

New directions: Exploring Google Play mobile app user feedback in terms of perceived ease of use and perceived usefulness

by

Dallas Hermanson
Master of Education, University of Victoria, 2014

A Project Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF EDUCATION

in the Education faculty, department of Curriculum and Instruction

© Dallas Hermanson, 2014
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

Supervisory Committee

New directions: Exploring Google Play mobile app user feedback in terms of perceived ease of use and perceived usefulness

by

Dallas Hermanson
Master of Education, University of Victoria, 2014

Supervisory Committee

Dr. Jason Price, (Department of Curriculum and Instruction)
Supervisor

Dr. James Nahachewsky, (Department of Curriculum and Instruction)
Co-Supervisor

Abstract

Supervisory Committee

Dr. Jason Price, (Department of Curriculum and Instruction)

Supervisor

Dr. James Nahachewsky, (Department of Curriculum and Instruction)

Co-Supervisor

The primary aim of this research project is to describe the presence of two highly validated technology acceptance constructs within the user-generated feedback text found on Google Play's mobile application website. Though this is an exploratory study, I hypothesized that the two constructs perceived ease of use and perceived usefulness would not be prevalent in the user reviews. After collecting and coding the data I found that only 3% of 13 099 reviews collected from Google Play contained information regarding perceived usefulness and less than 1% of reviews discussed perceived ease of use. I believe knowing this presence will help give teachers, researchers, and Google developers a new insight into Google Play's app store feedback mechanisms. In all, I would like to see these mechanisms improved in order to bring meaningful information to teachers adopting mobile technologies.

Table of Contents

Supervisory Committee	ii
Abstract	iii
Table of Contents	iv
List of Tables	v
List of Figures	vi
Dedication	vii
Introduction	1
The changing classroom	2
Potential benefits of mobile learning environments	3
Mobile applications	4
Identifying applications	5
Project Aims	7
Literature Review	8
Biases on the Google Play app store	8
Star Ratings	10
Written comments and feedback forums	12
“Newest” and “Rating” filtering	14
“Helpfulness” filtering	15
The influence of word-of-mouth	16
An alternative method	19
The Technology Acceptance Model developed by Davis (1989)	19
Perceived ease of use and perceived usefulness	20
External validity	22
Limitations	23
Data collection	23
Researcher limitations	25
Assumptions	26
Sample representation	26
Face validity	26
Methodology	28
Quantitative Approach	28
Qualitative approach	30
The dataset	31
Data analysis	33
Applying perceived usefulness to reviews	35
Applying perceived ease of use to reviews	37
Results	39
Conclusions and Discussion	41
References	44

List of Tables

Table 1 - Word lists used to isolate user feedback.....	35
Table 2 - Examples of feedback labelled as containing the construct perceived usefulness	Error! Bookmark not defined.
Table 3 - Examples of feedback labelled as containing the construct perceived ease of use	39

List of Figures

Figure 1 - A typical section of a mobile application's download page on Google Play website	9
Figure 2 - Users can see more reviews by clicking on a review or the arrow	9
Figure 3 - Reviews can be sorted by "Newest", "Rating", and "Helpfulness"	10
Figure 4 - Technology Acceptance Model (Davis, 1989).....	20
Figure 5 - Formula used to determine my sample size (n)	35
Figure 6 - The original items for perceived usefulness on the TAM instrument proposed by Davis (1989).....	34
Figure 7 - The original items for perceived ease of use on the TAM instrument proposed by Davis (1989).....	34

Dedication

I would like to acknowledge several people for helping me along my academic path and for making this project possible. First and foremost, my mother, Nancy Hermanson always supported me and helped me to get passed stress and strife during critical moments of my academic career. Secondly, my partner, Jennifer Muranetz is a true inspiration and fills my heart full of love and hope during times of despair. Third, I wish to thank Jason Price, my graduate supervisor who continues to guide me on matters of importance inside and outside of school and James Nahachewsky, who has never faltered in lending a critical mind and a kind word. Fourth, I wish to recognize Marcello Lins for his freely offered coding lessons and constant technical support. Finally, I would like my family and friends to know that they are an irreplaceable safety net that relaxes and relieves the pains of the academic system.

Introduction

Researchers estimate that by 2015, 80% of users accessing the Internet will connect through mobile devices (Johnson, Smith, Levine, Willis, Levine, & Haywood, 2011). In formal educational contexts, Norris, Hossian, and Soloway, (2011), have projected that by 2016 “*every child in every grade in every K-12 classroom in America will be using a mobile learning device*” (p. 25). As mobile technologies enter the classroom, more expectations will be placed on teachers by various stakeholders to utilize these technologies to adapt their pedagogical practices. One emerging resource for teachers to turn towards is the Google Play android application store. This online market place provides free and paid access for mobile users to over a million mobile applications; also referred to as “mobile apps”. Previous research has shown that utilizing an effective application in the classroom setting can provide numerous teaching and learning benefits to both educators and students (Chiong & Shuler, 2010; Goodwin, 2012; Ho, Hung, & Chen, 2012; Huang, Kuo, Lin, & Cheng, 2008; Lai, Yang, Chen, Ho, and Chan, 2007; Lehner & Nosekabel, 2002; Mac Callum & Jeffrey, 2010).

However, there are some barriers to finding appropriate apps on the Google Play mobile application market. Within this online market several sorting biases have been identified that have the potential to influence user assessment of applications (Ganu, Elhaded, & Marian, 2009; Hu, Pavlou, & Zhang, 2006; Wu & Huberman, 2010). Also, new applications are being created and presented to users at an alarming rate, around 1000 per day (AppBrain, 2014). In addition to this rapid mobile application growth, user generated ratings and reviews are also being created rapidly. To give educators a chance at making sense of this information, new feedback mechanisms may need to be created that bring useful information to teachers.

This project examines an alternative method of sorting through the vast amount of user-generated feedback. In this study, I have aggregated 13 099 user reviews from 1 283 “educational” mobile applications and have analyzed their content to determine the presence of sentiments around (1) perceived ease of use and (2) perceived usefulness, both of which have been previously described as significant constructs regarding the adoption of new technologies by educators in the classroom and other domains (Chuttur, 2009; Davis, 1989). Following this section, I describe the state of the changing classroom in British Columbia, Canada, how mobile applications can enhance the learning environment, how mobile applications fit within these environments, and the major barriers to finding these applications. I then detail how I looked at user reviews on Google Play and discuss the implications for future development of major mobile application distribution websites.

The changing classroom

In British Columbia, Canada, the Ministry of Education (2013) recognizes that teaching and learning across the province are changing. In the official education plan it states:

B.C. leads the country on internet connectivity – 85% of British Columbians use the internet on a regular basis. BC’s Education Plan will encourage smart use of technology in schools, better preparing students to thrive in an increasingly digital world. Students will have more opportunity to develop the competencies needed to use current and emerging technologies effectively, both in school and in life. Educators will be given the supports needed to use technology to empower the learning process, and to connect with each other, parents, and communities. Schools will have increased Internet connectivity to support learners and educators. (p. 7)

As Internet and mobile technologies continue to proliferate, most schools will be asked to evolve and accommodate. Students, parents and administrators will have expectations for teachers to adapt. For those who do choose to use mobile technologies in their classrooms, various support systems will be needed to help educators during this pedagogical transition.

Potential benefits of mobile learning environments

Through the use of mobile technologies, teachers can create a mobile learning environment. Mobile learning or “m-learning” has been described as “any service or facility that supplies a learner with general electronic information and educational content that aids in the acquisition of knowledge regardless of location and time” (Lehner & Nosekabel, 2002, p. 103). One of the biggest differences between these learning environments and traditional environments is the shift of interplay between the student, teacher, and knowledge. Students can access their teacher, the course content, as well as other expertise found online (Burden & Maher, 2014) at any time and in any location. This shift has the potential to create a “seamless” learning space (Chan et al., 2006) that provides more in-depth discussions and new motivations to explore educational content (Chiong, 2009; Chiong & Shuler, 2010; Huang, Kuo, Lin, and Cheng, 2008; Rau, Gao, & Wu, 2008). Additionally, this new capability has been shown to increase student engagement and collaboration as lessons can become more personalized and interactive (Kearney, Schuck, Burden, & Aubusson, 2012; Traxler, 2009; Valstad & Rydland, 2010). For example, in their study of two classes of fifth graders, one with mobile technology and one without, Lai, Yang, Chen, Ho, and Chan (2007), found mobile learning students retained more enriched conceptual knowledge of course content compared to traditional classrooms by personalizing their experiences through the use of various multimedia platforms. In another study of K-12 classroom students, where iPads were introduced as a mobile

supplement, Valstad and Rydland, (2010), concluded that “there is a growing realization of the limitations of traditional teaching with their emphasis often on the institutional delivery of curriculum content... I believe in a more collaborative learning environment together with game-based learning technologies have the potential to enhance student learning” (p. 81). Though there are many other examples of mobile learning environments enhancing traditional learning spaces (Evans, 2008; Fernandez, Simo, & Sallan, 2009; Lu, 2008; Motiwalla, 2007), detailing them is outside the scope of this project. It should be noted that the success of these environments reflects a combination of factors including the teaching theory adopted, the context of the classroom, the mobility of the students, the use of the technology over time, the crossover between informal and formal social circles, as well as the ownership and control of the technologies (Naismith, Sharples, Vavoula, & Lonsdale, 2004). As mobile applications are one of the primary tools for using the technology, the following sections will describe how they are defined and give some examples of the tasks they can help teachers perform.

Mobile applications

To effectively utilize and create electronic information and educational content on mobile technologies, users most commonly work with mobile software applications known as “mobile apps”. Mobile apps have been shown to be a growing medium for “providing educational content to children, both in terms of their availability and popularity” (Shuler, Levine, & Ree, 2012, p. 3). Most commonly, apps are developed for smart phones, tablet PCs, and other mobile devices and are defined by their simplicity, their interactive interfaces, and their individualistic capabilities (Lu, Chen, & Chen, 2012).

These defining qualities make mobile apps useful supplements to teachers because they tend to be intuitive and easy-to-use (Goodwin, 2012) and they have the capability of enhancing

and extending traditional pedagogical practices. For instance, some have been shown to provide guidance for learners trying to consolidate abstract course concepts (Lai, Yang, Chen, Ho, & Chan, 2007; Subramanya & Farahani, 2012), while others have been shown to provide differentiated learning by catering to different learning styles through the use of multimedia (Song, Wong, & Looi, 2012). Additionally, mobile apps have been found to support assessment by recording student feedback (Coulby, Hennessey, Davies, & Fuller, 2011). Some mobile applications have also provided visible motivations for students to participate in classroom activities (Callaghan, Harkin, McGinnity, Woods, & Harrison, 2006) while allowing students to create content and experiment with sharing information with others all the while remaining fun to use (Goodwin, 2012; Mac Callum & Jeffrey, 2010). Overall, well-designed mobile applications have been found to be great supplements to the classroom because they are intuitive, easy to use, interactive, customizable, flexible, and mobile (Chiong & Shuler, 2010; Lu, Chen, & Chen, 2012). However, teachers currently wanting to use mobile devices successfully in their classrooms invariably need to invest a large amount of time in identifying and evaluating mobile apps because the technology is developing at a rapid rate (Seipold, Pachler, Bachmair, & Honegger, 2013).

