

Valtteri Vuori

# **IMPROVING MOBILE APPLICATION FEEDBACK ACQUISITION IN APP STORES**

Faculty of Information Technology and Communication Sciences  
M. Sc. Thesis  
October 2020

# ABSTRACT

Valtteri Vuori: Improving mobile application feedback acquisition in app stores  
M.Sc. Thesis  
Tampere University  
Master's Degree Programme in Computer Science  
October 2020

---

With the ever increasing amount of mobile devices and their users worldwide, mobile application reviews have become an increasingly interesting topic within the industry. These reviews can be used to acquire never-before-seen insight to how users use and react to mobile applications, and how the reviews can be used within software development processes to further improve applications. The goal of this thesis is to study whether it is possible to increase the amount of reviews available for developers to use.

The thesis examines the psychology behind providing reviews, how reviews are currently collected, examines how reviews are already used within software development processes and provides a study to measure what potential changes could be made to the existing application marketplace ecosystems to further increase the amount of data available. By introducing three new scenarios for the participants, the study shows that users are more motivated to write reviews if they benefit financially from the review activity – particularly when receiving monetary rewards. The results of the study can be used to further improve existing application marketplaces.

Keywords: mobile application, review collection, psychology, software development

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

# Contents

<b>1. Introduction .....</b>	<b>1</b>
<b>2. User Review.....</b>	<b>3</b>
2.1. What is a User Review?	3
2.2. Mobile Application Review	4
2.2.1. Types of Mobile Application Reviews	6
2.3. Review Validity	8
<b>3. Psychology Behind Reviews.....</b>	<b>10</b>
3.1. Hennig-Thurau's Study	10
3.2. Wu's Study	13
3.3. Literature Summary	14
3.4. Collecting User Reviews and Feedback	15
3.4.1. Bug Tracking Systems	16
3.4.2. Online Forums	17
3.4.3. Integrated Feedback Channels	18
3.4.4. Third-Party Review Collection Sites	19
3.4.5. Application Marketplaces	20
3.4.6. Summary	21
<b>4. Review Usage in Software Development .....</b>	<b>23</b>
4.1. Crowdsourcing in Software Development	23
4.2. Crowdsourcing Reference Model	26
4.2.1. The crowd	26
4.2.2. The crowdsourcer	26
4.2.3. The crowdsourcing task(s)	27
4.2.4. The crowdsourcing platform	28
4.3. Crowdsourcing Mobile Application Reviews	28
4.3.1. Comparison to Crowdsourcing Reference Model	29
4.3.1.1 The crowd – Application Users	30
4.3.1.2 The crowdsourcer – The developers, users & app store owners	30
4.3.1.3 The crowdsourced task – App Reviews	31
4.3.1.4 The crowdsourcing platform – App Stores	31
4.3.2. CRISTAL	32

4.4. Summary	34
<b>5. Study and Methodology .....</b>	<b>35</b>
5.1. Research Design	35
5.1.1. Research Participants	36
5.2. Questionnaire	36
<b>6. Results and Discussion .....</b>	<b>39</b>
6.1. Current Motivations	39
6.1.1. Social benefits	39
6.1.2. Self-enhancement	40
6.1.3. Financial benefits	40
6.1.4. Altruism	41
6.1.5. Advice seeking	42
6.1.6. Emotional outlet	42
6.1.7. Summary	43
6.2. Reaction to Proposed Changes	44
6.2.1. Scenario 1 – Monetary Reward	44
6.2.2. Scenario 2 – Discount on In-App Purchases	45
6.2.3. Scenario 3 – App Store Points	46
6.2.4. Comparison of the Scenarios	46
6.3. Threats to Validity	47
<b>7. Conclusions and Future Work .....</b>	<b>48</b>
<b>References.....</b>	<b>49</b>

## 1. Introduction

With the introduction of the Internet, online shopping and software distribution has grown to be a part of our everyday lives. In 2015 two-thirds of consumers trust other customers' opinions posted online [Nielsen, 2015]. In addition, 91% of shoppers read online reviews before deciding on a purchase [P. Wu, 2019]. Reviews usually include praise, complaint or problems encountered with the product [Zha, 2019]. This information can be valuable to both customers and the manufacturers of the product. From a customers' point of view, the reviews can be seen as 'electronic word of mouth'. The difference between traditional word of mouth and electronic word of mouth is that next to the real-life friends, family and co-workers, a customer can see recommendations from a much wider selection of sources such as social media status updates and blogs. From a manufacturers' point of view, reviews can be seen as suggestions to improve functionality as well as a promotional channel [Rensink, 2013]. Online reviews can have an impact on product sales as an increase in the average rating, or a positive rating can increase the likelihood of a customer purchasing the specific item [Chevalier, 2018]. Online reviewing is primarily a type of prosocial behaviour that benefits other users or consumers and is costly to the reviewers themselves, since the ones who review rarely directly benefit from their effort. Research has shown that these prosocial behaviours are fuelled by motivations such as enjoyment and altruism (intrinsic), as well as reputation and career advancement (extrinsic) [P. Wu, 2019].

Alongside the increased usage of online platforms for various activities, mobile phones became an everyday accessory for many of us. These smart devices contain a plethora of features and applications that people use on a daily basis. A new industry was born when Apple released their App Store in 2008 as a distribution channel for applications developed by third-party developers [Vasa *et al.*, 2012]. During the same year Google released their Play Store that distributes applications for devices running on Google's Android platform [Shankland, 2008]. These marketplaces provided users finer control over the apps they have installed on their phones as well as a platform to express their thoughts on the apps they are using via a review function.

In software development user reviews can be used to elicit new requirements for the product. For example, Apple's App Store has a store page for each of the applications served on the marketplace. Once downloaded and used by the user, they can then give their review on the app. This review includes a star (numeric, 1 – 5) rating and an optional written message. From these active, heterogeneous and real-life user reviews the development team of the application can source new requirements to improve their application. This approach is described as *crowdsourcing* [Khalid *et al.*, 2015].

In order to crowdsource more and better requirements from an application marketplace for a software, the amount of reviews must be increased. For users to give reviews they need to have an incentive to do so, whether it is about sharing an experience, or to give a better buying experience for others. This thesis aims to find out what currently drives users to write reviews for mobile applications, and whether providing a reward in exchange for one's review would cause users to feel more inclined to write reviews, thus increasing the raw amount of reviews available for developers to use. The research questions presented in the study are:

**RQ1:** *How do users currently feel like giving reviews for mobile applications?*

**RQ2:** *How would users react to financial benefits in exchange for their reviews?*

To answer these questions a study is conducted within the thesis. This study is performed as an online questionnaire and seeks to answer the research questions outlined above. In order to find out the users' current motivations, one must first determine what motivations exist when writing reviews. The study will then measure the impact of each motivation to answer the first research question. The thesis will also find that financial rewards are a rare occurrence when reviewing or providing similar information. In order to measure how such an event would be received by the users, the participants in the study will be presented with a variety of such scenarios.

This thesis contains six chapters. Chapter 2 introduces user review as a concept. Chapter 3 examines the motivations behind writing a review and goes over examples of review acquisition methods. Chapter 4 discusses how reviews are used within software development. The study is presented in more detail in Chapter 5, and the results are discussed within Chapter 6. Chapter 7 concludes the thesis.

## 2. User Review

This chapter goes over the definition of a user review, how reviews are collected, can reviews be trusted and what the general psychology is behind giving a review. Mobile application reviews are a form of user review and come in various categories. These categories are also explained in this section.

### 2.1. What is a User Review?

According to the definition of Cambridge Dictionary [2020] a user review is:

“A report about a product written by a customer on a commercial website to help people decide if they want to buy it.”

The review is given to a public service or a released product such as a movie, electronic device or a subscription to a service. The review is often presented publicly at either the marketplace or at an online forum and is oftentimes given voluntarily. According to Trustpilot [2020] there are two categories in which reviews fall into; organic and invitational reviews. Organic reviews are of the user’s own initiative and not requested in any form. Invitational reviews are a result of the manufacturer/developer directly asking customers for their opinions on their products. Typically, a review contains a rating on a numeric scale and a review message [Pagano & Maalej, 2013]. The review message can contain either a generic message or it can comment on a specific part of the product in review, such as a problem with the product. Some systems can also employ mechanisms that ensure that the quality of the reviews is consistent. These systems can be for example votes (helpful vs. unhelpful) on a per-review basis [Vasa, 2012].

According to Pagano and Bruegge [2013] user involvement in software has changed significantly in the last three decades. The paradigm has shifted from programmers and technical staff to virtually anyone. For example, when shipping a piece of software, the *general* target audience can be identified during development (i.e. a banking software will probably not be used by children), but the real user base can differ. The consequence is that post-deployment feedback such as bug reports and feature requests have become even more important than before.

Typically, reviews are given to a specific revision of a product or a specific version of a software. The relational model of a review is then usually as follows:

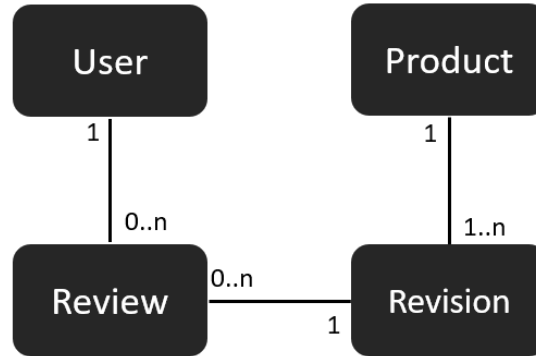


Figure 1: Relational model of a review  
[Vasa, 2012]

As seen in Figure 1, a single user can have multiple reviews. These reviews can be for the same product, or another product completely. A product consists of versions or *revisions*. Each revision tries to improve over the previous iteration and these revisions can attain different reviews than the previous iterations based on how well or bad the changes implemented were received by the users.

User reviews are a form of word-of-mouth (WOM) for the users taking part in the reviewing process as either the reviewer or the reader. WOM is recognized as an effective form of information transmission, particularly when expressing the users' experience on said product. Reviews communicate a "*user perceived quality*" that can help others in their purchase decisions.

## 2.2. Mobile Application Review

A mobile application is a piece of software that was specifically designed to be used on mobile devices, such as smartphones, tablets or smart watches [Hoehle & Venkatesh, 2015]. These applications can either be pre-loaded on the devices by the manufacturer or they can be downloaded from mobile application marketplaces such as Apple's App Store or Google's Play Store for free or at a cost. For their ever-increasing popularity, ease of sale and large communities of registered users, the platforms are very attractive for software development organizations. [Pagano & Maalej, 2013]

As mentioned earlier, Apple's App Store is a marketplace for various applications for iOS, iPadOS & watchOS platforms. While App Store only provides applications for Ap-

ple’s devices, Google Play Store is Google’s own marketplace which provides applications for Android-based devices. Combined, these two marketplaces have over 4 million applications available [Statista, 2020]. According to App Annie’s State of Mobile [2019] there were nearly 200 billion worldwide app downloads in these two marketplaces during the year 2019. These platforms (oftentimes called *app stores*) provide a user feedback feature. It is particularly interesting from a requirements engineering and software engineering perspective, since users who have downloaded an application (either free or a paid one) are able to give the app in question a rating (1 to 5 stars) and a *review message*. Applications that are highly rated are also featured in each app stores’ own top listings section, which in turn increases their visibility on the storefront [Maalen, 2013]. Agile development methodologies, which encourage fast reaction times to changes [Whitaker, 2009] can use these reviews to quickly elicit new requirements, and with the flexibility of the Agile approach the developers can also begin work on them quickly. [Khalid *et al.*, 2015]

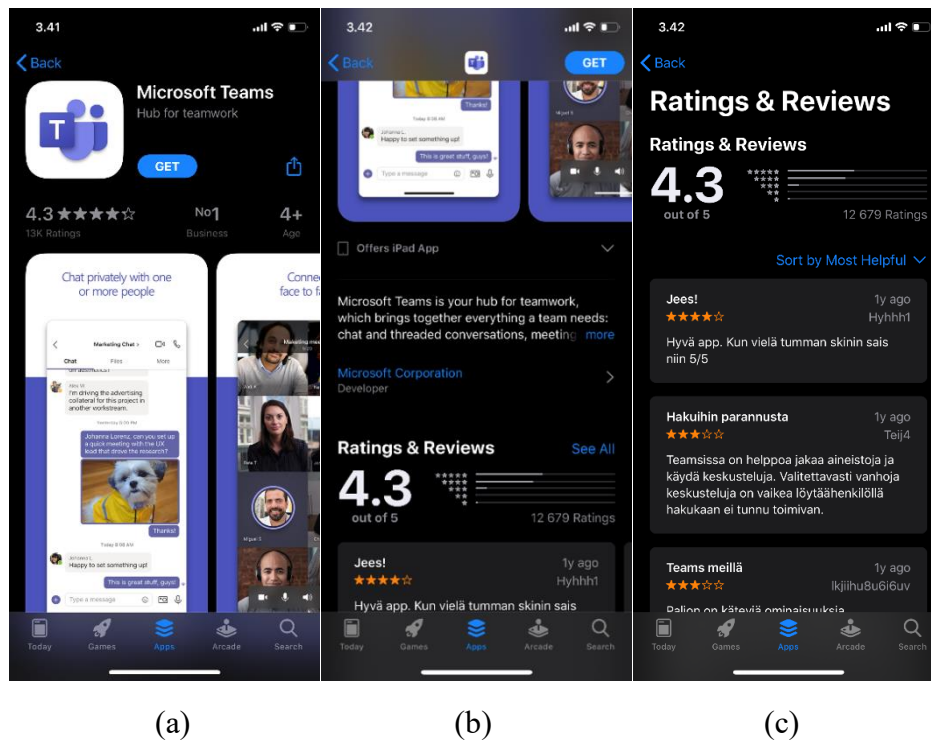


Figure 2: Microsoft’s Teams app page on App Store

Figure 2a depicts the application page of Microsoft’s Teams application on Apple’s App Store. Mobile application reviews contain the same key elements as any other review: a rating and an optional message field. When scrolling down the store page (Figure 2b) users can see the section “*Ratings & Reviews*”. Under this section is the applications’ current numeric score, the number of reviews the app has received and the general spread of reviews. The section also contains a horizontally scrolling list of reviews. When

tapped, the “*See All*”-button navigates the user to a page that contains all of the reviews the app has received (Figure 2c).

