Stockholm School of Economics
Bachelor's Thesis in Marketing
Center for Consumer Marketing
Spring 2015

The App Jungle

– A Quantitative Study Of Ratings And Icon Design Within The App Store

Abstract:

Since the introduction of app stores, the market for apps has grown exponentially. The Apple App Store announced in 2014 that it had reached 85 billion downloads. The growth of the market has made the landscape highly competitive and only a small percentage of these apps gain virality. This study aims to analyze the influence different factors have on an app in App Store. The chosen factors are the rating of the app and the design of the app icon. A quantitative research was made in which the respondents were shown different versions of an app and thereafter asked to state their perception of it. In an attempt to generalize the results, the experiment was conducted with two different app categories. The results indicated that rating do not have any significant impact on consumers' perception. However, the icon strongly influences the consumers' perception, where a picture icon is preferable to a letter icon. The implication of our findings is that app developers do not need to pay as much attention to the rating of their app. However, when designing the app icon, the app developer should preferably use an image instead of a letter icon.

Keywords:

App, App Store, Rating, Icon

Supervisor:
Jonas Colliander

Writers:
Vendela Lilliestråle (22864)
Camilla Lundin (50255)

Examinator:
Anna Nyberg

Thank you!

Anna Nyberg

For guidance and information

Jonas Colliander

For support and advice

Hanna Berg

For guidance in the SPSS jungle

Our Respondents

For kindly taking their time for this experimental study

Table of Contents

1. Introduction

1.1 Background

The aim of this paper is to explore how certain factors within the Apple App Store (hereby referred to as “App Store”) affect a consumer’s perception. There are a myriad of factors that may influence why some apps become successful. We have decided to focus on the medium where an app is downloaded; an app store. Within App Store, there are several factors that could influence the consumer. This report focuses on only two of these; the rating and the app icon. The rating in App Store is visualized through the number of stars where five stars is the maximum and one star is the minimum. To explore the impact of icons, we have chosen two different icons; one picture icon and one letter icon.

The aim among app developers is often to make the app go viral and many developers rely solely on word of mouth. Apple chooses some apps with potential to be exposed on special pages on App Store. However, it is hard to be featured among so many apps that all wants the attention (Wooldridge, 2011). In order for a user to use the app they need to be triggered to download it. Prior studies have not made it clear for the app developers how they can affect that. We want to give concrete recommendations of what aspects to focus on when developing an app to make consumer’s perception as favorable as possible.

1.1.1 Pilot study

To reach out to customers and make them download a specific app, app developers need to be aware of what makes people download one app instead of another. App developers were interviewed prior to the study to find out what they believed were the most important factors to succeed. We started out very broad and included factors that were outside the App Store. The companies have chosen to be anonymous and the apps they produced ranged from different categories. The common view seemed to be that it was hard to define ”the secret sauce” that made an app successful. Most of them believed it were important to have a user-friendly product and that testing the app on users was needed to get market approval. They thought that the graphics of the app did not matter

much as many popular apps have started without any user face that looked professional. However, they did not seem to have any empirically tested information about specific aspects of the app and how that can affect the likeliness of a user wanting to try the app. After the interviews, we chose to focus on the factors related to App Store since these are the ones in immediate relation to a potential download.

1.1.2 App stores

Apps are becoming one of our major platforms for information gathering. People are doing everything from social networking to bank errands through apps in their smartphones. Simultaneously, the landscape for apps is becoming increasingly competitive. There are today several different app stores such as App Store, the Google Play Store and the Blackberry World Store. Furthermore, there are smaller app stores that are specialized on a specific area or region (Khalid, 2015). The global app market is constantly growing and predicted to grow to $52 billion by 2016 (Kai-Chun & Chun Heng, 2013).

App Store was established in July 2008 and since then over 1.6 million apps have been uploaded. The number of apps that gets uploaded has an exponential growth. In July 2014 it was estimated that 60 000 apps were uploaded each month. In June 2014, there were 1.2 million apps available on App Store (Adjust, 2014). This makes it the second largest app store after the Google Play Store. In October 2014, the cumulative number of downloads from App Store was 85 billion (Statista, 2015).

1.2 The importance of ratings

Customer reviews and ratings are getting more frequent online. It is one of the most powerful tools to create online word of mouth (Duan et al., 2008). It can affect the purchase decision to a large extent for various products and services that provide that information. The users usually seek to lower uncertainty, although it can never be completely eliminated (Mudambi & Schuff, 2010).

1.2.1 Prior studies about rating systems

Research has been made regarding what effect rating has on consumers' decision-making when buying products such as books and movies. A previous study about the online store *Amazon* showed that rating was an important part of the purchasing decision. The result from this past study was that the majority found rating of a specific product helpful. A large fraction still thought that it did not give any help in the purchase-decision process. Furthermore, the results showed that whether or not the review was perceived helpful depended on if the product was considered an experience good or a search good. Experience goods require that the product is purchased by a consumer to be fully able to evaluate a product. If it is a search good on the other hand, the consumer can get a perception of the product quality prior to purchase (Mudambi & Schuff, 2010).

Chevalier & Mayzlin (2006) empirically studied how sales were dependent on online book reviews. The sample was taken from Amazon.com and Barnes&Nobles.com. The result was that better reviews on a book led to a relative increase in sales on that particular book. They also found out that the negative impact of a one star review was larger than the positive impact on a five star reviews. However these results are not consistent when comparing studies. Chen & Yoon (2004) got results that indicated that consumer ratings are not related to sales of books. Furthermore Duan et al. (2008) found that higher ratings did not lead to higher sales of movies.

1.2.2 Prior studies about rating systems in app stores

A user that is not satisfied with the experience of an app can leave a rating to express this and the lowest one is one star. If the user is highly satisfied with the app, five stars might be considered which is the highest rating possible. The average of the ratings is calculated by aggregating ratings from all users. App users can also choose to review the app and explain their rating by leaving comments. The reviews can contain useful information about the user's experience. It can both be interesting for developers, app store owners and users (Khalid et al., 2015). The two most common things to complain about are crashes and bugs. This implies that it is important to improve the quality of the app to get a better rating (Khalid & Hassan, 2014). Moreover, product quality is a key factor in consumers purchase decisions. The uncertainty of an app is relatively high due

to that it is hard for app developers to establish a strong brand identity. A fast growing market like the app market makes new brands visible for the consumer at a faster rate. Rating could therefore reduce that uncertainty.

Applications platforms are relatively new but there have been made some studies on the subject. However, the results differ between them regarding if the rating of applications is correlated with frequency of downloads.

In a study about user feedback in App Store it was tested if the rank and the ratings of the applications were independent. The conclusion was that the two were not independent (Pagano & Maalej, 2013). That implies that the ratings of apps could be linked to how often the application is downloaded. Furthermore, Harman et al. (2012) came to the conclusion that customer rating and the number of downloads were strongly correlated. The study has been made regarding how price, downloads and ratings are connected. The sample was taken from Blackberry app store and there was in fact a strong, statistically significant correlation between the ratings and downloads. The result applied both within almost all categories and to the whole app store. The correlation for all categories was 0.79.

