, A., Guzman, E., Visaggio, C. A., Canfora, G., & Gall, H. C. (2015). How Can I Improve My App? Classifying User Reviews for Software Maintenance and Evolution. In *2015 IEEE International Conference on Software Maintenance and Evolution ICSME* (pp. 281–290). doi:10.1109/ICSM.2015.7332474

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, *64*(1), 12–41.

Parasuraman, A., Zeithaml, V. A., & Malhotra, A. (2005). ES-QUAL: A multiple-item scale for assessing electronic service quality. *Journal of Service Research*, *7*(3), 213–233. doi:10.1177/1094670504271156

Parthasarathy, S., & Daneva, M. (2021). A Requirements Engineering Framework for Software Startup Companies. *Journal of Database Management*, *32*(3), 69–94. doi:10.4018/JDM.2021070104

Prahalad, C. K., & Ramaswamy, V. (2004). *The Future of Competition: Co-Creating Unique Value With Customers*. Harvard Business Review Press.

Sarstedt, M., & Mooi, E. (2019). A Concise Guide to Market Research: The Process, Data, and Methods Using IBM SPSS Statistics (Springer Texts in Business and Economics) (Softcover reprint of the original 3rd ed. 2019). Springer.

Shams, R. A., Hussain, W., Oliver, G., Nurwidyantoro, A., Perera, H., & Whittle, J. (2020). Society-oriented applications development: Investigating users' values from Bangladeshi agriculture mobile applications. In *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: Software Engineering in Society*, ICSE-SEIS '20 (pp. 53–62). Association for Computing Machinery. https://doi.org/doi:10.1145/3377815.3381382

Statista. (2022, August 11). Number of apps available in leading app stores Q2 2022. https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/

Statista. (2022a, July 27). Google Play: Number of available apps 2009-2022. https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/

Tan, F. B., & Chou, J. P. (2008). The relationship between mobile service quality, perceived technology compatibility, and users' perceived playfulness in the context of mobile information and entertainment services. *International Journal of Human-Computer Interaction*, *24*(7), 649–671. doi:10.1080/10447310802335581

Thomke, S., & von Hippel, E. (2002). Customers as innovators: A new way to create value. *Harvard Business Review*, *80*(4), 74–81. https://evhippel.files.wordpress.com/2013/08//hbrtoolkitsaspub.pdf

Vaniea, K., & Rashidi, Y. (2016). Tales of software updates: The process of updating software. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 3215–3226). doi:10.1145/2858036.2858303

Villarroel, L., Bavota, G., Russo, B., Oliveto, R., & Di Penta, M. (2016). Release planning of mobile apps based on user reviews. In *2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE)* (pp. 14–24). doi:10.1145/2884781.2884818

Vu, P. M., Nguyen, T. T., Pham, H. V., & Nguyen, T. T. (2015). Mining user opinions in mobile app reviews: A keyword-based approach. In *2015 30th IEEE/ACM International Conference on Automated Software Engineering*, (pp. 749–759).

Wiggins, N. (2015). The difference between a bug, error and feature. Available at: https://www.webigence.com/blog/the-difference-between-a-bug-error-and-feature

Xie, K. L., Zhang, Z., & Zhang, Z. (2014). The business value of online consumer reviews and management response to hotel performance. *International Journal of Hospitality Management*, *43*, 1–12. doi:10.1016/j.ijhm.2014.07.007