User reviews are a form of co-value creation, and so are mobile application reviews. The users can add value in multiple ways [Vasa, 2012]:

- Users provide feedback to the developers as well as the community.
- If an application has positive reviews, it will encourage others to try the application. More users mean more potential reviews for the application, which benefits *both* the developer and other customers.
- Negative reviews inform the developer of potential problems. In addition, negative reviews can help other customers steer clear of said application when the potential defects are documented via reviews.

The value given to an application depends on the perspective. A negative review does not seem to add value from a developers’ point-of-view, but from a potential future customers’ point of view, the review can hold much value so that the user finds something more fitting. An unfortunate paradox is that resulting decrease in download numbers will steer customers away from the product [Vasa, 2012], therefore causing less reviews and less possibilities for the developer to fix any problems experienced by the prior reviewers. This highlights how important listening to change and reacting to it is. Davis [1989] also points out that user reviews can communicate the perceived usefulness and ease of use of the product under review.

In an ideal world the reviews contain well formatted, easily understandable and more importantly instantly recognizable feedback. Platzer [2011] states that the review can contain abbreviations, colloquial expressions and non-standardized spelling. The reviews’ length will also vary. Vasa [2012] noted that when a review criticizes an application it has a higher character count than a review that expresses content. Additionally, in their paper Pagano and Bruegge [2013] came to the conclusion that different users provide feedback in different ways. A customer is more likely to provide information ad-hoc while a more professional user would be more elaborate in their feedback.

### **2.2.1. Types of Mobile Application Reviews**

Due to the iterative processes used in Agile methodologies [Whitaker, 2009] and the usage of reviews within software development and requirements engineering processes, it has become increasingly important to recognize the category of feedback received from the users. According to Guzman *et al.*, [2015] there are seven distinct categories for mobile application reviews. These categories are presented in Table 1.

<i>Category</i>	<i>Description</i>
<b><i>Bug reports</i></b>	Reviews which contain information about a problem the user encountered while using the application. A problem could be a faulty feature of an app, or crashes on certain devices.
<b><i>Feature strength</i></b>	Reviews which point out a specific feature of the application in which the users are satisfied with.
<b><i>Feature shortcoming</i></b>	Reviews which point out a specific feature of the application in which the users are not satisfied with.
<b><i>User request</i></b>	Reviews which request for new functionality, feature or other content to be added. The review can also request improvements to an already existing feature.
<b><i>Praise</i></b>	Reviews where the users express their general satisfaction. In contrast to feature strength, a praise is focused on the application as a whole and not to a specific part or feature of the application.
<b><i>Complaint</i></b>	Reviews where the user expresses their general dissatisfaction to the application. As with praise, this information is focused to the application and does not target a specific part of the application.
<b><i>Usage scenario</i></b>	Reviews where users describe use cases, scenarios and workarounds involving the application.

Table 1: Mobile Application Review Categories [Guzman *et al.*, 2015]

In their study Guzman *et al.* [2015] also studied the distribution of reviews for each category within Google’s Play Store and Apple’s App Store. The study was formed around a set of six mobile applications, three for each platform. This set formed the truth set used in the study. The classification was done manually by human annotators whom systematically assessed user reviews according to an annotation guide provided. The distribution of the review categories seen in the truth set is as follows: The majority of reviews consists of *praise* (27.7 %) and *feature shortcoming* (20.8 %). Following these two comes *bug reports* (16.1 %), *feature strength* (10.5 %) and *usage scenarios* (9.6 %). Finally, at the bottom of the spectrum comes *user requests* (6.6 %) and *complaints* (4.5 %). In addition, 4.3 % of the reviews were classified as *noise* by the annotators. Reviews

marked as noise were either written in a foreign language, contained only non-character symbols or did not make sense to the annotators.

Depending on the category of the review, it can be then used to extract new information for the software project. For example, reviews labelled as “praise” can show the development team that whatever they are doing pleases the audience but does not provide much information regarding where to take the application next. User reviews which provide feedback as to what to fix or improve can be labelled as more valuable than those which do not.

### 2.3. Review Validity

Since user reviews can have a significant impact on consumers’ purchase decision, some manufacturers and sellers want to exploit this and falsely generate reviews that help their product [P. Wu, 2019; Y. Wu *et al.*, 2019]. This behaviour can express itself via explicitly asking users for reviews in exchange for material or immaterial compensation [Martens & Maalej, 2019]. These reviews that are non-spontaneous, rewarded and requested are referred as *fake reviews* in this chapter. Fake reviews are a threat to the integrity of app stores. If real users stop trusting the reviews, they will also refrain from reading and writing new reviews themselves. [Martens & Maalej, 2019]. According to Yuanyuan Wu and colleagues’ [2019] study, in 2012 approximately 10% of online products were subjected to review manipulation.

App stores rely on user generated content (UGC) in the form of reviews to communicate quality to customers and having a communication channel between the users and the developers. Having this functionality tainted by a shadow of doubt will have troublesome consequences to not only customers, but developers as well. Numerous researches have studied app reviews but according to Martens & Maalej [2019], none of the works they reviewed had taken fake reviews into account.

In addition to gaining fake positive reviews on their applications, some entities create fake negative reviews on their competitors’ apps. These reviews can include information such as false issues, which in turn can lead to waste of development resources in order to fix the aforementioned false issue. [Martens & Maalej, 2019; Y. Wu *et al.*, 2019]

Both Apple and Google have explicitly forbid developers from amassing fake reviews to boost their app’s popularity. For instance, Apple [2020b] states in their app store review guidelines that:

“If we find that you have attempted to manipulate reviews, inflate your chart rankings with paid, incentivized, filtered, or fake feedback, or engage with third party services to do so on your behalf, we will take steps to preserve the integrity of the App Store, which may include expelling you from the Developer Program.”

Google has issued a similar statement that highlights the problems within false reviews, and explicitly asked developers to not purchase or use services that grant them fake reviews. Google also asked users not to participate in this kind of behaviour. [Google, 2018]

Martens & Maalej [2019] initially assumed that fake reviews would be short but found out that on average a fake review consists of roughly 120 characters. In comparison, regular reviews usually consist of 110 characters. In addition, fake reviews tend to be more complex in structure when compared to their real counterparts [Y. Wu *et al.*, 2019].

Another problem that developers and researchers can stumble upon is that according to Chen *et al.* [2014] only 35% of reviews posted for applications contain *useful* information. In addition, the sheer quantity of some applications is an extremely daunting aspect. For example, Facebook application on Google’s Play Store gets over 2000 reviews *daily*.

### 3. Psychology Behind Reviews

Ryan & Deci [2000] define being motivated as *to be moved to do something*. When a person feels inspired or energized to achieve a goal he is characterized as motivated, whereas someone who lacks impetus or inspiration is considered unmotivated. Motivation can be split into two distinct categories; intrinsic and extrinsic. The foremost refers to doing something due to it being enjoyable or interesting while the latter is to obtain a separable outcome [Rensink, 2013]. For a person to write a review they therefore must be motivated to do so.

#### 3.1. Hennig-Thurau's Study

Traditional word-of-mouth (WOM) played a major role in customer's purchasing decisions before the Internet introduced electronic word-of-mouth (eWOM) as an alternative [Hennig-Thurau *et al.*, 2004]. As mentioned in the introduction, eWOM amplifies traditional WOM by taking into account not only the people you talk to in person, but friends and even strangers' opinions online. Hennig-Thurau and colleagues [2004] studied the motivations behind WOM and eWOM by compiling results from three researches. The researches were performed by Dichter [1966], Engel *et al.* [1993] and Sundaram *et al.* [1998] respectively.

Dichter depicted four motives in his study (Table 2). The motives are of *positive* communication. While the work was deemed as plausible and prominence, one key weakness in his work was the lack of information regarding the development of his typology. [Hennig-Thurau *et al.*, 2014]

<i>Motivation</i>	<i>Description</i>
<b><i>Product-involvement</i></b>	Behind this motive was the idea that a customer can feel so strongly about a product that a pressure builds up inside, resulting in a recommendation to other customers to reduce the tension caused by the consumption experience.
<b><i>Self-involvement</i></b>	The product served as a means through which the person can gratify certain emotional needs.
<b><i>Other-involvement</i></b>	Word-of-mouth activity that addresses the need to give something to the receiver.
<b><i>Message-involvement</i></b>	Refers to discussion, which is stimulated by advertisements, commercials or public relations.

Table 2: Summary of Dichter’s study in 1966 [Hennig-Thurau *et al.*, 2004]

Engel *et al.* (Table 3) modified this typology and renamed the categories. It is important to notice that unlike in Dichter’s work, Engel *et al.* also have a negative motive present in dissonance reduction. Dissonance reduction is often referred as *buyer’s remorse* [Hennig-Thurau *et al.*, 2004; Kowol, 2008].

<i>Motivation</i>	<i>Description</i>
<b><i>Involvement</i></b>	Level of interest or involvement in the topic under consideration serves to stimulate discussion.
<b><i>Self-enhancement</i></b>	Recommendations allow for a person to gain attention, show connoisseurship, suggest status, give the impression of possessing inside information and asset superiority.
<b><i>Concern for others</i></b>	A genuine desire to help a friend or a relative to make a better purchase decision.
<b><i>Message intrigue</i></b>	Entertainment resulting from talking about certain ads or selling appeals.
<b><i>Dissonance reduction</i></b>	Reduces doubts following a major purchase decision.

Table 3: Summary of the study by Engel *et al.* in 1993 [Hennig-Thurau *et al.*, 2004]

The most comprehensive study presented on Hennig-Thurau’s paper was done by Sundaram *et al.* (Table 4). They identified eight motives for customer WOM communication with four depicting positive communication and four depicting negative communication.

<i>Motivation</i>	<i>Valence</i>	<i>Description</i>
<i>Altruism</i>	<i>Positive</i>	The act of doing something for others without anticipating any reward in return.
<i>Product involvement</i>	<i>Positive</i>	Personal interest in the product, excitement resulting from product ownership and product use.
<i>Self-enhancement</i>	<i>Positive</i>	Enhancing images among other consumers by projecting themselves as intelligent shoppers
<i>Helping the company</i>	<i>Positive</i>	Desire to help the company.
<i>Altruism</i>	<i>Negative</i>	To prevent others from experiencing the problems they had encountered.
<i>Anxiety Reduction</i>	<i>Negative</i>	Easing anger, anxiety and frustration.
<i>Vengeance</i>	<i>Negative</i>	To retaliate against the company associated with a negative consumption experience.
<i>Advice seeking</i>	<i>Negative</i>	Obtaining advice on how to resolve problems.

Table 4: Summary of the study by Sundaram *et al.* [Hennig-Thurau, 2004]

Hennig-Thurau’s study also suggested that customers desire equitable and fair exchanges. If, for example, a consumer feels like he or she received “more” from a purchase than they paid for, they might be inclined to help the manufacturer/seller by providing a (in this case) positive review. If the consumer feels as if they received *less* value from their purchase than what they paid for, they want to balance the scales by giving out a negative review. This theory is referred as *equity theory*.

While the studies described before were focused on WOM-motives, eWOM has some additional incentives for users to provide feedback. Hennig-Thurau *et al.* point out that through feedback they (users) can become a part of a virtual community. Affiliation with a virtual community can provide social benefit for a consumer via social integration and sense of identity. This virtual community can be for example an online forum where the consumers exchange their thoughts on the product. Virtual communities can also hold collective power over companies. Due to the reviews’ long-term availability, and the accessibility of them, negative reviews can influence a company’s image inadvertently. In a Kotaku article by Nathan Grayson [2015] he describes how some users group up to

*review-bomb* a game on Steam. Steam has user reviews front-and-centre. Even when casually looking through potential games, users can see a consensus of the game under its name. Review-bombing is the act of multiple users (hundreds, thousands even) giving a negative review to a game in very quick succession. This can cause the game's *user score* to drop tremendously, and potentially affect sales of said game. According to the article, the motivations behind review bombing are similar to a protest on the streets; when the people in charge refuse to listen the people (in this case the users) feel like they have no other option but to publicly express their dislike in hopes of the developer wanting to douse the starting fire. In order to combat this kind of behaviour, Steam updated their review system to show a game's score over time [Brightman, 2016].

### 3.2. Wu's Study

In Philip Wu's [2019] paper he identified four key motivation categories for users to write reviews on Amazon. The study consisted of highly active reviewers on Amazon, as well as less active reviewers. The categories are:

**Enjoyment** – In this category users expressed their enjoyment towards writing a review. For them, writing reviews is a form of self-expression and the reward comes from the possibility to express their own opinions on a large public platform. Especially the newer reviewers interviewed seemed to be most motivated by this category.

**Material reward** – An external reward such a free product sample or a discount coupon can play an important role in online reviewing. Some users' motivations lie in receiving these external incentives. On Amazon's forums heated discussion emerged regarding the validity of these pseudo-paid reviews. Some users commented that they have only seen 5-star reviews on products received for free, and another user commented that they have given plenty of 2 and 3-star reviews for products they've received for free.

**Direct reciprocity** – Reciprocity is the act of exchanging things with others for *mutual* benefit. General reciprocity is a common motivation for knowledge contribution, but when interviewed the interviewees did not feel obliged to review due to general reciprocity. Instead, the reviewers' sense of obligation emerged toward the person who sent them products for review. One interviewee shared their story of a self-published book they accepted to review. They couldn't bring themselves to give the book a bad rating even though it deserved one. The reviewee had socially interacted with the author, and the author was "so sweetly sincere". This effect is called *socially induced reciprocity*.

**Ranking/status recognition** – Amazon provides a ranking system for users who review products. Multiple categories exist such as "Top 100" or "Top 10". This ranking is

defined by Amazon [2020b] as “A reviewer's rank is determined by the overall helpfulness of all their reviews, factoring in the number of reviews they have written”. A quantified recognition such as a reviewer’s rank is an important motivator for reviewers. In order to keep one’s rank, they must continue to provide quality reviews. In the study an interviewee claims that “a negative vote is worth 4–7 positives.”.

It is important to note that especially the last motivation discussed in this study largely applies only to Amazon and its reviewing system. From these four motivations, all but enjoyment are extrinsic motivations.

### **3.3. Literature Summary**

To summarize, according to both Wu’s and Hennig-Thurau’s study, users have a wide variety of motivations to give reviews. These motivations revolve around helping others, enhancing the reviewer’s image of self, expressing dislike and following social expectations to provide feedback. Each reviewer has their own motivations and it is impossible to pinpoint exactly what causes users to write reviews. From these two literatures we can summarize that the following motivational areas are key in user review motivation and are used to measure user motivations within the study presented in the thesis:

#### **1. Social benefits**

Social benefits such as seeing one’s review amidst other reviews and indirectly communicating with other members of a virtual community can help create a sense of belonging for the reviewer. Being affiliated with such a community provides the user with social benefits. Seeing others gain benefit from one’s reviews can also attribute towards this motivation.

