Norwegian School of Economics

Bergen, Fall, 2017

Electronic Word-of-Mouth and Consumers' Intention to Purchase

A study of the effects of valence and argument quality of online reviews

Ingrid Rabben & Vilde Instanes Larsen

Supervisor: Einar Breivik

Master Thesis, Economics and Business Administration,

Marketing and Brand Management

NORWEGIAN SCHOOL OF ECONOMICS

This thesis was written as a part of the Master of Science in Economics and Business Administration at NHH. Please note that neither the institution nor the examiners are responsible − through the approval of this thesis − for the theories and methods used, or results and conclusions drawn in this work.

Abstract

As technology has developed and phenomenon such as the participative web has emerged, the opportunities of communication and sharing has expanded simultaneously. The participative web represents a revolution where users are empowered to develop, collaborate, distribute and customize Internet content and applications. Word-of-mouth (WOM) has been known for decades to be of profound influence in consumers purchasing behavior and as consumers share their experiences and opinions, the need for advertisements decreases. Electronic word-of-mouth (eWOM) has enabled users to share information and opinions with others within seconds giving them the power to influence other consumers' behavior. Although previous research on eWOM has shown effects of valence and argument quality, no studies have explored these two concepts combined. For this reason, the purpose of this thesis is to examine the relationship between valence and argument quality of online consumer reviews on consumers' purchase intention.

By conducting an online experiment and thereafter analyzing the data from 155 respondents, we discovered two main results. Firstly, we found that valence had a direct effect on consumers' purchase intention, where the effect of negative online consumer reviews exceeds the effect of positive reviews. Secondly, we found that argument quality did not moderate the effect of valence on consumers' intention to purchase. Our findings suggest that consumers focus more on the valence of an online review rather than the relevance of the content. Furthermore, we discovered that the use of online reviews is widespread and a key factor in consumer decision-making. Due to this, we suggest that both researchers and companies implement a greater focus on eWOM in the future.

Preface

This thesis is written as a part of our Master of Science in Economics and Business Administration within the Major Marketing and Brand Management at the Norwegian School of Economics (NHH).

Online consumer behavior is a topic of great interest for both of us, and was important when choosing the subject for this thesis. Working with this thesis has been a challenging process, but overall very interesting and educational. We both feel we have acquired valuable experience conducting this research.

We would like to thank our supervisor, Professor Einar Breivik, for his valuable feedback and guidance. We have greatly appreciated his expertise, as well as his quick responses.

Finally, we would like to thank our family and friends for great support and encouragement throughout the work of this thesis.

Bergen, December 2017

Ingrid Rabben and Vilde Instanes Larsen

Contents

List of Figures

List of Tables

1. Introduction

The Internet has transformed the way we gather and provide information, as well as communicate with others. The accessibility of information is arguably one of the most important ways the Internet has changed traditional ways of shopping. Ready access to information helps consumers use less time on deciding on which product that best fits their needs and preferences. On the other hand, unlimited sources of information and the increasing number of options for consumers, is making it more and more difficult to be confident that you are making the right purchase decision. The digitalization has opened for communication between consumers, and from this derives the term electronic word-of-wouth, commonly referred to as eWOM (Hennig-Thurau et al., 2004). With eWOM, consumers have the opportunity to spread their opinions and experiences with brands, products or services on various platforms such as social media, review websites and retailer websites. They can reach thousands of other consumers all over the world within seconds. In addition, the information is available for an indefinite period of time. Furthermore, considering the power of eWOM being greater than advertisements or other market-generated information, it is even more important for managers and marketers to pay attention to this development (Bickart and Schindler, 2001). eWOM content is also not possible for companies to control. Thus, eWOM can be crucial for the reputation of businesses. Within minutes, a business or product could be branded positively or negatively in the minds of consumers in just a click.

Online reviews in particular, have proven to be of great importance in consumer decision-making (Bickart and Schindler, 2001). Consumers turn to online reviews for help in decisions regarding almost all products or services available to buy. From deciding where to eat, where to travel or stay, to the choice of carpenter or painter. The Fan and Fuel survey (2016) from December last year show that 97 % of consumers implement consumer reviews into their purchase decision and 92 % are hesitant to make a purchase if there are no consumer reviews available. In addition, the Local Consumer Review Survey of 2017 shows that 85% of consumers trust online reviews as much as family and friends (BrightLocal, 2017). This emphasizes the importance of online reviews.

A problem that arises when reading online reviews, is whether the source can be considered credible and trustworthy or not. Given that there is no standard format of online

reviews and the communicator often has the possibility of writing anonymously, it is difficult to be certain that what you are reading is relevant and stems from a credible source (Schindler and Bickart, 2005). In a perfect world, all online reviews would provide objective information containing arguments based on the functions and performance of the product. There would be no incentives to write online reviews other than to help other consumers in making the perfect decision amongst the wide range of options. This indicates that fake reviews and reviews based on emotions or monetary incentives would not exist. Unfortunately, this is not the case. People can have a variety of different motives to write an online review, which is hidden from the public eye, making it hard for consumers to distinguish between helpful and irrelevant reviews (Metzger, 2007). In 2014 it was reported that every minute 26.380 reviews were posted on the review site Yelp alone (DMR, 2017). Controlling this vast amount of reviews is impossible, and it is common knowledge that there are many fake and irrelevant reviews posted online. A recent example of fake reviews from the review site Tripadvisor is a restaurant in London accomplishing being ranked as the best, without even existing (DN, 2017).

According to the 2017 Global Online Consumer Report, 92 % of reviews shared online are positive, and only 2 % are negative (KPMG, 2017). Regardless, if a product has received only one negative review, this can be enough for 35 % of consumers to refrain from purchasing the product (Fan&Fuel, 2016). This shows that negative reviews hold great power over consumers in a purchase decision.

Previous researchers have investigated several aspects regarding different characteristics of eWOM and online reviews. A widespread finding is the negativity effect, showing that negative reviews have a stronger impact on consumer behavior than positive reviews (Park and Lee, 2009; Lee and Lee, 2009; Lee and Youn, 2009; Sen and Lerman, 2007). In addition, argument quality of online reviews has been shown in several studies to be an important influential factor (Cheung et al., 2009; Cheung et al., 2008). To the best of our knowledge, the negativity effect has not been addressed in conjunction with the quality of the reviews and we therefore want to investigate this further on Norwegian consumers. The aim of this study is to contribute to previous studies on the effect of online reviews on purchase intention, and with basis in the previous discussing, we propose the following research question:

***RQ**: What effect does online reviews have on purchase intention, and what is the impact of valence and argument quality of the online reviews?*

To answer this research question, we will start with reviewing existing relevant theory and literature, which will form a basis for developing three hypotheses and a conceptual model, given in Section 2. Furthermore, we will elaborate on choices regarding research design and data collection. We will continue by performing the analyses required to answer our hypotheses and present the results. Thereafter, we will discuss our findings and connect this to both theoretical and managerial implications. Finally, we will elaborate on the limitations of our study and provide suggestions for future research, before closing the thesis with a conclusion.

2. Theory

In this section we will elaborate on relevant theory and provide a review of previous literature on electronic word-of-mouth. We will start with presenting theory about the purchasing process, followed by an overview of theory on the concept eWOM. Finally, we will present a literature review.

2.1 Purchasing Process

The purchasing process includes everything a consumer does in the context of a purchase, from discovering a need to an actual purchase and following usage (Solomon et al., 2011). Traditionally, the belief was that when consumers face a purchase decision problem, they would carefully collect all information needed about different products and weigh positive and negative attributes against each other to make a satisfactory decision that will lead to no regret. Later studies on consumer behavior has shown that this is not the case and that the motivation to exert an effort in information search depends on the importance of the decision. Kahneman (2013) introduced two concepts that categorize our processing depending on how much effort we put into a situation. He characterizes System 1 processing as fast, instinctive and emotional elaboration, while system 2 is slower, more deliberative and more logical reasoning. System 2 requires attention and concentration compared to system 1, which is characterized by impulse. Kahneman argues that because of limited attention, system 2 forces a consumer to focus on the aspects that he believes are important and thus, other aspects get less attention or no attention at all. This often leads to consumers being affected by heuristics and biases in their decision-making process, for instance the negativity effect which will be elaborated on later.

The consumer decision-making process can be divided into five steps, depicting the different stages a consumer engages in when faced with a purchase decision (Solomon et al., 2011). Whether the consumer go through all the steps or not, depends on the level of involvement, i.e. the use of system 1 or 2, where a higher perceived risk leads to higher involvement (Hoyer et al., 2012). Factors that may reinforce the risk is price, complexity of the product, frequency of buying or social visibility. The first step in the decision-making process is problem recognition, which will be apparent when there is a perceived difference between the actual and the ideal state. The second step is information search, where the

consumer gathers information to solve the recognized problem (Solomon et al., 2011). In the third step, evaluation of options, the consumer will judge the alternatives found in step two. The fourth step is the product choice, and where the consumers will complete a purchase. The fifth and final step is post-purchase evaluation, and the consumers will in this stage evaluate on whether the product chosen met their expectations or not. The five stages are illustrated in Figure 1. We will elaborate on steps two and three in the following sections.

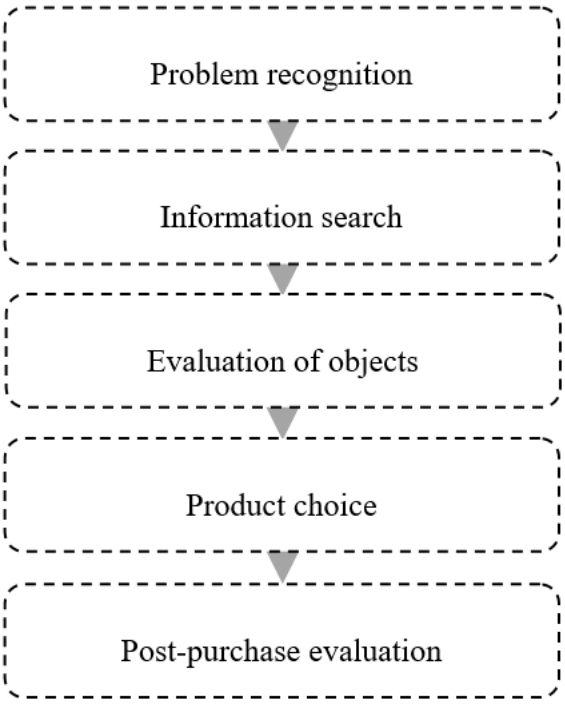


Figure 1: The five steps in the consumer decision-making process

2.1.1 Information Search

To be able to solve a decision problem, the consumer needs to gather information about different options (Solomon et al., 2011). Typically, an internal search, processing information from the consumer's memory, is the first part of this step (Hoyer et al., 2012). The consumer's evoked set of alternatives consists of the top-of-mind brands and corresponding information and associations. If the information recalled from the consumer's memory is not sufficient, the consumer will engage in an external search for additional information. When consumers have little knowledge of a brand or product, they look for other consumers' opinions. Sources of information might be advertisements, media, family

and friends, social media, retailers, product websites or other sources of information from the Internet. The Internet contains loads of information that might be overwhelming for the consumer. Most consumers gather information from solely a couple of different websites, as searching the Internet for information can be time consuming and difficult because of the amount of information available.