Identifying applications

As the Google Play mobile app market expands, with over 1 400 000 apps available as of October 30, 2014 (AppBrain, 2014), the task of identifying and evaluating mobile applications has become challenging. The most rigorous review analysis available online, conducted by Children's Technology Review, was only able to study 14% of applications in 2012 while 56% went completely unanalyzed by any expert (Shuler, Levine, & Ree, 2012). If I were to use these same numbers to estimate the potential number of unstudied applications on Google Play I

would have over 700 000 applications that were not analyzed by any outside source. However, as a means of bringing some analysis information to users, Google Play's application market offers various details for every application provided by developers as well as user generated ratings and reviews. Yet, as sections of this paper will illustrate, the ratings and reviews provided are typically biased and are poor resources for teachers assessing mobile applications. An alternative method of analysis will be discussed in the following sections of this paper.

Project Aims

The primary aim of this research project is to highlight for educators the limitations of user-generated feedback on Google Play's website for assessing mobile applications by exploring the presence of two highly validated technology acceptance constructs within the feedback text of "Education" mobile applications. Though this is an exploratory study, I hypothesize that the two constructs - perceived ease of use and perceived usefulness - will not be prevalent in the user reviews as previous studies have demonstrated a multitude of other topics covered in similar online markets (Fu et al., 2013; Ha & Wagner, 2013). By testing this hypothesis, I am hoping to help give teachers, researchers, and Google developers a new insight into some of Google Play's app store feedback mechanisms for "Education" applications. In all, I would like to see these mechanisms improved in order to bring the most meaningful information to teachers.

Literature Review

When making the decision about how to help teachers trying to utilize mobile technologies, I knew that mobile apps would be a major component of any mobile learning environment. Consequently, as opposed to looking at individual applications due to their higher degree of contextual factors, I decided to analyze at a larger scale and look at the distribution of applications. In the end I analyzed Google Play because it is one of the biggest mobile app distribution websites for multiple mobile devices and it offers user reviews as public information. The following sections review the literature with regards to common rating and review biases found on the Google Play app store, as well as, the impacts of review content on innovation acceptance and adoption, and finally the validity and reliability of the two constructs' perceived usefulness and perceived ease of use as developed by Davis (1989).

Biases on the Google Play app store

On the Google Play app store (GPAS) website, users can choose from over a million mobile apps for various mobile devices. A mobile app, as mentioned previously, is a small piece of software that is used in conjunction with a mobile device to offer ways of manipulating content. On the GPAS website mobile apps are sorted into 34 categories including "Education". Once a user has downloaded an "Education" app they are permitted to leave feedback and a star rating which is added to that mobile app's download page for other users to see. Alongside user reviews, the download page may also contain developer descriptions and details about version changes, the network allowances and storage permissions (see Figure 1).

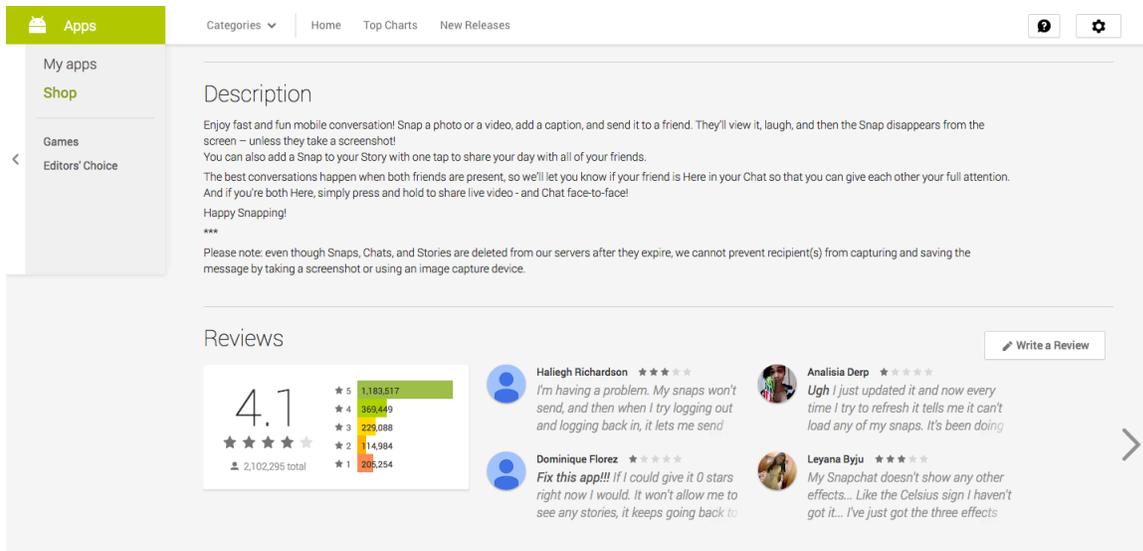


Figure 1 - A typical section of a mobile application's download page on Google Play website

By simply scrolling down the page, a user can see an average star rating representing the mean of all user star ratings. Users can also see a feedback forum that contains textual reviews written by other users. If a user then decides to see additional reviews they can click on a review or the navigation panel on the right side of the page (see Figure 2). The Google Play website will then bring into view more reviews all sorted by “Helpfulness” (see Figure 3).

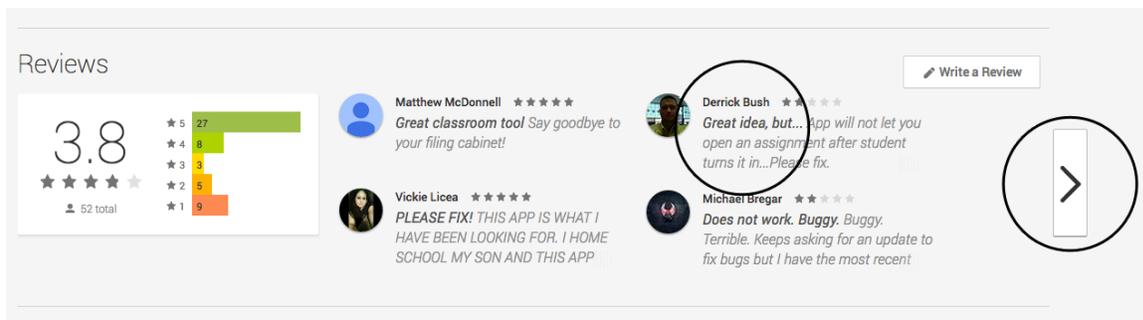


Figure 2 - Users can see more reviews by clicking on a review or the arrow

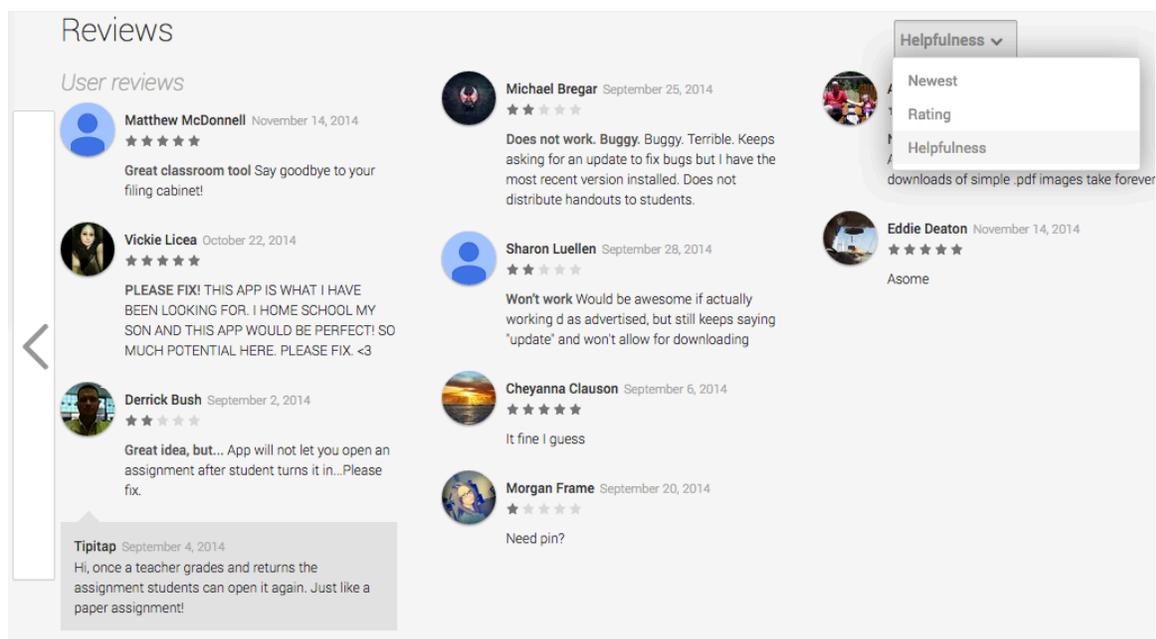


Figure 3 - Reviews can be sorted by "Newest", "Rating", and "Helpfulness"

If a user wants to re-organize the reviews presented, they can change the sorting near the top right corner of the user review page to “Newest” or “Rating” (see Figure 3). The following sections will discuss the limitations of these different information sorting mechanisms starting with star ratings and then progressing through the other sorting methods of textual comments.

Star Ratings

On the GPAS website, users can provide a general 5-star rating about the app they have downloaded. These individual ratings are then combined and the average rating is posted beneath the app description (see Figure 1 and Figure 2). The literature regarding the use of these simplistic types of rating systems as indicator of product value is detailed below.

Though some researchers have reported that star ratings are comparable to expert review scores and therefore are accurate measures of user sentiment (Shuler, Levine, & Lee, 2012), a larger number of studies have found that these ratings do not accurately capture the complex and detailed sentiments expressed by users (Ganu, Elhaded, & Marian, 2009; Hu, Pavlou, & Zhang,

2006; Riedl, Leimeister, Blohm, & Kremer, 2010; Wu & Huberman, 2010). In essence, if the product or service being reviewed is multifaceted and complex, generalized mean scores are not appropriate for capturing user sentiments. Additionally, Hu, Pavlou, and Zhang (2006) found that a single measurement of the mean product rating was unreliable because most ratings stemmed from extreme opinion holders while moderate opinion holders did not usually leave feedback. Furthermore, a single aggregate mean score is also problematic because it does not reflect any recent changes in ratings (Leberknight, Sen, & Chiang, 2011).

When star ratings are combined with user comments on the same page, the rating system becomes more susceptible to bias (Cosley, Lam, Albert, Konstan, & Riedl, 2003). Due to the nature of the interface, if user reviews are not isolated but are part of the rating page, as found on the GPAS website, star ratings may reflect the difference between the user's experience and the experience described by others (Wu & Huberman, 2010). For example, if a user reads a review describing a quick load time and the application is not "quick" it may receive a lower star rating. It is also possible that the reviews could influence star ratings in the opposite way. Several positive reviews might influence new users posting in a more positive manner when scoring star ratings (Ganu, Elhadad, & Marian, 2009).