Chen and Liu (2011) also researched about this phenomenon. The result in the study was that the customer ratings were not for sure correlated with the highest ranked paid applications. The reason why the effect of ratings is inconsistent across studies could be explained by that the quality of the app is causing an increase in both the rating and downloads. It implies that the quality effect is hidden, and that there is not a correlation between the rating of the app and how frequent it has been downloaded (Dellarocas et al., 2008).

1.3. The importance of icon design

A central part of the graphical interface of the app market is the icon design. An app icon is required for all applications on App Store. It is viewed on the home page and a necessary attribute in order to start the app after it has been downloaded (Cederberg & Sjöström, 2012). Moreover, it gives the first impression and if it is not well designed, both sales and reviews could potentially be affected. The design of the app icon can also

affect how easily it is discovered among all the other apps on the market. If the app icon is interesting enough in combination with the screenshots and the app's name, it can trigger curiosity that makes the user want to see more what the app has to offer. It is also important that it is memorable and unique as it represents the app's brand identity (Wooldridge, 2011).

A common way to decide on an icon is to develop some alternatives and evaluate the usability. It has driven a demand for concrete recommendations regarding this that can spare cost and time. Apple (1996) recommends the icon to be strongly connected to the app's brand and emotionally connect with the people using it. It is furthermore recommended to have a design that is unique, uncluttered, engaging and memorable. Lastly, it is important that the icon looks good in different conditions such as the background chosen for the smartphone. Research from other studies claim that it is essential to consider ambiguity, uniqueness and dominance when designing an icon (Goonetilleke et al., 2001). Apple further recommends that the icon should be understood in many different countries.

1.3.1 Prior studies about icon designs

It has been shown that consumers prefer icons with miniaturized designs of real goods (Hou et al., 2013). There has also been a study that explores how recognition varies depending on how abstract the icon is. Results show that the semi-concrete icons had the highest number of recognition rate. An app that has a too abstract or too concrete icon might therefore not be ideal when trying to maximize the recognition (Kim, 2005). It is more common that the icons are based on illustrations rather than on photography. The reason for that is that photographic icons can look too cluttered in the size of an app icon (Wooldridge, 2011).

A study has been made that describes different types of icons which can include images or text. The abstractness of the icon can be divided into three broad categories. *Representations icons*, with representative images of the object, *Abstract icons* which tries to capture a phenomenon closely related to the object and lastly, *Arbitrary icons* which do not have a clear connection to what it intends to express. The study continues with analyzing how apps that use images compare to those who use text in their icons. Apple has in its book recommended not to use text in the icon to avoid confusion for the user (Apple, 1996). However the recommendations regarding this subject are not

conclusive. Several other studies show that text and graphics used together can be more effective. It has been proven to increase memorability. Moreover, there are some text in languages that are more graphical than others, such as Chinese and Japanese. It is therefore harder to distinguish the effect on users between graphics and text (Goonetilleke et al., 2001).

1.4 Purpose

The area that this study aims to investigate is rather unexplored. Regarding the icon design there have been studies about the abstractness or concreteness of icon designs, while no clear comparison between the usages of images versus letters in the icon has been made. Regarding the ratings, the most similar study did an extensive research about the correlation between ratings and downloads covering all categories in an app store. However, the study only included a sample from the Blackberry World Store and apps with non-zero prices. The price factor can highly affect the result and they recommended further research to be made for other app stores and prices. We have therefore identified a gap in the research.

Our purpose is the following:

To study the influence of ratings and icon design on users' perception of an app.

From this purpose, we hope to be able to derive two findings.

1. If an app developer should focus on improving the ratings in order to get the most positive perception of the app.

2. If an app designer or an app developer should choose an image or a letter in the design of the icon to get the most positive perception of the app.

1.5 Delimitations

We want our results to be universal and relevant for all kinds of app stores. Due to time limit, and an identified gap in prior research, we chose to focus on App Store. However, some implications may be relevant for other app stores with a similar layout.

Moreover, we used a mock app in the study to avoid that consumer's prior relation to the brand would affect the results. In many cases when consumers download an app, they have prior knowledge about it. However, we wanted to exclude the effect of such variables and chose to design the apps by ourselves.

Lastly, we only conducted the experiment in Sweden and the results may be different if the same study is being done in another country.

1.6 Expected contributions

We hope that the conclusions from this report can help app developers to prioritize what to focus on in order to receive the most favorable perception from potential users. Furthermore we hope it can contribute to clarify what impact ratings and icons have on perception as the results from previous studies have been diverse. Moreover we want to open up for further research regarding this booming market based on our conclusions.

2. Theoretical framework

Studies can be found that have analyzed ratings and icons separately. In our theoretical framework we have decided to apply three theories that give different perspectives to our research problems.

2.1 Online word of mouth

The rating of stars on App Store could be interpreted as an online version of word of mouth. There are some people who claim that word of mouth is ”the world’s most effective, yet least understood marketing strategy” (Misner, 1999). Compared to traditional ads, the trustworthiness of the messages from word of mouth is often perceived higher as there are no intentions to sell something (Sun et al., 2006). Another benefit of using online word of mouth as a marketing strategy is that it can manage to overcome consumers’ resistance at a relatively low cost and high speed (Trusov et al., 2009).

By contrast to traditional word of mouth, online word of mouth is opinions that are communicated through written words or symbols (Sun et el., 2006). Another main difference between traditional word of mouth networks compared to the relatively newly established online feedback mechanism is the scale of them (Dellarochas, 2003). The traditional word of mouth is usually limited to a local network, while online word of mouth can reach people all over the world by the accessibility of Internet (Chen & Xie, 2004).

According to Litvin et al. (2007) online word of mouth can be defined as:

“All informal communications directed at consumers through Internet-based technology related to the usage or characteristics of particular goods and services, or their sellers. This includes communication between producers and consumers as well as those between consumers themselves – both integral parts of the word of mouth flow, and both distinctly differentiated from communications through mass media.”

According to Nyilasy (2006) there are four different ways to study word of mouth which is presented in Figure 1.

Unit of analysis	Main focus of study: *Antecedents to word of mouth (causes)*	Main focus of study: *Consequences of word of mouth (effects)*
Receiver of communication (input word of mouth)	**Q1:** Why do people listen?	**Q2:** The power of word of mouth
Communicator (output word of mouth)	**Q3:** What makes people talk?	**Q4:** What happens to the communicator after the word of mouth event?

Figure 1. The different ways to study word of mouth

In this study we chose to focus on the power of word of mouth and hence how consequences of word of mouth are related to the receiver of communication. Prior studies have shown that word of mouth can influence several different aspects beyond behavioral intention and purchase behavior such as perceptions, expectations and attitudes. Word of mouth can have an impact on these conditions during the phase when consumers search for information as well as the evaluation process prior to when the decision is made. It can both influence decision in a positive way and in a negative way. Some studies have shown that negative word of mouth has a stronger effect on consumers than positive word of mouth (Buttle, 1998).