Internet sites including product information, price comparisons, consumer reviews and product ratings are helpful for the consumers as they can get answers to many of their questions in one place and thus save time. Such information, not controlled by marketers, are often seen as more credible and has a great influence on the consumer's purchase decision (Bickart and Schindler, 2001). Consumers' information search on the Internet can be different based on product type (Hoyer et al., 2012). With experience goods, consumers spend more time on each website thoroughly reviewing the product specifications and reading consumer reviews to get a better understanding of the product. This as opposed to information search on search goods, where consumers most likely visit multiple websites, but with a shorter time span. Research show that for hedonic or symbolic products interpersonal sources are important, as consumers want to minimize the social risk. When consumers are faced with a decision they consider risky, they will do a greater effort in external search for information than decisions with low risk. Thus, their involvement is higher when the risk is higher.

2.1.2 Evaluation of Options

When the consumer considers the information search satisfactory, the next step in the process is to evaluate the options (Solomon et al., 2011). The consumer will judge the different alternatives based on expected performance or attributes and narrow down the options. When consumers evaluate a product, they often use an anchoring and adjustment process (Hoyer et al., 2012). This entails that the consumer has an initial judgement from memory or personal values that can be adjusted when receiving additional information about the product. A negative first judgement, or anchor, is more difficult to change compared to an initial positive judgement which is likely to stay positive. The confirmation bias also implies that consumers value higher judgements that are in accordance with their own beliefs, and may ignore information in contrast to these. The Internet provides a vast amount

of information and it is difficult for the consumer to sort out which information is helpful (Schindler and Bickart, 2005). This often leads to the use of "rules of thumb" by focusing on familiar brands, salient options or other cues the consumer finds relevant (Solomon et al., 2011). Solomon et al. refer to this type of behavior as limited problem solving, indicating that the consumer is somewhere between the two extremes habitual decision making and extended problem solving. The phenomenon called the negativity effect, that consumers weigh negative information more heavily than positive information, is a possible outcome of this because negative information is categorized as more salient and diagnostic than positive information (Skowronski and Carlston, 1989). Each consumer can have their own individual evaluative criteria, expressing which factors are important to choose amongst the alternatives (Hoyer et al., 2012). These criteria can be for instance style, price or performance. Consumers also differ in the evaluative step whether they process information by brands or by attributes. Some consumers evaluate from a compensatory perspective, where good attributes can compensate for bad ones, while others from a non-compensatory perspective, eliminating an option with bad attributes.

Determined by the attitude towards the product after buying it, the consumer will experience either satisfaction or dissatisfaction based on whether the product met the consumer's expectations or not (Solomon et al., 2011). Both the information search process and evaluation of the options is important steps to reduce what we call cognitive dissonance after purchase. This concept refers to the feeling of anxiety or regret on whether the right decision was made and often occurs when more than one alternative is attractive, and the decision is important. If the product does not meet the expectations, the dissatisfaction can lead to negative word-of-mouth. It is therefore important for marketers to portray the expectations of a product correctly to avoid a dissatisfied customer.

2.2 User-Created Content

User-created content (UCC), also referred to as user-generated content (UGC) or consumer-generated content (CGC), has emerged from the concept of the "participative web" also referred to as Web 2.0 (OECD, 2007). This term encompasses the increasing participation and interaction between Internet users, making the Internet a platform to communicate and express themselves. Further, the participative web represents a revolution where the Internet

has adopted new technologies where users are empowered to develop, collaborate, distribute and customize Internet content and applications. There are several different definitions of UCC. Tabbane and Debabi (2015) define it as “Any form of original content, available online and published by users whose motivations are devoid of commercial purposes.” OECD (2007) use three characteristics to describe what UCC is and what it is not. Firstly, the content must be published in some context, excluding content sent by for instance email or instant message. In addition, a certain amount of creative effort must be put into making a content their own, meaning merely copying and publishing existing content is not UCC. Lastly, the creation of the content should take place outside of any professional routine, excluding content created for commercial purposes.

2.3 Electronic Word-of-Mouth

Word-of-mouth (WOM) has been known for decades to be of great influence in consumers purchasing behavior and is often considered as a critical aspect of brand building because as consumers share their experiences and opinions, the need for advertisements decreases (Keller, 2013). As Internet usage has increased and phenomenon such as the participative web and UCC has emerged, the opportunities and development of communication has expanded simultaneously. Electronic peer-to-peer communication has enabled users to share information and opinions with others in a much easier way than previously (Hennig-Thurau et al., 2004). In turn, this has given consumers the opportunity and power to influence other consumers through posting content with reference to consumption experiences of products and services. Electronic peer-to-peer communication is often referred to as eWOM, and Hennig-Thurau et al. (2004, p. 39) describe it as: “Any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet.”

There are many different platforms consumers can use to publish statements about a product, brand or company. The most common are personal blogs (e.g. through blogg.no), discussion forums (e.g. kvinneguiden.no), review websites (e.g. epinions.com, yelp.com or tripadvisor.com), retail websites (e.g., amazon.com) and social networking sites (e.g. facebook.com).

2.3.1 WOM vs eWOM

Despite the similarities between WOM and eWOM, there are several important factors that differentiates them as two different concepts. Firstly, the eWOM communication network is considerably greater than that of traditional WOM (Hennig-Thurau et al., 2004; Park et al., 2007). In comparison to face-to-face communication, a published statement on the web can reach far beyond the local community and personal connections, making the potential impact on consumers substantially greater. Because of this, information can easily reach numerous of consumers in a short amount of time. Sharing on social media is a fitting example of how easy it is today to create awareness or "buzz" around a topic or opinion. In addition, the size of the network of eWOM contributes to a larger volume of statements compared with traditional WOM.

Another key difference is the communication form (Park et al., 2007). While traditional WOM is made orally, eWOM is done in writing, making it observable, measurable and possible to access for an indefinite period of time. Since the information is stored online, the amount of accessible information is much larger for eWOM than for WOM. Being that WOM is shared orally, you need to be present at the exact time the information is being shared to get access to the information. eWOM has disrupted the geographical and temporal constraints, making the messages more persistent and accessible.

An important difference between WOM and eWOM is the familiarity of the messenger (Park et al., 2007). Conversations done face-to-face are often private in nature, making the tie strength and familiarity between the information senders strong. This is in contrast with eWOM where most of the communication occur between strangers and can usually be described as one-way communication. eWOM has the power to create communities of information exchange that is so large that most of the users are unknown to each other. However, this could make it more difficult for the consumer to filter out relevant information and be able to compare oneself with the messenger. Characteristics that are important for one consumer, may not be important at all for another. Furthermore, eWOM can be published anonymously, which lowers the threshold of stating your opinion because the social risk is lower.

In summary, the potential that eWOM has to damage or improve a company's reputation is significantly higher than for WOM.

2.3.2 eWOM Characteristics

Cheung and Thadani (2012) have classified the most common eWOM characteristics into argument quality, recommendation framing, recommendation sidedness, number of reviews, review type, recommendation rating, recommendation consistency, review rate and sales volume. The following section will elaborate on argument quality and recommendation framing.

As of today, there has not been established a consensus on the different components determining argument quality. Park et al. (2007) defines argument quality from the perspective of information characteristics, i.e. relevance, understandability, sufficiency, and objectivity. Further, they argue that the quality varies from subjective and emotional to objective and logical, with the latter being arguments of high-quality. Thus, a high-quality argument supports the arguments with reasons based on specific facts about the product and a low-quality argument contrasts with this offering only vacuous and often irrelevant information. Regarding studies done on the topic of eWOM, argument quality has received much attention over the years (Cheung and Thadani, 2012) and findings show that reviews of higher quality are perceived as more credible (Cheung et al., 2009), resulting in a more favorable attitude towards the product (Lee et al., 2008) and consequently higher scores on purchase intention (Park et al., 2007).

Recommendation framing, also called valence, categorize the argument based on the information being positive, negative or neutral (Cheung and Thadani, 2012). Studies have shown that consumers weigh negative information more heavily than positive information in decision-making tasks across a wide range of domains (Baumeister et al., 2001; Rozin and Royzman, 2001; Kanouse, 1984). Skowronski and Carlston (Skowronski and Carlston, 1989) argue that the mechanisms behind this negativity effect can be explained by using the category diagnosticity approach. Consumers use informational cues available to categorize other individuals, or in this case products. Information is perceived as diagnostic if it helps the consumer assign a product to a cognitive category, for instance high or low quality. When forming overall product judgements, negative information tends to be weighed more heavily than positive information because consumers perceive negative product information as more diagnostic than positive information. In essence, the performance of a high-quality product must be good most of the time to retain that categorization, while a low-quality product only needs to perform badly in some cases to be perceived as low quality. Positive

information is often more ambiguous because any product can have some positive attributes, and it is thus difficult to categorize a product as high quality only by evaluating positive information (Herr et al., 1991; Bone, 1995). In a study by Weinberger and Dillon (1980), subjects were given either positive or negative ratings for a set of unbranded goods and services and afterwards asked to indicate their intention to purchase the item. The results show that the valence did have an impact on purchase intention and that in general, unfavorable product ratings had a greater impact on purchase intention than favorable ratings. In more recent years several studies have been done on the negativity effect regarding eWOM on credibility (Cheung et al., 2009), attitude towards the product (Sen and Lerman, 2007) and intention to purchase (Park and Lee, 2009; Lee and Youn, 2009) and results show that the effect is constant.

2.3.3 Motives for Reading eWOM

It is common to distinguish between two types of behavior regarding consumer engagement in eWOM communication: a transmission behavior of eWOM and an exposure behavior to eWOM (Hennig-Thurau and Walsh, 2003; Hennig-Thurau et al., 2004). The latter describes factors relating to why consumers choose to expose themselves to eWOM messages and in the following we will elaborate on this behavior.

Hennig-Thurau and Walsh (2003) did research on eight different motives for reading eWOM and categorized them into five factors as follows: obtaining buying-relevant information, social orientation through information, community membership, remuneration and to learn how to consume a product. The first factor represents motives involving reduction of both risk and time. Today there are a wide variety of products available both online and offline and the information and specifications given by the seller is often long and tiresome. eWOM help consumers collect information necessary for them to make a purchase decision in a fast and inexpensive way. When consumers face decision where the risk is perceived to be moderate to high, they are more motivated to search for information. Perceived risk for the consumer depends on several factors and in general it can be said to be higher when little information is available, when the product is new, technologically complex or expensive, when brands differ substantially in quality, when the consumer is

likely to be judged based on the acquisition, usage or disposition of the product (Hoyer et al., 2012).

The four remaining factors will subsequently be explained briefly. The second factor contains items from determination of social position and dissonance reduction (Hennig-Thurau and Walsh, 2003). This involves evaluating the product in terms of prestige or to reduce cognitive incongruence that can arise when there are conflicting information concerning alternative offers and from different sources. The third factor relates to elements regarding affiliation to a virtual community and acquiring knowledge of which products are new in the marketplace. The fourth factor, remuneration, refers to situations where consumers are being offered monetary incentives for reading and evaluating consumer reviews of a product or service. The research show that this motive has no impact on buying behavior and can have a negative impact on the consumers' interest in the information itself. The last factor refers to learning about new products and how to consume them. According to Park et. al. (2007) and Godes et. al. (2005), consumers rely more on eWOM as product characteristics become more complex and technical.