#### **2. Self-enhancement**

By projecting themselves as an intelligent shopper via reviews, some users can enhance their own image in their eyes. This self-related consumer need can only be gratified through social interaction and having an electronic communication form that transcends traditional WOM communication avenues provides an excellent source of motivation to provide one’s opinion online. In addition, having a quantifiable recognition for one’s status drives some users to continue writing reviews.

### **3. Financial benefits**

Obtaining financial benefits – such as free products or discounts – can have be a positive motivation for some users. Instead of wanting to write reviews in exchange for free products, in some cases users receiving a financial benefit can cause the user to feel obliged to provide a review.

### **4. Altruism**

Some users' motivations lie in the selfless concern for other users' well-being. Users can for example want to help others avoid a mistake they did in the past, or want to share a great experience they had to their peers.

### **5. Advice seeking**

For some, posting reviews is a source of information. The user might have encountered a problem and a review can be used as a discussion platform for said problem. In some platforms the product provider can reply to reviews, providing a two-way discussion between the reviewer and the product provider.

### **6. Emotional outlet**

Consumers who have had negative or positive consumption experiences might want to express their emotions towards the company behind the product. A mobile application can act as a *face* for a corporation and focusing one's frustration or joy to this entity can help discharge this intense emotional state.

## **3.4. Collecting User Reviews and Feedback**

In software development, many developers encounter problems with acquiring and analysing reviews for their products in an effective way. In addition to reviews, the users can provide valuable feedback via other sources and methods. Traditionally, software review and feedback collection has been focused on bug-reporting repositories such as Bugzilla, online forums and even e-mails. [Chen *et al.*, 2014]. As mentioned in Subsection 2.2.1, both Apple's App Store and Google's Play store provide a very straightforward and easy way for users to provide their reviews for mobile applications, as well as provide the developers an easy, centralized location which hosts a majority of their reviews. In addition to the *app stores*, some mobile applications have a self-contained feedback section within them similarly to what some websites provide. Different approaches also cater to different motivations discussed in previous chapters. Not all approaches gather reviews, but user feedback in general. Both reviews and feedback share the same core principle: this is what the user thinks. This section showcases some of the systems currently in place to gather user reviews and feedback, and what motivations they harness to gather the data.

### 3.4.1. Bug Tracking Systems

Bug tracking systems or *repositories* such as JetBrains' YouTrack [2020] play a critical role in development. Before long, the number of bugs in a software project will reach a level where it is no longer efficient to keep track of them manually in a spreadsheet. These bug trackers provide a way for multiple people to provide input on any bugs found on a system. In addition to the project's developers', the users can also report bugs on these systems. This user-interaction can be considered as a form of feedback, similarly to a review which reports a defect in the product in hand. Having a dedicated defect-tracking system will allow for the team to keep track of outstanding bugs, issues or problems that arise in the system. YouTrack was developed by the JetBrains team and is free to use in an open source project and paid in a team larger than ten people. [JetBrains, 2020]

Bug tracking systems provide opportunities for a multitude of motivations presented in Section 2.3. Users can gain *social benefits* and *self-enhancement* via seeing their reports being referenced and voted by other users. Figure 3 exhibits how in YouTrack votes are presented at the right-hand side of the screen. Reporting bugs and defects can be a form of *advice seeking*, and users can *express their altruism* by reporting any bugs they come across. Bugs and problems with the product are objective rather than subjective, so the user cannot use bug reporting as a source of emotional outlet to show vengeance against the service provider. Some websites and bug-trackers provide *financial incentives* for users who seek out critical vulnerabilities in their systems [Bugcrowd, 2020]. While these vulnerabilities usually require extensive amounts of IT-security knowledge, they can be still classified as feedback and the tasks are motivated by significant rewards based on the vulnerability's severity.

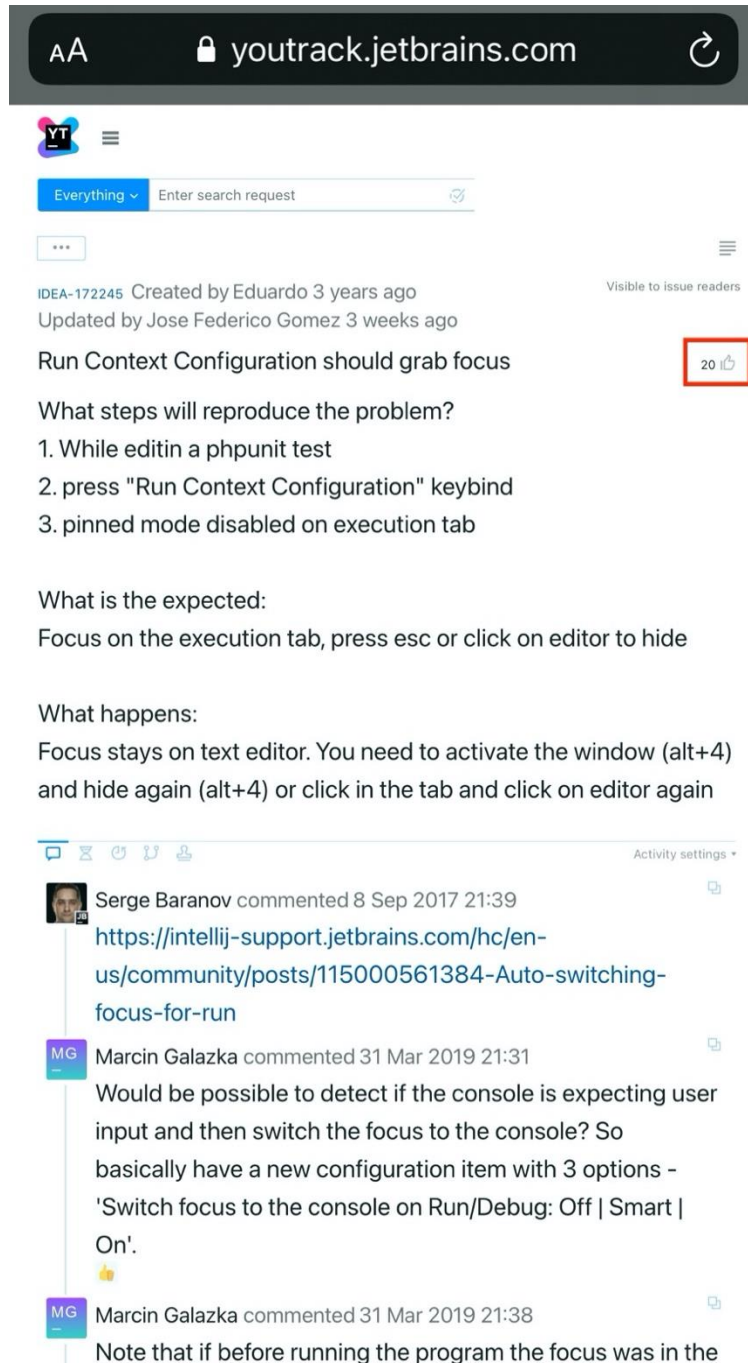


Figure 3: An issue in YouTrack

### 3.4.2. Online Forums

Many online marketplaces have a feedback forum attached to them. This forum located at the retailer's website provides a place where the consumers and the retailers can directly interact with each other [Lee and Lee, 2006]. For example, the digital video game marketplace GOG [2020] operates their own *wishlist* feature which they describe as:

”The Community Wishlist is the place to suggest, discuss and vote on all the things you would love to see on GOG.com.”

The users of the service can provide their wishes on which games they’d like to see in the marketplace and which features they’d like to be implemented. Other users can freely comment and vote on other users’ submissions, and all of the suggestions in each category are ranked on a weekly basis. As of September 16th 2020, there has been a total of 7.5 million votes cast, and 2.2 million of them have been fulfilled. The service is a great way to collect feedback from active users. This phenomenon is later examined in Chapter 4.

Users have multiple motivations to provide feedback within online forums. When examining GOG’s wishlist, the user can gain *social benefit* and *self-enhancement* from receiving votes from other users in their suggestion. Since the users cannot report bugs or express their unhappiness via this interface, no motivation for advice seeking or social outlet is in place. The suggestion-based approach does not motivate users acting out of altruism, since a request specifically asks for something to be added/changed which the person reporting wants. Online forums generally do not provide any financial incentives for users to provide feedback.

### **3.4.3. Integrated Feedback Channels**

Some websites provide the possibility of providing instant feedback during your visit. Figure 4 depicts one instance of this behaviour.

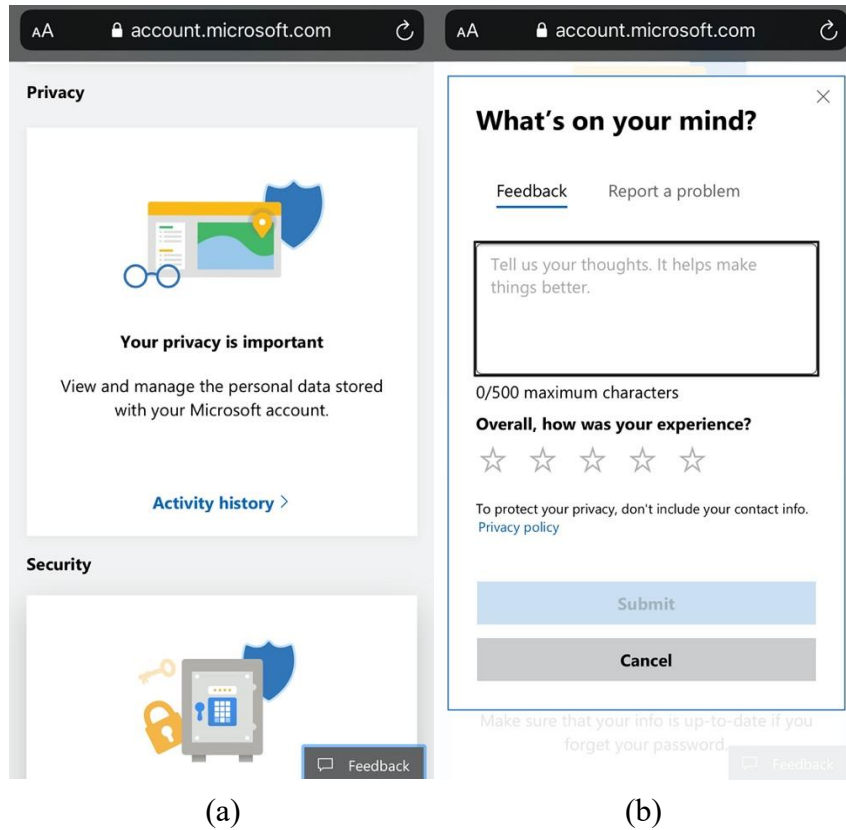


Figure 4: Microsoft's account page feedback option

In this figure while the user is anywhere on the Microsoft's account management page (Figure 4a), there is a button on the lower right-hand side of the page labelled as "Feedback". By interacting with this button, a modal window appears (Figure 4b) in which the user can write up to 500 characters worth of feedback, and provide a star rating in response to the question "Overall, how was your experience?". In addition to providing feedback, the user can also switch to another mode by interacting with the grey "Report a problem"-section. Within this section the user can choose from a predetermined set of problems one that suits their needs and provide any additional information they have regarding the problem. This information is then sent to Microsoft. [Microsoft, 2020]

This approach does not provide any *social benefit or self-enhancement* for the user since all of the feedback provided is strictly for the internal use of Microsoft and is therefore not shareable with other feedback providers. Microsoft does not reward the user for providing feedback. This leaves the user with any of the three motivations to provide feedback: emotional outlet, altruism and advice seeking.

#### 3.4.4. Third-Party Review Collection Sites

In addition to forums hosted on the software's or the product's own website, there are various review aggregation sites such as Trustpilot [2020]. Trustpilot operates by having

profiles for either companies or pieces of software. Users can then freely leave their review of this entity to a public *wall*. This wall is visible for anyone – even visitors and contains review posted by the users chronologically. According to Trustpilot they have two kinds of reviews hosted within the platform: organic and reviews by invitation. Organic reviews are such that the user has navigated to the site by themselves with their own incentive or motivation, while invitational reviews are ones in which the system explicitly requests users to provide feedback. This request is typically sent via e-mail and can be automated. [Trustpilot, 2020]

Users providing reviews can gain *social benefit* by providing their voice as a part of a community in a review. Other users can vote whether a certain review is useful, which in turn provides *self-enhancement* for the original reviewer. Since users can post traditional reviews on the sites, the user can *seek advice* via reviews. The profiles within Trustpilot are either for products or companies. The latter allows for the user to use the site as an *emotional outlet*. Writing a review on Trustpilot is a completely voluntary activity. If a user has had a bad experience with a product, they might want to warn others not to encounter the same fate they had experienced. This in turn provides motivation for *altruism*. No *financial incentives* are in place at Trustpilot.

### 3.4.5. Application Marketplaces

Similar to a review collection site, mobile application marketplaces such as Apple’s App Store and Google’s Play Store provide a rich review system for users and developers alike to use. Unlike Trustpilot, the app stores mentioned require for the user to download or purchase the application in order to review it. This eliminates *some* of the malicious activity mentioned in Section 2.3. In order for the user to review an application on Apple’s App Store, they must first navigate to the application page within their respective store (as shown in Figure 5). Then, users can interact with the “*Write a review*”-button shown in Figure 5a. This activates a new window (Figure 5b) in which the user can leave their review. Users provide a fitting title and a star rating (Numeric, 1 to 5 stars) for their review. Optionally, the user can also provide textual feedback to further explain or comment the review. After the review activity has finished by the user interacting with the “*Send*”-button located at the upper right-hand corner of Figure 5b, the window collapses and the user is navigated back to the application page which they left from. [Apple, 2020a]

When providing reviews to mobile application marketplaces, the users can gain *social benefit* by being a part of a review their voice as a part of a community in a review. Since the reviews can be replied to by the developers, writing a review provides a channel for advice seeking for the user when a problem occurs. Users can also provide reviews out of altruism, to warn others to steer clear of said product or to recommend the product

so that others can share their excitement. While there is no explicit ranking system present, the self-enhancement motivation is present by having one's reviews visible on a public forum which other users can vote to be either helpful or unhelpful. This can help the reviewers project themselves as advanced users in the eyes of other users if their review is seen as helpful by others. Currently no financial incentives are in place within app stores. As mentioned in Section 2.3, exchanging reviews for financial gains is strictly prohibited by Apple, and users are not ranked by their reviewing habits.