After studying the use of rating scales by 349 participants submitting almost 16 000 ratings, Riedl, Blohm, Leimeister and Kremer (2010) found that online rating mechanisms are important aspects to consider when collecting product or innovation ideas from users. For instance, a more complex multi-attribute scale can prompt users to contribute ideas on traits that they had not previously considered. Overall, they concluded:

...the established practice of Internet-based rating which proposes that rating scales should be as simple as possible to avoid user drop-out, our research finds

that very simple scales [thumbs-up, thumbs down, or five-star scales] lead to near-random results. Consequently, more complex scales should be used, accepting higher drop-out rates but improving rating accuracy...Effective and accurate design of mechanisms for collective decision making is critical to harness the wisdom of the crowds. If the design is ill-fitted to the desired task, outcomes can be misleading or simply wrong (p. 16).

Likewise with the Google Play mobile app website interface, the star-rating system is inaccurate because textual data is difficult to measure numerically (Lim, Ortiz, & Jun, 2012). Fortunately for teachers, scaled ratings are not the only information generated by mobile application users, as written comments are also available to gauge application effectiveness. However, as described in the section below, the GPAS website feedback mechanisms tend to sort user-generated information in problematic ways.

Written comments and feedback forums

With regards to written comments, previous studies have found that online feedback forums generally suffer from two main weaknesses including the lack of structure within feedback language and the scale of comments. Vasa, Hoon, Mouzakis and Noguchi (2012) found that one of the major flaws with the feedback they studied was the lack of information as short sentences were common and when Petz et al., (2013) compared discussion forums, blogs, microblogs, product reviews, and social networking sites, they found that user generated feedback tended to contain many spelling mistakes, grammar mistakes and slang. They also found that these characteristics made reviews harder to understand and harder to categorize using a systematic mechanism. Another issue with review content is the diversity of topics covered. As review websites ask open-ended questions, most review forums tend to contain a

multitude of topics (Khalid, Shihab, Nagappan, & Hassan, 2014; Platzer, 2011). Sometimes, without a focus, feedback forums become discussion forums and potentially pertinent information for teachers gets lost within the conversation between users. Finally, content is heavily dependent on the feedback mechanism being employed. When comparing different interface designs on social media websites, blogs, discussion forums, and product reviews, McGlohon, Glance, and Reiter (2010) concluded that different rating displays resulted in changes in feedback for identical products. They recommend that more research is needed with regards to improving review content as well as mechanisms used for sorting content.

A further exacerbating factor in the difficulty of trying to access pertinent review information is that reviews on distribution websites similar to Google Play are collected at a massive scale. To put this scale into perspective, in 2012, on the Apple app store website alone, there were 8.7 million reviews provided by over five million users (Vasa, Hoon, Mouzakis & Noguchi, 2012). There are no public figures for the number of Google Play app store reviews as of 2014. However, during my study I found over 106 000 reviews for only one category of mobile apps out of 34 available. Furthermore, it has been found that the amount of information available for applications is so large that some users experience “information overload” (Park & Lee, 2008). As defined by Park and Lee (2008), information overload is “the phenomenon of too much information overwhelming a consumer, causing adverse judgmental decision making... this can result in confusion, cognitive strain, and other dysfunctional consequences” (p. 388). As a means of mitigating information overload and giving users an opportunity to organize information in different ways, visitors to the GPAS website can filter the user reviews by “Newest”, “Rating” and “Helpfulness” (see Figure 3). The following sections detail the studied problems of these various content-filtering techniques.

“Newest” and “Rating” filtering

When a user filters the reviews by “Newest”, the Google Play app store organizes the user comments by the date and time that the review was posted starting with the most recent post. If a user decides to sort the reviews by “Rating” then the list of comments is automatically arranged by the star-rating associated with the feedback beginning with the highest ratings (i.e. comments left by users who also gave a five star-rating would be listed first, then four star-rating comments, then three stars, and so on). The problem with these two filtering methods is that they are prone to bias.

As mentioned earlier, one of the major problems with relying on user star-ratings is that they are typically bi-modally distributed (Hu, Pavlou, & Zhang, 2006) and ill-suited to fit with text-based reviews (Riedl, Leimeister, Blohm, & Kremer, 2010). Typically, this means that users who choose to leave feedback are leaving extremely opinionated feedback and are assigning their comments an extreme star-rating. Subsequently, moderate opinions are not commonly accounted for when users sort comments by the “Rating” filter.

Similarly, when user generated content is organized by date, a different bias tends to occur, similar to “Early Bird Bias” described by Liu, Cao, Lin, Huang, and Zhou (2007) “...reviews posted earlier are exposed to users for a longer time. Therefore, some high quality reviews may get fewer users’ votes because of later publication” (p. 336). Essentially, if reviews are organized by the date that they are posted, the newest reviews will receive more visual attention because they are exposed to new website visitors on the main page for a longer period of time. Though it is not found in the literature, I believe this problem is exacerbated by the removal of older reviews to make visual room for the new reviews. At a basic level the problem with these filtering systems seems to be the use of an automated filtering system which is based on numbers for a database made up of text and sentiment-based information. To

counter-act these automatic filtering systems, the GPAS website has a “Helpfulness” filter. Unlike the previous two filters, this sorting mechanism lets the *users decide* on which comments should be presented as containing helpful information.

“Helpfulness” filtering

As a means of trying to get “helpful” reviews to users and reduce automatic biases or feedback manipulation (Wan & Nakayama, 2014), Google Play has incorporated a “Helpfulness” filter that sorts reviews based on helpfulness-votes given by other users. However, similar to “Newest” and “Rating” this filtering system has also been shown to inaccurately present content. Liu et al. (2007) and Jurca, Garcin, Talwar, and Faltings (2010) commonly found that reviews containing more words were typically ranked as more helpful regardless of content. Additionally, Liu et al. (2007) also found reviews with higher helpfulness-ratings attracted more helpfulness votes in a positive feedback loop. In a separate study of over 50 000 reviews ranked on helpfulness, Cao, Duan, and Gan (2011) found that helpfulness votes were typically influenced by a number of semantic characteristics including the quantity of four-letter words or the easier to read tended to be voted more helpful, quantity of cons or number of negative words in a sentiment can influence helpfulness votes, smaller titles led to more helpfulness votes and extremeness of the reviews (high deviation from the norm) was noticed more often and voted helpful more often. In a different study looking at the subjective aspects of helpfulness, Danescu-Niculescu-Mizil, Kossinets, Kleinberg, and Lee (2009) point out that helpfulness-ratings were determined by their conformity (how close is the reviewers star rating to the consensus rating), their agreement (voted as helpful if the user reading the review agrees with the review), if they are seen as brilliant but cruel (negativity is seen as intelligence, competent, and expert), and present objectivity (people are seen as reacting

naturally to the review content). In another study of 1 600 product reviews, Mudambi and Schuff (2010) found that review helpfulness for experiential products, (i.e. something that needed to be experienced to evaluate product quality) were significantly influenced by word count and negative sentiments. Essentially, the helpfulness mechanism tends to be a subjective rating that relies on the user's definition of helpful topics. Although bias has been found within user-generated feedback, it is still an important information source. User-generated feedback has been shown to be an influential part of marketing and innovation adoption, and it is worth exploring further.

The influence of word-of-mouth

Though there is no standardized definition of user-generated content (UGC), it is usually considered to be any broadcasted content that is created by a user of the broadcast system as opposed to the owners or developers of that system (Balasubramaniam, 2009). Currently, this type of content is created and distributed across electronic channels including blogs, social media networks, forums and user reviews. These channels have been found to present new informal forms of traditional word-of-mouth networks (Cheung & Thadani, 2012; Jansen, Zhang, Sobel, & Chowdury, 2009) and are similarly influential.

Before becoming digital, word-of-mouth (WOM) networks were identified as highly influential sources of information towards adopting innovations. In the 1950s, two commonly cited empirical studies were conducted to assess the influence of WOM on the consumption of products and both of them found that WOM was a significant influential factor in product adoption (Whyte, 1954; Katz & Lazarsfeld, 1955). In the 1960s, several more major studies found a similar result that WOM was a primary influencing factor of consumers' purchasing decision making process (Arndt, 1967; Engel, Blackwell, & Kegerreis, 1969; Feldman &

Spencer, 1965). During the 1970s and 1980s, more studies were conducted and similar results were found for buying paper (Martilla, 1971), adopting a new exporting processes (Lee & Brasch, 1978), adopting a new telephonic banking system (Horsky & Simon, 1983), and the development and use of non-intelligent data terminals (Moriarty & Spekman, 1984) among many others.

Then in the 1990s, with the advent of the Internet, the original WOM networks started appearing as electronic WOM (e-WOM) networks with relatively similar influence on user choices (Martin & Lueg, 2011; Trusov, Bucklin, & Pauwels, 2009). These online networks are continually found to be valuable resources of information to modern consumers (Bailey, 2005). For example, as Goldsmith and Horowitz (2006) reported with their study of 309 consumers, the common reasons online word-of-mouth networks are utilized include “[reducing the risk of choosing a poor product], because others do it, to secure lower prices, to get information easily, by accident (unplanned), because it is cool, because they are stimulated by off-line inputs such as TV, and to get pre-purchase information” (p. 3). Goldsmith and Horowitz (2006) also found that in a second sample of consumers, there was a strong indication that e-WOM networks would be used continually in the future and that e-WOMs provided information that is more important than advertising. Li, Lin, and Lai (2010) who studied 4573 reviews on Epinions.com also found that consumers tend to rely on online reviews over traditional marketing because online reviewers are perceived as more trustworthy. Other reasons that users utilize e-WOM include the market transparency, the increased voicing of concerns, and the encouragement to being part of the value chain (Rezabakhsh, Bornemann, Hansen, & Schradar, 2006). In essence, the users that post information are perceived as less biased towards a product than the producers of that product and e-WOM channels provide a place for them to voice their views.

However, the connection between user review content and influence on innovation adoption is not fully understood. Several studies have found that the manipulation of the review feedback system may change the impact of the reviews (Chevalier & Mayzlin, 2006; Petz et al., 2013). For example, Ghose and Ipeirotis (2011) found that a large amount of subjective sentiments on Amazon.com correlated with an increase in product sales, whereas, a mixture of objective sentiments with subjective sentiments is negatively associated with product sales. Similar impacts have also been found to change depending on information organization and presentation for book sales (Chevalier & Mayzlin, 2006). In their study comparing various online feedback systems including blogs, social media websites, product review websites and discussion forums, Petz et al., (2013) found differences in posting length, the use of internet slang, grammatical correctness, and the amount of subjective information all had differing effects on feedback influence.

In regards to the Google Play app store, one example of the impacts of review feedback manipulation would be the amount of visual room provided to show reviews. If an informative and well-written comment is removed from the main page, to make room for new posts, it becomes useless to all site users who have not seen it. This is especially problematic for two main reasons. First, researchers have found that on average, users typically do not read all the reviews available and most commonly may only read 10 reviews or less (Anderson, 2014; Park & Lee, 2007). The second difficulty is, as mentioned in the introduction, user-generated feedback is constantly being created as on average, seven to 37 reviews are posted everyday (Pagano & Maalej, 2013). With most feedback systems, this usually means older reviews are pushed down the list and with Google Play it can mean off of the main page and onto a secondary page resulting in a type of feedback manipulation.