Mudambi (2010) have in a study described how the power of word of mouth depends on whether a particular good is an *experience good* or a *search good.* He further describes the key attributes of *experience goods* as "subjective or difficult to compare, and there is a need to use one's senses to evaluate quality.". The key attributes for *search goods* are "objective and easy to compare, and there is no strong need to use one's senses to evaluate quality.".

Previous studies have also shown that the power of word of mouth is stronger than ads that are paid for. However, less is known why that is the result. In studies about word of

mouth it is common to distinguish between *weak tie sources* and *strong tie sources.* The strength of an interpersonal tie has been defined by Granovetter (1973):

"The strength of a tie is a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie"

Some studies have indicated that word of mouth often is spread through so called *strong tie sources,* which often is family and friends.

2.2 Peirce's theory of signs

Goonetilleke et al. (2001) describes that icon evolved from sign as a concept. Pierce's Theory of Signs attempts to conceptualize what parts a sign consists of. Pierce claim that signs consist of three parts; a sign, an object and an interpretant. Pierce has in his research about semiotics concluded that these three parts in a sign interact in a process he named semiosis (Gatsou et al., 2012).

The Sign

The sign, or icon in this case, is a signifier of the object it attempts to describe. It is therefore just an element of the sign as a whole.

The Object

There are some characteristics of the object that is represented by the sign. The object therefore has conditions regarding its representation to be successfully signified. An object can be represented by many signs.

The Interpretant

The sign can only signify if it is being interpreted by a viewer. Simultaneously, a sign can be interpreted in many different ways. The interpretant can by processing what the signs attempt to describe gain a deeper understanding about the original object. The interpretant therefore has a central role in the content of signs (Atkin, 2005).

Pierce's theory further deepens the concept of signs by dividing it in three categories; icons, index and symbols. Symbols and index have according to Pierce a more abstract

and indirect relation between the sign and the object compared to icons. The icon is the simplest sign among the three categories. An icon should consist of features that physically resemble the object it attempts to describe (Gatsou et a., 2012).

2.3 Dual Coding Theory

To further investigate the relationship between the interpretant and the sign we have chosen Dual Coding Theory (DCT) which is proposed by Allan Paivio (1986). DCT was evolved by imagery variables that were compared to verbal ones to draw conclusions how they impacted learning processes and memorability (Canadian Journal of Psychology, 1991).

Paivio (1986) claims that: "Human cognition is unique in that is has become specialized for dealing simultaneously with language and with nonverbal objects and events.". He continues by stating that "Any represential theory must accommodate this dual functionality.". DCT describes that humans have two cognitive subsystems. One of the systems processes images and the other one processes language. These two systems are interacting but simultaneously also functions independent from one another (Thomas, 2014)

The theory has been applied to many areas such as learning and problem solving. One relevant application that derives from the DCT is the beneficial effect images have on memory compared to words.

The system which processes visual context generally specialize on more concrete information which also includes emotions and sounds. The system which processes verbal information stores more of the relatively abstract information such language. The beneficial effect of the visual system is that it allows activation of connections between the images and words. The impression of a specific image is therefore stored in both of the systems.

2.4 Hypotheses

We want to investigate how the two factors ratings and icon design influence user perception. The most important part we want to emphasize regarding the perception is

the probability to download. In case there would not be any significant results regarding that variable we chose to include some more aspects that are connected to the perception of the app.

Previous studies about ratings in the app market showed a correlation between ratings and downloads. However, the study did not view ratings as a dependent variable which affect the perception of the app. There has been a gap in studies regarding this topic that we want to analyze further.

Based on that information, we have formulated the following hypothesis.

Hypothesis 1: Apps regardless of category with five stars will, compared to apps with one star, have a higher mean in the following variables;

a) Probability to download
b) Likelihood to recommend the app
c) Belief of others' opinion
d) Perception of quality
e) Belief of trustworthiness of app developer
f) Expected usage frequency

Apple has some own guidelines regarding the app icon, which is written on their websites. According to the website it is not uncommon for people to form their initial perception about the application's purpose, quality and reliability only by looking at the app icon. This implies that the app icon can play a significant role regarding forming the first impression about the app. Hence, it could be an important factor in the decision-making process of downloading an app.

If we assume that the icon makes a large impact on the consumers' willingness to download it, it would be interesting to further investigate how to design it in the most appropriate way. It is common that app developers choose to either have a basic, simplified image or text icon. It could either be text expression such as a letter, characters or the name of the brand. According to a study, people's ability to identify images is better than their ability to identify text (Shephard, 1967). This fact in

combination with Apple's guidelines regarding using images in the icon give reasons to believe that other beneficial effects can be derived from using images instead of text in the icon. However, there has been a gap in studies regarding this topic that we want to analyze further.

We have based on that information formulated the following hypothesis:

Hypothesis 2: Apps regardless of category with an image icon will, compared to apps with a letter icon, have a higher mean in the following variables;

g) Probability to download
h) Likelihood to recommend the app
i) Belief of others' opinion
j) Perception of quality
k) Belief of trustworthiness of app developer
l) Expected usage frequency

Apart from probability to download we want to analyze how triggered they feel to use word of mouth and how they perceive others' opinions about the app. The quality variable and trustworthiness of the app developer were chosen to analyze how authentic and reliable the app seemed. The last variable is another indicator of the perception of the app and to what extent the user thinks the app will be used in the future.

Our assumptions

1. A higher mean of the dependent variables for one group compared to another signifies a more positive perception of the app.

2. A lower mean of the dependent variables for one group compared to another signifies a more negative perception of the app.

3. Methodology

3.1 Approach

We chose a deductive approach to our study. This means that we began by exploring a theory which we found compelling and then we tested its complications with data. We moved from a general to a more specific level, starting with exploring what others had done and then tested our own hypotheses (Blackstone, 2012).

3.2 Experiment design

The method chosen for the experiment is a quantitative one. A quantitative method proved to be the most appropriate since our hypotheses were narrow, our description focused and our type of research conclusive. We also wanted to make sure that our results were representable for the whole population which made a quantitative method suitable. (Cohen et al., 2000)

In the research, we have tried to avoid the existence of extraneous variables. The questionnaires for the groups are identical except for the independent variables we wanted to examine. The different independent variables were randomly assigned to our respondents to further eliminate the risk (Malhotra & Birks, 2006).

3.3 Manipulation of independent variables

We wanted to explore what impact rating and icon design have on potential customers of an app. We designed apps with either five stars or one star which is the highest and lowest amount of stars, respectively. We designed equally as many apps with either a picture describing the app activity (for example cutlery if it was a restaurant app) or an icon with the first letter of the activity (for example an R if it was a restaurant app). We chose to do the experiment with two different app categories; one exercise app and one restaurant app.

We took a screen shot of a list for the specific category in App Store and replaced the first app with the new mock app we had designed. We blurred the screen shot except for the spot where we had placed the new app. We did this to prevent that the respondent would compare or in other ways be influenced by the other apps.