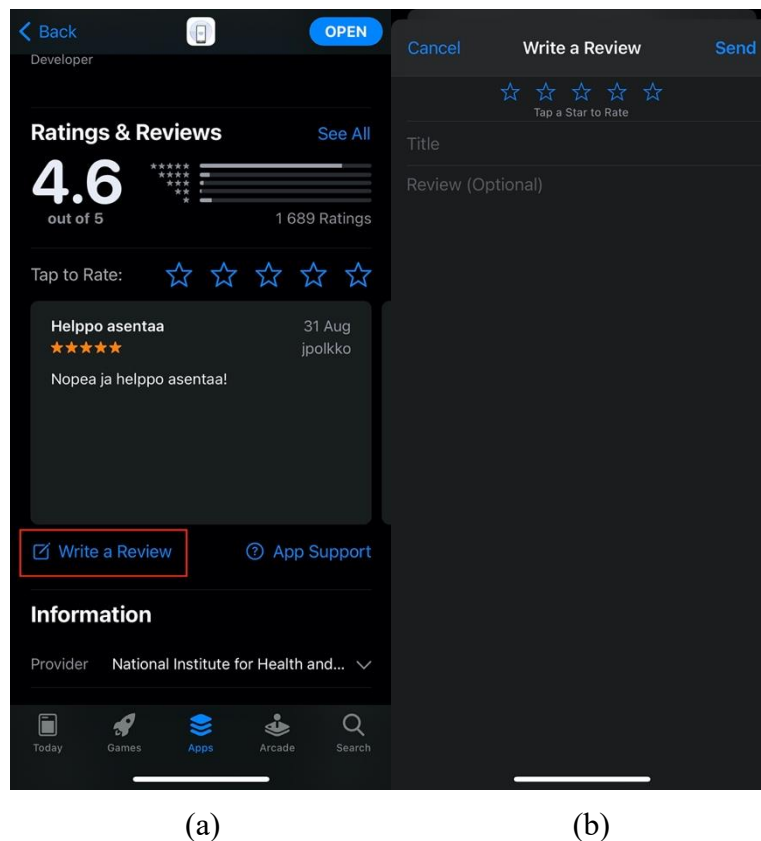


Figure 5: Application review process in Apple's App Store

### 3.4.6. Summary

There are multiple approaches in user feedback and review acquisition. User generated content such as reviews, bug reports and feature suggestions provide an exciting insight to the minds of the users. Different systems employ different mechanisms to extract this information from the users. Table 5 portrays the distribution of motivations available (found motivations represented by a green box) within the showcased systems. Not many systems provide financial benefits for users to provide feedback as can be seen in the table. The only one present in this thesis is of such high stakes that the organizations

behind these systems (bug bounties) want to ensure that there is a motivation large enough to assess the issues outlined. As mentioned before, normal bug reporting systems do not have financial benefits systems in place.

A study by van Dooren *et al.* [2018] claim that receiving rewards can increase a person's motivation in various tasks such as video games. In the study, it was noted that receiving rewards in exchange for playing a game can increase the effort put into the activity. As financial benefits are such a rare occurrence, the study presented in the thesis aims to measure what kind of an impact implementing such a functionality would have within the motivations of the users within mobile application marketplaces.

	Social benefits	Self-enhancement	Financial benefits	Altruism	Advice Seeking	Emotional outlet
Bug-tracking systems	Green	Green	Green	Green	Green	Grey
Online forums	Green	Green	Grey	Grey	Grey	Grey
Integrated feedback channels	Grey	Grey	Grey	Green	Green	Green
Third-party review collection sites	Green	Green	Grey	Green	Green	Green
Application marketplaces	Green	Green	Grey	Green	Green	Green

Table 5: Motivational matrix

## 4. Review Usage in Software Development

In software development reviews can be used to extract useful information and opinions from the users. This chapter examines *crowdsourcing* as a phenomenon, explains how it can be used to elicit new requirements for development teams to use and what pitfalls to avoid when using reviews as a requirements engineering source.

### 4.1. Crowdsourcing in Software Development

Crowdsourcing is a sourcing model in which organizations or individuals obtain useful information or goods – such as ideas – from a large, relatively open and quickly evolving group of users. This group of users typically exists in an internet-based service. Some examples of crowdsourced domains include *Wikipedia* and *Waze*. These platforms provide a way for users to provide useful data to them. When using Waze [2020] users can report possible road accidents or other hazards other users might encounter on their journey, which other users can then observe and plan their journey accordingly. The term was coined in 2006 and comes from the portmanteau of “Crowd” and “Outsourcing”. [Hosseini *et al.*, 2014; Estellés-Arolas *et al.*, 2012]

Requirements engineering is the process of defining, documenting and maintaining requirements. These requirements focus on satisfying the needs of the projects’ stakeholders. These stakeholders can be for example the customers or the client [Hosseini *et al.*, 2014]. Requirements elicitation is a sub-part of requirements engineering which revolves around discovering the requirements for the system in hand. This can be done in various ways such as interviews, brainstorming or scenarios [Sommerville, 2006].

Crowdsourcing can be used to solve various requirements engineering problems. These include but are not limited to:

1. **Stakeholders’ Discovery** – In a complex system it can be increasingly difficult to identify the set of stakeholders and what their role is in the system. Lim *et al.* [2010] propose that a small set of identified stakeholders are put into their own respective groups. These stakeholders can then suggest other stakeholders who might fit into the category they are in themselves.
2. **Requirements Identification** – The user base of a modern cloud-based app or mobile app can be very large and diverse. Instead of using an expensive set of elite users as the base for user stories, Hosseini *et al.* [2015] suggest that the developers could harness the power of the already diverse group of users to identify requirements.

**3. Feedback-Based Requirements Engineering** – In order to better understand the requirements for the next release of the software, user feedback can be acquired from various sources (i.e. explicitly via forums or implicitly via user usage pattern monitoring) [Hosseini *et al.*, 2015].

In their paper Hosseini *et al.* [2015] explain that crowdsourcing can be used to elicit new requirements from users. Eliciting the requirements from users has traditionally been a very manual process and done in the design phase. With crowdsourcing the requirements engineering process can continue alongside the release of the project for example via user feedback. Another benefit of crowdsourcing is that the users whom the data is sourced from are actual users of the product. The data gained can complement already established user stories. Systems which are highly interactive and used by a wide variety of users in different contexts are particularly suitable for requirements elicitation via crowdsourcing.

According to Prpić *et al.* [2015] there are four approaches to crowdsourcing. These four approaches are divided by two axes: *contributions* and *content* as shown in Figure 6.

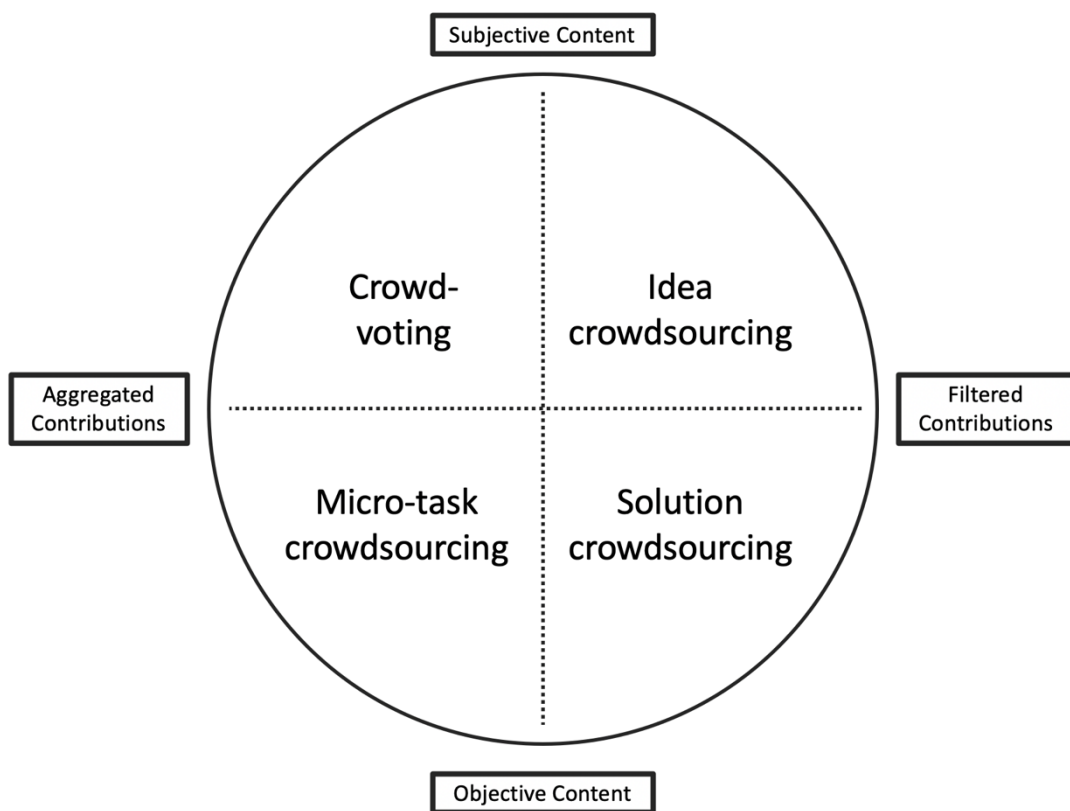


Figure 6: Crowdsourcing approaches [Prpic *et al.*, 2015]

The crowdsourcing style of *crowd-voting* focuses on aggregating answers for pre-made choices. The authors show the TV-show *American Idol* as an example where the organization presents  $N$  contestants for the crowd to choose from. From these contestants the crowd then votes for their favourite, and after the deadline for voting has ended the resulting votes would then be aggregated together to see which contestant won. [Prpić *et al.*, 2015]

As with the previous style, the *micro-task crowdsourcing* style focuses on data aggregation. These tasks are often unachievable through normal methods due to their sheer size and complexity. An example is Google's reCAPTCHA which is a system that allows web hosts to differentiate humans from software-based visitors (*bots*). The system has seen numerous revisions, but the original goal was to help in book digitization. In order to enter a site (or when the site suspects the visitor is non-human) the site prompts the user to type out the two words presented for them in a picture format. One of the pictures is known, and the other is unknown. With enough submissions for a certain word, the crowd had provided an answer to the previously uncertain word by aggregating the answers. [Ahn, 2011; Prpic *et al.*, 2015]

Another approach is to evaluate the answers on an atomic level via *idea crowdsourcing*. The crowdsourcers can for example seek creativity from the crowd since crowds are inherently large and contain a diverse set of individuals. An online artist community called *Threadless* asked the crowd to design T-shirts. From these crowdsourced designs the team internally chose their favourites which were then produced for sale. Contrary to the previous examples, this approach does not aggregate the results but evaluates them from a subjective point of view independently.

On a more objective approach we have *solution crowdsourcing*. In this approach the crowdsourcing organization or individual seeks an answer to a usually well-defined and idiosyncratic problem from the crowd. The crowd can then provide their answers which the organization or individual then objectively tries to measure whether the solution actually solves their problem. One of the most well-known video streaming platforms Netflix issued a competition in which they wanted to improve their predictive accuracy in their personalized suggestions on what to watch next. Based on past data, the company tested the solutions received, and the most effective solution won. [Prpic *et al.*, 2015]

Prpic *et al.* [2015] also note that these approaches can also be implemented simultaneously as well. Starwood Hotels & Resorts wanted to find a new marketing campaign. First, they implemented an idea-crowdsourcing activity from which they obtained multiple marketing ideas. From these ideas they had a crowd-voting phase where the employees voted on their favourite ideas.

## 4.2. Crowdsourcer Reference Model

Prior to the study done by Hosseini *et al.* [2014] there was not a general agreed upon terminology for crowdsourcing and its' activities. In their study they referenced papers from various domains in which crowdsourcing was used in and came up with the crowdsourcer reference model. They found that crowdsourcing consists of four key parts, or *pillars*. These pillars constitute the entire crowdsourcing operation. Each pillar also has its own features. As per Hosseini *et al.* [2014] the pillars and their respective features are the crowd, the crowdsourcer, the crowdsourcing task(s) and the crowdsourcing platform.

### 4.2.1. The crowd

The crowd is the group of people who take part in the crowdsourcing activity. Generally, this group has the following features: *Diversity*, being the quality or state of being different than another participant. Diversity can be categorized for example via age, gender or expertise. *Unknown-ness*, which is the condition of being anonymous. In crowdsourcing the unknown-ness of the crowd is measured in two ways; the crowd does not know the crowdsourcer, or the crowd does not know other members of the crowd. *Largeness* is the measure of the size of the group. In crowdsourcing, largeness is achieved when the crowd is large enough to complete the task successfully. *Undefined-ness* portrays the randomness of the crowd. A crowd is undefined when the crowd is chosen without any constraints. If a crowdsourcing activity targets only students living in the Tampere region, the crowd is not undefined. *Suitability* is the fit of the crowd for performing the specified activity. If a crowdsourcing task requires users to solve third-degree equations, a certain mathematical skill is probably required.

### 4.2.2. The crowdsourcer

The crowdsourcer is the entity who gains knowledge or wisdom from the crowdsourcing activity. This entity can be an institution, a company or an individual. Their primary goal is to seek the completion of a task through the use of the crowd. The crowdsourcing entity has the following features: In crowdsourcing, the crowdsourcer may provide the crowd with incentives as a stimulus or encouragement. Three main categories for incentives exist: financial (monetary rewards), social (peer recognition) and entertainment (enjoyment and/or fun). This is referred to as *incentives provision*. The crowdsourcer may provide an *open call* so that anyone can participate. This is synonymous to undefined-ness presented in the previous subsection, but from the crowdsourcers' perspective. *Ethicality provision* is if and how the crowdsourcer conforms to moral and ethical standards in the industry.

For example, the crowdsourcer must ensure that the crowd has an opt-out option so that they can stop participating in the crowdsourcing activity at any time. *Privacy provision* is a feature that describes how well the crowdsourcer handles personal or identifiable information. The information should not be given to other participants, organizations or other entities. In addition, all data gathered will not be shared without consent from the participating crowd.