In addition to the five factors above, studies have shown that people also read eWOM for fun and entertainment (Schindler and Bickart, 2005).

2.3.4 Impact of eWOM

As discussed in section 2.3.1 *WOM vs eWOM*, the most important characteristics that make eWOM a considerably more powerful tool than traditional WOM, is the potential reach and the amount of accessible information at any given point in time. The impact that eWOM can create, give rise to new opportunities for both consumers and businesses and we will elaborate on this in the following.

The greatest benefit for consumers involves the aspect of information and knowledge. Because of the extensive network size, consumers can search for information about almost anything and find answers. Because of this, consumers can with less effort find products that matches their needs and thus make a more informed buying decision, especially if the consumers are novices in the product category (Chen and Xie, 2008). Novices, or unsophisticated users, may avoid making a purchase decision if only seller

created information is available, because they are unable to process the information correctly and thereby select a suitable product.

Bickart and Schindler (2001) suggests that compared to regular marketing, information on Internet forums should be more influential on consumer behavior because of credibility, relevance and empathy. Consumers often feel that traditional marketing and advertising are manipulating and not concerned about the consumer's wellbeing. In contrast, research have shown that both WOM and eWOM is perceived as more credible and trustworthy because the authors are fellow consumers and thus have no underlying motives of persuading others to buy a specific product. However, there have been several cases where corporations have been caught publishing fake consumer reviews to either boost their own reputation or impair the reputation of competitors. In addition, compared to seller created information, opinions of fellow consumers are expected to be of greater relevance and therefore lead to increased persuasive influence because of greater similarity between the reader and the informant. Information from consumers is often performance related and can contain both positive and negative sides of the product, in contrast with seller created information where they emphasize positive sides and conceal the negative sides. Regarding empathy, many of the contributions in online platforms consist of stories from personal experience and therefore contain elements of both entertainment and education. These stories have an ability to enchant the reader making them empathize with the author's feelings and experience. Empathy can indirectly affect the behavior of the consumer because the feelings generated by the story can be transferred to the actual product. The advantages explained above can relate to both WOM and eWOM. The opportunities of eWOM thus needs to be viewed with the advantages of eWOM in mind, namely the reach and written form.

Firms can also benefit from eWOM. Firstly, it is a cost-effective tool for establishing product awareness as well as acquiring and retaining customers (Mayzlin, 2006; Dellarocas, 2003). Dellarocas also states that eWOM can assist a company in product development by learning about how consumers react to its products and about their likes and dislikes in the product category. Bickart and Schindler (2001) found through their research that online discussions is more powerful than seller-generated information regarding stimulating product category interest and suggest that focusing on relational elements can generate product interest for many consumers. In addition, online feedback may help identify problems faster and thereby collect resources needed to solve the problem more rapidly. For new products, eWOM may serve as a risk reduction tool that is more effective than

traditional advertising. If little information exists about the product, talking to other consumers might help to simplify complexity and increase their confidence in buying the product (Berger, 2014).

2.3.5 Challenges of eWOM

Although there are a variety of opportunities and benefits to extract from eWOM communication as mentioned in section 2.3.4 *Impact of eWOM,* there are challenges and drawbacks as well that we need to address.

A factor widely discussed in previous literature is regarding the reliability of the source and the content of the information given in eWOM (Schindler and Bickart, 2005; Chatterjee, 2001; Lee and Youn, 2009; Dellarocas, 2003). In traditional WOM settings people rely on social cues such as personal relations, facial expressions and personality traits, but the anonymous nature of eWOM prevent consumers from gaining such knowledge and are faced with the task of evaluating information from complete strangers. Because of this, it is difficult for consumers to assess the author's background and expertise on the given topic (Schindler and Bickart, 2005; Lee and Youn, 2009). Given that the writers are publishing their opinions to complete strangers, ethical aspects are not a large concern and there is a greater possibility that the information is inaccurate or irrelevant. In reality, there are in most cases no regulation ensuring standards of reliability of the content accessible to the public (Metzger, 2007). It is thus exceptionally important for consumers to evaluate eWOM carefully. Further, the content of eWOM has no standardized form and therefore it ranges from simple statements containing only an expression of disapproval to highly nuanced assessments of both the benefits and problems regarding the product. Making the evaluation process even more difficult, anonymity facilitates strategic manipulation of the source. Firms, or consumers on behalf of firms, can cheat the system and boost their own reputation with positive eWOM of their own products, or worse, dishonest negative feedback of competitors' products (Dellarocas, 2003; Lee and Youn, 2009; Chatterjee, 2001). As argued in section 2.3.2 *eWOM Characteristics,* the accessibility and range of eWOM makes the potential impact on both products and companies immense (Park et al., 2007). The negativity effect and the category diagnosticity approach can make the impact even greater and it is thus clear that negative information can disrupt a brand or product image

(Dellarocas, 2003). Consumers spreading bad words is out of the company's control and to make the matter worse, the statements made by a consumer might not necessarily reflect the characteristics or quality of the product. As suggested by Hennig-Thurau et al. (2004), negative articulations online can be used as an instrument of power to try to hurt the company's reputation. Thus, the information given could be false and for instance given in spite to the company after a bad experience with them.

2.4 eWOM and Purchasing Process

eWOM is strongly related to the consumer decision-making process and the associated stages. Need recognition could be said to be the least related to eWOM, but even at this stage there exist possibilities of consumers discovering a need when exposed to some sort of eWOM. At the information search stage, however, eWOM plays an important role (Lee and Lee, 2009). Online reviews in particular, is frequently used by consumers in order to make a judgement on a product. With almost unlimited options, and information overload, the online reviews can guide consumers through a at times overwhelming set of alternatives. There are several sources for online reviews, such as review sites, social media, online stores or product sites. The credibility of eWOM is considered high by consumers (Cheung et al., 2009). Negative reviews are unlikely to derive from marketers or promoters of the product and they are often written by consumers sharing their personal experiences to warn others. As most consumers want to avoid taking risks, they would stay clear of products with negative reviews. Overall, consumers have greater trust in others experiences than ads or seller-created information (Bickart and Schindler, 2001). Elaborating on consumer reviews can thus be an essential part of evaluating the options available, also described as stage three in the purchasing process. This further form the basis of the consumer's purchase decision. eWOM or online reviews could in turn be derived from the consumers' post purchase evaluation of a product or service. If the product does not meet their expectations, the dissatisfaction can lead to negative word-of-mouth. The valence of the reviews will depend on whether the product met the consumer's expectations. The online reviews will again be visible for consumers contained at the information search step in the consumer decision-making process, evaluating the information available on whether to buy or not. Prior consumers' reviews are thus a part of deciding future consumers' purchase intentions.

eWOM is therefore an important part of consumers overall decision-making process, as it is present both in information search, evaluation and decision, as well as in the post purchase evaluation. Figure 2 illustrates this.

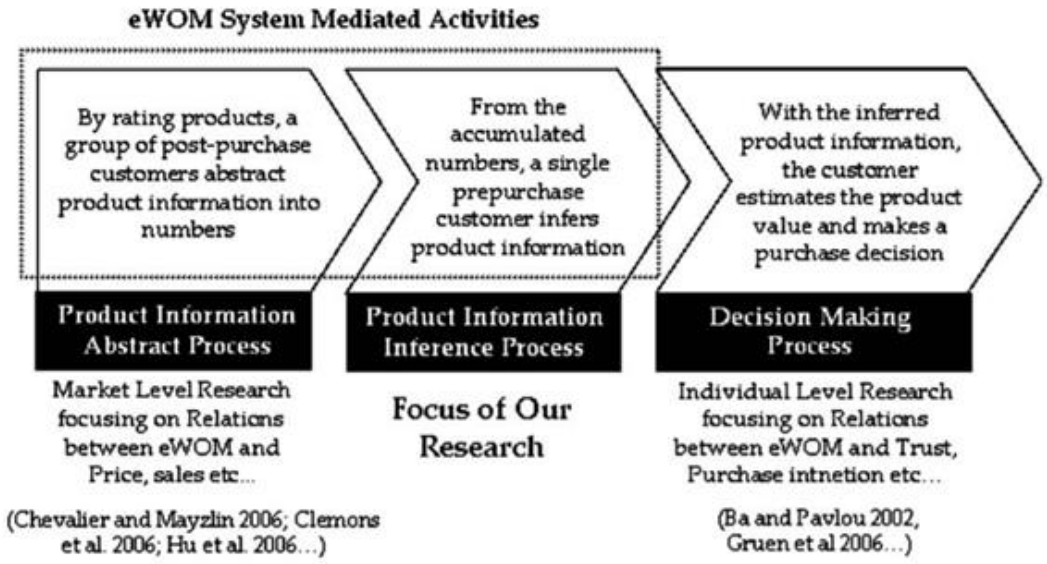


Figure 2: eWOM activities (Lee and Lee, 2009)

2.5 Literature Review

This section will elaborate on prior individual-based research done on eWOM and summarize common themes and findings.

The effect of WOM has been recognized in the literature for decades, while the concept of eWOM only has been apparent since the early 2000s. Most articles on eWOM is experimental studies where four main elements are considered: the main effect (response), the content (stimulus), the readers (receivers) and the writers of eWOM (communicators) (Cheung and Thadani, 2012). The most frequent main effect studied are purchase, purchase intention and attitude towards a product or brand. Other effects studied are for instance information usefulness and credibility of the reviews. There are several aspects of online reviews, i.e. the stimulus, that can be investigated. The most common aspects are valence, orientation of the reviews and quantity. In addition, it is possible to investigate different aspects of consumers who expose themselves to online reviews (receivers) or consumers who write online reviews (communicators). The most frequent factors investigated regarding

receivers are involvement and consumer expertise, and for communicators, source credibility is most common.

Regarding purchasing behavior and purchase intention, previous research on both traditional WOM and eWOM suggest that consumers use online feedback mechanisms as an influential factor in purchase decisions (Arndt, 1967; Bickart and Schindler, 2001; Dellarocas, 2003). Studies focusing on valence have shown that the negativity effect is prominent (Park and Lee, 2009; Lee and Lee, 2009; Lee and Youn, 2009; Sen and Lerman, 2007). Regarding studies investigating argument quality, the common findings indicate that high argument quality is more influential than low argument quality (Cheung et al., 2008; Cheung et al., 2009) and the effect is greater when involvement is higher (Lee et al., 2008; Park et al., 2007). Park et al. also showed that the volume has a significant effect on purchase intention, but involvement did not play any significant role here. The same effect has been shown for attitude towards the product (Lee et al., 2008) and for book sales (Chevalier and Mayzlin, 2006). Park and Kim (2008) combined volume with consumer expertise and found that a higher number of reviews was a more important factor for novices rather than experts regarding purchase intention. Regarding source credibility of the communicator, Chu and Kamal (2008) combined perceived blogger trustworthiness with argument quality and found that argument quality has a greater impact on consumers brand attitudes when perceived blogger trustworthiness is high. The fact that stronger source credibility leads to stronger attitudes is consistent with other findings on information adoption (Zhang and Watts, 2008) and intention to book a hotel (Xie et al., 2011).