3.4 Design of questionnaire

We printed the pictures of the screen shot with the app and the interview questions. The respondent was supposed to take a look at the app on the picture and thereafter answer some questions. The question part was divided into two subparts. The first part consisted of six questions regarding the respondent's opinion about the app. These six questions are also our dependent variables.

Variables	Questions
a)	How likely is it that you would download the app?
b)	How likely is it that you would recommend the app?
c)	What do you think other people think of the app?
d)	How do you perceive the quality of the app?
e)	What do you think of the trustworthiness of the app manufacturer?
f)	How often do you think you would use the app?

Table 1: Dependent variables

The second part consisted of demographical questions. We asked these to see if our sample was representative to the population as a whole.

Variable	Question
a)	What is your gender?
b)	What is your age?
c)	In which city do you live?

Table 2: Demographical questions

We gave the questionnaire to people at the Stockholm Central Station. The people receiving our questionnaire were sitting down and we chose not to give it to anyone standing up. We thought they might be in a hurry and therefore not pay as much

attention to the questions as necessary. The questionnaire took approximately three minutes to complete and the respondent was being left alone when completing it.

3.5 Design of questions

All of the questions measuring our dependent variables had a scale ranging from 1 to 7 where 1 was the least positive, e.g. “very low”, and 7 was the most positive, e.g. “very high”. In the part covering the demographical questions the respondent just stated their age, gender and city and there were no intervals.

In all of the questions, the respondent had the choice to answer “I don’t know” in which the respondent wrote “0”.

3.6 Variables

3.6.1 Dependent variables

We had six different dependent variables which are stated below.

Download probability

We asked the respondents how likely it is that they would download the app.

Likelihood to recommend the app

We asked the respondents how likely it is that they would recommend the app to others.

Other people’s opinion

We asked the respondents what they believe other people thought of the app.

Perceived quality

We asked the respondents how they perceived the quality of the app.

Trustworthiness

We asked the respondents how trustworthy they perceived the app company.

Estimated frequency of usage

We asked the respondents to estimate how often they would use the app.

3.6.2 Independent variables

We had three different independent variables which are stated below.

Rating

We showed the respondent an app that had either five stars or one star.

Icon design

We showed the respondent an app that was either a picture icon or a letter icon.

App category

We showed the respondent an app that was either an exercise app or a restaurant app.

In total, we showed eight different app versions.

3.7 Collection of data

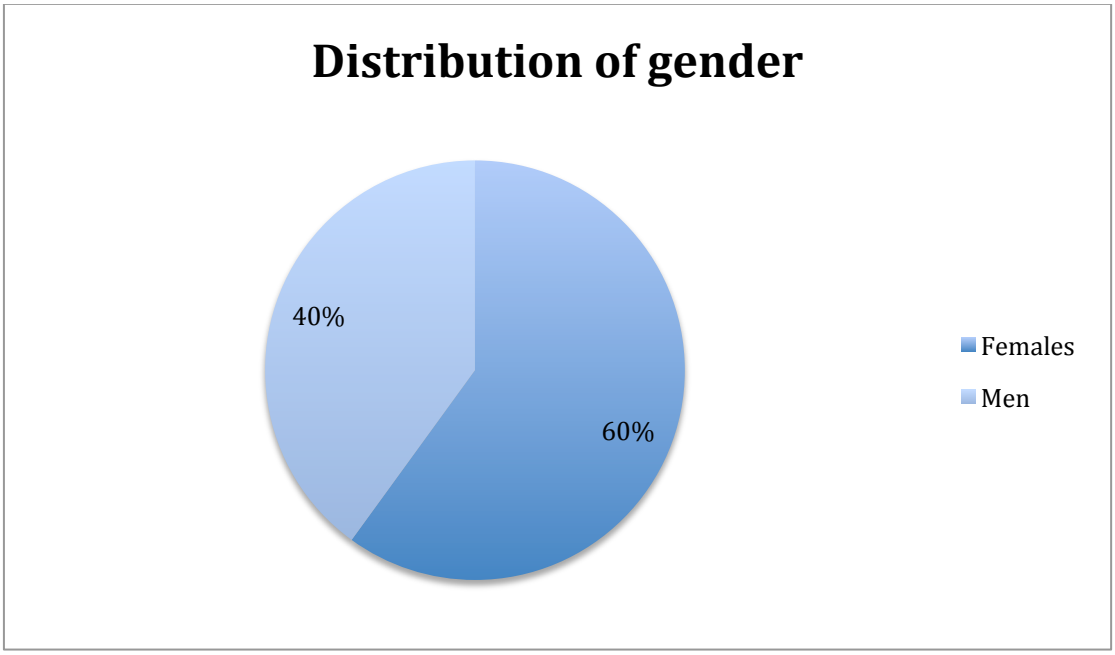
3.7.1 Sample

Since we did not have a specific target group for the apps we showed, we wanted a sample that was representable for the Swedish population. We choose not to do an online questionnaire since we wanted to avoid having a sample with people too similar to us regarding different demographical variables. We decided to give the questionnaire to people at the Stockholm Central Station. People at train stations are often from many different cities. We therefore hoped that this would maximize our effort of achieving a representable sample.

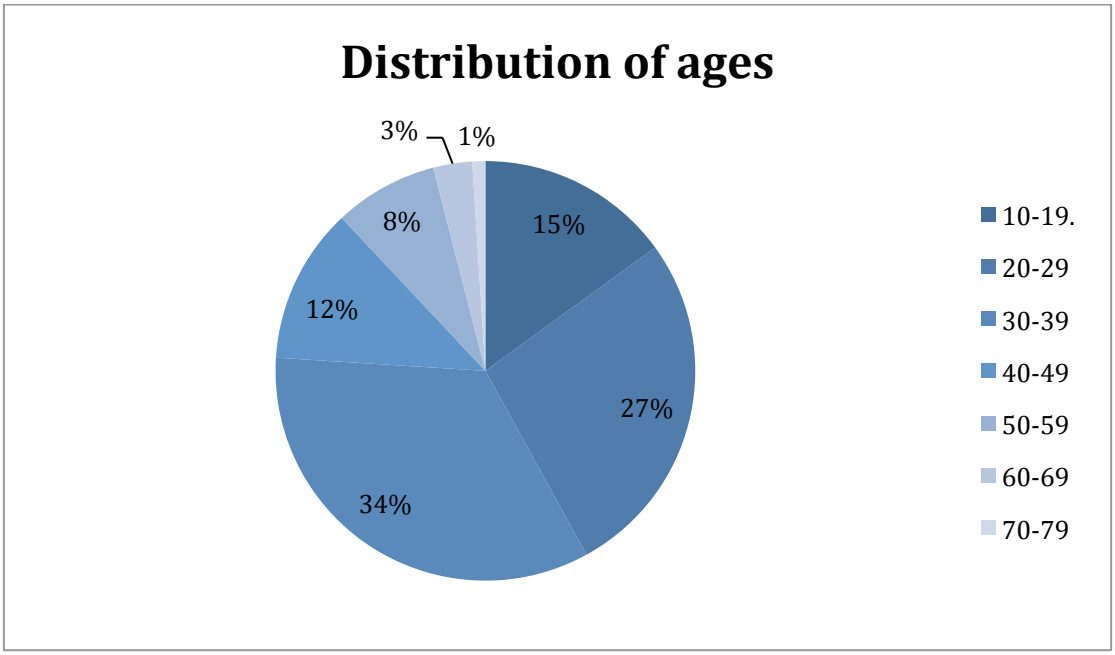
We needed at least 30 respondents per group in order for our data to be significant with t-tests and ANOVA tests. We aimed at having at least 60 respondents in each of our eight groups since we wanted to compensate for losses of people who answered “I don’t know” or who did not answer the questions at all.