#### **4.2.3. The crowdsourcing task(s)**

The crowdsourcing tasks are the activities the crowd takes part in. The activity is provided by the crowdsourcer and can take different forms. The task might be in the form of a data collection, fundraising, problem solving or new idea collection. Usually the task requires something from the crowd, whether it is their knowledge, skills or money. The crowdsourcing task has the following features: *Traditional operation*, which describes whether the task is doable in a traditional way. If the task were not crowdsourced, it would be instead done by the organizations' employees, or outsourced to someone who could complete the task. *Outsourcing task* describes whether the task would otherwise be outsourced. A crowdsourced task can be non-atomic; it can be broken down to smaller steps. These micro tasks can be then completed by the crowd resulting in completed task and achieving modularity. This is referred to as *Modularity*. *Complexity* measures how complex the task in hand is. A complex task is usually split into multiple smaller ones (in line with *modularity*). If this is not possible, the task is then considered complex. *Solvability* measures the task's capability to be solved. Usually the crowdsourced task is easy enough for us humans to understand, but too complex for computers. *Automation characteristics* define how difficult or expensive to automate the task is. If either of the constraints exist, the task is open for crowdsourcing. Otherwise the task would be solved by a computer instead of crowdsourcing it. A *user-driven* activity is something that is controlled or powered by users. If a task is crowdsourced, it is then user driven. These tasks are divided to three distinct sub-categories: Problem solving, innovation or generating ideas or designs. *Contribution type* defines how the crowdsourced task gathers information. A crowdsourced task expects results from the crowd, and the crowd contributes their results to the task as a group or as an individual.

#### **4.2.4. The crowdsourcing platform**

The crowdsourcing platform is the location where the crowdsourcing task happens. This location can be either a physical place or an online platform. Most of the time the platform is the latter. These online platforms are for example websites, forums or marketplaces. A crowdsourcing platform has four features which are:

*Crowd-related interactions:* These interactions are provided by the platform between the crowd and the platform. These interactions include but are not limited to providing authentication & enrolment mechanisms for users, assigning tasks to the crowd members and providing a mechanism which crowd members can use to submit their results.

*Crowdsourcer-related interactions:* As with the previous entry, these interactions are provided by the platform and are between the crowdsourcer and the platform. Some examples are providing a verification mechanism which the crowdsourcer can use to verify that the results given by the crowd are valid, and providing a broadcasting mechanism via which the crowdsourcer can communicate their crowdsourced task to the crowd

*Task-related facilities:* These facilities are provided by the platform to act between the task and the platform. Again, some examples are providing an aggregation mechanism which will aggregate the results of the task and send the result to the crowdsourcer. In some cases, the crowd might receive an aggregation of their generated results as a form of feedback. Additionally, the platform can provide privacy to each crowd members' results. Only the individual or group who provided the answer should be able to see it. This also ensures that the opinions of others do not infer with the opinions of others.

*Platform-related facilities:* These facilities exist within the platform. Some examples are the possibility to moderate the platform to reduce misuse. An example of a moderation system could be the ability to delete submissions that do not contain information related to the activity in hand. In addition, the platforms usually provide an easy and intuitive user interface for both the crowd and the crowdsourcers to use.

#### **4.3. Crowdsourcing Mobile Application Reviews**

Groen *et al.* [2015] suggested that user reviews could be used in addition to regular requirements engineering techniques when eliciting new requirements. One key benefit of this approach is that the need for co-presence between the stakeholder and the analyst during the requirements engineering process would be eliminated. Instead, the users can provide their feedback whenever they have time for it, and the analyst can analyse the feedback when enough data has been collected. Therefore, mobile application marketplaces such as Google Play Store and Apple App Store can be used as crowdsourcing platforms. These platforms amass millions of reviews and these reviews can be used to

elicit new requirements from them by for example fixing reported bugs and fulfilling community requested features. While mobile application reviews are filled with useful information and reflect the needs of the crowd excellently, many developers are not relying on them due to the large amount of review that would need to be analysed. Additionally, the developers might not realize the hidden potential buried within the reviews. [Khaled *et al.*, 2015; Palomba *et al.*, 2018]

#### 4.3.1. Comparison to Crowdsourcing Reference Model

In their paper Khalid *et al.* [2015] propose that various kinds of information can be extracted from user reviews such as bug reports and feature requests. From this information new requirements can be elicited. Figure 2 outlines the crowdsourcing process for mobile application reviews

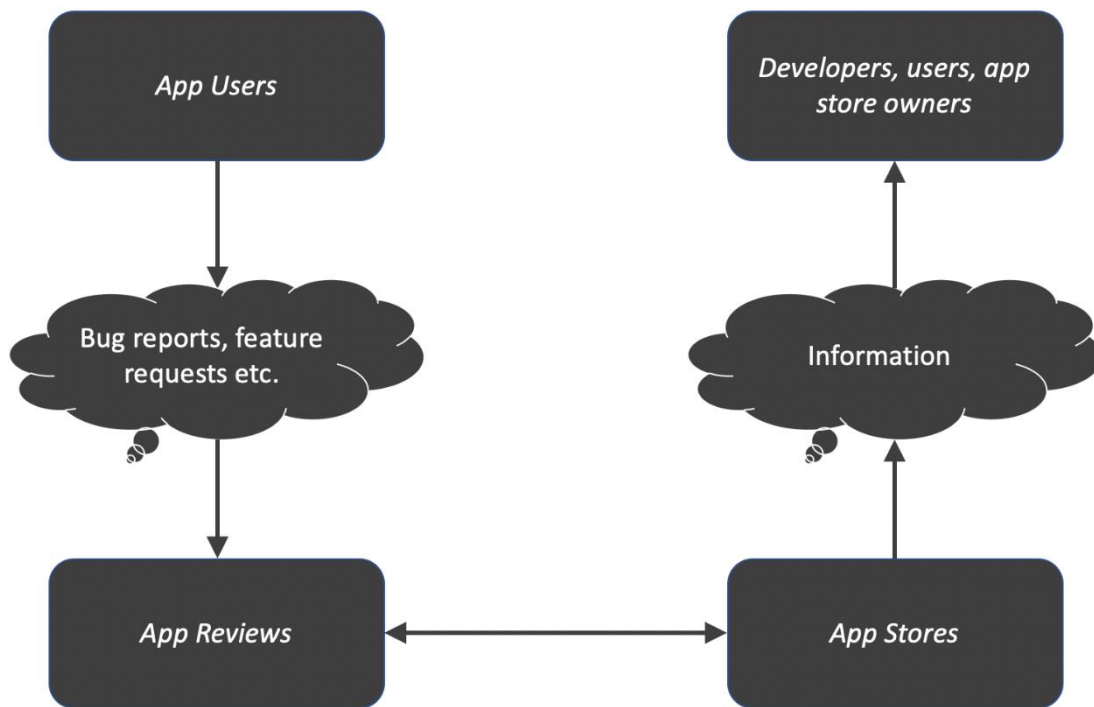


Figure 7: Mobile application review crowdsourcing process [Khalid *et al.*, 2015]

The process in Figure 7 uses the crowdsourcing reference model introduced in Section 4.2.

#### **4.3.1.1 The crowd – *Application Users***

The crowd is the group of people who perform in the specified task. In case of mobile application reviews the crowd is consisted of the users who use the application and provide reviews for it. Since anyone can download an app, use it and review it, the group can consist of various users of differing gender and age. The group can be therefore labelled as *diverse* and is considered *undefined*. As mentioned in the reference model, it should be noted that the diversity of the group cannot be guaranteed on applications with low amounts of reviews. The group is also *suited* to provide feedback on the app since there are no right or wrong answers when providing feedback. Since the reviews are posted on a public forum where anyone can inspect them, and the reviewer's username is visible, the crowd does not fulfil the *unknown-ness* feature. [Khalid *et al.*, 2015]

#### **4.3.1.2 The crowdsourcer – The developers, users & app store owners**

The crowdsourcer is the entity who benefits from the crowdsourcing activity. In case of mobile application reviews there are multiple stakeholders who can gain benefit from the reviews. Khalid *et al.* [2015] outlined four different stakeholders in their paper:

*Users* can gain benefit from the crowdsourced activity by having an understanding of the quality of the application before using it. Finkelstein *et al.* [2014] found that a higher rating correlates to higher download numbers for that particular application.

*The development team* benefits from the crowdsourcing activity by obtaining useful information from the reviews. This information can be used to determine which features to pursue in the next release, or if there's something wrong with the app that the team was not aware previously. [Khalid *et al.*, 2015]

*App store owners* such as Apple and Google gain benefit from the reviews by getting feedback on malicious and/or dangerous applications. [Khalid *et al.*, 2015]

*Researchers* can gain benefit by having data for their studies. These studies can help the researchers to for example better understand users' likes and dislikes. [Pradhan *et al.*, 2016]

In the case of mobile application reviews the crowdsourcer can incentivize the crowd to provide reviews. Khalid *et al.* [2015] claim that some applications they observed offer users some entertainment incentives in exchange for 5-star reviews. The crowdsourcers also provide an *open call* for anyone to join the reviewing process. The only requirement is that the user has downloaded the application, so that they are able to review it. The ethical constraints present only affect the app store owners, since they run the service. The users have a possibility to stop the reviewing process at any time, as well as have the option to retroactively delete the reviews they have provided. [Khalid *et al.*, 2015]

#### 4.3.1.3 The crowdsourced task – *App Reviews*

The crowdsourced task is the activity the crowd takes part in. In the case of mobile application reviews the activity is collecting the reviews through the app store platform. Within the task, the users who participate share their opinions and experiences within the app. The review also contains a quantifiable *rating* the users provide for the app. The task itself cannot be fulfilled without the crowd of users, so there isn't a possibility for a *traditional operation* outlined in the reference model. The task cannot be directly outsourced either, since the task consists a large, undefined and be outsourced to a individual or organization. The task itself cannot be split into smaller tasks (*micro tasks*) but is still *modular*: each individual can perform the task independently. Reviewing is also a very simple task and due to the nature of the task, not automatable. As a task, writing reviews for an application is purely user-driven. [Khalid *et al.*, 2015]

#### 4.3.1.4 The crowdsourcing platform – *App Stores*

The crowdsourcing platform is the place in which the task performed at. Again, in the case of mobile application reviews the locations are the app stores for each platform respectively. The platforms provide *crowd-related interactions* in multiple ways such as via authentication services and result submission. The app stores do not inherently assign tasks to users, but the users can choose which tasks – or to be more specific, which applications – they want to review. Additionally, the app stores do not provide the users with feedback when they are performing the task.

*Crowdsourcer-related interactions* are also present. While the app store is free to browse for anyone seeking information, the developers are required to perform an enrolment activity if they want to have their application on sale at the app store. Depending on the stakeholder they have different options available to inspect the application. From the stakeholders mentioned the app store owners and developers of a specific application have the most information available to them. This information includes information such as in which countries the app is used at, what the usage demographic is and the total number of active installs. For regular users and researchers, the information is restricted to a much more limited view that includes all of the reviews and the star rating average for each application. As mentioned earlier, the users choose when and which apps to review, therefore there are no broadcasting mechanisms in place for the app stores. [Apple, 2020a; Google, 2020; Khalid *et al.*, 2015]

Both platforms provide *task-related facilities*. The platforms form an aggregated score for each application based on the reviews received. The platforms do not hide the results of others from view: anyone can observe the posted reviews. In addition, since the

sole reason of reviews it to share one's experience with others it would contradict the basic principles of reviews to hide them from others. [Khalid *et al.*, 2015]

Since the platforms are located online, they inherently provide *platform-related facilities* to use. The platforms are an online-based facility via which users can do various activities, including writing reviews. As mentioned in Section 2.2, review validity is a problem which both platforms take seriously. Tools and systems are in place to combat false and/or malicious reviews. The users are also not required to pay for the usage of the platforms, so writing reviews is a completely free activity. [Apple, 2020a; Google, 2020; Khalid *et al.*, 2015]

From this analysis we can see that mobile application reviews – and especially the collection of them – can be described as a crowdsourcing activity. Khalid *et al.* [2015] outlined that there are four different categories in which the users can provide crowdsourcable information for the stakeholders. These categories are also supported by the paper by Guzman *et al.* [2015]. The kinds of information extracted from user reviews as shown by Khalid *et al.* [2015] are:

*Feature requests and suggestions:* These reviews contain explicit requests for the developers regarding a specific part of their application. For example, a user could wish that the mobile game they play would notify them when a special event occurs in-game. Combining these feature requests en masse from the reviews can help developers outline some highly requested features and focus future development on them.

*Recommendations for Users:* Users can provide valuable feedback for other users in their reviews. Reading these reviews can make the download or purchase decision of an application easier for a new user. For developers, this information can be used to elicit a general picture of how the crowd reacts to the application. Fu *et al.* [2013] proposed a system called WisCom which is used to extract user sentiment from reviews. If there are more positive toned reviews posted, it can be said that the application is well received.

*Problem Spotting:* Sometimes users encounter problems while using the application. While a software project should have a quality assurance unit [Whitaker, 2009], sometimes bugs slip into production. These bugs can require very specific conditions to occur, and some users might encounter these bugs in their usage. Reviews can be used as an alternative bug reporting channel, and developers should be aware of reviews containing this kind of information. An example could be that an alarm clock application does not function well when travelling from one time zone to another.

### 4.3.2. CRISTAL

Palomba *et al.* [2018] studied the usage of mobile application reviews in software development. The study was aimed to answer to what extent developers can leverage

crowdsourcing mechanisms to plan for future changes, and how the changes are received by the users after implementing them. The study consisted of linking mobile application reviews on open-source projects to source code changes in the applications' public repositories. The impact of each change was measured with the success (i.e. rating) of the application after the change was applied, as well as an interview with the developers themselves. To aid the research team and the project team of each application, the team developed a new approach to monitor whether and how the user reviews have been implemented to the target application. The approach was dubbed as **CRISTAL** (**C**rowdsour**C**ing **R**ev**E**ws to **S**upport **A**pp **e**vo**L**ution). Figure 8 shows an overview of the system. In addition, the team conducted a survey aimed at mobile application developers. The survey collected these developers' opinions on their usage of reviews when developing applications.