3. Model and Hypotheses

In this section we will present the conceptual model and its corresponding hypotheses. As discussed in section 1 *Introduction*, the objective of this study is to investigate consumers' information processing, specifically the effect of online reviews on purchase intention, leading us to the following research question:

RQ: *What effect does online reviews have on purchase intention, and what is the impact of valence and argument quality of the online reviews?*

3.1 Conceptual Model

This study is based on Model 1 in the PROCESS computational tool by Hayes (2013) as shown in Figure 3. Figure 4 gives a visual representation of the concepts this study will investigate and the proposed relationship between them. We propose that valence has a direct effect on intention to purchase, while argument quality moderates this effect. To answer the aforementioned research question, the authors propose four hypotheses which in turn will be elaborated on in the following section.

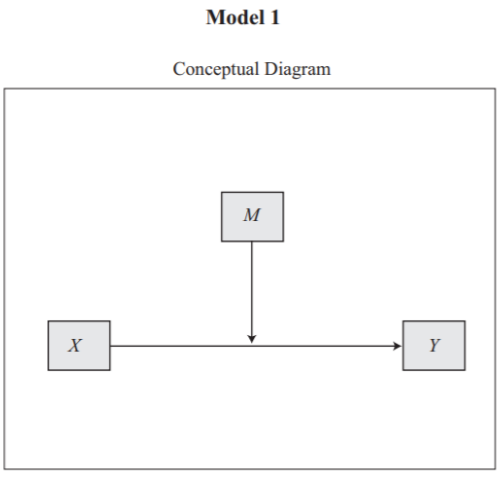


Figure 3: Model 1 (Hayes, 2013)

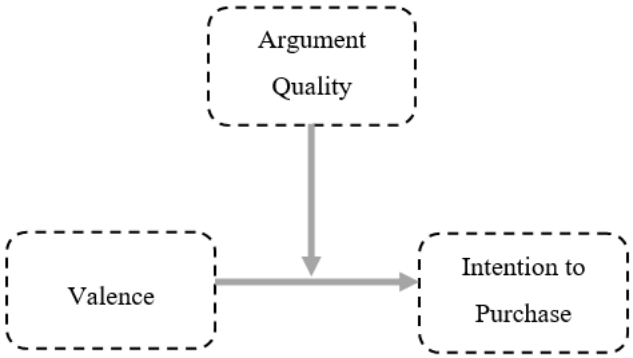


Figure 4: Conceptual model with moderation

3.2 Hypotheses

As mentioned in section 2.3.2 *eWOM Characteristics*, positively framed eWOM highlights the strengths of a product, while negatively framed eWOM concerns the weaknesses and problems of a product (Dellarocas, 2007). Chevalier and Mayzlin (2006) found through their study of two Internet retail websites that consumer reviews have an impact on book sales and thus the purchasing behavior of consumers. In addition, several researchers have investigated the effect of valence on purchase intention, and found that this effect is significant (Park and Lee, 2009; Lee and Youn, 2009). In view of the previous arguments, we believe that people view consumer reviews as a factor directly influencing the purchasing decision process. Thus, the valence of eWOM contribute to respectively encourage or discourage consumers to adopt the product and we hypothesize that:

H1: Valence of online reviews has a direct effect on consumers' purchase intention

Consumers encounter both positive and negative reviews while browsing products online. A finding that has been well recognized in previous research is that negative information has a stronger influence on consumer behavior than positive information (Park and Lee, 2009; Lee and Lee, 2009; Lee and Youn, 2009; Sen and Lerman, 2007). The influence of the negativity effect on purchasing behavior has previously been studied regarding UCC, both offline and online, and the results indicate that the bias is applicable to

several aspects. Wright (1974) found that when under time pressure, consumers are more influenced by negative than positive information in product evaluations. Park and Lee (2009) found through their study that negative eWOM has a greater influence on purchase intention than positive eWOM for both search goods and experience goods. Weinberger and Dillon (1980) found that when provided with consumer-dominated ratings, unfavorable product ratings tended to have a greater impact on purchase intention than did favorable ratings. This effect was prominent regarding both products and services. Sen and Lerman (2007) found that their experiments support the existence of a negativity bias for utilitarian products, but not for hedonic products. Lee et al. (2009) found through their study that extremely negative reviews had a greater impact on brand attitude than both moderately negative reviews and extremely positive reviews. We have already well established through this paper that consumers consider negative information as more diagnostic and informative and is thus more likely to be affected by negative consumer reviews than positive consumer reviews in a potential purchase decision. Thus, we hypothesize that:

H2: Negative online consumer reviews have a stronger effect on consumers' purchase intention than positive online consumer reviews.

The category diagnosticity approach introduced by Skowronski and Carlston (1989), stresses the importance of the diagnostic characteristics of the available information. Skowronski and Carlston assume that people categorize objects by interpreting the informational cues that are available, where some cues are regarded as more helpful than others. The category-diagnosticity, and thereby the influence on the consumer, will be higher when attributes are easier to place in a category. As discussed in section 2.3.2 *eWOM Characteristics*, high-quality online reviews contain arguments that affiliates with the features and performance of the product, while a low-quality online review is recognized by ambiguous and irrelevant information that is not related to the performance of the product (Park et al., 2007). It is therefore more difficult to understand why the reviewer does or does not recommend a product when the argument quality is low. The interpretation of this is that information of high quality is easier to place into a category and is thus more diagnostic than low-quality information.

Research has shown that using clear reasons that relates specifically to the product's performance have been shown to be more persuasive than low-quality online reviews (Lee et al., 2008) and has been shown to increase purchase intention (Park et al., 2007). In addition, Jiménez and Mendoza (2013) found that the intention to purchase a cell phone was higher when the level of detail in the reviews were high. This effect was mediated by the credibility of the review indicating that the diagnosticity was higher for detailed reviews. Based on the previous arguments and research findings, we hypothesize that:

H3: The effect that valence of online consumer reviews has on consumers' purchase intention, is stronger when it is moderated by argument quality.

4. Methodology

In this section we will elaborate on research design and make decisions regarding the approach of answering the research question and the following hypotheses. As the aim of this study is to investigate the causal relationship between valence, argument quality and consumers' intention to purchase by testing numeric data statistically, it is characterized as explanatory research with a quantitative approach (Saunders et al., 2015). We will study a particular phenomenon at a particular time and the research will thus be classified as cross-sectional. Figure 5 illustrates the main steps in our research methodology.

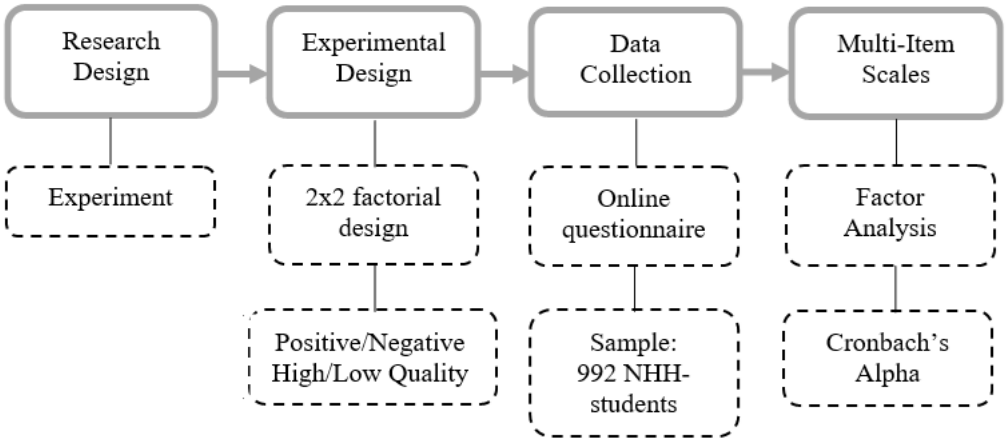


Figure 5: Flow chart of the methodological choices

4.1 Research Design

The purpose of this thesis is to explore the effect of online reviews on consumers' intention to purchase. More specifically, we want to test for changes in consumers' intention to purchase by manipulating the valence and quality of online reviews. Based on this, the preferred research method is experiment. An experimental approach allows us to reduce the number of possible explanations for an observed effect on the dependent variable, and in this manner, be able to assign more trust in the experimental treatments being responsible for the measured effects (Breivik, 2017a). The variation measured on the dependent variable will thus be connected to the manipulations, and not other unobservable variations within the sample.

Furthermore decided to conduct the experiment by employing a self-completed online questionnaire. A questionnaire makes it possible to collect great amounts of data and generalize to the population (Saunders et al., 2015). The drawbacks of using questionnaire as data collection method is that it is challenging to be as wide ranging as with other strategies. One of the reasons for this is that there is a limit to the number of questions the questionnaire can contain such that the participants do not feel overwhelmed or bored and thus answers truthfully in every question. Other obstacles could be the difficulty of gaining additional information from participants if necessary, challenges regarding misunderstandings and few respondents. However, we found that a questionnaire was the best option to collect sufficient amount of data to statistically test our hypotheses based on the time frame and resources available.

4.1.1 Experimental Design

For this study, we will apply the experimental approach known as between-subjects design. We did not want the respondents to be aware of the exact purpose of this study, i.e. measuring the effects of online reviews on purchase intention. In addition, it is important to avoid that the respondents would compare the different treatments against each other which is likely to happen in a within-subjects design. We developed a 2x2 factorial design, where the first dimension is valence and the second dimension is argument quality. In addition, we created a control group without exposure of reviews to be able to measure if the treatments have an effect. Combined, this gives four different experimental treatment groups as illustrated in Figure 6.

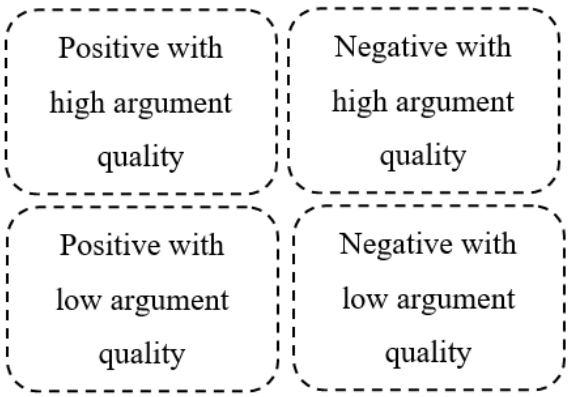


Figure 6: Factorial design

The Product

The first step of constructing the experiment was choosing a product. We determined the product based on criteria adapted from former studies. Firstly, it should be a product that the participants would use and purchase, a product that they would express their opinion of to others and would be interested in other consumers' opinion of as well (Mizerski, 1982). As we will elaborate on in section 4.5 *Data Collection*, the target group of this research is students, and it was thus important to choose a product that students would use and purchase, and in addition where the opinion of others has an impact on their evaluation of the product. Previous brand knowledge can play a major part in consumers purchase decisions (Hoyer et al., 2012). Over time consumers develop attitudes and preferences for brands they like and dislike, transferring into brand loyalty for those brands that they have the best attitude towards. Brand loyalty drives consumers to purchase products from them repeatedly because they have formed a belief that this product is greater than that of competitors. In such cases, measuring purchase intention for products is a problematic task and loyalty thus becomes a rival explanation if a well-known brand is included in the experiment. It is therefore beneficial to avoid product categories where strong brand loyalty clearly has emerged amongst consumers even though this strategy possibly will make the respondents less influenced by the treatments. Furthermore, since measurements of attitudes are not included as a component in this study, creating a hypothetical brand will help control for attitudes as a rival explanation for our results because their attitudes are perceived as neutral. Examples of products to avoid are mobile phones (Apple vs. Samsung) or cars (BMW vs. Audi). In addition, to collect as many responses as possible, we found it necessary that both genders could use the product.