An alternative method

Regardless of what content is actually found useful, the information on Google Play's website is presented in a systematized way and this system is subject to bias and potential manipulation (Hu, Bose, Koh, & Liu, 2012). Additionally, allowing the users to organize information by "Helpfulness" is also problematic because users decide helpfulness based on a variety of semantic and subjective reasons. I propose an alternate method of organizing information to help innovation adopters in a professional setting would be to categorize reviews around sentiments concerning perceived ease of use and perceived usefulness as these constructs have been repeatedly shown to predict technology acceptance and adoption.

The Technology Acceptance Model developed by Davis (1989)

The Technology Acceptance Model (TAM) developed by Davis (1989) was originally intended to provide a valid and reliable measure for predicting user acceptance of a computer technology system. The TAM focused primarily on the constructs perceived usefulness and perceived ease of use as these two constructs had previously been shown to have high correlations with self-perceived innovation acceptance (Robey, 1979; Schultz & Slevin, 1975). However, as Davis (1989) noted, very few high quality instruments had been previously tested to measure the major determinants of user acceptance; of those that did, very few reported consistent correlations across studies. Alongside improving the measures for key theoretical constructs, Davis outlines the practical reasons for creating the TAM as helping vendors who would "like to assess user demand for new design ideas", and for "information systems managers within user organizations who would like to evaluate these vendor offerings" (p. 319). As seen in Davis' model (Figure 4) it is hypothesized that a person's intention to use a technology in the workplace is determined by their beliefs about using that technology, more specifically, the two key constructs are perceived usefulness and perceived ease of use.

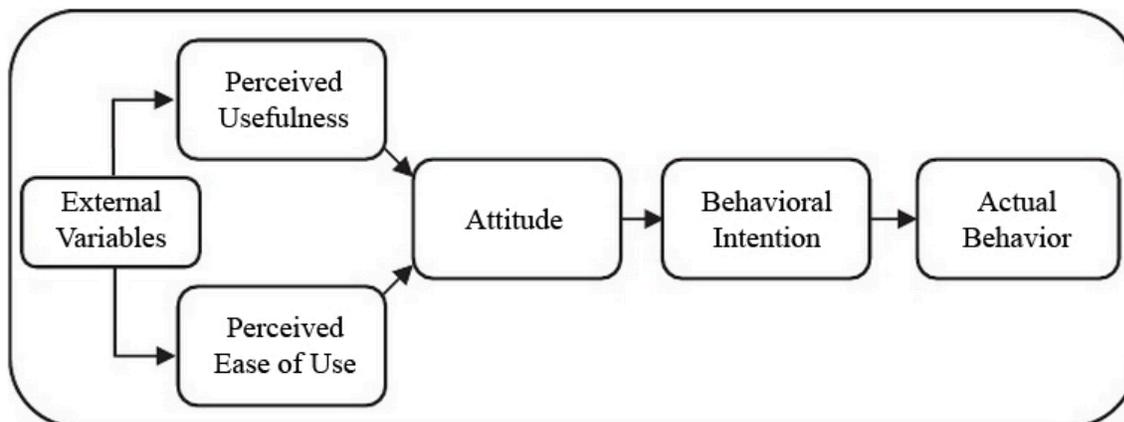


Figure 4 - Technology Acceptance Model (Davis, 1989)

Perceived ease of use and perceived usefulness

As defined by Davis (1989), perceived usefulness is “the degree to which a person believes that using a particular system would enhance his or her job performance” (p. 320) and perceived ease of use refers to “the degree to which a person believes that using a particular system would be free of effort” (p. 320). When Davis (1989) was developing the Technology Acceptance Model it was noted that perceived ease of use and perceived usefulness were obviously not the only variables of interest with regards to the decision to use information technology, but they appeared “likely to play a central role” (p. 323). When Davis (1989) was developing the items for the construct perceived ease of use, the definition of “ease” was used for guidance: “freedom from difficulty or great effort” (p. 320). Davis was assuming that users would accept an application that was perceived to be easier to use than another. Davis’ (1989) theoretical foundations stemmed from Bandura’s (1982) research on self-efficacy and the judgments of the self to execute an action required to handle a certain situation as well as a behaviourist cost-benefit paradigm that involves a decision strategy comparing effort of use (cost) with the quality of the benefit for making the decision (benefit). When applying this theoretical framework to the adoption of technology, Davis closely followed Rogers and

Shoemaker's (1971) "Complexity" definition of "the degree to which an innovation is perceived as relatively difficult to understand and use" (p. 154). Though originally tested and included with perceived usefulness, Davis (1989) decided to isolate perceived ease of use instead of merging the two constructs into perceived usefulness, because these two dimensions had been previously shown to have different loadings in a previous study by Larcker and Lessig (1980).

For the construct perceived usefulness, Davis built instrument items based on the work of Shultz and Slevin (1975). From Schultz's and Slevin's analysis of 67 questionnaire items and seven dimensions, Davis decided to utilize the "performance" dimension for the Technology Acceptance Model and test items originally constructed in a study by Robey (1979). The wording utilized in Davis' (1989) scaled instrument stems from Swanson's (1987) model of "channel disposition" where the items "important", "relevant", "useful", and "valuable" loaded strongly to a value dimension perceived usefulness and "convenient", "controllable", "easy", and "unburdensome" loaded strongly with an access dimension perceived ease of use (p. 322). Davis (1989) claimed to keep the instrument measuring *perceived* instead of *actual* use because measuring actual use would require the use of test facilities and heavy instrumentation.

Originally developed with the theoretical framework noted above, as Davis (1989) notes, the TAM started with 14 items, based on 47 previous studies, for each construct totaling a 28-item instrument. Pre-test interviews removed four items each bringing the total to 10 items per construct. This 20-item instrument was then tested with 112 users and two innovations and was found to have a Cronbach Alpha reliability of 0.97/0.97 for perceived usefulness and 0.86/0.93 for perceived ease of use. Davis (1989) also found that there was a significant convergent validity for the two scales, as well as, a high discriminant validity and good factorial validity.

After a post-test was conducted with the 112 users, all correlations between perceived use and actual use were found to be significant at the 0.001 level. However, as it was found that some of the items within the instrument reported poor convergent and discriminant validity, Davis decided to remove these items, bringing the total number to six items per construct.

Davis (1989) then decided to test the streamlined instrument with a new group of 40 participants who were unfamiliar with two different systems. For this second study, Cronbach Alpha was 0.98 for perceived usefulness and 0.94 for perceived ease of use. Similar to the first study, factorial, convergent, and discriminant validity were all supported. In Davis' (1989) discussion about these constructs, it is stated that "the new scales were found to have strong psychometric properties and to exhibit significant empirical relationships with self-reported measures of usage behavior" (p. 333). Davis (1989) also noted that perceived usefulness had a significantly stronger relationship to usage than perceived ease of use and that perceived ease of use may in fact be an antecedent to perceived usefulness. Finally, Davis notes that the question of external validity remained to be tested by future studies.

External validity

Since the initial studies conducted by Davis, researchers have found that perceptions around perceived usefulness and perceived ease of use are still significant predictors of the intention to use new technologies (Davis, 1989; Ho, Hung, & Chen, 2012; Teo, 2012). These two instrument constructs have been tested numerous times for internal consistency by researchers over the last two decades. When comparing previous studies using the TAM constructs perceived ease of use and perceived usefulness, Legris, Ingham, and Collette (2003), found that Cronbach's Alpha coefficient was consistently high as the 22 studies identified reported an Alpha of 0.83 or higher. In addition to the constructs perceived usefulness

and perceived ease of use, numerous other constructs have been developed for several other models since the TAM. These constructs include: subjective norm, extrinsic motivation, intrinsic motivation, perceived behavioural control, job-fit, complexity (reversed), long-term consequences, affect toward use, social factors, facilitating conditions, relative advantage, result demonstrability, trialability, visibility, image, compatibility, voluntariness, outcome expectations, self-efficacy and affect (Venkatesh, Morris, Davis & Davis, 2003). However, no constructs have been tested as rigorously as the original constructs perceived usefulness and perceived ease of use (Hess, McNab, & Basoglu, 2014; King & He, 2006). Within this project, as I knew that these two constructs had been shown to be highly reliable and valid constructs with regards to the attitudes towards the intention to adopt new innovations, I was compelled to examine their presence in mobile app feedback. The following sections detail my limitations and assumptions as well as the steps I took to conduct my project analysis.

Limitations

Throughout this research project there have been a few limitations and assumptions and the subsections below have separated these areas. My limitations involved my data collection timeframe and restrictions I imposed on my collection program to accommodate for Google's requests about interacting with their servers. My other limitations involved my coding capabilities, researcher's analysis, and linguistic expertise.

Data collection

I was limited in two main ways with regards to my data collection methods. The first way was direct access to the reviews as I was not able to collect the entire population of user reviews for one main reason: my collection program was restricted to collecting around 50 applications per day due to Google's request that I keep my traffic to a minimum. To collect the entire

population at this rate, my collection timeline would have been over six years. As I did not have this option, I had to draw a sample of reviews from the Google Play website. In the end I believe that this may have been beneficial because it kept my study manageable and current. However, as Wiedemann (2013) notes, the choices on sample selection and algorithm can have a significant influence on results in a computer-assisted analysis of textual data.

Another limitation worth noting were the limits on my type of data collected. I was unable to gather other information about each application including the star-rating, the developer's information, and other sources of information about the applications that might have provided a deeper context to the feedback. For example, some of the reviews I sampled were short and lacked specific information. This could have happened because the writer was only adding information already provided by the application developer. However, as I was unable to collect multiple data sources, I could not infer this connection and subsequently could not add these reviews to my final results.

Another major limitation was my delimited use of Google Play website reviews which is only one of the major online applications stores. As stated previously, millions of applications are also available on the iTunes app store and on the Amazon marketplace. Though the Google Play store offers more diverse applications for more diverse mobile technologies, it may not cover all the types of "educational" applications. Furthermore, it may be the case that the online culture for Google Play does not foster a community concerned with perceived usefulness and perceived ease of use whereas the other online stores do. This limited my analysis and ultimately the generalizability of my results to those teachers and developers using the Google Play app store.

Researcher limitations

In addition to my data collection limits, I was also limited in linguistic research expertise. Because my study was valuable to helping teachers and possibly the future development of m-learning, I continued to pursue this project though it was multidisciplinary and contained aspects of the education domain as well as the linguistic and computer science domains. The drawback in doing so limited my ability to fully and exhaustively extract and code all the reviews in my database validly and reliably. To counter these limitation I chose to keep my coding schemes as simplistic as possible in order to maintain consistency and my timeline relatively short in order to preserve my external validity. However, as the results show, there was a large portion of the reviews that did not get analyzed. This could be an indication of poor internal validity. Alternatively, it is possible that other technology acceptance constructs were reported more frequently than the ones that I chose. It may be useful for future researchers with more expertise in this field to test alternative constructs. As I was unable to locate a prior study that utilized the TAM model provided by Davis (1989) in the method I proposed for this study, I was limited in guidance and modeling. The closest studies I could find in this area involved the use and adoption of the mobile device but not the applications. However, it should be noted that Davis originally used the TAM to study the adoption of software, not just hardware.