In order to link application reviews to source code changes, the system must first collect and filter reviews from a mobile application marketplace. In this study, the team decided to use Google's Play Store as their marketplace of choice. From the reviews submitted on the marketplace for each application in the study, the team filtered out *non-informative* reviews; reviews that do not contain a reason for the emotional (positive or negative) reaction. For example, a review message "*this app is terrible!*" does not provide any information as to *why* the user sees the application as terrible. *Informative* review messages such as "*The app crashes when sharing photos*" provide a reason for the review. The team used the AR-Miner classifier by Chen *et al.* [2014] to aid with the filtering. The second phase consists of mapping *informative reviews* to issues and commits within the code repository. Any issue that was created on or after the date of the review in hand is considered to be linked to the review. In addition, any commits created within the same time period are considered to be linked to the review as well. Third phase of the process uses the *Link Identifier* part to analyse both the reviews and the issues and commit messages and determines whether there is a link present. After all informative reviews have been analysed, the system generates a report for various stakeholders to use. After deploying CRISTAL to analyse 100 Android applications the system managed to reach an average precision of 77% and recall of 73% when observing issues. When observing commit notes the precision rose to 80%.

According to Palomba *et al.*, [2018], most of the developers take reviews carefully into account when working on a new release of their application. Among the 100 applications they observed, on average 49% of informative reviews were addressed by developers in a new release. The survey results also support these findings. The respondents claimed to often rely on user reviews to plan the next release of their application (49% very often, 38% often). The main areas of interest are bug reports and new feature suggestions. The majority of the respondents indicated that only a small amount (25%) of

user reviews are informative, thus useful in the development of the application. Additionally, the study found that applications that implement requested features and fix reported bugs are rewarded by their users with higher positive average rating than those that do not.

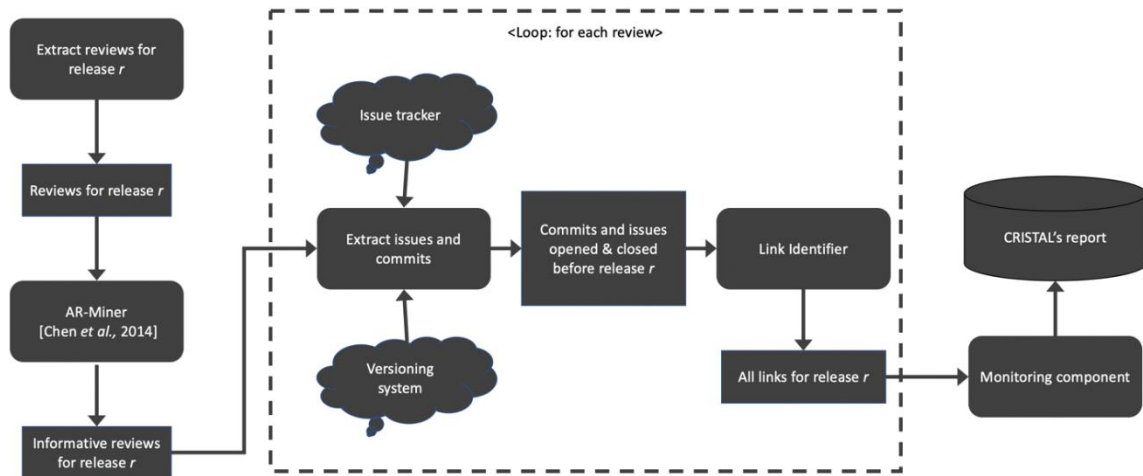


Figure 8: Simplified overview of CRISTAL [Palomba et al., 2018]

#### 4.4. Summary

Mobile application review writing can be categorized as a crowdsourcing activity, and the app stores provide a platform to host this activity in. The reviews generated by the review writing activity can be used to elicit new information for the applications, and developers strongly feel that reviews should be used in development. While the raw amount of reviews available for some applications can be overwhelming, there have been developments both in the industry and in scientific research to further enable developers to better harness the data available within reviews. A key drawback in review usage in software development is the vagueness of reviews, since only roughly a third of reviews contain useful information that can be used to elicit new features or discover defects within the application. As the number of smartphone users worldwide increases by roughly by 300 million annually [Statista, 2020] it is increasingly important that the developers and the scientific community continue to develop new and improved methods to use reviews to enhance software development processes.

## 5. Study and Methodology

This chapter outlines the study presented in the thesis. The thesis aims to answer the two research questions presented in the introduction and is performed as an online questionnaire. The following chapters will further explain the research design, the methods used and how the data was collected and analysed.

### 5.1. Research Design

From the information presented in the previous chapters we can clearly see that mobile application reviews can be a very valuable asset within software development and harnessing the power of the crowd to collect bug reports, feature requests and overall feelings towards the application can help in the development of the application. In order to use reviews as a data source, there must be reviews for the application. By increasing the raw amount of reviews available in general, it would lead to more data available for developers and other stakeholders – such as researchers – to use. The hypothesis presented in this research is that users would be more inclined to write reviews if they were rewarded for them.

Subsection 3.4.1 outlined which motivations are currently in place within mobile application marketplaces. In order to better understand this particular topic, we want to measure how important these motivations are in the current application marketplace review collection system. This will answer our first research question, and it will be accomplished by forming questions around the motivations summarized in Section 3.3. It is important to note that as mentioned in Section 2.3, currently providing a reward in exchange for reviews is forbidden within the application marketplaces. However, this attitude can change in the future were Apple or Google change their stance on the topic. To answer the second research question, we will query whether the users would like to receive rewards for their reviews, as well as what kind of reward would suit them the best.

Financial benefits can come in various forms. In order to measure what the best option would be, the thesis provides the participants with three different options already implemented (or similar to) as in other services found online; monetary reward, discount and *App Store Points*. Monetary rewards – such as Amazon Mechanical TURK [Amazon, 2020a] – reward users for their manual work. In this case, the user would be *monetarily rewarded* by their time spent on writing the review. Providing *discounts* in exchange for reviews is an activity that previously took place at Amazon [P. Wu, 2019] before Amazon forbid the activity. In this scenario users would either receive discounts after their review or write reviews after acquiring a targeted discount from a manufacturer [P. Wu, 2019].

*App Store Points* can be categorized as a form of gift balance. Some e-commerce marketplaces such as WooCommerce [2020] provide the store owners the option to reward users with a set amount of points for example in exchange for their reviews. Having a scientific basis on which approach to take in such an event that financial rewards become available could be beneficial for a large community.

### **5.1.1. Research Participants**

A mobile application can be downloaded, used and reviewed by anyone. The user base can be described as a *crowd*. In order to have a realistic representation of the crowd to answer the questionnaire, it was sent to friends and family of all ages, accompanied with a request to forward it onwards to their social circles as well. The questionnaire was answered 37 times. Due to the anonymous nature of the questionnaire, no log-in or user identification features were set. Additionally, in order to avoid potential misuse, the questionnaire was not sent to any public forum due to not being able to identify malicious answers that might occur.

## **5.2. Questionnaire**

The research is conducted as an online questionnaire via Google Forms. The questionnaire contains two sections, one for each research question. Each section is presented on a single page, and the user can only enter and see the section on the second page only after they have answered all of the questions in the first page and pressed the “next”-button. An answer is required for all fields on the questionnaire.

Table 6 demonstrates the first section. In it we have a total of six statements based on the motivations outlined in Section 3.3. The interviewees are asked to answer the statements in this section on a scale of 1 to 5, with 1 representing “Strongly disagree” and 5 representing “Strongly agree”. From this data we can calculate the average and median scores for each, resulting in a ranking system which indicates which motivations users view the most relatable when writing mobile application reviews.

<i>Motivation</i>	<i>Statement</i>
<i>Social benefits</i>	Writing a review brings me a sense of belonging to an online community
<i>Self-enhancement</i>	Writing a review helps me portray myself as an advanced user
<i>Financial benefits</i>	I would like to receive a reward in exchange for my reviews
<i>Altruism</i>	Writing a review helps me share my positive or negative experiences with the application with others
<i>Advice seeking</i>	Writing a review can help me get a solution to a problem I have with the application
<i>Social outlet</i>	Writing a review allows me to express my happiness or unhappiness towards the developer of the application

Table 6: Statements presented in the questionnaire

The second section of the questionnaire (Table 7) contains three proposed scenarios. The number of scenarios was chosen to provide a clear comparison between different approaches and was chosen purposefully. The scenarios presented all provide the reviewer with some kind of financial benefit. Each scenario is accompanied with a short description of the scenario in the questionnaire. The same ranking system is used within the first section is used in this section as well.

In the first scenario (monetary reward) the user is rewarded with a set amount of money for each review they write. The reward is automatically deposited into the payment option linked to their account, or in case of no payment option into their application marketplace account's balance.

In the second scenario the user is rewarded with discounts for various in-app purchases within the application they review. For example, a video streaming service's subscription fee could be lowered temporarily in exchange for one's review.

The third scenario proposes a new type of currency – App Store Points. In this scenario, by writing reviews the user will be rewarded with App Store Points. These points can be used to make purchases within the application marketplace alongside regular payment options.

<i>Scenario</i>	<i>Question</i>
<b><i>Monetary Reward</i></b>	I would feel motivated to write a review for an application if I was rewarded monetarily for it
<b><i>Discount on In-App Purchases</i></b>	I would feel motivated to write a review for an application if I was rewarded with a discount for it
<b><i>App Store Points</i></b>	I would feel motivated to write a review for an application if I was rewarded with App Store Points for it

Table 7: Questionnaire Scenarios

## 6. Results and Discussion

This chapter goes over the results of the study presented in the thesis. The chapter is split into two sections, one for each research question.

### 6.1. Current Motivations

To answer **RQ<sub>1</sub>**, a set of statements was generated (Table 6) to measure how each motivation outlined in Section 3.3 was perceived by the participants. Each statement is represented as a bar chart. Table 8 contains the average and median value for the answers received for each statement.

<i>Motivation</i>	<i>Median</i>	<i>Average</i>
<i>Social benefits</i>	2	2,38
<i>Self-enhancement</i>	3	2,70
<i>Financial benefits</i>	3	3,21
<i>Altruism</i>	4	4,03
<i>Advice seeking</i>	3	3,05
<i>Social outlet</i>	4	3,89

Table 8: Median and average scores for motivations in the study

#### 6.1.1. Social benefits

In order to measure how strongly social benefits affect the users' motivations for writing reviews, the statement claims that writing a review can help the writer to gain a sense of belonging to an online community. As seen in Figure 9, a majority of participants disagree that by writing a review in mobile application marketplaces can they gain a sense of belonging to an online community. Only a handful of answers were above 3 (neither agree nor disagree) with zero participants agreeing strongly with the statement. From the answers we can see that with an average score of 2,38, this motivation is not often present when writing a review. A reason for this result can be the lack of any distinct social features present in the app stores, so the feeling of belonging to an online community must come from within for those who feel so.

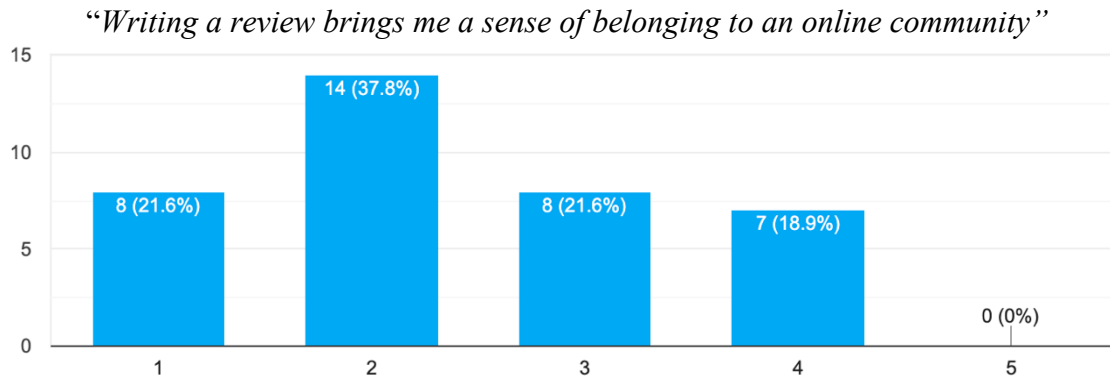


Figure 9: Answers for *social benefits*-statement

### 6.1.2. Self-enhancement

To measure the impact of *self-enhancement* as a motivation the questionnaire provided the participant with a statement claiming that when writing a review it can help them to portray themselves as advanced users. As discussed in Section 3.3, by portraying one as an intelligent shopper can enhance their own image in their eyes. The answers (Figure 10) show that while some participants view self-enhancement as an important motivation, a small majority of the answers is on the *disagree*-side of the spectrum. An average of 2,70 and a median of 3 also further promote how the answers were spread out.

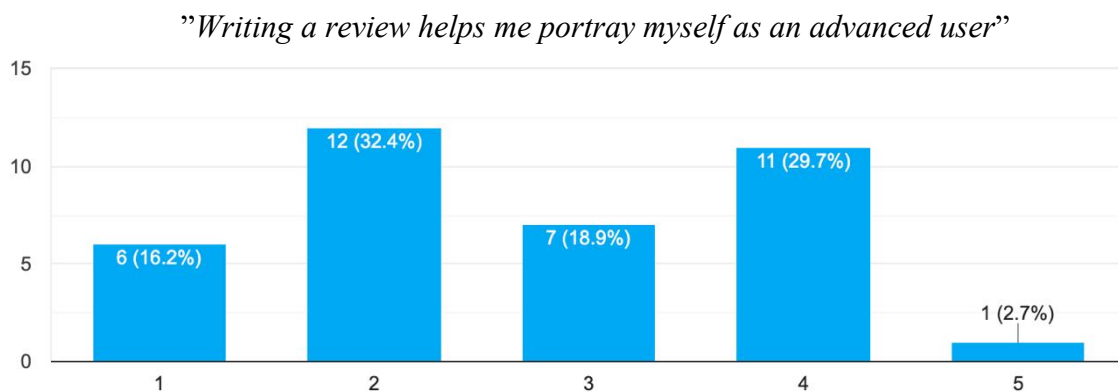


Figure 10: Answers for *self-enhancement* statement

### 6.1.3. Financial benefits

Both platforms currently prohibit providing financial rewards in exchange for one’s reviews. Still, obtaining financial benefits in exchange for one’s reviews is present on other reviewing platforms such as Amazon [P. Wu, 2019]. In order to measure what kind of an

impact this approach would have within mobile application marketplaces if implemented, the questionnaire presents the participants with a statement claiming that they would like to receive rewards in exchange for their reviews. Looking at the results (Figure 11), we can see that a majority of participants tend to agree with the statement, but only a third of participants who agree did agree strongly with the statement. Clearly, for some participants receiving a reward in exchange for their reviews is unwanted. Reasons for this can be various, such as feeling increased pressure to provide a better review than one intended [P. Wu, 2019].