Searching the Internet for inspiration, we found that amongst gender neutral objects, electronic products received the most consumer reviews. Park et al. (2007) stated in their study that electronic products are often complicated, and it is therefore difficult to determine which products are of good quality without buying one. Of this reason, consumers often seek information from other consumers before they make a purchase decision when the product is technology-intensive (Chen and Xie, 2008). Drawing upon the previous criteria, we chose a portable DAB radio as the product of this experiment. Digital audio broadcasting has become an important phenomenon in Norway the last couple of years because the Norwegian parliament decided to close down FM broadcasting in the country by the end of 2017 (Kulturdepartementet, 2017). For this reason, we assumed that DAB radio is a product

that Norwegian consumers have some knowledge about, but still would search for others' opinions about because of complex features, and in addition would be willing to purchase in the near future.

Even though a brand logo can disturb the validity of the results in the experiment, we believed that a brand name was important regarding the authenticity of the set-up. We therefore created a fabricated brand name and kept this constant in all the treatments as well as the control group. In order to find a brand name, we browsed for inspiration online in the electronic commerce category. Eventually, we decided to go for the Latin word for sound, Canetis.

Furthermore, we determined a suitable price for the product based on the same argument as with brand name. Again, we went online looking for DAB radios carrying similar features in order to create a uniform price. We wanted to go for a price in the middle of the range that we had discovered, neither too expensive nor too cheap. We did not want the price to be an indication of the product's quality in any direction. To ensure the authenticity of the price, we made five phone-call interviews where we asked if the price we had chosen was within what they could spend on a DAB radio, and whether they found this to be expensive, fairly priced or cheap. These interviews confirmed our choice of price as somewhat average of a DAB radio with similar specifications. None of these five individuals participated in the actual experiment.

4.1.2 Pretest

The next step in creating the manipulations was constructing online consumer reviews with different valence and quality in accordance with our factorial design. We developed and conducted a pretest before deciding on reviews to portray in the manipulation for the experiment. The purpose of the pretest was to confirm the choice of DAB radio as a product and to ensure that the fabricated consumer reviews were perceived as intended. We collected consumer reviews from websites and made some adaptations for the purpose of the experiment, including the most recurring attributes and comments about DAB radios. The reviews were adjusted to have approximately the same lengths, about three to four lines, in accordance with Park et al. (2007), to ensure that the length did not influence the perceived quality of the review. In addition, the reviews within the same level of argument quality

were kept as similar as possible, but with contradictory descriptive adjectives. The purpose of this is to reduce potential bias that could occur due to more dominant and persuasive word selections and thus reduce the possibility of false results (Weinberger and Dillon, 1980).

We started the pretest by asking the participants two questions about their knowledge of DAB radio, and subsequently presented them with consumer reviews of high or low quality, and negative or positive valence. Prior knowledge of the product, either too low or too high, can contaminate the evaluation and it is therefore beneficial to include a measure of this (Lee et al., 2008). All 19 respondents knew of the product, but none expressed that they had expert knowledge of DAB radio. The mean for product knowledge on a scale from 1 to 5 was 3.16, indicating a neutral standing (see Appendix A.1). In total all respondents evaluated 16 reviews in terms of positivity, objectiveness, understandability, credibility, clarity and quality on a scale from 1 to 5 (Park et al., 2007). The 16 reviews consisted of 4 reviews within each category of the factorial design.

From the descriptive statistics, see Appendix A.2, the consumer reviews with the best mean rating within each category were retained, i.e. high mean for high quality and low mean for low quality. We removed the rest of the reviews. In total, the remaining reviews constituted eight reviews. The means of both levels of valence and quality were compared and tested using a paired t-test. The t-test on valence resulted in a significant difference between the groups, with means of 1.27 for negative reviews and 4.54 for positive reviews. Regarding argument quality, the t-test was also significant with means of 2.11 for low quality reviews and 4.01 for high quality reviews. The results are given in Appendix A.2. These results together make it possible to accept the differences as sufficient for further implementation into the experiment.

4.1.3 Manipulations

We created the manipulations for the experiment with similarities to the study of Park et al. (2007) where the respondents were presented a fabricated page from a generic retailer website. The website displays a DAB radio, with the fictitious brand name and price we previously decided on, and a short advertorial description of the features. We added some extra features to make the description of the product more complete. In addition, we presented the four treatment groups with the chosen online consumer reviews from the

pretest. The largest issue regarding the manipulation is to make the illustration as realistic as possible. For this reason, each treatment contained three reviews, where two were manipulated (negative/positive and high/low quality) and one was neutral, solely a reproduction of some of the specifications given in the description of the product. We named the manipulations based on the valence and quality of the reviews. The five manipulations for the experiment are thus: 1. Positive High Quality (PHQ), 2. Negative High Quality (NHQ), 3. Negative Low Quality (NLQ), 4. Positive Low Quality (PLQ) and 5. Control (CON). Illustrations of the five different manipulations are provided in Appendix B.1.

4.2 Designing the Questionnaire

For designing the questionnaire, we used the research and experience software Qualtrics. We created the five different treatment groups and a setting in Qualtrics allowed for random assignment of respondents to each group evenly. Randomly assigning the participants to the different treatments help control for individual differences and selection, and do not need to be measured within the experiment (Saunders et al., 2015). As an opening of the questionnaire we included a short introduction that would be the first page the respondents would be redirected to once they had received a link to participate. In the introduction, we included a short description of the questionnaire being a source of data collection for a master thesis, that participation was voluntary and information about the possibility of winning a gift card. We also stated that by progressing from the introduction, the respondents gave their consent of participation. Respondents could only take the online experiment once, it was not possible to change their answers once they had pushed forward in the experiment and all questions had to be answered for completion, i.e. we checked for forced responses. The experiment uses closed questions exclusively, adopted by previous research on the same topic. The five treatment groups received identical questions before and after the manipulation, with an exception of a question measuring the moderating variable which was not displayed to the control group. As an alternative to manipulation checks, we added a setting in Qualtrics that allowed us to log the time spent on the manipulation. This makes it possible to investigate and perhaps remove answers where the treatment illustration is skipped after only a few seconds. All measures of dependent, independent and moderating variable are in Likert-scale format with seven points. The order and the number of the

response categories are kept constant throughout the experiment to avoid confusing the respondents (Saunders et al., 2015). The rating questions vary between two types of ratings: agreement and likelihood, and we clearly indicate the type of rating used before the question. At the end of the questionnaire we included a short debrief, stating that the product page they had been presented was fictive and only created for research purposes. Furthermore, we informed about the overall purpose without mentioning online reviews specifically. Finally, the respondents were given the possibility of participating in a drawing of gift cards. We created an independent questionnaire for this, ensuring the anonymity of the respondents. A full overview of the online experiment is found in Appendix B.2.

4.3 Measurement

In the following section, we will elaborate on the measurements used for the independent, dependent, moderator and control variables, in addition to the manipulation checks.

4.3.1 Dependent Variable

Intention to Purchase. For research purposes it might be difficult to measure an actual purchasing behavior. Purchase intention is found to be statistically related to actual buying behavior and has been shown to be affected by eWOM (Ismagilova et al., 2017). For this thesis, the aim is to measure the effect of online consumer reviews on purchase intention. The attitudes towards the brand or product could be favorable to measure, but is not emphasized in this paper. Attitudes and behavioral intentions are not fully correlated, and for the purpose of this thesis, purchase intention will be used to measure consumers' probability of purchase (Spears and Singh, 2004). Purchase intentions are the closest construct to actual purchase behavior and Spears and Singh found that attitudes towards the brand and purchase intentions exist as independent, but correlated, constructs. In addition, research done on eWOM and purchase intention has focused on the direct effect, and not the mediating effect of attitudes (Park et al., 2007; Park and Lee, 2008).

Spears and Singh (2004) found the most common and suited items for measuring purchase intentions amongst the scales used by previous researchers, which resulted in eight items. From them we adopted probability of buying or not buying the item to this study.

When measuring purchase intentions, two or three seven-point items should be provided. Therefore, we present two questions measuring the purchase intention of the radio immediately after the respondents have been exposed to one of the treatments. This is important in order to avoid contamination from other questions (Saunders et al., 2015). We adopted the two questions from the study of Park et al. (2007) and asked the participants to rank the probability of buying and recommending the radio to their friends on a scale of 1-7.

4.3.2 Independent Variable

Valence of consumer reviews. We developed consumer reviews of either negative or positive valence through online research and pretest. The respondents received an illustration of either negative or positive online reviews as a manipulation, together with the moderating variable as explained in section 4.1.3 *Manipulations*.

4.3.3 Moderating Variable

Argument quality of consumer reviews. As elaborated on in section 3 *Model and Hypotheses*, we propose that argument quality of online reviews moderates the effect that valence has on the consumers' intention to purchase. To be able to measure the moderating effect of argument quality, the online reviews presented to the respondents are of either high or low quality.

4.3.4 Control Variables

There might exist other factors that can be rival explanations of our predicted relationship between the independent and the dependent variable (Saunders et al., 2015). For that reason, we included several measures in the experiment to be able to hold these effects constant and in that way, improve the internal validity.

Firstly, we considered it necessary to include a measure of how often the participant read information about a product online before a purchase to ensure that the questions were relevant for the participant and thus can answer all the questions. We placed this question

immediately after the introduction. Hereafter, we refer to this variable as *Information Search.*

Secondly, prior knowledge of the product can contaminate the evaluation of the dependent variable, and it is therefore beneficial to include a measure of this as a control variable (Lee et al., 2008). We measured the prior knowledge of DAB radio in the question: "Do you know what a DAB radio is?". We placed this question before the manipulation to get an honest evaluation from the participant. We will examine carefully participants who check "No" on this question, and if necessary remove them from the analysis. From this point on, we refer to this variable as *Prior Knowledge*.

We presented a question addressing whether the participant or someone in their household had previously bought a DAB radio. We included this question to further get an idea of the participants' previous experience with the product and to control for the possible effect previous purchase could have on the intention to purchase the same type of product in the future. Hereafter, we call this variable *Prior Purchase*.