Another researcher-based limitation to my study was potential researcher bias. As all the reviews were coded manually, there is a high level of researcher bias. This is due to the subjective interpretation of the review content. Without multiple coders working with the same data or an additional post-project re-coding to check for consistency, it is very hard to be confident that these interpretations were entirely objective. For example, in the assessment of the term “useful”, it was sometimes difficult to discern if a comment about a teaching task would be sufficient to influence a users perception of mobile app’s usefulness. Some teachers

may describe “...understanding the shape of their letters” (see Table 2, example BA) as just a description whereas I coded all examples containing a word list indicator as well as a task description as relating to a construct. Future studies in this domain will need to look at their definition of perceived usefulness and perceived ease of use with regards to its fit with mobile application reviews.

Assumptions

The following sections detail the major assumptions made during this research project. Again, similar to the cause of the limitations, assumptions had to be made in order to maintain a workable timeline and scope. My major assumptions included the representativeness of my sample as well as the validity and reliability of the procedures that I chose during the data analysis stage.

Sample representation

Due to the size of my sample database I was unable to manually code all of the reviews collected. Subsequently, I had to assume that my sample collection of cases was adequate and captured a good representation of the population. Though I took steps to collect a random sample and identify all the possible identifying words used within the sample, it is possible that the word lists (see Table 1) missed important word slang and nuances. For example, it may be the case that a word that is not necessarily a synonym or antonym to “easy” was used as such. Though I was able to identify all the words that normally apply to these categories, it is well known that online language is not always used in this manner (Petz et al., 2013).

Face validity

One of my biggest assumptions alongside sample representation was a form of face validity. Essentially, face validity refers to a measure of validity that is done by someone who is

not an expert in that field. As a student researcher I am not an expert in content analysis nor linguistic analysis and therefore the assumed validity of this research is only what appears to be valid. This is not to be confused with content or construct validity, which I did not measure. As I used words from the original construct as guidance for coding reviews, I may have introduced error regarding the actual content validity (i.e. how my coding scheme captured all the potential reviews available regarding the original constructs). This was my greatest assumption because the context used for the original TAM may not be relatable to the context I was applying it. Furthermore, when coding the reviews, internal validity was only measured at face value, as there was a high variance of language and only a single coding iteration was performed. Ultimately, future research will be needed in this area with experts in this field to determine the extent of lost construct validity due to mixed context interpretation.

Methodology

The data I was interested in exploring was user reviews written about “Education” applications found on the Google Play app store website over the months of September and October 2014. As the mobile app market is rapidly growing, I decided to limit my scope to strictly user reviews to make this project manageable and timely. Due to the fact that my hypothesis was quantitative in nature and my data was natural language and required a qualitative interpretation, I utilized a mixed methodology. It should be noted that all the data collected from the Google Play website was verified by the Google legal team as public information being used for research, and as such, there was no ethical restrictions imposed on this research. As an extra precaution, all data identifiers including user names and timestamps were removed from the data in order to preserve user anonymity. The only restriction imposed with regards to data collection was the amount accessed. The Google development team asked that I keep my server requests low so as not to overburden their system. The following sections detail my methodology approaches, dataset, data analysis, results, and conclusions.

Quantitative Approach

Due to the size of my database and the fact that my research hypothesis was concerned with the frequency of construct sentiments within the user reviews, I knew my methodology would involve the use of a content analysis software package in order to reduce researcher error (Wiedemann, 2013). This approach, which also utilized data mining, is commonly referred to as computer assisted text analysis (CATA) and usually involves the task of breaking down large sources of textual content into a quantifiable database. These sources can range from web logs to journal articles to product reviews as well as many other content sources. Once a database of

content has been created it can be analyzed and used to potentially explain phenomenon or create models for further analysis of other content caches. As defined by Mergenthaler (1996):

The preliminary task [of CATA] is the compilation of a glossary or dictionary, often consisting of a collection of [...] word forms which are assigned to different categories. The categories themselves constitute a system including either the facets to a special topic or the aspects of a more general complex of topics. The vocabulary of a dictionary can be derived either *inductively* from a text or *deductively* from more general constructs whose consequences can be detected in the choice of categories. The computer's task is to examine a text word for word and to compare it to the dictionary. If a word form is found, the number of entries counted for the corresponding category is increased by one. The resulting frequency distribution can also be relativized according to the text for the purposes of comparison. (p. 4)

As previous research has shown, larger databases are difficult to code manually (Wiedemann, 2013). As I ended up extracting 13 099 entries, I decided to utilize computer software to help me describe the frequency distribution of my categories. After this decision has been made, researchers usually have two main options for a computer assisted analysis: they can let the software do the entire analysis, usually n-gram frequency analysis, factor analysis, or agglomerative clustering (Cavnar & Trenkle, 1994; Hillard, Purpura, & Wilkerson, 2008) also known as unsupervised CATA, or, they can interpret and modify the dictionary, database, or algorithm, during the process of a

supervised CATA (Brier & Hopp, 2011). I chose to conduct a supervised CATA because I was not primarily concerned with exhaustively identifying all categories within my database because I knew the results would be scattered in a similar way to other studies (Hoon, Vasa, Schneider, & Grundy, 2013; Hoon, Vasa, Schneider, & Mouzakis, 2012). Also, I was more interested in exploring already established categories from the literature which is better suited for supervised text analysis (Grimmer and Stewart, 2013). Additionally, if I had chosen an unsupervised approach to classify and categorize the entire database there would have been a higher risk of improper classification due to missed nuances in sentiment and meaning (Hillard, Purpura, & Wilkerson, 2008) as well as a categorical mismatch that may not have fit properly with the context of future studies because of the different word databases used (Wiedemann, 2013; Lowe, 2003).

Qualitative approach

Due to the reasons mentioned above, I decided to conduct a supervised computer assisted text analysis. As the supervision was to be conducted by a student researcher, all coding procedures and category definitions were to be kept within a manageable scope. Defining this scope involved using a poststructuralist framework for my CATA analysis.

In a structuralist theoretical framework, the researcher presumes that “[human interrelations] constitute a structure, and behind local variations in the surface phenomena there are constant laws of abstract culture” (Blackburn, 2008). In the domain of linguistics, this means that there is an underlying structure to the language we use, regardless of context, and this language - in terms of its character - can be used to identify these structures. However, since the 1950s, most linguists have dropped linguistic structuralism as a framework for studying language and have moved towards a poststructuralist framework because of several

challenges made during the 1950s and 1960s (Koster, 1996). In response to structuralism, some poststructuralists claimed that the structuralists who study language are unable to separate themselves from the domain they are studying. They argued that language is an interpretation and cannot be separated from the current context it is being understood. Thus as a researcher using a poststructuralist framework, I am aiming to avoid the problem of presupposing the knowledge (structure) I am trying to understand. In terms of methodology, this means using a combination of words found in the content of my data as part of my categorical dictionary as opposed to strictly using only predefined words from previous literature. By combining my word sources I can be more confident that my analysis is not presuming structure and missing contextual meaning (Pecheux et al., 1995, p. 65). Consequently, though my analysis was going to be computer assisted it was also going to be heavily supervised to ensure the results addressed these contextual meanings.

The dataset

For this research project I decided to look strictly at the textual reviews from Google Play because, as I have mentioned, star ratings are not entirely accurate at conveying sentiments about mobile apps and other details are outside the scope of this project. I chose Google Play because it offers the widest range of applications for the widest range of mobile devices (VentureBeat, 2013) and it categorizes mobile apps as “Education” oriented. This categorization was important for my selection because I wanted to look at the most likely source of mobile apps that would encourage teaching and learning. Though other categories might contain these types of applications, I assumed the “Education” category would be one of the most obvious starting points for educational app exploration.

In the “Education” category on the Google Play website, the total number of mobile apps I found was 106 880. Due to the complexities of analyzing such a large database of natural language reviews with software I decided to analyze a smaller random sample. This meant that I needed to decide on a sample size that was sufficient for making inferences to the population and at the same time be small enough to manually code in a timely manner. As I was using a proportion of the population as my sample, I used the population proportion formula offered by Moore (2004) to estimate an appropriate sample size.

$$n = \left(\frac{z^*}{m} \right)^2 p^* (1 - p^*)$$

Figure 5 - Formula used to determine my sample size (n)

For this estimate, I chose a standard Normal critical point (z^*) of 2.576 for a 99% level of confidence, an approximate margin of error (m) of 0.05 and a highly conservative sample proportion (p^*) of 0.5. These approximations and high confidence level resulted in a sample size of 664 mobile applications. However, after several weeks of web scraping and collecting the reviews and data from Google Play, I managed to gather the data for 1 283 applications. This larger sample size reduced my approximate margin of error from 0.05 to 0.036 or 3.6%.

With regards to web scraping, I constructed a program in the programming language Python to gather the names of all the “Education” apps on the Google Play store and then add them to a Microsoft Excel database. I then used Excel to assign each application name a random number between 0 and 1 using Excel’s RAND() function. Once every application had a number I copied the values and then sorted them in ascending order. I then collected all the reviews from the top 1 283 application names on my list. The collection stage involved using another program written in Python to communicate with the Google Play app store server to

locate the application and then download the entire set of user reviews for that application into a new Excel spreadsheet. In total, this process took several weeks and resulted in a final database of 13 099 reviews for analysis.

Data analysis

One of the first steps that conducting a computer assisted text analysis includes is a data cleaning stage also known as data preprocessing (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). This process essentially involves removing noise and empty data fields. Though preprocessing has been shown to remove potential useful data, the benefits of removing irrelevant items tends to outweigh the loss of information as it provides a more accurate sample of the data (Cooley, Mobasher, & Srivastava, 1997). In my study, the process of removing noisy and empty reviews was relatively insignificant as only 18 reviews (0.0014%) contained no characters or they contained a majority of indistinguishable symbols. Subsequently, I chose to keep as much data as possible and did not remove these reviews from my database. Another preprocessing task I could have tried was to find and replace spelling mistakes and other improperly formatted text. I decided this additional step was unnecessary after studying the word frequencies within my database. The list of individual word frequencies constructed by NVivo 10 (e.g. “and”, “be”, “cd”, “hello”) showed a total number of unknown n-grams as only 0.06% of the total database (excluding single characters). In the end, I did not remove unknown n-grams, reviews containing unidentifiable characters or reviews containing no characters and my database total remained at 13 099 reviews.

My next task involved trying to apply Davis’ constructs to my dataset. After skimming over some of the data, I noticed immediately that this was going to be a challenge. When I compared the original TAM construct instrument wording (see Figure 6 and Figure 7) to some

of my sample data I noticed that the construct wording did not easily match the feedback wording (see Table 2 and Table 3).

Perceived Usefulness

	Using CHART-MASTER in my job would enable me to accomplish tasks more quickly.							
likely								unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	
	Using CHART-MASTER would improve my job performance.							
likely								unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	
	Using CHART-MASTER in my job would increase my productivity.							
likely								unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	
	Using CHART-MASTER would enhance my effectiveness on the job.							
likely								unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	
	Using CHART-MASTER would make it easier to do my job.							
likely								unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	
	I would find CHART-MASTER useful in my job.							
likely								unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	

Figure 6 - The original items for perceived usefulness on the TAM instrument proposed by Davis (1989).