3.7.2 Experiment execution

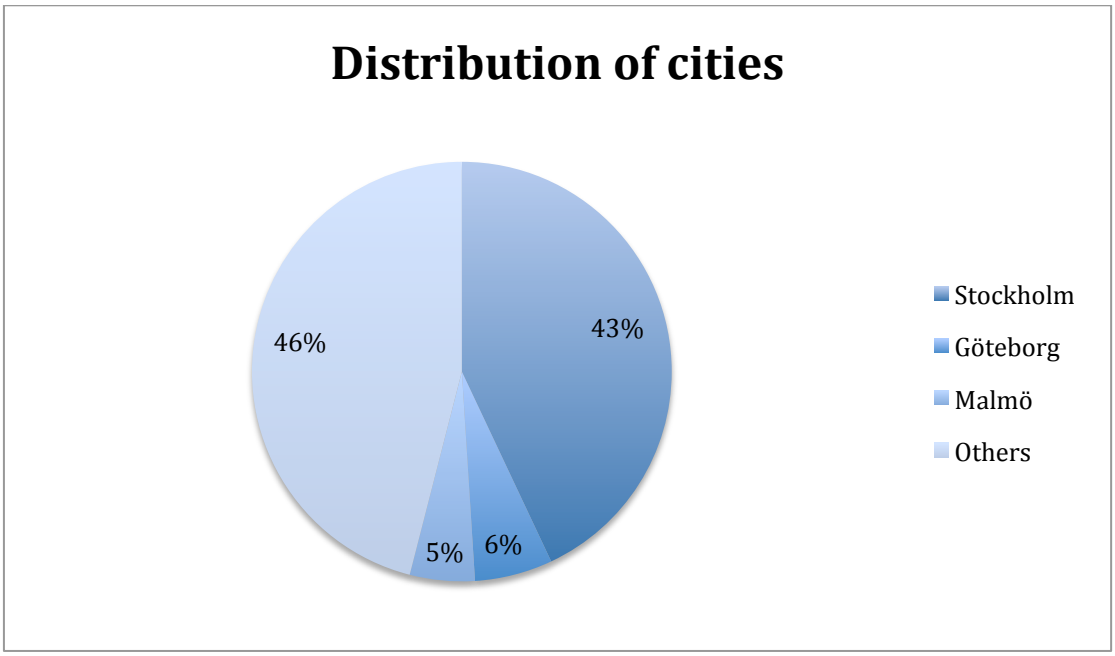
The interviews took place the days between the 10th and the 15th of April. In total, 541 people answered our questionnaire which means that there were almost 70 people in each of our eight groups. Approximately 40 % were men and 60 % were women.



The mean age was 33 years old with a span from 13 to 72 years of age. The majority of our respondents, approximately 60 %, were in the ages 20-39 years of age.



The majority, 54 %, of our respondents lived in one of the three cities Stockholm (43 %), Göteborg (6 %) and Malmö (5 %). The rest lived in other cities in Sweden. We did not interview people living outside of Sweden.



3.8 Quality of data

3.8.1 Validity

Validity as a term is used to describe to what extent a report is measuring what it is supposed to measure (Malhotra & Birks, 2006). We wanted to avoid random sampling errors and wanted our sample to represent the Swedish population. We therefore only interviewed people at the Stockholm Central Station and gave the questionnaire to a relatively high number of respondents. We also checked for different kinds of validity which are explained below.

Firstly, internal validity concerns if the independent variables are the only ones giving the results (Malhotra & Birks, 2006). When designing the pictures with the mock apps, we made sure to blur the rest of the picture so that the mock app was the only thing visible. We only changed the variables we wanted to study. For example, in the two exercise apps with different numbers of stars, we only changed five stars to one star. The

rest of the picture was intact. We therefore believe that the internal validity in our test is high.

Another type of validity we have evaluated is external validity. If a study has a high external validity, the results can be used to draw conclusions on a larger population (Malhotra & Birks, 2006). In order to increase our external validity we have asked people on several occasions. The people we asked were random people in the Stockholm Central Station which we believe gives us a relatively representable sample. What could be improved is the sample size and to ask people in different cities. In conclusion, we believe the external validity in our experiment is medium.

Lastly, we wanted to examine the study's ecological validity. If the results are applicable to real life situations, the report has a high ecological validity (Bryman & Bell, 2011). This means that the methods, materials and setting must be similar to the situation which is examined. In our study, there are two factors that could have affected the ecological validity. Firstly, our respondents did not view the app in its natural environment, i.e. an app store. Secondly, the apps we showed do not exist and customers often have previous knowledge about a brand when evaluating it. In conclusion, we believe our ecological validity is relatively low.

3.8.2 Reliability

Reliability implies whether the experiment could be done again and give the same results. If the results from the assessment tool are consistent, the test has a high reliability (Manion & Morrison, 2000). One way to improve reliability is to do the same test on a different occasion. This is called test-retest reliability (Phelan & Wren 2005-06). Unfortunately, we did not have time to do this. However, to increase the reliability we interviewed people on different days.

To avoid different interpretations by different respondents we did not use terms such as "often" or "sometimes". If people interpret these words differently, it is hard to achieve objectivity (Söderlund, 2005). We also avoided questions that could be perceived as arbitrary and we only asked closed ended questions. The only exception to this was the questions where the respondent typed their age, gender and city.

4. Results

4.1 Control questions

In the first four groups, the only variable changing was the number of stars. We did this with two different apps, one exercise app and one restaurant app. Our goal was to see if the number of stars correlated with the mean of our dependent variables. To start with, we asked the respondent if he or she thought five versus one star was high or low. The result showed that our respondents thought five stars on average were high and one star low. If we did not obtain these results, we would not be sure if the potentially different means from the apps were because of the number of stars or some other factors.

4.2 Overview of results

We did a Post Hoc test in ANOVA to compare the means of the eight groups. The results showed that the respondents preferred picture icons compared to letter icons, regardless of app category. However, the number of stars does not seem to affect the dependent variables as much. In some dependent variables and in some app categories, the mean is higher for a five star app and in some cases it is higher for a one star app. All our results are significant (<0.05).