We do not know whether the students in our experiment are interested in purchasing a DAB radio or not. We believe that if the respondents are not interested in this type of product, this could be a rival explanation of our results, and we thus need to control for this aspect. We therefore included a question on a 7-point Likert scale asking how likely it is that they would buy a DAB radio in the future. The question was placed before the treatment took place to make sure the answer is not affected by the experimental condition (Pallant, 2007). From this point on, we refer to this variable as *Probability of Future Purchase.*

To control for the credibility and authenticity of the manipulation set-up, we added a question of overall perceived usefulness together with the measures of *Intention to Purchase* with one question using a 7-point Likert scale (Park et al., 2007). We measured this in the question: "If I had to make a decision now, I would find the given information about this DAB radio useful". We hereafter address to this variable as *Perceived Usefulness.*

It is expected that consumers have some previous attitude towards consumer reviews and it might be necessary to include a variable as a covariate if it is proven that attitude is interfering with the results. To measure and control for personal attitude towards consumer reviews, five measures were adopted from the study of Park et al. (2007) using a 7-point Likert scale (Questions AR1-AR5). The five questions address the following: how often the

respondents read consumer reviews in a decision-making process, if consumer reviews are considered helpful in a decision-making process, if consumer reviews make the respondents more confident in purchasing a product, if the respondents find consumer reviews irritating and lastly, if not reading online reviews before a purchase would make them worry about their decision. We placed the measures after the manipulation to prevent contamination of the results. We believe that if we placed the measures before the manipulation, the participants would speculate in which factors the experiment is examining and because of this have an excessive focus on the consumer reviews in the manipulation. Hereafter, we call this variable *Review Attitudes.*

Lastly, we controlled for age and gender by question *Gender* and *Age* at the end of the questionnaire, as encouraged by the guidelines of Saunders et al. (2015).

4.3.5 Manipulation Checks

The measurement of argument quality of the reviews is an adaption from Park et al. (2007), where this study uses five of the six statements and the scale is expanded to a 7-point Likert scale, as with the dependent variable. The five statements used were objective, understandable, credible, clear and of high quality. The reason for including only five of the measures is that when translating the statements from English to Norwegian, we find that using six measures are excessive. In addition, we included a question that addressed whether the respondents perceived the online reviews as positive or negative to check that the independent variable was perceived as intended. We placed this question in the same matrix as for argument quality. In the following, we refer to these variables as *Quality* and *Valence*, respectively.

4.4 Pilot Test

Prior to distributing the questionnaire, testing with a small number of respondents similar to the sample is beneficial (Saunders et al., 2015). This makes it possible to detect if the respondents have any difficulties in completing the questionnaire or if some of the questions are misunderstood. Pilot testing is done to ensure the quality of the questionnaire, secure the validity and reliability and thus avoid problems in the final data collection. In addition, the

supervisor of this thesis provided comments and suggestions for improvement prior to the pilot testing. The pilot test was conducted by friends and family, mainly consisting of people from within the ages 18-26. The pilot test had 11 respondents, which is considered acceptable. From the responses it was possible to measure how long time they spent on the questionnaire (less than five minutes), which would later be used to inform future possible respondents in the final data collection. We had a small interview or discussion with all of the pilot testers to receive their feedback. An error on two of the manipulations (NHQ and NLQ) was detected, were the same word appeared twice in the matrix on the question where they should rate the quality of the reviews. No further errors were communicated, and all of the pilot testers had no difficulties in understanding or completing the questionnaire.

4.5 Data Collection

The first step of collecting data for the research is choosing a population. The ideal population to answer our research question would be Norwegian consumers who use Internet and are responsible for purchasing products in a household. Ideally the sample should be generalizable to this population. Due to time and resource limitations in this thesis, the target population is chosen to be Norwegian students with NHH-students as the sampling frame. We believe that based on the topic of this study and the product chosen, that NHH-students will not differ to a large extent from other Norwegian students. We consider DAB radio as a neutral product that can be relevant for anyone. In addition, the product category does not require a specific expertise. What might distinguish NHH-students from other Norwegian students is that they have specific interest for economics and business, and by studying these topics they will also have greater knowledge of this. However, regarding measuring the effects of eWOM on consumers' purchase intention, we believe that field of study will not be of significant importance. With basis in this discussion, a sample consisting of NHH-students should be representative to the target population, Norwegian students.

The study administration at NHH provided 992 student email addresses belonging to Norwegian students from both the bachelor and the master programs at NHH. The only request we made was that the sample should only include Norwegian students, since the questionnaire was in Norwegian. We therefore assume that the probability of being included in the sample was equal for all Norwegian NHH-students, indicating that the sample was drawn using probability sampling. We distributed the questionnaire to the provided student

emails from Qualtrics with an invitation to participate in the study and a following link to the questionnaire. The recipients of the invitation could choose themselves whether they wanted to participate or not. To increase the response rate, an incentive (gift card to the cinema) was included with registration at the end of the questionnaire with the goal of having fewer incomplete responses.

The invitation sent by e-mail included information about the possibility of winning a gift card, explained the purpose of the questionnaire and informed the possible respondents about completion time. Our goal was to collect minimum 150 responses in order to meet the minimum required number of 30 responses in each treatment group (Pallant, 2007).

4.6 Multi-Item Scales

Before we can start the analysis of the data collected, it is important to justify the convergent and discriminant validity of the multi-item scales (Breivik, 2017b). The best way to do this is to run a factor analysis and check for high cross loadings. An assumption made for factor analysis to be an appropriate model, is correlation among two or more variables. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO), including Bartlett's Test of Sphericity is a possible measure for this purpose (Tabachnick and Fidell, 2012). Values above .6 is considered acceptable and the Bartlett's Test of Sphericity should be significant. The KMO value is .704 and the Bartlett's Test of Sphericity is strongly significant (ρ=.000) and we thus accept the assumption. This is shown in Appendix C.1, Table 6. A method of deciding the number of factors suitable for the data, is examining the eigenvalues of the components. The eigenvalues are a measure of the variance explained by the factor out of the total variation (Bartholomew et al., 2008). Components with eigenvalues below 1 should be excluded. We see from the table showing total variance explained (see Appendix C.1, Table 7) that three factors have eigenvalues above 1, which is the result we are aiming for. We then continue by examining the factor loadings in the rotated solution, also called the pattern matrix. Loadings can range from -1 to 1, where values close to -1 or 1 indicate that the factor strongly affects the variable. We see from table 1 that we are provided with three factors, where the factors are in correspondence with our multi-item scales. Factor 2 depicts the strongest grouping with loadings around .9, followed by the first factor with the lowest loading being .636. In the third factor, four out of 5 variables have acceptable loadings, but the question measuring whether people find consumer reviews irritating (AR4_1) has a loading below .4.

Before we safely can group the variables as is suggested by the factor analysis, it is beneficial to run a test for the reliability of the factors created (Breivik, 2017b). This is done by using Cronbach's alpha as a measure. The value should be as close to 1 as possible and a minimum of .7 is recommended (Saunders et al., 2015). The Cronbach's alpha values from the reliability analysis for *Intention to Purchase, Review Quality* and *Review Attitudes* is .860, .825 and .714 respectively for each scale measure (shown in Appendix C.2). Thus, we approve all factors and can continue with our analysis.

Pattern Matrix[a]

	Component		
	1	2	3
AR1_1			.719
AR2_1			.745
AR3_1			.646
AR4_1			
AR5_1			.764
IP1		.892	
IP2		.926	
Objectivity	.636		
Understandability	.823		
Credibility	.790		
Clearness	.756		
Quality	.767		

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 8 iterations.

Table 1: Pattern matrix from factor analysis

5. Analysis and Results

In this section we will provide the statistical analyses and the corresponding results. We have used SPSS Version 24 to conduct the analyses. The accepted significance level in this thesis is the traditional level of ρ=.05. We will start with presenting descriptive statistics. Thereafter, we will answer each hypothesis and lastly supplement with additional analyses.

5.1 Descriptive Statistics

From the sample consisting of 992 students at NHH, 190 started the questionnaire while the final number of complete responses was 155. The response rate of the questionnaire is thus 15.63 % (155/992 = 0.15625). We removed 35 responses that were incomplete responses from the dataset. From the 35 incomplete answers, only 14 had progressed until the point of the treatment. The other incomplete responses had answered none of the questions. See Appendix D.1, Table 9, for the distribution amongst the treatments for the incomplete answers.

We set Qualtrics to randomly assign the respondents evenly to the five different treatments. However, due to the 14 incomplete responses that Qualtrics did not account for in the randomly assigning, the distribution among the different treatments of complete responses are not completely equal. As depicted in Table 2, positive high-quality reviews (PHQ) and negative high-quality reviews (NHQ) are just below the required number of 30, missing respectively one and two responses to meet the requirements (Pallant, 2007). Nevertheless, we consider this deviation small, and overall the distribution between the treatments are relatively equal.

Treatment

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	PHQ	29	18.7	18.7	18.7
	NHQ	28	18.1	18.1	36.8
	NLQ	33	21.3	21.3	58.1
	PLQ	33	21.3	21.3	79.4
	CON	32	20.6	20.6	100.0
	Total	155	100.0	100.0	

Table 2: Distribution of respondents in treatments

The distribution of genders are 49.7 % male and 50.3% female, as shown in Table 3. The distribution of gender between the treatments are not as equal as we would have hoped for, especially for treatment groups Positive with High Argument Quality and Negative with High Argument Quality. For an overview of the distribution, see Appendix D.1, Table 10. However, we choose to accept this and provide an additional analysis assessing the possibility of differences between the genders in section 5.7 *Additional analyses*.

The age distribution has an overweight of responses from the age group 18-24 with the second largest group being 25-34, as presented in Table 4. There were no responses from the age groups under 18 and over 45, and a few responses from the group 35-44.

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	77	49.7	49.7	49.7
	Female	78	50.3	50.3	100.0
	Total	155	100.0	100.0	

Table 3: Distribution of genders

Age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	18 - 24	110	71.0	71.0	71.0
	25 - 34	42	27.1	27.1	98.1
	35 - 44	3	1.9	1.9	100.0
	Total	155	100.0	100.0	

Table 4: Distribution of ages

The majority of the respondents stated that they use the Internet to search for information when planning to buy a product. 95.5 % of the respondents answered “Yes” to the question on whether they knew what a DAB radio was. This meaning that 7 respondents did not have any prior knowledge of DAB radio. 63.2 % answered that they, or someone in their household, previously had purchased a DAB radio, 30.3 % answered no this question while 6.5 % was uncertain. On the question of the probability of future purchase of DAB

radio, 34.2 % answered 1-3 (unlikely), 18.1 % were neutral, while 37.7 % answered 5-7 (likely). Charts and tables are available in Appendix D.1.

Lastly, we will include some descriptive statistics of the five questions measuring *Review Attitudes.* Even though we have merged them together to one variable, we believe each question also gives valuable information individually. The three first question measure how often the respondents read consumer reviews in a decision-making process (AR1), if they are considered helpful in a decision-making process (AR2) and if consumer reviews make the respondents more confident in purchasing a product (AR3). On these questions less than 10 % answered 1-3 (disagree) while 70-80 % answered 5-7 (agree). The fourth question was inverted and measured whether the respondents find consumer reviews irritating (AR4). On this question, 85 % answered 1-3, i.e. that they do not find consumer reviews irritating. The last question measure whether not reading online reviews before a purchase make them worry about their decision (AR5). On this question, 53 % answered 1-3 (disagree), while only 28 % answered 5-7 (agree). A full overview of the distributions is found in Appendix C.2.3, Table 8.