Perceived Ease of Use

	Learning to operate CHART-MASTER would be easy for me.							
likely								unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	
	I would find it easy to get CHART-MASTER to do what I want it to do.							
likely								unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	
	My interaction with CHART-MASTER would be clear and understandable.							
likely								unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	
	I would find CHART-MASTER to be flexible to interact with.							
likely								unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	
	It would be easy for me to become skillful at using CHART-MASTER.							
likely								unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	
	I would find CHART-MASTER easy to use.							
likely								unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	

Figure 7 - The original items for perceived ease of use on the TAM instrument proposed by Davis (1989).

Instead of using phrases like “enabled me to accomplish” or “increased my effectiveness” (Davis, 1989, p. 340) the feedback contained phrases similar to “I can practice” and “...beginning to read music”. After reading through more reviews I noticed that some of the reviews discussed application usefulness in a manner similar to Davis’ while others varied slightly using alternative language. To proceed with my analysis, I decided to use all of the descriptive words in Davis’ instrument construct (e.g. “enable”, “accomplish”, “improve”, “performance” etc.) and combine them with any other synonymous or antonymous language I

Table 1 - Word lists used to isolate user feedback

Perceived Usefulness Identifiers			Perceived Ease of Use Identifiers		
Davis (1989)	Syn. / Ant.	Stems	Davis (1989)	Syn. / Ant.	Stems
Accomplish	Appropriate	Effect*	Clear	Difficult	Clear*
Effectiveness	Convenient	Function*	Easy	Ease	Difficult*
Enable	Functional	Help*	Interact	Frustrating	Eas*
Enhance	Practical	Learn*	Interaction	Intuitive	Frustrat*
Improve	Practice	Perform*	Flexible	Hard	Hard*
Increase	Properly	Practic*	Learning	Simple	Interact*
Job	Handy	Produc*	Operate		Learn*
Task	Helping	Proper*	Skilful		Intuit*
Performance	Teaches	Study*	Understandable		Operat*
Productivity	Utility	Task*	Use		Simpl*
Quickly	Useless	Teach*	Using		Skill*
Using		Use*			Understand*
Useful					Use*

found within the sample word frequency list to create my final category dictionaries.

Applying perceived usefulness to reviews

After looking over some of the data, I found that perceived usefulness was being described in user reviews but with a variety of terms and phrases. Knowing that the database

was going to have a high degree of word variance, I decided to isolate all of the words in the entire sample and then manually collect every word I could identify that related to usefulness as well as words that were synonymous, antonymous and word stems of those words. An example of this would be the word “useful” found in the wording of item six on Davis’ instrument. Synonyms and antonyms of this word would be identified using multiple online thesauruses (e.g. collinsdictionary.com, thesaurus.com, etc) and all mentioned words would be added to my final coding scheme. Additionally, the word stems of these words would also be included in my coding wordlist (e.g. “useful” would give the word stem “use*”, the synonym “helpful” would give the word stem “help*”, etc.) in case the exact wording did not fully match. After reviewing the entire word bank found within my database, I isolated all the words that pertained to the descriptors found in Davis’ (1989) instrument (see Table 1).

Once I had this list of words, I used the query function in NVivo 10 to extract all the reviews from my sample that contained a single utterance of any word or word stem in the final list. After this extraction, I was left with 2048 reviews out of 13 099 (15.6%) that contained a matched word or word stem for the construct perceived usefulness. After looking over some of these reviews, I decided to manually code the data to determine how many actually described perceived usefulness in a similar manner to the wording found on the usefulness scale (Davis, 1989). When I compared some of the reviews to the items, I noticed that Davis tended to discuss a specific teaching task, action, or “job performance” within the items (see Figure 6) and many of the reviews collected by NVivo only included a single word or a short phrase (e.g. “helpful”, “this helped me a lot”, “this app is a VERY useful app” etc). Though these reviews had triggered NVivo’s extraction, I did not feel that they were a sufficient match with Davis’ items, nor were they sufficient information for a teacher to assess the mobile application’s

usefulness with regards to a teaching task or action. As the reviews contained a variety of sentiments, I decided to manually code the reviews instead of relying on a computer analysis because my skill and proficiency with computer assisted textual analysis was limited and unreliable. During this coding stage, if a comment contained information regarding perceived usefulness using a keyword and contained information about a teaching job or task, I included it in my final results (see Table 2 for examples).

Perceived Usefulness Reviews	
Identifier	Example
BA	A win! Absolutely love this, using it to help my dyslexic kiddos get a better <u>understanding of the shape of their letters</u>
BB	Amazing! Currently using this to <u>study for the MCAT</u> the different types of spaced repetition are amazing, and the sync is incredible. I am incredibly amazed!
BC	Excellent app with great function <u>to take attendance</u> . Great job with taking attendance. (I don't use the grade book feature at my school uses PowerSchool), but love the attendance feature and tracking options.
BD	Good app I love this app. I am learning <u>to read music</u> and am having trouble remembering the notes. This trainer lets me learn in practice mode where it shows the note and the coordinating letter. This is perfect for beginners.
BE	Just what I wanted. Learning piano and this app is great for <u>practicing my site reading while traveling!</u>
Bold – Indicator word found within the review <u>Underlined</u> – Teaching task mentioned	

Table 2 - Examples of feedback labelled as containing the construct perceived usefulness

As example BA illustrates, the review contains a descriptor (in bold) as well as a job related task (underlined) and thus would be included in my results.

Applying perceived ease of use to reviews

When deciding what constitutes a review that adequately describes Davis' items for the perceived ease of use scales, I followed a similar logic as in my perceived usefulness coding scheme. Again, due to the variety of words and word stems found within the entire sample, I

used Davis' wording of nouns, adjectives, and verbs as a guide for my coding scheme and then added synonyms, antonyms and word stems of these words found from online thesauruses.

After, looking back over the entire sample's word list I pulled out as many words as I could identify that fit my final coding scheme (see Table 1). Once I had this completed list I again used NVivo 10 to isolate all the reviews that contained one instance of a word or word stem in my list. After this extraction, I was left with 1 113 reviews (8.5%) that contained a match. Again, after looking over the reviews, I found many were short and general statements that contained little information about how the mobile application was being used. Subsequently, I looked back towards Davis' (1989) instrument to determine if there were other identifying terms or phrases I could use to find more appropriate reviews. I found that the perceived ease of use construct included word statements such as "Learning to operate...", "get CHART MASTER to do...", "...interact with..." and "...become skillful at using..." (p. 340). When I compared these statements to the user reviews containing ease of use sentiments I found some reviews also contained statements regarding using the application for a specific teaching task. At this point, I decided to code a review as discussing perceived ease of use if it contained a word from my word list and contained a statement about a teaching task or skill (see Table 3).

Perceived Ease of Use Reviews	
Identifier	Example
CA	LOVE THIS APP!! DON'T PONDER JUST BUY! This by far in my opinion, is the best app available for kids to learn their letters and sounds. I heard about this app from my sons elementary school where they use this app in all of their kindergarten classes. It's one of his favorite apps and is now using the starfall reading app for his next progressive step. I have also incorporated it into my own daycare program and all of the children utilize this app on android tablets. I am soooooo happy this app is now available! It is very interactive, easy enough for a 1.5 year old to use and beneficial for all ages who are <u>learning their ABCs</u> .
CB	Intuitive and playable learning game, <u>teaches letter strokes</u> decisively without becoming frustrating or tedious
CC	Amazing and easy to use IDE This is a wonderful IDE on the go, <u>for</u>

	learning <u>C++</u> or everything you want!. But, in my opinion, it could be better if its have more highlighting with the standard library, not only the “core of the language”. But, anyways, I recommend you buy this app, It is simple for novice C++ developers.
CD	Stellarium Mobile Sky Map I have used this app on the PC for a few years now. This mobile version is in my opinion the best sky map app available on the play store, and is totally worth it’s rather small price. Great if you have a passing interest in astronomy and even better if you have a telescope as it makes <u>finding planets/deep-sky objects</u> much much easier. Controls are easy and intuitive , setup and location setting very simple, just choose your country and then the town you live in and that’s all.
CE	Brilliant way for early learning! This game is awesome. The concept is great. This game <u>teaches kids numbers, letters, shapes, colors, animals, toys, fruits</u> and more. My kids simply adore it. Very easy to use and fun. This is a very well designed app with fantastic graphics. Well worth the money! I am looking forward to seeing your new apps.
Bold – Indicator word found within the review Underlined – Teaching task mentioned	

Table 3 - Examples of feedback labelled as containing the construct perceived ease of use

As example CA illustrates, the review contains a descriptor as well as a statement about using the application and thus would be included in my results.

Results

As stated previously, the aim of this study was to look at feedback about “Education” applications and the presence of the constructs perceived ease of use and perceived usefulness as described in Technology Acceptance Model proposed by Davis (1989). As previous studies have found user review forums to contain a high degree of variance (Guzman & Maalej, 2014; Khalid, Shihab, Nagappan, & Hassan, 2014; Platzer, 2011), I hypothesized that their presence would not be substantial. After analyzing the data using NVivo 10, I found that only 407 of 13 099 reviews (3.107%) discussed the construct perceived usefulness in relation to a specified teaching task and only 122 reviews out of 13 099 (0.93%) contained a mention of perceived ease of use. When deciding on my sample size I chose a confidence level of 99% and this gave

us a confidence interval of +/- 3.6% giving perceived usefulness an interval between 0% and 6.7% and ease of use an interval of 0% to 4.53%. Overall, my findings are consistent with previous studies as well as my hypothesis.

Conclusions and Discussion

Similar to many other online feedback systems, the Google Play app store utilizes several filtering and sorting mechanisms to organize and display user-generated content for users. As my results indicate, important constructs that have the potential to aid in the analysis of educational mobile applications are underutilized. As the categorization of “Education” applications does not provide a way of efficiently gathering information about a user’s perceived ease of use and perceived usefulness with regards to a specific teaching task, users will remain confounded and influenced by known information display biases. Ultimately, I believe that this misused and misconstrued information will negatively impact the adoption and use of mobile learning technologies by teachers during a time of continual mobile growth. Though this project was only exploratory, I believe my findings are enough to warrant further exploration of the Google Play feedback mechanisms; especially those used with “Education” app review forums. As the aim of this project is to help improve the Google Play website, I am going to discuss some future steps with regards to research and development in this area.

As previous studies have also looked at the content within application reviews and have found a large degree of variance in semantics and subjectivity (Cao, Duan, & Gan, 2011; Danescu-Niculescu-Mizil, Kossinets, Kleinberg, and Lee, 2009) future development should be focused on improving the organizing mechanisms for user-generated content. This could be done at the user input stage, at the display stage, or both in order to enhance the information delivered to new users by giving users a clear focus. For example, when users are presented an opportunity to leave feedback on the Google Play website, topic suggestions or tags could be offered to help label and classify comments. Using this categorization with textual content has

already been implemented on many major websites (e.g. twitter.com, stackoverflow.com, ebay.com etc.) and adoption should be relatively straightforward.