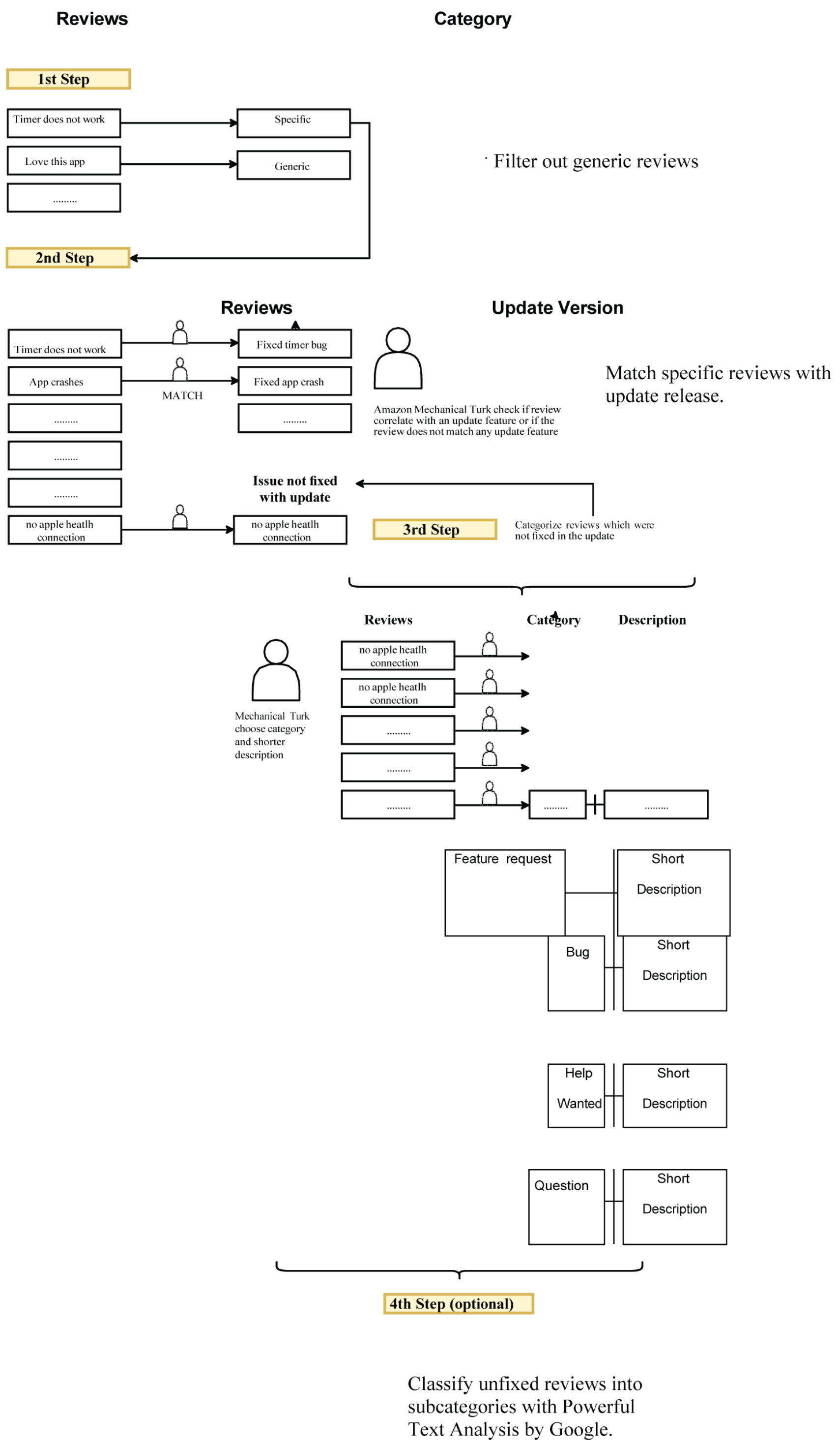
Zeithaml, V. A. (2002). Service excellence in electronic channels. *Managing Service Quality*, *12*(3), 135–139. doi:10.1108/09604520210429187

Zeithaml, V. A., Parasuraman, A., & Malhotra, A. (2000). *A conceptual framework for understanding e-service quality: implications for future research and managerial practice* (Vol. 115). Marketing Science Institute.

Zhang, J., Wang, Y., & Xie, T. (2019). Software feature refinement prioritization based on online user review mining. *Information and Software Technology*, *108*, 30–34. doi:10.1016/j.infsof.2018.12.002

APPENDIX: FRAMEWORK TO CATEGORIZE REVIEW

Figure 2. Reviews Category



Jiang Zhiying is the Head of Master of Digital Marketing Programme at Singapore University of Social Science. Before joining Singapore University of Social Science, she is an assistant professor at the Erasmus University of Rotterdam. Her research interests include consumer analytics, pricing and sharing economy.

Vanessa Liu is currently Associate Professor at the School of Business at the Singapore University of Social Sciences (SUSS). She earned her PhD in Information Systems (IS) from the City University of Hong Kong. Prior to joining SUSS, she acquired academic experience with the the Hong Kong Polytechnic University and the New Jersey Institute of Technology. She has published in renowned peer-reviewed journals in the information systems (IS) discipline such as the Journal of the Association for Information Systems (JAIS). Her research interests include online consumer behavior and sustainability management.

Miriam Erne earned her MS in Economics and Business from Erasmus University Rotterdam. Since then, she has specialized in online marketing, applying her extensive knowledge of economics and business to the digital domain. Her innovative approaches to online marketing have contributed significantly to her field, making her a sought-after expert for driving growth and understanding evolving digital trends.