Since we wanted to investigate how the two chosen variables affect consumer perception regardless of app category, we merged some groups. In this way, a potential pattern regarding rating and icon could be revealed, regardless of app category. We started with exploring this for the apps with different amount of stars. Two different t-tests were made. In the first one, all six dependent variables were put in and compared over the independent variable “Rating”. The results showed a higher mean for all dependent variables in the five star groups. However, the results were not significant in any of the groups, which may imply that the differences were too small.

The same tests were conducted with the icon groups. Again, two different t-tests were made. In the first one, all six dependent variables were used and then compared over the independent variable “Icon design”. The results showed a higher mean for all the dependent variables in the picture icon group. All the results were significant.

Dependent variable	1	2	Significance
a) Download probability	3.935	3.592	0.145
b) Likelihood to recommend	3.569	3.375	0.404
c) Belief of other people's opinion	4.294	4.184	0.597
d) Perceived quality	3.248	3.191	0.828
e) Trustworthiness of app company	4.381	3.970	0.059
f) Expected usage frequency	3.595	3.336	0.304

Table 3: Means of the independent variable "Rating".

The same tests were conducted with the icon groups. Again, two different t-tests were made. In the first one, all six dependent variables were used and then compared over the independent variable "Icon". The results showed a higher mean for all the dependent variables in the picture icon group. All the results were significant.

Dependent variable	1	2	Significance
a) Download probability	5.195	3.322	0.000
b) Likelihood to recommend	5.373	3.254	0.000
c) Belief of other people's opinion	5.517	3.669	0.000
d) Perceived quality	5.229	3.475	0.000
e) Trustworthiness of app company	5.453	3.583	0.000
f) Expected usage frequency	5.237	0.000	0.000

Table 4: Means of the independent variable "Icon".

4.3 Results in relation to hypotheses

As stated earlier, we had two different hypotheses and each hypothesis had six sub hypotheses. We now explore if the results support these.

4.3.1 Probability to download

Five star apps had a higher mean compared to one star apps. However, the result was not significant. Picture icon apps had a higher mean than letter icon apps. The result was significant. Thus, hypothesis 1 a) is not supported while hypothesis 2 a) is supported.

4.3.2 Probability to recommend

Five star apps had a higher mean compared to one star apps. However, the result was not significant. Picture icon apps had a higher mean than letter icon apps. The result was significant. Thus, hypothesis 1 b) is not supported while hypothesis 2 b) is supported.

4.3.3 Other's opinion of the app

Five star apps had a higher mean compared to one star apps. However, the result was not significant. Picture icon apps had a higher mean than letter icon apps. The result was significant. Thus, hypothesis 1 c) is not supported while hypothesis 2 c) is supported.

4.3.4 Quality perception

Five star apps had a higher mean compared to one star apps. However, the result was not significant. Picture icon apps had a higher mean than letter icon apps. The result was significant. Thus, hypothesis 1 d) is not supported while hypothesis 2 d) is supported.

4.3.5 Trustworthiness of app manufacturer

Five star apps had a higher mean compared to one star apps. However, the result was not significant. Picture icon apps had a higher mean than letter icon apps. The result was significant. Thus, hypothesis 1 e) is not supported while hypothesis 2 e) is supported.

4.3.6 Estimated usage frequency

Five star apps had a higher mean compared to one star apps. However, the result was not significant. Picture icon apps had a higher mean than letter icon apps. The result was significant. Thus, hypothesis 1f) is not supported while hypothesis 2f) is supported.

5. Discussion and implications

5.1 Discussion of results

The intention with this study was to explore what kind of factors within App Store that affect a consumer's perception of an app. We found that the most important factor that increases these variables is the design of the icon. When comparing ratings, neither of the means of the variables were significant if we neglected app category. While our results support the hypothesis that picture icon is superior to a letter icon, the results do not support the hypothesis that a higher rating is superior to a lower one.

5.1.1 Discussion about first hypothesis

The insignificant result of the rating may appear surprising and against intuition. The paper by Harman et al. (2012) showed that there exists a positive correlation between ratings and downloads and similar results were expected. There are several reasons that could explain why a higher rating did not correlate with a higher perception in our study.

Firstly, the aforementioned paper did solely their research on apps that the user pays for. Even though we did not state in our survey if the app was for free or not, the majority of the apps in App Store is for free (Statista, 2015). Moreover, the apps that do charge a fee often state that clearly. We therefore believe that our respondents thought our app was for free of charge. When an app is free, the user may perceive rating of less importance since the risk of trying that good is low. If the user would not be satisfied with the app, he or she could simply delete it from their phone. The rating of the app may therefore not have affected the respondent's willingness to try it. Other factors that connect the user to emotionally wanting to try the app might be more relevant with this reasoning. The perception of the app as an experience good may explain why this is the case.

As Mudambi & Schuff (2010) showed, rating is of different importance depending on whether the purpose of the app is experience or search. If an app belongs to the former category, the consumer will have a harder time judging the quality of the app without having tested it. Search goods on the other hand are easily evaluated in terms of quality by looking at factors such as rating. It is possible that our respondents found our two

app categories to belong to the experience category more than they belonged to the search category. The user may therefore have felt that they have to try it to form an opinion of the app. Observing the result, a large number of people answering "I don't know" on the question regarding the quality of the app can be identified. Since quality was hard to determine for the respondent, the apps shown may belong more to the experience category instead of the search category. Furthermore, an app can be viewed as a rather subjective good which also confirms that it might be categorized as an experience good. The theory of word of mouth states that different goods require different information that is provided to the consumers. Our findings indicate that perception of apps which potentially are categorized as experience good might not be affected of ratings. However, perception of apps that are categorized as search goods might be influenced to a higher extent. An example of an app of that kind could be a mobile bank application.

As mentioned in the part about word of mouth, the source of information can be divided into strong ties and weak ties. Since rating in App Store can be affected by anyone, the source could be interpreted as weak ties. Rating did not seem to have any impact on the consumer's perception and therefore it may be the case that weak tie sources have less power on the consumer when he or she considers downloading an app. These findings indicate that an app company might want to focus on making the app viral by spreading it through strong tie sources instead of weak tie sources. An example of that would be to get users to recommend the app to family and friends either in an online or an offline context.

Lastly, the app displayed was on top of a list that featured several apps. A possible explanation to why rating did not show to be important to the consumer's perception may be because the consumer saw the top placement as a substitute for rating. If the app was without context, the respondents may have felt a need to depend more on rating.

5.1.2 Discussion about second hypothesis

The support of our second hypothesis, that a picture icon is superior to a letter icon, was in line with prior research. The images on the icons can be characterized as miniaturized designs of real goods which previously has been proven to be preferred to text. Furthermore we can draw the conclusion that users can experience difficulties in interpreting icons with abstract signifying elements such as the letter on the verbal icon.

This is in line with a previous study about the importance of mobile interface icons (Gatsou et al., 2012). Additional to confirming the importance of the images, the experiment successfully managed to distinguish how the dependent variables were related to images and letters respectively. The experiment clearly compared verbal to visual elements in the app icon with significant results, which have not been done before under the same circumstances.