Also, as another immediate first step, the issues of slang, spelling, and grammar need to be addressed so that current information is at least readable and subsequently understandable to users and researchers. Not only does improper spelling and grammar hinder feedback it will also hinder any attempts at automatically sorting reviews in an effective way. Unlike the previously mentioned adjustment, this modification would require a greater investment and development of a comment dictionary that links topic classification with comment words. Though this is a relatively difficult task, some progress has been made with regards to movie reviews, product reviews, twitter posts and a number of other similar text forums (Abbasi, Chen, & Salem, 2008; Go, Bhayani, & Huang, 2009; Pang & Lee, 2008).

Furthermore, as this study has shown, there are numerous other constructs that are more readily commented on that need to be studied. Though it was outside the scope of this project, numerous constructs have been added to the TAM since 1989 including - but not limited to - privacy, security, subjective norm, perceived behavioural control, social factors and compatibility (Legris, Ingham, & Collette, 2003; Venkatesh, Morris, Davis & Davis, 2003). Further researcher will need to explore additional constructs covered in other studied frameworks in order to better understand potential ways of structuring the feedback system.

In conclusion, this research and development comes at a crucial time as mobile technologies expand into the classroom impacting both teaching and learning. My final and urgent recommendations, as discussed above, mainly concern exploring and developing the focus of user-generated content and information. As more pertinent information about learning and teaching apps would support the adoption and adaptation towards mobile teaching

pedagogy, I believe that the users of the Google Play website would benefit from improved feedback mechanisms. Most importantly, constructs and topics used as focus areas should be underpinned by education research as opposed to user trends. As illustrated in this M Ed project's research and that of others, an open review feedback forum does not provide a sufficient means of finding pertinent information because it is highly subjective and fails to adequately address the challenges and potential inherent in the classroom use and adoption of mobile technology applications. Overall, I hope this project will help educators and developers understand that improvements are needed within the feedback mechanisms on Google Play in order to bring meaningful information to teachers.

References

- Abbasi, A., Chen, H., & Salem, A. (2008). Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. *ACM Transactions on Information Systems (TOIS)*, 26(3), 12.
- Anderson, M. (2014, July 7). 88% of consumers trust online reviews as much as personal recommendations. Retrieved from <http://searchengineland.com/88-consumers-trust-online-reviews-much-personal-recommendations-195803>
- AppBrain. (2014). *New android apps per month*. Retrieved from <http://www.appbrain.com/stats/number-of-android-apps>
- AppBrain. (2014). *Number of android applications*. Retrieved from <http://www.appbrain.com/stats/number-of-android-apps>
- Arndt, J. (1967). Role of product-related conversations in the diffusion of a new product. *Journal of marketing Research*, 291-295.
- Bailey, A. A. (2005). Consumer awareness and use of product review websites. *Journal of Interactive Advertising*, 6(1), 68-81.
- Balasubramaniam, N. (2009). User-generated content. In *Proceedings of business aspects of the internet of things, seminar of advanced topics* (pp. 28-33). Zurich: ETH.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American psychologist*, 37(2), 122-147.
- Blackburn, S. (2008). *Oxford Dictionary of Philosophy*, second edition revised. Oxford: Oxford University Press, ISBN 978-0-19-954143-0.
- Brier, A., & Hopp, B. (2011). Computer assisted text analysis in the social sciences. *Quality & Quantity*, 45(1), 103-128.
- British Columbia Ministry of Education. (2013). *BC's education plan*. Retrieved from http://www.bcedplan.ca/assets/pdf/bc_edu_plan.pdf
- Burden, K., & Maher, D. (2014). 15 Mobile technologies and authentic learning in the primary school classroom. *Teaching and Learning with ICT in the Primary School*, 171.
- Callaghan, M. J., Harkin, J., McGinnity, T. M., Woods, D. N., & Harrison, B. (2006, October). Technology Supported Pedagogy in a Flexible, Mobile, Learning Environment. In *Proc. Association of Pacific Rim Universities Conference on Online Distance Learning and Internet Tokyo*.
- Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the “helpfulness” of online user reviews: A text mining approach. *Decision Support Systems*, 50(2), 511-521.

- Cavnar, W. B., & Trenkle, J. M. (1994). N-gram-based text categorization. *Ann Arbor MI, 48113*(2), 161-175.
- Chan, T. W., Roschelle, J., Hsi, S., Kinshuk, Sharples, M., Brown, T., Patton, C., Cherniavsky, J., Pea, R., Norris, C., Soloway, E., Balacheff, N., Scardamalia, M., Dillenbourg, P., Looi, C.K., Milrad, M., & Hoope, U. (2006) One-to-one technology-enhanced learning: an opportunity for global research collaboration. *Research and Practice in Technology Enhanced Learning*, 1, 3-29.
- Cheung, C. M., & Thadani, D. R. (2012). The impact of electronic word-of-mouth communication: A literature analysis and integrative model. *Decision Support Systems*, 54(1), 461-470.
- Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, 43(3), 345-354.
- Chiong, C. (2009). Can video games promote intergenerational play and literacy learning. In *The Joan Ganz Cooney Center at Sesame Workshop*. Retrieved from http://www.joanganzcooneycenter.org/wpcontent/uploads/2010/03/intergen_final_021210.pdf
- Chiong, C., & Shuler, C. (2010). Learning: Is there an app for that? Investigations of young children's usage and learning with mobile devices and apps. In *New York: The Joan Ganz Cooney Center at Sesame Workshop*. http://pbskids.org/read/files/cooney_learning_apps.pdf.
- Chuttur, M. (2009). Overview of the technology acceptance model: Origins, developments and future directions. In *Working papers on information systems* (Vol. 9). Sprouts, USA: Indiana University.
- Cooley, R., Mobasher, B., & Srivastava, J. (1997, November). Web mining: Information and pattern discovery on the world wide web. In *Tools with Artificial Intelligence, 1997. Proceedings., Ninth IEEE International Conference on* (pp. 558-567). IEEE.
- Cosley, D., Lam, S. K., Albert, I., Konstan, J. A., & Riedl, J. (2003, April). Is seeing believing?: how recommender system interfaces affect users' opinions. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 585-592). ACM.
- Coulby, C., Hennessey, S., Davies, N., & Fuller, R. (2011). The use of mobile technology for work-based assessment: the student experience. *British Journal of Educational Technology*, 42(2), 251-265.
- Danescu-Niculescu-Mizil, C., Kossinets, G., Kleinberg, J., & Lee, L. (2009, April). How opinions are received by online communities: a case study on amazon.com helpfulness votes. In *Proceedings of the 18th international conference on World wide web* (pp. 141-150). ACM.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340.

- Engel, J. F., Kegerreis, R. J., & Blackwell, R. D. (1969). Word-of-mouth communication by the innovator. *The Journal of Marketing*, 15-19.
- Evans, C. (2008). The effectiveness of m-learning in the form of podcast revision lectures in higher education. *Computers & education*, 50(2), 491-498.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37.
- Feldman, S. P., & Spencer, M. C. (1965). *The Effect of Personal Influence on the Selection of Consumer Services* (pp. 440-452). Center for regional studies.
- Fernandez, V., Simo, P., & Sallan, J. M. (2009). Podcasting: A new technological tool to facilitate good practice in higher education. *Computers & Education*, 53(2), 385-392.
- Fu, B., Lin, J., Li, L., Faloutsos, C., Hong, J., & Sadeh, N. (2013, August). Why people hate your app: Making sense of user feedback in a mobile app store. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1276-1284). ACM.
- Ganu, G., Elhadad, N., & Marian, A. (2009). *Beyond the stars: Improving rating predictions using review text content* (Doctoral dissertation). Retrieved from WebDB.
- Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *Knowledge and Data Engineering, IEEE Transactions on*, 23(10), 1498-1512.
- Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1-12.
- Goldsmith, R. E., & Horowitz, D. (2006). Measuring motivations for online opinion seeking. *Journal of interactive advertising*, 6(2), 2-14.
- Goodwin, K. (2012). Use of tablet technology in the classroom. *NSW Department of Education and Communities*. Retrieved from http://rde.nsw.edu.au/files/iP-ad_Evaluation_Sydney_Region_exec_sum.pdf
- Goodwin, K. (2012). Use of tablet technology in the classroom. *NSW Department of Education and Communities*.
- Ha, E., & Wagner, D. (2013, January). Do android users write about electric sheep? examining consumer reviews in google play. In *Consumer Communications and Networking Conference (CCNC), 2013 IEEE* (pp. 149-157). IEEE.
- Hess, T. J., McNab, A. L., & Basoglu, K. A. (2014). Reliability generalization of perceived ease of use, perceived usefulness, and behavioral intentions. *MIS Quarterly*, 38(1), 1-28.
- Hillard, D., Purpura, S., & Wilkerson, J. (2008). Computer-assisted topic classification for mixed-methods social science research. *Journal of Information Technology & Politics*, 4(4), 31-46.

- Ho, L. H., Hung, C. L., & Chen, H. C. (2012). Examining the Acceptance Behavior of Using Mobile Phone Messaging as a Parent-Teacher Medium Based on TAM and UTAUT Models. *AISS: Advances in Information Sciences and Service Sciences*, 4(1), 254-263.
- Ho, L. H., Hung, C. L., & Chen, H. C. (2012). Examining the Acceptance Behavior of Using Mobile Phone Messaging as a Parent-Teacher Medium Based on TAM and UTAUT Models. *AISS: Advances in Information Sciences and Service Sciences*, 4(1), 254-263.
- Hoon, L., Vasa, R., Schneider, J. G., & Grundy, J. (2013). *An Analysis of the Mobile App Review Landscape: Trends and Implications*. Tech. rep., Faculty of Information and Communication Technologies, Swinburne University of Technology, Melbourne, Australia, 1-23.
- Hoon, L., Vasa, R., Schneider, J. G., & Mouzakis, K. (2012, November). A preliminary analysis of vocabulary in mobile app user reviews. In *Proceedings of the 24th Australian Computer-Human Interaction Conference ACM*, 245-248.
- Horsky, D., & Simon, L. S. (1983). Advertising and the diffusion of new products. *Marketing Science*, 2(1), 1-17.
- Hu, N., Bose, I., Koh, N. S., & Liu, L. (2012). Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision Support Systems*, 52(3), 674-684.
- Hu, N., Pavlou, P. A., & Zhang, J. (2006, June). Can online reviews reveal a product's true quality?: empirical findings and analytical modeling of Online word-of-mouth communication. In *Proceedings of the 7th ACM conference on Electronic commerce* (pp. 324-330). ACM.
- Huang, Y. M., Kuo, Y. H., Lin, Y. T., & Cheng, S. C. (2008). Toward interactive mobile synchronous learning environment with context-awareness service. *Computers & Education*, 51(3), 1205-1226.
- Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American society for information science and technology*, 60(11), 2169-2188.
- Johnson, L., Smith, R., Willis, H., Levine, A., & Haywood, K., (2011). *The 2011 Horizon Report. Austin, Texas: The New Media Consortium.*
- Jurca, R., Garcin, F., Talwar, A., & Faltings, B. (2010). Reporting incentives and biases in online review forums. *ACM Transactions on the Web (TWEB)*, 4(2), 501-527.
- Katz, E., & Lazarsfeld, P. F. (1970). *Personal Influence, The part played by people in the flow of mass communications*. Transaction Publishers.
- Kearney, M., Schuck, S., Burden, K., & Aubusson, P. (2012). Viewing mobile learning from a pedagogical perspective. *Research in Learning Technology*, 20.
- Khalid, H., Shihab, E., Nagappan, M., & Hassan, A. (2014). What do mobile app users complain about? A study on free ios apps. *Software, IEEE*.