5.2 Control Variables

To explore the relationship between the control variables and the dependent variable, we used Pearson correlation as measure (Pallant, 2007). If there are any significant correlations, we need to assess the relationship further because the variable might be a rival explanation of the results. The results, given in Table 5, show that the only variables with significant correlation against *Intention to Purchase* is *Probability of Future Purchase* (ρ=.001, r=.265) and *Information Search* (ρ=.047, r=.160). However, Pallant state that with large sample sizes (N > 100), even small correlation coefficients can reach statistical significance. Because of this, it is better to focus on the shared variance between the two variables, a measure done by squaring the r-value. *Information Search* give a shared variance of only 2.56 % which we consider small. *Probability of Future Purchase* and *Intention to Purchase* has a shared variance of 8 %. This percentage might not be large either, but with a significance level of ρ=.000 we believe it is large enough to potentially be included as a covariate for our analysis. We assess the assumptions regarding covariates in section 5.5 *Assumptions of ANCOVA.*

Correlations[c]

	Information Search	Prior Knowledge	Prior Purchase	Probability of Future Purchase	Gender	Age	Perceived Usefulness	Review Attitudes
Intention to Purchase	.160	.120	-.093	.284	.069	.021	.120	.032
	.047	.137	.248	.000	.394	.795	.136	.694

c. Listwise N=155

Table 5: Pearson Correlations of control variables

5.3 Manipulation Checks

To make sure that the intended manipulation did work, we conducted manipulation checks. After answering questions regarding the dependent variable, *Intention to Purchase*, the respondents answered questions about the argument quality and the positivity of the online consumer reviews. The control group did not receive those questions since the treatment for this group did not include any online reviews. To assess if the manipulations in our experiment worked on the respondents, we collected the means from each stimuli, valence and argument quality. The means indicate that both *Valence* and *Quality* are perceived as intended (see Appendix D.2). To validate this interpretation, a one-way ANOVA was run in order to check if the difference between the groups was significant, which was confirmed (ρ=.000 for both groups). Tables from the ANOVAs are available in Appendix D.2.

In Qualtrics we recorded the time respondents spent on the treatment. This as a way of detecting possible careless responses in addition to checking the dataset for habitual answering. Some of the questions would naturally have the same answer, but except for this we detected no obvious habitual answering. Further, we applied the explore-option in SPSS to get an overview of the timing. The Stem-and-Leaf Plot (see Appendix D.2, Table 15) shows that there were 22 responses where the time spent was less than 10 seconds, which we consider as a minimum to be able to get a grasp of the information given. Moreover, we detected 7 extreme values, with more than 95 seconds spent on the site, by interpreting the box plot depicted in Figure 7. Given these findings, it was desirable to test whether these responses were significantly different from the ones with an acceptable time use and thus have an impact on the further analysis and results. To ensure that these responses were acceptable and would not interfere with the results, we ran an ANOVA to compare the responses over and under 10 seconds, and those over and under 95 seconds (see Appendix D.2, Table 17 and 19). The ANOVA results show that there was no significant difference

between the groups on the dependent variable and we therefore could safely keep all the responses in the following analysis.

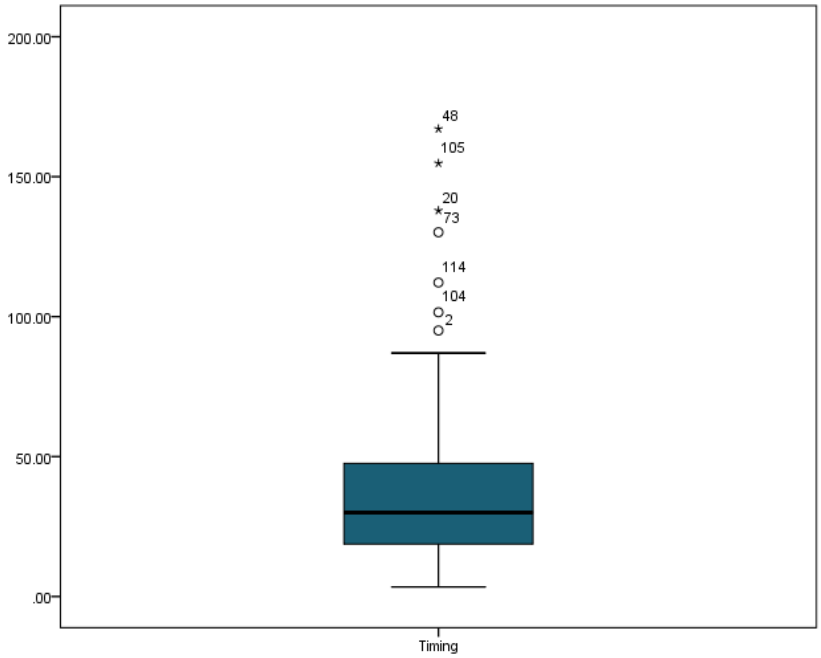


Figure 7: Box plot of timing variable

5.4 Assumptions of ANOVA

Analysis of variance or ANOVA is used to test whether three or more groups are different (Saunders et al., 2015). For parametric measures such as ANOVA, there are several assumptions that should be met before continuing the analysis (Pallant, 2007). We will elaborate on each assumption and appropriate measures will be performed where necessary.

The first assumption regards independent observations (Pallant, 2007). Observations are assumed to be independent if each observation or measurement is not influenced by each other. It is thus beneficial that the responses are not collected in a group setting and that the experimental design does not require some kind of interaction with each other. In this study, the sample was drawn randomly from the population and the invitation to participate was sent by email to the whole sample. Each participant was randomly assigned to one of the experimental conditions. Furthermore, the introduction of the experiment stated that all answers are anonymous and stressed the importance of completing the experiment alone and

without communication with others. Based on this, we can conclude that the first assumption is met.

The second assumption concerns the distribution of the variables, which preferably should be normal (Pallant, 2007). The distribution of the data can be investigated by looking at the skewness and kurtosis of each variable. Skewness is a measure of the symmetry of the distribution and can be seen as a clustering of scores either towards the low or the high end of the scale. Positive skewness indicates clustering towards the left and with an extended tail towards the right. A negatively skewed distribution is indicated by a peak shifted to the right and the tail extending towards the lower scale on the left. Kurtosis measures the peakedness of the distribution where a highly peaked distribution is given by positive values, and a flat distribution is indicated by negative values. For both measures, values ranging from -1 to 1 is considered as an indication of normal distribution. Table 6 presents descriptive statistics including skewness and kurtosis of all the variables. Variables with values outside of the acceptable interval on skewness and/or kurtosis is highlighted and needs to be addressed before continuing the analysis. It is shown that *Prior Knowledge, Prior Purchase* and *Age* are positively skewed which is expected because the majority of the participants have knowledge of what a DAB radio is, have purchased a DAB radio previously and are within the age group 18-24. *Information Search* is negatively skewed towards most people using the Internet to search for information, which is also expected. Regarding kurtosis, it is shown that *Information Search* and *Prior Knowledge* have a peaked distribution, while *Probability of Future Purchase*, *Gender* and *Positivity* has a flatter distribution. As we can see, some variables break the assumption of normality, but Pallant (2007) states that with large enough sample sizes (>30), the violation of this assumption should not cause any major problems. Even though two of the groups contains less responses than 30, the sizes are fairly close to 30 and considered acceptable.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Information Search	155	1	7	5.50	1.208	-1.168	.195	1.400	.387
Prior Knowledge	155	1	2	1.05	.208	4.424	.195	17.798	.387
Prior Purchase	155	1	3	1.43	.614	1.115	.195	.207	.387
Probability of Future Purchase	155	1	7	4.35	2.072	-.170	.195	-1.240	.387
Gender	155	1	2	1.50	.502	-.013	.195	-2.026	.387
Age	155	2	4	2.31	.504	1.286	.195	.604	.387
Intention to Purchase	155	1.0	7.0	2.835	1.4695	.338	.195	-.850	.387
Perceived Usefulness	155	1	7	5.21	1.606	-.885	.195	.184	.387
Positivity	123	1	7	3.98	2.473	-.064	.218	-1.736	.433
Review Quality	123	1.0	7.0	4.180	1.1626	-.248	.218	.133	.433
Review Attitudes	155	2.2	6.8	5.010	.9116	-.508	.195	.246	.387
Valid N (listwise)	123								

Table 6: Skewness and Kurtosis

The third assumption of ANOVA encompasses homogeneity of variance (Pallant, 2007). Pallant states that samples must be obtained from populations of equal variances in order for this condition to hold. A Levene's test can be used to verify this assumption. This test is a part of the output when an ANOVA or t-test is run in SPSS and the aim is to get a result that is not significant ($\rho > .05$). The test shows no significance for all of the analyses of variance done in this study. The conclusion is thus that all assumptions for ANOVA holds and the analysis may proceed. The corresponding Levene's test to each ANOVA is found in Appendix D.3.1.

5.5 Assumptions of ANCOVA

In addition to the aforementioned assumptions, ANCOVA has several additional assumptions that needs to be ensured before the analysis can continue (Pallant, 2007). Several of the assumptions addresses the reliability of the chosen covariates. In section 5.2 *Control Variables,* we assessed the potential correlations between the control variables and the dependent variable and discovered that *Probability of Future Purchase* was the only variable with sufficient significance. We believe that *Probability of Future Purchase* is an indicator of the participants willingness to purchase the proposed DAB radio. If the participant is not interested in purchasing this type of product in the future, there is reason to believe that this attitude could affect the answer given on *Intention to Purchase*. We will in

the following consider all assumptions regarding covariates before we make a conclusion on whether to include *Probability of Future Purchase* in the analysis or not.

Regarding the covariate, there are three assumptions that we need to confirm (Pallant, 2007). The first assumptions say that we need to make sure that the experimental condition does not affect the covariate. We placed the variable before the treatments took place and we thus accept this assumption as true for *Probability of Future Purchase*.

Secondly, we need to assure the reliability of the covariate (Pallant, 2007). Pallant suggests using well established questions and measures, performing tests of Cronbach's alpha when several measures are clustered together and performing a pilot test of the experiment. We have used scales and measures that have been used in previous studies as well as performing a pilot test before conducting the experiment. With basis in section 4.3 *Measurement*, 4.4 *Pilot Test* and 4.6 *Multi-Item Scales*, we consider all aspects of this assumption met.

Thirdly, there should be a significant correlation between the covariate and the dependent variable (Pallant, 2007). Our investigation of the control variables, indicate that *Probability of Future Purchase* is the only variable with sufficient correlation with *Intention to Purchase*. As shown in section 5.2 *Control Variables,* we approve of the significant relationship between *Probability of Future Purchase* variable and *Intention to Purchase* and accept the third assumption.

The last assumption regards homogeneity of regression slopes (Pallant, 2007). The relationship between the dependent variable and the covariate should represent all the groups, meaning that there should not be an interaction between the covariate and the experimental treatment. It is possible to assess this assumption statistically by exploring if there is a statistically significant interaction between the treatment and the covariate. The results, see Table 23 in Appendix D.3.2, show that there is not a significant relationship between *Probability of Future Purchase* and *Intention to Purchase* (ρ=.076) and we therefore consider this assumption met.

As a conclusion, we accept all the assumptions of ANCOVA and can safely conduct the analysis and include *Probability of Future Purchase* as a covariate.