The results are also in line with the predictions by using Peirce's Theory of Signs. As mentioned, the icon is the simplest of the three categories of a sign. The icon should attempt to depict the object. (Gatsou et al., 2012) The icons in our research all depicted the activity undertaken in the app which is in line with the theory. The process of interpreting the icon is often subconscious but according to Pierce's Theory of Signs a vital part of a sign. This leads to the conclusion that in order for an app company to get a positive perception of their app, it can help to use images that evoke the right feelings. The users might not be aware of these consequences themselves but by analyzing our results it can make a significant difference to design the icon in a certain way. The findings should remind the app companies that the design influence the user's perception to a relatively large extent.

As the Dual Coding Theory explains, pictures are processed in a way in our brain that increases memorability (Canadian Journal of Psychology, 1991). The pictures of the activity were therefore probably easier for people to remember. The visual system processed the information that was provided from the icon and created associations in the mind. Connections could be drawn between the images and verbal information to store in different systems. The results indicated that the icon was processed in two different ways according to the Dual Coding Theory. The visual system might have processed the information so a higher recognition rate was achieved, leading to a more favorable perception of the icon that contained an image.

5.2 Practical implications

An increasing number of companies realize the importance of having their services available through an app. The results in this paper are of interest to any company that releases an app on App Store. Our report highlights the importance for these companies to put effort on factors that could determine whether an app becomes successful or not.

The result show that the icon of an app plays a big role on the consumer's perception while rating has no significant impact. Companies should therefore choose a picture icon rather than a letter icon while we find no evidence that rating is of importance for the consumer's perception.

We do want to stress however that factors within an app store are certainly not the only ones that will decide whether one app will gain more downloads than another. In many cases, the customers have heard about an app through advertisements or word of month and thereafter search for the app. The importance of building a strong brand and an app that people want to use will most probably increase the chances of creating a successful app.

5.3 Limitations and sources of error

The different versions of the app were made in two categories. The expansion to two categories was made in an attempt to generalize our results. Choosing only one category, one can not be sure if the results were specific for that category. However, if we had more time we would have done the experiment with more than two categories.

Furthermore, when people download an app, they do it through the context of a smartphone where you can browse through demo pictures and obtain more information through reading texts describing the app. You can also find other people's opinions where they write short texts about what they think of the app. When people were to answer our survey, they only saw a screenshot from a specific page in App Store and they saw it on a physical paper. These are not normal circumstances when downloading an app and it may therefore have been hard for the respondents to form an opinion about the app. If we had the resources, we could have giving our respondents a smartphone through which they could view the app in the questionnaire.

A source of error can be the small and perhaps non-representable sample of people receiving the inquiry. As seen earlier, there was an overrepresentation of females and of people living in big cities.

Another possible source of error was due to the fact that we chose to show the app in context of a list with other blurred apps. As discussed previously, the appearance of such

a list may perhaps have been more important than the actual rating. If we only showed the app without the being placed on a list, rating may have been more relevant.

5.4 Further research

As mentioned as a limitation, it would be interesting to expand the experiment and include more app categories to see if the results can be generalized for all apps. What also could be explored further, is if some variables in App Store affect the app categories in different ways. It would for example be interesting to see if the results are different between hedonic and non-hedonic apps. To do this, the respondents could by themselves state how hedonic they perceive the apps.

It would be interesting to see if the results would be the same if the respondents answered the questionnaire after viewing the app in its natural setting, i.e in App Store. The unnatural setting probably made it hard for respondents to evaluate some of the variables. The researcher who does this must be aware of though that it is hard to eliminate other non-relevant variables in App Store which may decrease internal validity. Moreover, further research is encouraged to investigate if these results are valid for all kinds of app stores and not only within App Store.

6. References

6.1 Literature

Apple Computer Inc. 1996, Newtons 2.0 User Interface Guidelines. Reading, MA: AddisonWesley

Atkin, A. 2005, Peirce On The Index and Indexical Reference. *Transactions of The Charles S. Peirce Society*, pp. 161–188

Buttle, F. A. 1998, Word of mouth: Understanding and managing referral marketing. J*ournal of Strategic Marketing*, 6, pp. 241–254

Canadian Journal of Psychology 1991, Dual Coding Theory: Retrospect and Current Status 45(3), pp. 255-287

Cederberg, J. & Sjöström, J. 2012, Uppgångsappen - En rapport om utvecklingen av en mobilapplikation. Södertörns Högskola

Chen, L. & Liu, X. 2011, Predicting popularity of online distributed applications: iTunes app store case analysis

Chen, P., Wu, S., & Yoon, J. 2004, The Impact of Online Recommendations and Consumer Feedback on Sales. Proceedings of the 25th International Conference on Information Systems

Chen, Y. & Xie, J. 2004, Online consumer review: A new element of marketing communication mix. Working paper, University of Arizona, Tucson

Chevalier, J.A. & Mayzlin, D. 2006, The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research* (JMR) (43:3), pp. 345-354

Cohen, L., Manion, L. & Morrison, K. 2000, *Research Methods in Education*, 5th ed. Routledge Falmer

Dellarocas, C., Xiaoquan Z., & Neveen, A. 2008, Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures, *Journal of Interactive Marketing*

Dellarocas, C. 2003, The Digitalization of Word-of-Mouth: Promise and Challenges of Online Reputation Systems. Boston University

Duan, W., Gu, B. & Whinston, A.B. 2008, Do Online Reviews Matter? - an Empirical Investigation of Panel Data, *Decision Support Systems*

Gatsou, C., Polotos, Anastasios, P. & Zevgolis, D. 2012, The Importance Of Mobile Interface Icons On User Interaction. *International Journal of Computer Science and Applications.* pp. 92-107

Goonetilleke R. S., Martins Shih H., Kai on H., & Fritsch J. 2001, Effects of training and representational characteristics in icon design, *International Journal of Human-Computer Studies*, vol. 55, no. 5, pp. 741–760

Granovetter, M. S. 1973, "The Strength of Weak Ties", *American Journal of Sociology*, Vol.78, May, 1360-1380

Harman, M., Jia, Y. & Zhang, Y. 2012, App store mining and analysis: MSR for app stores. Proceedings of the 9th Working Conference on Mining Software Repositories (MSR '12), Zurich, Switzerland

Kai-Chun H. & Chun-Heng H. 2013, Aug. A Preliminary Study on Aesthetic of Apps Icon Design. *The 5th International Congress of International Association of Societies of Design Research* (IASDR 2013)

Khalid, H. 2014, On The Link Between Mobile App Quality And User Reviews. Thesis (Master, Computing), Queen's University

Khalid, H., Mcllroy, S., Ali, N. & Hassan, E. A. 2015, Analyzing and automatically labeling the types of user issues that are raised in mobile app reviews. School of Computing, Queen's University

Litvin, S. W., Goldsmith, R. E. & Pan, B. 2008. Electronic word-of-mouth in hospitality and tourism management Tourism Management, pp. 458–468