- Koster, J. (1996). Saussure meets the brain. In R. Jonkers, E. Kaan, and J. K. Wiegel, (Eds.), *Language and Cognition 5. Yearbook 1992 of the Research Group for Linguistic Theory and Knowledge Representation of the University of Groningen*. Groningen: University of Groningen.
- Lai, C. H., Yang, J. C., Chen, F. C., Ho, C. W., & Chan, T. W. (2007). Affordances of mobile technologies for experiential learning: the interplay of technology and pedagogical practices. *Journal of Computer Assisted Learning*, 23(4), 326-337.
- Lai, C. H., Yang, J. C., Chen, F. C., Ho, C. W., & Chan, T. W. (2007). Affordances of mobile technologies for experiential learning: the interplay of technology and pedagogical practices. *Journal of Computer Assisted Learning*, 23(4), 326-337.
- Larcker, D. F., & Lessig, V. P. (1980). Perceived Usefulness of Information: A Psychometric Examination*. *Decision Sciences*, 11(1), 121-134.
- Leberknight, C. S., Sen, S., & Chiang, M. (2012). On the volatility of online ratings: An empirical study. In *E-Life: Web-Enabled Convergence of Commerce, Work, and Social Life* (pp. 77-86). Springer Berlin Heidelberg.
- Lee, W. Y., & Brasch, J. J. (1978). The adoption of export as an innovative strategy. *Journal of International Business Studies*, 85-93.
- Legris, P., Ingham, J., & Collette, P. (2003). Why do people use information technology? A critical review of the technology acceptance model. *Information & management*, 40(3), 191-204.
- Lehner, F., & Nosekabel, H. (2002). The role of mobile devices in E-Learning first experiences with a wireless E-Learning environment. In *Wireless and Mobile Technologies in Education, 2002. Proceedings. IEEE International Workshop on* (pp. 103-106). IEEE.
- Lehner, F., & Nosekabel, H. (2002). The role of mobile devices in E-Learning first experiences with a wireless E-Learning environment. In *Wireless and Mobile Technologies in Education, 2002. Proceedings. IEEE International Workshop on* (pp. 103-106). IEEE.
- Li, Y. M., Lin, C. H., & Lai, C. Y. (2010). Identifying influential reviewers for word-of-mouth marketing. *Electronic Commerce Research and Applications*, 9(4), 294-304.
- Lim, C. U., Ortiz, P., & Jun, S. W. (2012). *Automatic star-rating generation from text reviews*. Retrieved from <http://enellepi.googlecode.com/files/report.pdf>
- Liu, J., Cao, Y., Lin, C. Y., Huang, Y., & Zhou, M. (2007). Low-quality product review detection in opinion summarization. In *EMNLP-CoNLL*, 334-342.
- Lowe, W. (2003, April). The statistics of text: New methods for content analysis. In *Midwest Conference*, 1-17.
- Lu, J., Chen, Q., & Chen, X. (2012). App interface study on how to improve user experience. *Proceedings of the Computer Science & Education (ICCSE), 2012 7th International Conference, Melbourne, VIC, 726-729*. doi: 10.1109/ICCSE.2012.6295176

- Lu, M. (2008). Effectiveness of vocabulary learning via mobile phone. *Journal of computer assisted learning*, 24(6), 515-525.
- Mac Callum, K., & Jeffrey, L. (2010). Resistance to the Inclusion of Mobile Tools in the Classroom: The Impact of Attitudes and Variables on the Adoption of Mobile Learning. *Mobile Learning: Pilot Projects and Initiatives*, 143.
- Martilla, J. A. (1971). Word-of-mouth communication in the industrial adoption process. *Journal of Marketing Research*, 173-178.
- Martin, W. C., & Lueg, J. E. (2013). Modeling word-of-mouth usage. *Journal of Business Research*, 66(7), 801-808.
- McGlohon, M., Glance, N. S., & Reiter, Z. (2010). Star quality: Aggregating reviews to rank products and merchants. in *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*.
- Mergenthaler, E. (1996). Computer-assisted content analysis. *Nachrichten Spezial*, 3-32.
- Moore, D. S. (2004). *The basic practice of statistics*. New York, NY: W. H. Freeman and Company.
- Moriarty Jr, R. T., & Spekman, R. E. (1984). An empirical investigation of the information sources used during the industrial buying process. *Journal of Marketing Research*, 137-147.
- Motiwalla, L. F. (2007). Mobile learning: A framework and evaluation. *Computers & Education*, 49(3), 581-596.
- Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon. com. *Management Information Systems Quarterly*, 34(1), 185-200.
- Naismith, L., Sharples, M., Vavoula, G., & Lonsdale, P. (2004). *Literature review in mobile technologies and learning*. NESTA Futurelab series, Report 11, retrieved from http://www.nestafuturelab.org/research/lit_reviews.htm#lr11
- Norris, C., Hossain, A., & Soloway, E. (2011). Using smartphones as essential tools for learning: A call to place schools on the right side of the 21st century. *Educational Technology*, 51(3), 18-25.
- Pagano, D., & Maalej, W. (2013, July). User feedback in the appstore: An empirical study. In *Requirements Engineering Conference (RE), 2013 21st IEEE International* (pp. 125-134). IEEE.
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2), 1-135.

- Park, D. H., & Lee, J. (2009). eWOM overload and its effect on consumer behavioral intention depending on consumer involvement. *Electronic Commerce Research and Applications*, 7(4), 386-398.
- Park, D. H., & Lee, J. (2009). eWOM overload and its effect on consumer behavioral intention depending on consumer involvement. *Electronic Commerce Research and Applications*, 7(4), 386-398.
- Pêcheux, Michel; Hak, Tony & Helsloot, Niels (1995). *Automatic discourse analysis*. Amsterdam: Rodopi.
- Petz, G., Karpowicz, M., Fürschuß, H., Auinger, A., Štríteský, V., & Holzinger, A. (2013). Opinion Mining on the Web 2.0—Characteristics of User Generated Content and Their Impacts. In *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data* (pp. 35-46). Springer Berlin Heidelberg.
- Platzer, E. (2011). Opportunities of automated motive-based user review analysis in the context of mobile app acceptance. *Proc. CECIIS*, 309-316.
- Rau, P. L. P., Gao, Q., & Wu, L. M. (2008). Using mobile communication technology in high school education: Motivation, pressure, and learning performance. *Computers & Education*, 50(1), 1-22.
- Rezabakhsh, B., Bornemann, D., Hansen, U., & Schrader, U. (2006). Consumer power: a comparison of the old economy and the Internet economy. *Journal of Consumer Policy*, 29(1), 3-36.
- Riedl, C., Blohm, I., Leimester, J.M., & Krcmar, H. (2010). Rating Scales for Collective Intelligence in Innovation Communities: Why Quick and Easy Decision Making Does Not Get it Right. *Proceedings of Thirty First International Conference on Information Systems (ICIS '10)*, St. Louis, MO.
- Robey, D. (1979). User attitudes and management information system use. *Academy of Management Journal*, 22(3), 527-538.
- Rogers, E. M., & Shoemaker, F. F. (1971). *Communication of Innovations; A Cross-Cultural Approach*. New York: Free Press.
- Schultz, R. L., & Slevin, D. P. (1975). Implementation and organizational validity: An empirical investigation, in *Implementing Operations Research / Management Science*, R.L. Schultz and D.P. Slevin (Eds.), New York: American Elsevier, 153-182.
- Seipold, J., Pachler, N., Bachmair, B., & Döbeli-Honegger, B. (2013). Mobile learning. Strategies for planning and implementing learning with mobile devices in secondary school contexts. In: Leask, Marilyn; Pachler, Norbert (Eds.), *Learning to teach using ICT in the Secondary School*, 3, London: Routledge.
- Shuler, C., Levine, Z., & Ree, J. (2012, January). iLearn II: An analysis of the education category of Apple's app store. In *New York: The Joan Ganz Cooney Center at Sesame Workshop*.

- Song, Y., Wong, L. H., & Looi, C. K. (2012). Fostering personalized learning in science inquiry supported by mobile technologies. *Educational Technology Research and Development*, 60(4), 679-701.
- Subramanya, S. R., & Farahani, A. (2012). Point-of-view article on: Design of a smartphone app for learning concepts in mathematics and engineering. *International Journal Of Innovation Science*, 4(3), 173-184.
- Swanson, E. B. (1987). Information channel disposition and use. *Decision Sciences*, 18(1), 131-145.
- Teo, T. (2012). Examining the intention to use technology among pre-service teachers: an integration of the technology acceptance model and theory of planned behavior. *Interactive Learning Environments*, 20(1), 3-18.
- Traxler, J. (2009). Learning in a mobile age. *International Journal of Mobile and Blended Learning (IJMBL)*, 1(1), 1-12.
- Trusov, M., Bucklin, R. E., & Pauwels, K. (2009). Effects of word-of-mouth versus traditional marketing: findings from an internet social networking site. *Journal of marketing*, 73(5), 90-102.
- Valstad, H., & Rydland, T. (2010). iPad as a pedagogical device. *Norwegian University of Science and Technology*.
- Vasa, R., Hoon, L., Mouzakis, K., & Noguchi, A. (2012, November). A preliminary analysis of mobile app user reviews. In *Proceedings of the 24th Australian Computer-Human Interaction Conference* (pp. 241-244). ACM.
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.
- VentureBeat. (2013). *Comparing apples and googles: The App Store vs. Google Play (infographic)*. Retrieved from <http://venturebeat.com/2013/07/17/comparing-apples-and-googles-the-app-store-vs-google-play-infographic/>
- Wan, Y., & Nakayama, M. (2014). The reliability of online review helpfulness. *Journal of Electronic Commerce Research*, 15(3), 179-189.
- Whyte Jr, W. H. (1954). The web of word of mouth. *Fortune*, 50(1954), 140-143.
- Wiedemann, G. (2013, May). Opening up to Big Data: Computer-Assisted Analysis of Textual Data in Social Sciences. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (Vol. 14, No. 2).
- Wiedemann, G. (2013, May). Opening up to Big Data: Computer-Assisted Analysis of Textual Data in Social Sciences. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* 14(2). Retrieved from <http://www.qualitative-research.net/index.php/fqs/article/view/1949/3552>

Wikipedia. (2014). List of mobile software distribution platforms. Retrieved from http://en.wikipedia.org/wiki/List_of_mobile_software_distribution_platforms

Wu, F., & Huberman, B. A. (2010). Opinion formation under costly expression. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 1(1), 5.