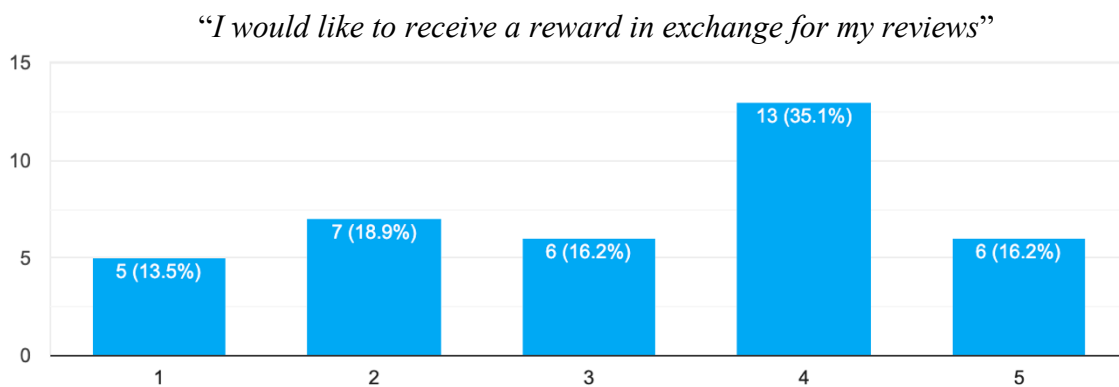


Figure 11: Answers for *financial benefits* statement

#### 6.1.4. Altruism

Based on the literature reviewed in the thesis, one of the most prevalent reasons to provide reviews is altruism; the act of doing something without expecting a reward in return. Sharing one’s positive or negative experiences with one’s peers does not provide the person with any direct benefit and is ultimately a selfless act when the reviewer’s time is consumed by the review process. The results (Figure 12) for this statement reinforce this notion. A clear majority of participants agree with the presented statement. Only a handful of participants did disagree with the statement. With an average score of 4,03 altruism is clearly one of the most important motivations for users to review applications in mobile application marketplaces.

*“Writing a review helps me share my positive or negative experiences with the application with others”*

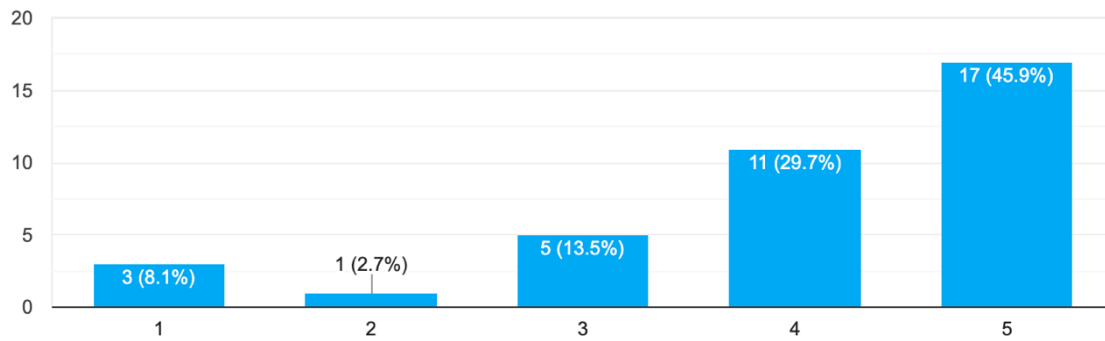


Figure 12: Answers for *altruism* statement

### 6.1.5. Advice seeking

Sometimes when using an application the user can encounter problems within the application. This can cause the user to seek solutions to their problem. In some cases posting a review detailing the problem can lead to a solution for the user, for example via a developer’s reply. To measure this motivation, the questionnaire presented the participants with a statement claiming that writing a review can help them to get a solution to a problem they are experiencing. With an average score of 3,08 and the spread of the answers, advice seeking can be categorized as varying. Different persons have differentiating opinions on whether feel like writing a review can help them with their problem.

*“Writing a review can help me get a solution to a problem I have with the application”*

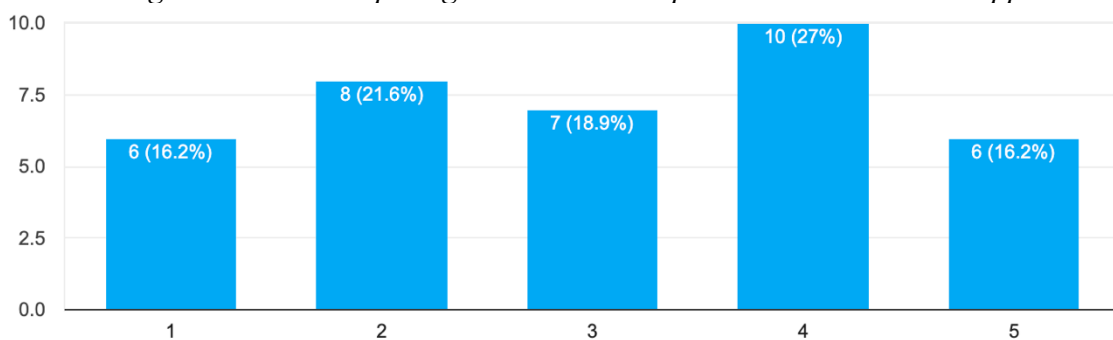


Figure 13: Answers for *advice seeking* statement

### 6.1.6. Emotional outlet

As discussed in Section 3.3, consumers who have had emotional consumption experiences may want to express these feelings towards the developers behind the product. By providing a review, the reviewer can express these feelings towards the developers on a public forum. The questionnaire presented the participants with a statement claiming that

writing a review can help the participants to express their happiness or unhappiness towards the developer of the application. These reviews are then directed not at the application itself. Looking at the results (Figure 13) we can see that a vast majority of participants agree with the statement; a review *is* a good way to express one’s feelings towards the developers behind the application. A cause for this behaviour could for example be that a company has caught bad publicity and is now being reprimanded publicly by customers via reviews [Brightman, 2016].

“*Writing a review allows me to express my happiness or unhappiness towards the developer of the application*”

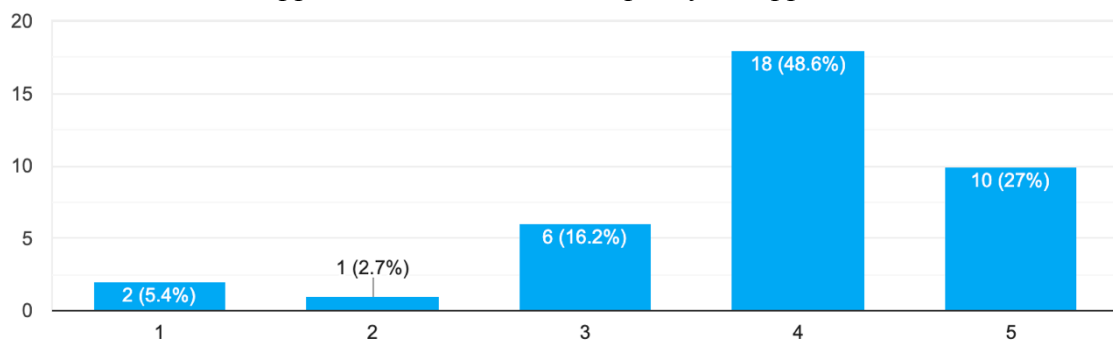


Figure 13: Answers for *social outlet* statement

### 6.1.7. Summary

When looking at the data available, it can be said that currently *altruism* and *emotional outlet* are the leading motivations for users to write reviews. These two motivations do not require anything special from the application marketplaces and the motivations are present in nearly all reviewing activities found online. On the other hand, *social benefits* and *self-enhancement* are the least important motivations for users to write reviews. It can be argued that the lack of any significant social features within the application marketplaces currently can cause this effect. When asked whether the participants would like to receive financial benefits in exchange for their reviews, a minority liked the idea, but others had differing opinions. It could be argued that some users could view the idea of receiving a reward in exchange for one’s review to affect the review itself, rendering it untrustworthy. Further studies could be performed to determine if - and in such case how much - rewards skew the review.

## 6.2. Reaction to Proposed Changes

The second research question (**RQ<sub>2</sub>**) in the study asks how users would react to being rewarded in exchange for their reviews. As seen in Subsection 6.1.3, a majority of participants agreed with the statement that they would like to receive rewards in exchange for their reviews. The second section of the questionnaire proposes three new scenarios in which the reviewer is rewarded. Table 9 displays the average and median ratings for each scenario.

<i>Scenario</i>	<i>Median</i>	<i>Average</i>
<i>Monetary Reward</i>	4	4,08
<i>Discount on In-App Purchases</i>	4	3,70
<i>App Store Points</i>	3	3,13

Table 9: Median and average scores for scenarios in the study

### 6.2.1. Scenario 1 – Monetary Reward

The first scenario proposed to the participants is the *monetary reward* scenario. The participants were given the following description of the scenario:

“In this scenario you are rewarded with a set amount of money for each review you write. The reward is automatically deposited into the payment option linked to your account, or in case of no payment option into your application marketplace account's balance.”

The results (Figure 14) show that this scenario is very well received by the participants. Only a total of 8 participants were indifferent or disagreed with the scenario. From this we can see that even the ones who showed their disagreement for getting rewards in Subsection 6.1.3 would feel motivated to write reviews were they monetarily rewarded for them.

*”I would feel motivated to write a review for an application if I was rewarded monetarily for it”*

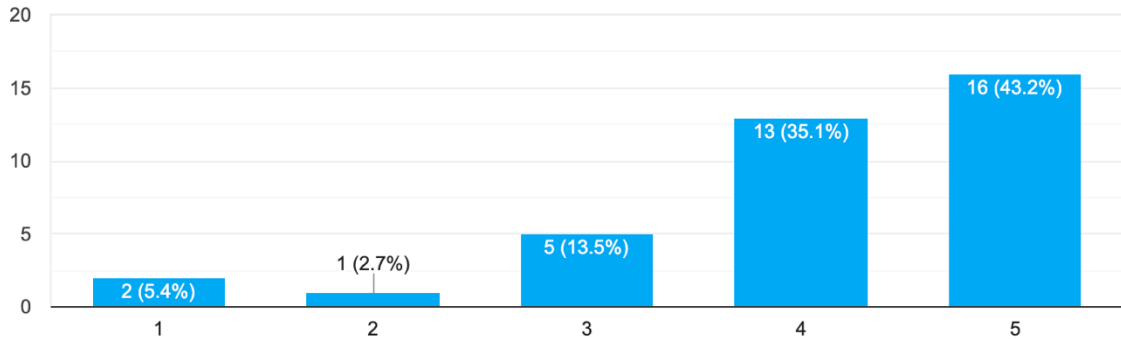


Figure 14: Answers for the first scenario

### 6.2.2. Scenario 2 – Discount on In-App Purchases

In the second scenario the participants were proposed with a scenario in which they receive discounts within the application they review. The scenario had the following description:

*“In this scenario you are rewarded with discounts for in-app purchases within the applications you review. For example, a video streaming service's subscription fee could be lowered temporarily in exchange for your review”*

This scenario was received within the participants. With an average score of 3.70 the scenario has a positive valence in the eyes of the participants as a majority of the participants agreed or strongly agreed with the statement presented in the question. This scenario saw a minor increase to answers on the disagree-end of the answer spectrum.

*“I would feel motivated to write a review for an application if I was rewarded with a discount for it”*

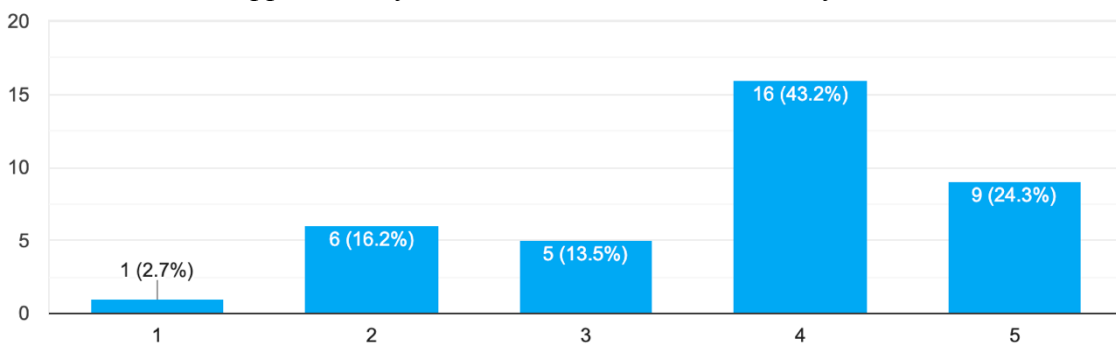


Figure 15: Answers for the second scenario

### 6.2.3. Scenario 3 – App Store Points

The third scenario presented in the questionnaire proposed that the reviewer would gain App Store Points upon reviewing applications within the application marketplace. The scenario was accompanied with the following description:

“In this scenario, by writing reviews you will be rewarded App Store Points. These points can be used to make purchases within the application marketplace alongside regular payment options.”

Based on the results (Figure 16) this scenario was the least liked scenario of the three. With an average score of only 3,13 it is nearly in the middle of the answer spectrum. Looking at the spread of the answers it can be seen that the answers were spread out over all of the answer options, signalling that this particular approach divides opinions. With 18 participants agreeing with the statement, and 14 disagreeing the scenario still tilts ever so slightly to positive valence.