Malhotra, N.K. & Birks, D.F. 2006, Marketing Research - An Applied Approach, 2nd European ed. 2 England: Prentice Hall

McIlroy, S., Ali, N. & Hassan, A. 2015, Analyzing and Automatically Labeling The Types of User Issues that are Raised in Mobile App Reviews. *Journal of Empirical Software Engineering,* pp. 2

Misner, I. 1999, The World's Best Known Marketing Secret: Building Your Business with Word-of-mouth. Bard Press

Mudambi, S. M. & Schuff, D. 2010, What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.Com. *MIS Quarterly* (34:1), pp. 185-200

Nyilasy, G. 2006, Word of mouth: what we really know - and what we don't. University of Georgia

Pagano, D. & Maalej, W. 2013, User feedback in the app store: An empirical study. Requirements Engineering Conference (RE), 2013 21st IEEE International, July 2013, pp. 125–134

Paivio, A. 1986, Mental Representations. New York: Oxford University Press

Shepard R. N. 1967, Recognition memory for words, sentences and pictures. *Journal of Verbal Learning and Verbal Behaviour*

Sun, T., Youn, S., Wu, G. & Kuntaraporn, M. 2006, Online Word-of-Mouth (or Mouse): An Exploration of its Antecedents and Consequences. *Journal of Computer-Mediated Communication*

Söderlund, M. (2005), Mätningar och Mått i Marknadsundersökarens Värld, Liber Ekonomi

Thomas, N. 2014, Dual Coding and Common Coding Theories of Memory. Stanford Encyclopedia of Philosophy

Trusov, M., Bucklin R. & Pauwels, K. 2009, Effects of word-of-mouth versus traditional marketing: Findings from an Internet social networking site. *Journal of Marketing*

Wooldridge D. and Schneider, M. 2011, Ch4 Your iOS App Is Your Most Powerful Marketing Tool / The Business of iPhone and iPad App Development . Berkeley, CA: Apress, pp. 61–96.

6.2 Electronic sources

Adjust, 2014, Birth, life and death of an app. [Online] Available: https://www.adjust.com/assets/downloads/AppleAppStore_Report2014.pdf (May 10, 2015)

Blackstone, A. 2012 ,Principles of sociological inquiry: Qualitative and Quantitative Methods.
[Online] Available:
http://catalog.flatworldknowledge.com/bookhub/reader/3585?e=blackstone_1.0-ch02_s03 (May 16, 2015)

Explorable, Quantitative and Qualitative Research, [Online] Available: https://explorable.com/quantitative-and-qualitative-research (May 16, 2015)

Phelan C., Wren J., Graduate Assistans, UNI Office of Academic Assessment (2005-06) [Online] Available: https://www.uni.edu/chfasoa/reliabilityandvalidity.htm (May 16, 2015)

Statista, 2015, Number of apps available in leading app stores,
[Online] Available: http://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/ (May 15, 2015)

Number of apps downloads from the apple app store.
[Online] Available: http://www.statista.com/statistics/263794/number-of-downloads-from-the-apple-app-store/
(May 5, 2015)

Number of free and paid mobile app store downloads worldwide from 2011 to 2017 (in billions).
[Online] Available: http://www.statista.com/statistics/271644/worldwide-free-and-paid-mobile-app-store-downloads/ (May 17, 2015)

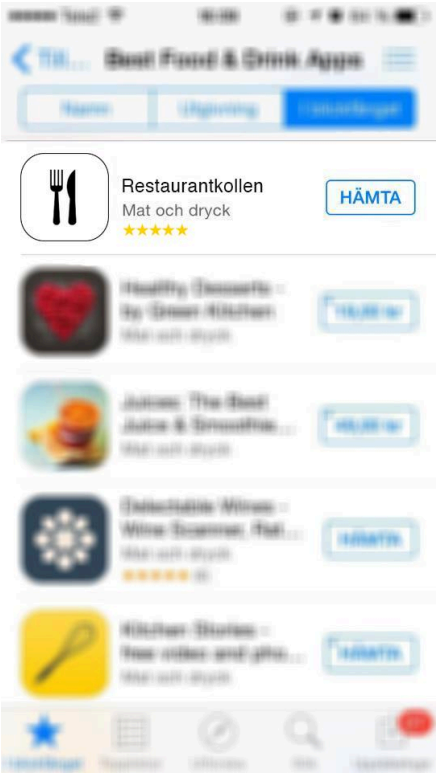
7. Appendix

Below are the different versions of the apps we showed to our respondents. Note that all other apps in the same screen shot are blurred in order for the respondent to focus solely on the app we are analyzing.

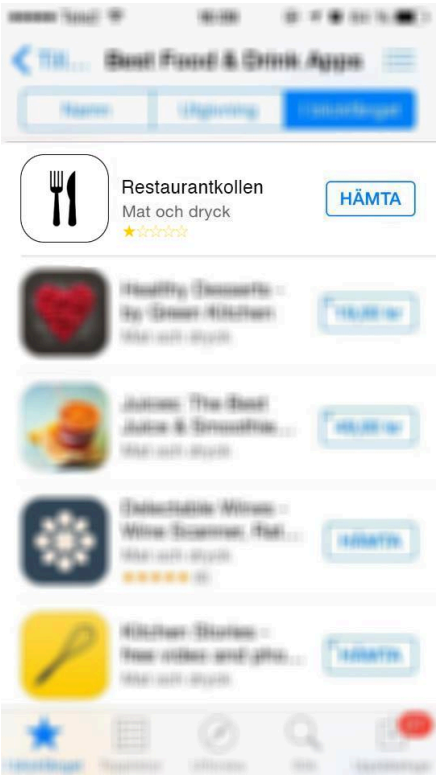
7.1 Five star exercise app

7.2 One star exercise app

7.3 Five star restaurant app



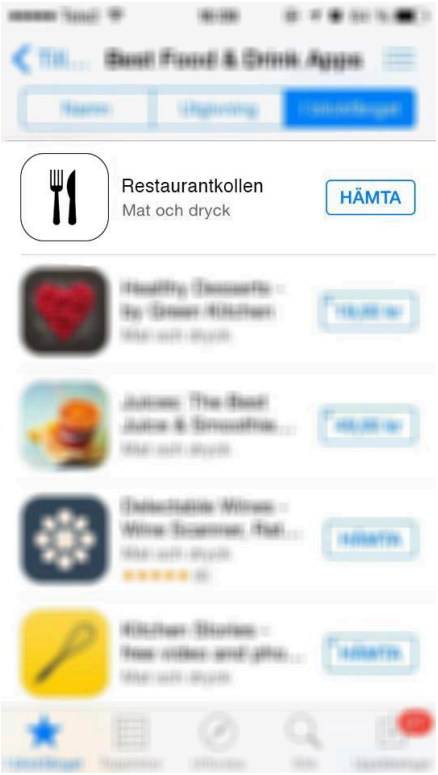
7.4 One star restaurant app



7.5 Picture icon exercise app

7.6 Letter icon exercise app

7.7 Picture icon restaurant app



7.8 Letter icon restaurant app