*“I would feel motivated to write a review for an application if I was rewarded with App Store Points for it”*

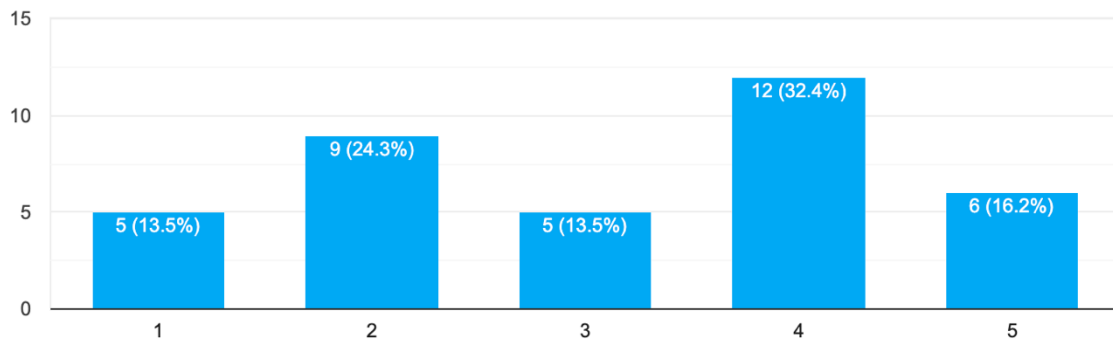


Figure 16: Answers for the third scenario

### 6.2.4. Comparison of the Scenarios

Looking at all of the results, it can be clearly said that the first scenario was favoured by the participants the most, with the second scenario and third scenario coming in second and third respectively. One reason for the great reception of the first scenario could be the ease of it. Once reviewed, the reward is automatically delivered to the reviewer’s bank account where they can use it as they see fit. When comparing this scenario to the App Store Points-scenario, both of the scenarios provide instantaneous rewards which the user can use to purchase something. With App Store points the purchasing is limited to the application marketplace ecosystem, rendering the points useless if the reviewer does not

intend to purchase something from the marketplace. On the other hand, obtaining “free” credits to use could drive the reviewer to acquire something from the marketplace they previously had no intent on doing. In the study by van Dooren *et al.* [2018] the authors measured the effectiveness of various reward schemes in exchange for the participants effort put into a video game. Their findings showed that the participants were most motivated by monetary rewards, similarly to the results shown in this thesis.

An aspect to consider is that the participants might not have had any specific ideas in mind when answering to the statement presented in Subsection 6.1.3. The scenarios in the second section of the questionnaire presented the participants with three different scenarios (some of which the participants might not have realized would be possible) thus providing more agreeable answers towards the statements in each of the scenarios.

### **6.3. Threats to Validity**

Regarding the study, one threat is the possible lack of answers. The 37 answers gathered does provide a good baseline for the study but with a bigger data set the differences could stand out more. Especially in such questions that divide the opinions of the participants (Scenario 3) an increased data set could reveal more information. Additionally, the participants age could be taken into account. By collecting more answers alongside the age of the participant, a study could contain age brackets which could help to outline which approach to choose when targeting a certain age group of users.

Another issue to consider is that the study does not take into account whether someone has previously written a review or not. Some participants might be heavily against writing reviews in general, thus skewing the results towards the disagree side of the answer spectrum. On the other hand, the study aims to find what motivates people to write reviews and therefore must target everyone, not just ones who have previously written reviews. This issue could be taken into account in future research as well.

## 7. Conclusions and Future Work

This thesis has examined what reviews are, what motivates users to write reviews, how reviews are collected and thereafter used within software development. A study was performed to seek questions to the two research questions presented in the thesis. The results of the study clearly show that currently the core motivations for users to write reviews are altruism and social outlets. In addition, obtaining financial rewards through the reviews process was something that the participants generally felt as being something they'd like. Upon providing the participants three reward scenarios currently present in other online services, the most prevalent scenario was one of monetary reward. Furthermore, some of the participants who disagreed with receiving financial rewards in exchange for their reviews felt that receiving a monetary reward *would* motivate them to write reviews.

The results of the study could be used to further improve application marketplaces. By implementing financial benefit systems within the review processes, the user base would be more motivated to provide reviews for applications. It has to be noted that while the *motivation* to write reviews was increased within the proposed scenario, the *quality* of reviews may not be uniform with current reviews. In their study, van Dooren *et al.* [2018] noted that participants who were rewarded monetarily rated the experience more positively than those who did not receive rewards. Upon receiving a reward, a reviewer might put more effort into the review, increasing its usefulness but it could also affect the general valence of the review.

## References

- Ahn, L. (2011). Massive-scale online collaboration. Retrieved from [https://www.ted.com/talks/luis\\_von\\_ahn\\_massive\\_scale\\_online\\_collaboration?language=fi-FI](https://www.ted.com/talks/luis_von_ahn_massive_scale_online_collaboration?language=fi-FI) Access date: October 5, 2020
- Amazon. (2020a). Amazon mechanical turk. Retrieved from <https://www.mturk.com/> Access date: October 1, 2020
- Amazon. (2020b). Amazon's top customer reviewers. Retrieved from <https://www.amazon.co.uk/gp/customer-reviews/guidelines/top-reviewers.html> Access date: August 29, 2020
- App Annie. (2019). State of mobile 2019. Retrieved from <https://www.appannie.com/en/go/state-of-mobile-2019/> Access date: Aug 25, 2020
- Apple. (2020a). App store connect. Retrieved from <https://developer.apple.com/app-store-connect/> Access date: Sep 11, 2020
- Apple. (2020b). App store review guidelines. Retrieved from <https://developer.apple.com/app-store/review/guidelines/> Access date: August 30, 2020
- Brightman, J. (2016). Steam updates user reviews to show a game's score over time. Retrieved from <https://www.gamesindustry.biz/articles/2016-05-04-steam-updates-user-reviews-to-show-a-games-score-over-time> Access date: Aug 28, 2020
- Bugcrowd. (2020). Bug bounty list - all active programs in 2020. Retrieved from <https://www.bugcrowd.com/bug-bounty-list/> Access date: Sep 17, 2020
- Chen, N., Lin, J., Hoi, S. C. H., 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 (ICSE 2014). ACM. 767–778. doi:10.1145/2568225.2568263

- Chevalier, J. A., & Mayzlin, D. (2018). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3), 345-354. doi:10.1509/jmkr.43.3.345
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340. doi:10.2307/249008
- Dichter, E. (1966, Nov 1,). How word-of-mouth advertising works. *Harvard Business Review*, 44, 147. Retrieved from <https://search.proquest.com/docview/1296466094>
- Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2), 189-200. doi:10.1177/0165551512437638
- Finkelstein, A., Harman, M., Jia, Y., Martin, W., Sarro, F., & Zhang, Y. (2014). App store analysis: Mining app stores for relationships between customer, business and technical characteristics. UCL Department of Computer Science.
- Fu, B., Lin, J., Li, L., Faloutsos, C., Hong, J., & Sadeh, N. (2013). Why people hate your app: Making sense of user feedback in a mobile app store. *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM. 1276-1284. doi:10.1145/2487575.2488202
- GOG. (2020). Wishlist - GOG.com. Retrieved from <https://www.gog.com/wishlist>  
Access date: September 16, 2020
- Google. (2018). In reviews we trust — making google play ratings and reviews more trustworthy. Retrieved from <https://android-developers.googleblog.com/2018/12/in-reviews-we-trust-making-google-play.html> Access date: August 30, 2020
- Google. (2020). Google play console. Retrieved from <https://play.google.com/console/about/> Access date: Sep 11, 2020
- Grayson, N. (2015). Steam 'review bombing' is A problem. Retrieved from <https://kotaku.com/steam-review-bombing-is-a-problem-1701088582> Access date: Aug 28, 2020

- Groen E.C., Doerr J., Adam S. (2015) Towards Crowd-Based Requirements Engineering A Research Preview. In: Fricker S., Schneider K. (eds) Requirements Engineering: Foundation for Software Quality. REFSQ 2015. Lecture Notes in Computer Science, vol 9013. Springer, Cham.  
[https://doi.org/10.1007/978-3-319-16101-3\\_16](https://doi.org/10.1007/978-3-319-16101-3_16)
- Guzman, E., El-Halaby, M., & Bruegge, B. (2015). Ensemble Methods for App Review Classification: An Approach for Software Evolution. 30th IEEE/ACM International Conference on Automated Software Engineering (ASE). 771-776. doi:10.1109/ASE.2015.88
- Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (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. doi:10.1002/dir.10073
- Hoehle, H., & Venkatesh, V. (2015). Mobile application usability: Conceptualization and instrument development. *MIS Quarterly*, 39(2), 435-472. doi:10.25300/misq/2015/39.2.08
- Hosseini, M., Phalp, K., Taylor, J., & Ali, R. (2014). The four pillars of crowdsourcing: A reference model. *IEEE Eighth International Conference on Research Challenges in Information Science (RCIS)*, 1-12, doi:10.1109/RCIS.2014.6861072.
- Hosseini, M., Phalp, K. T., Taylor, J., & Ali, R. (2014). Towards crowdsourcing for requirements engineering. *The 20th International Working Conference on Requirements Engineering: Foundation for Software Quality (REFSQ 2014)*, , 1-6. Retrieved from <http://ceur-ws.org/Vol-1138/et2.pdf>
- James F. Engel, Roger D. Blackwell, & Paul W. Miniard. (1994). *Consumer behavior* (8th ed.) Fort Worth: Dryden Press.
- Khalid, M., Shehzaib, U., & Asif, M. (2015). A case of mobile app reviews as a crowdsource. *International Journal of Information Engineering and Electronic Business*, 7(5), 39-47. doi:10.5815/ijieeb.2015.05.06

- Kowol, A. (2008). The theory of cognitive dissonance. Retrieved from <http://adamkowol.info/works/Festinger.pdf>
- Lee, S. J., & Lee, Z. (2006). An experimental study of online complaint management in the online feedback forum. *Journal of Organizational Computing and Electronic Commerce*, 13(1), 65-85. doi:10.1207/s15327744joce1601\_4
- Lim, S., Quercia, D., & Finkelstein, A. (2010). StakeNet: Using social networks to analyse the stakeholders of large-scale software projects. *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering - Volume 1*, ACM. 295-304. doi:10.1145/1806799.1806844
- Martens, D., & Maalej, W. (2019). Towards understanding and detecting fake reviews in app stores. *Empirical Software Engineering : An International Journal*, 24(6), 3316-3355. doi:10.1007/s10664-019-09706-9
- Microsoft. (2020). Microsoft account page. Retrieved from <https://account.microsoft.com/> Access date: September 16, 2020
- Nielsen. (2015). Global trust in advertising. Retrieved from <https://www.nielsen.com/uk/en/insights/report/2015/global-trust-in-advertising-2015/#>
- Pagano, D., & Brügge, B. (2013). User involvement in software evolution practice: A case study. *Proceedings of the 2013 International Conference on Software Engineering (ICSE '13)*. IEEE Press, 953–962. doi:10.1109/ICSE.2013.6606645
- Pagano, D., & Maalej, W. (2013). User feedback in the appstore: An empirical study IEEE. doi:10.1109/RE.2013.6636712
- Palomba, F., Linares-Vásquez, M., Bavota, G., Oliveto, R., Penta, M. D., Poshyanyk, D., & Lucia, A. D. (2018). Crowdsourcing user reviews to support the evolution of mobile apps. *The Journal of Systems and Software*, 137, 143-162. doi:10.1016/j.jss.2017.11.043
- Platzer, E. (2011). Opportunities of automated motive-based user review analysis in the context of mobile app acceptance. Graz, Austria: Evolaris Next Level GmbH.

Pradhan, L., Zhang, C., & Bethard, S. (2016). Towards extracting coherent user concerns and their hierarchical organization from user reviews. 17th International Conference on Information Reuse and Integration (IRI). 582-590. IEEE. doi: 10.1109/IRI.2016.85.

Prpić, J., Shukla, P. P., Kietzmann, J. H., & McCarthy, I. P. (2015). How to work a crowd: Developing crowd capital through crowdsourcing. *Business Horizons*, 58(1), 77-85. doi:10.1016/j.bushor.2014.09.005

Rensink, J. M. (2013). What motivates people to write online reviews and which role does personality play? Retrieved from [https://essay.utwente.nl/63536/1/Rensink\\_Maarten\\_-s\\_0176486\\_scriptie.pdf](https://essay.utwente.nl/63536/1/Rensink_Maarten_-s_0176486_scriptie.pdf)

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*; *Contemp Educ Psychol*, 25(1), 54-67. doi:10.1006/ceps.1999.1020

Shankland, S. (2008). Google announces android market for phone apps. Retrieved from <https://www.cnet.com/news/google-announces-android-market-for-phone-apps/> Access date: Sep 11, 2020

Sommerville, I. (2006). *Software engineering : (Update) (8th ed.)*. Old Tappan: Pearson Education UK.

Statista. (2020a). Number of apps in leading app stores. Retrieved from <https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/> Access date: Aug 25, 2020

Statista. (2020b). Smartphone users worldwide 2020. Retrieved from <https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/> Access date: Oct 13, 2020

Sundaram, D. S., Mitra, K., & Webster, C. (1998). Word-of-mouth communications: A motivational analysis. *Advances in Consumer Research*, 25, 527. Retrieved from <https://search.proquest.com/docview/1839802963>

- Trustpilot. (2020). How are reviews collected? Retrieved from <http://support.trustpilot.com/hc/en-us/articles/223402468> Access date: Sep 6, 2020
- van Dooren, Marierose M. M, Visch, V. T., & Spijkerman, R. (2018). Rewards that make you play: The distinct effect of monetary rewards, virtual points and social rewards on play persistence in substance dependent and non-dependent adolescents. 6th International Conference on Serious Games and Applications for Health (SeGAH). 1-7. IEEE. doi: 10.1109/SeGAH.2018.8401312.
- Vasa, R., Hoon, L., Mouzakis, K., & Noguchi, A. (Nov 26, 2012). (Nov 26, 2012). A preliminary analysis of mobile app user reviews. Paper presented at the 241-244. doi:10.1145/2414536.2414577 Retrieved from <http://dl.acm.org/citation.cfm?id=2414577>
- Waze. (2020). Waze livemap. Retrieved from <https://www.waze.com/fi/livemap> Access date: Aug 31, 2020
- Whitaker, K. (2009). Principles of software development leadership: Applying project management principles to agile software development. Boston: Course Technology.
- WooCommerce. (2020). WooCommerce points and rewards. Retrieved from <https://woocommerce.com/products/woocommerce-points-and-rewards/> Access date: Oct 13, 2020
- Wu Fei, P. (2019). Motivation crowding in online product reviewing: A qualitative study of amazon reviewers. *Information & Management*, 56(8). 1-9. Retrieved from <https://doi.org/10.1016/j.im.2019.04.006>
- Wu, Y., Ngai, E. W. T., Wu, P., & Wu, C. (2020). Fake online reviews: Literature review, synthesis, and directions for future research. *Decision Support Systems*, 132, 113280.1-15. doi:10.1016/j.dss.2020.113280
- Zha, R. (2019). Improving the usability of mobile application user review collection (Master's Thesis). Tampere University. Retrieved from <http://urn.fi/URN:NBN:fi:tuni-201903211400>