5.6 Testing Hypotheses

5.6.1 H1

Valence of online reviews has a direct effect on consumers' purchase intention

To test the first hypothesis, we run a one-way ANOVA to investigate the effect that the independent variable, *Valence*, has on the dependent variable *Intention to Purchase*. The results from the ANOVA is portrayed in Table 7, showing that the effect of Valence on *Intention to Purchase* is strongly significant at the 5% significance level (ρ=.000). The different means of the levels of valence is 2.0000 for negative reviews and 3.5484 for positive reviews (Table 24 in Appendix D.4). Based on these results, we accept H1.

Tests of Between-Subjects Effects

Dependent Variable: Intention to Purchase

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	73.718[a]	1	73.718	43.122	.000
Intercept	946.564	1	946.564	553.694	.000
Valence	73.718	1	73.718	43.122	.000
Error	206.855	121	1.710		
Total	1231.500	123			
Corrected Total	280.573	122			

a. R Squared = .263 (Adjusted R Squared = .257)

Table 7: One-way ANOVA

5.6.2 H2

Negative online consumer reviews have a stronger effect on consumers' purchase intention than positive online consumer reviews.

We tested the second hypothesis by running a one-way ANOVA with planned contrasts. Firstly, we explored the means of *Intention to Purchase* for both negative and positive reviews to determine if running the planned contrasts could give any significant results. We found in the analysis for H1 that the mean was 2.0000 for negative and 3.5484 for positive, showing a possible significant difference. We weighted the two groups with corresponding valences equally against the control group, making two different contrasts (see Appendix

D.5). Contrast 1 is measuring the effect of positive reviews (PHQ & PLQ), while Contrast 2 measures the negative (NHQ & NLQ). When assuming for equal variances the p-values are ρ=.074 and ρ=.000 for Contrast 1 and 2 respectively. These results imply that NHQ and NLQ are statistically different from the control group, while we did not find a significant difference from the control group for PHQ and PLQ. As mentioned earlier, we should take the variable *Prior Knowledge* into consideration when conducting the analyses because of possible interference with the results. Therefore, we ran the planned contrast analysis excluding the seven responses where the respondent had no prior knowledge of the product. The p-values for Contrast 2 stays at ρ=.000, but for Contrast 1 it changes to ρ=.047 making the difference between PHQ and PLQ and the control group significant as well, as shown in Table 8. The p-values from both analyses indicate that the statistical power of Contrast 1 is greater than Contrast 2 (.000>.074 and .000>.047). From these results it is possible to conclude that negative online consumer reviews have a stronger effect than positive online consumer reviews on consumers' purchase intention. Thus, we accept H2[1].

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
Intention to Purchase	Assume equal variances	1	.572	.2849	2.007	143	.047
		2	-1.054	.2856	-3.691	143	.000
	Does not assume equal variances	1	.572	.2961	1.932	63.306	.058
		2	-1.054	.2668	-3.951	48.271	.000

Table 8: Planned contrasts

5.6.3 H3

The effect that valence of online consumer reviews has on consumers' purchase intention, is stronger when it is moderated by the argument quality.

For the third hypotheses we conducted a two-way ANOVA, including both *Valence* and *Quality* to measure the interaction effect. The analysis show significance for *Valence* (ρ=.000), but neither *Quality* (ρ=.892) nor the interaction effect (ρ=.211), measuring *Valence*Quality*, is significant. Based on these results, we cannot accept H3.

[1] We assume equal variances given that the Levene's test is not significant (ρ=.442)

As done with the previous hypothesis, excluding the responses with no prior knowledge of DAB radio, could also change the results of the two-way ANOVA. However, the results including the filter are similar as the previous result, with significance for *Valence* (ρ=.000) and no significance for *Quality* (ρ=.628) and *Valence*Quality* (ρ=.111). Thus, we still do not have sufficient grounds for accepting the hypothesis.

Given that the results improved by filtering out respondents with no prior knowledge of DAB radio, we will continue the analysis of H3 using this filter. In section 5.4 *Assumptions of ANCOVA* we discuss including covariates to the ANOVA. In addition to the two previous ANOVAs, we therefore further test the hypotheses with an ANCOVA. Based on previous discussions, we include *Probability of Future Purchase* as a covariate in the model. The ANCOVA results (see Table 9) show that the effect of *Valence* is still significant (ρ=.000). For *Quality* the effect is not significant (ρ=.404), and the interaction effect between *Valence* and *Quality* is slightly approaching significance at the 5% level with significance at the 10% level (ρ=.065). The covariate is proven to be significant (ρ=.003), indicating that the results are in fact improved by controlling for this factor. The interaction effect is significant at the 10% level which indicates that there exists an effect, but not strong enough to accept the hypothesis. Figure 8 illustrates the difference in purchase intention for negative and positive reviews with low or high quality. The slope is steeper for the negative reviews, showing that quality has a stronger (negative) impact on *Intention to Purchase* compared to the positive line, where the slope is less steep. An overview of the outputs from the previous analyses on H3 is found in Appendix D.6.

To further get an understanding whether the effect of reviews is stronger when the quality is high, we ran a planned contrast analysis comparing PHQ and PLQ in Contrast 1, and NHQ and NLQ in Contrast 2. This will give an indication of the effect of argument quality. From the contrast tests none of the contrasts are significant as ρ=.425 and ρ=.141 for respectively Contrast 1 and 2 (see Appendix D.6, Table 25). However, these values indicate that quality has a stronger effect for negative reviews.

Overall, the analysis run to test H3 has shown that *Valence* has a significant effect, but neither *Quality* nor the interaction effect can be proven to have a significant effect. From these results, we reject H3. The figure below illustrates the final result (Figure 9).

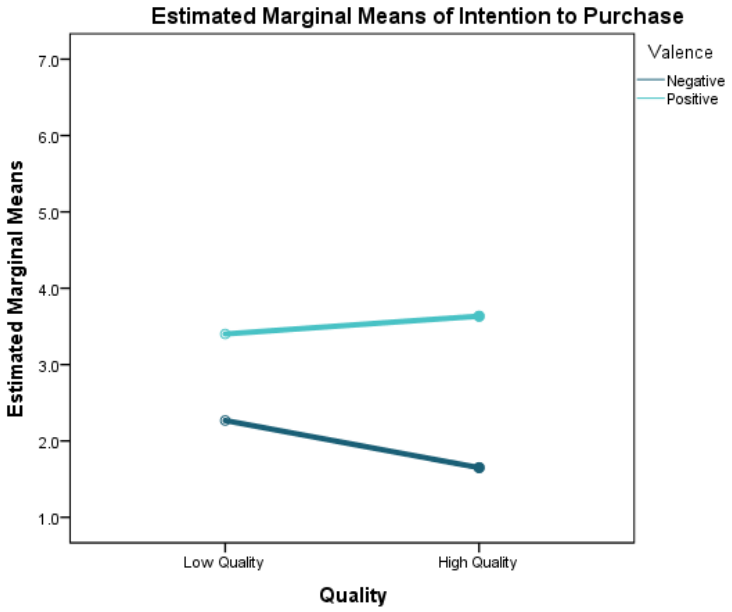
Tests of Between-Subjects Effects

Dependent Variable: Intention to Purchase

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	89.916[a]	4	22.479	14.768	.000	.341
Intercept	83.396	1	83.396	54.790	.000	.325
PBDAB_1	9.992	1	9.992	6.565	.012	.054
Valence	70.705	1	70.705	46.452	.000	.290
Quality	1.068	1	1.068	.702	.404	.006
Valence * Quality	5.291	1	5.291	3.476	.065	.030
Error	173.521	114	1.522			
Total	1164.750	119				
Corrected Total	263.437	118				

a. R Squared = .341 (Adjusted R Squared = .318)

Table 9: Two-way ANCOVA



Covariates appearing in the model are evaluated at the following values: Probability of Future Purchase = 4,54

Figure 8: Effects of Valence and Quality on Intention to Purchase

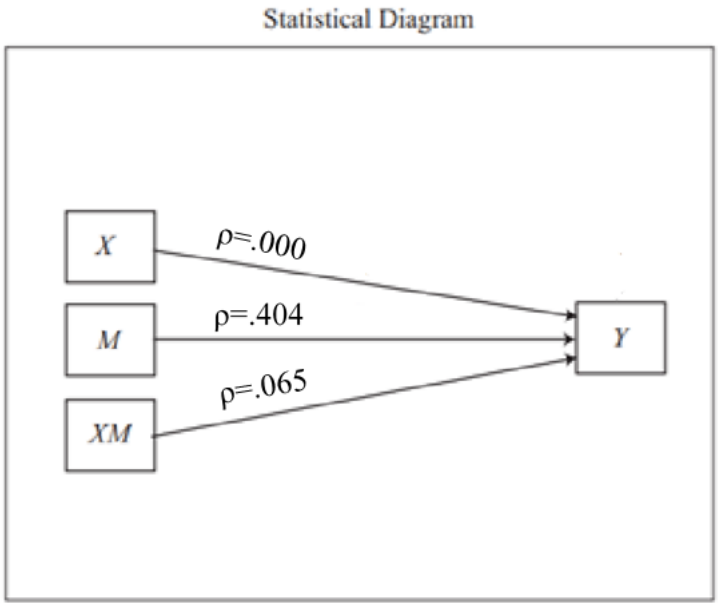


Figure 9: Overview of the results

5.7 Additional Analyses

5.7.1 The Negativity Effect

Research by Park et al. (2007) shows that the purchase intention of high-involvement consumers is affected by both review quantity and review quality, while low-involvement consumers are only affected by review quantity. Hence, argument quality requires more effort than more salient aspects (i.e. such as review valence). In addition, Lee et al. (2008) found that the effect of negative high-quality reviews on attitude towards an MP3 player was greater when consumers were highly involved. Considering that this study does not consider different levels of involvement, we want to check if the negativity effect will outshine the effect of argument quality, because argument quality requires more effort to elaborate on than valence.

Firstly, by looking at the means of all four groups, we see that the ranking proposed could hold true, but as of yet it is not possible to address the significance and therefore it is beneficial to run a post-hoc test and interpret the significance levels of the four different treatments against the control group. We chose to run the post-hoc test using Tukey HSD, a

procedure located in the middle of both statistical power and conservatism (Breivik, 2017b). In essence, what we are looking for is that the effect of NHQ > NLQ > PHQ > PLQ.

The results show that the ranking is as anticipated, with a significant effect of NHQ (ρ=.003) and NLQ (ρ=.049), while PHQ (ρ=.308) and PLQ was not significant (ρ=.769). See Appendix D.7 for the complete post-hoc test. From these results, it is possible to conclude that negative online reviews, with both levels of argument quality, have a stronger effect on purchase intention than positive reviews with both levels of quality.

5.7.2 Gender Differences

We mentioned in section 5.1 *Descriptive Statistics* that because of uneven distribution of the genders amongst the treatment groups, we want to address this issue by performing a t-test comparing men and women's intention to purchase. With means of 2.734 for men and 2.936 for women, the t-test shows that the difference between the genders is not significant (ρ=.394).

In addition, we discovered that the means of *Probability of Future Purchase* for men and women were quite different (3.94 for men and 4.76 for women). We therefore ran an ANCOVA with *Probability of Future Purchase* as a covariate. As found in the previous analyses, only *Valence* was significant. However, the plots show that women have a slightly higher intention to purchase than men when online reviews are positive, but the slope is less steep than that of males. The slopes are shown in the plots in Figure 10.

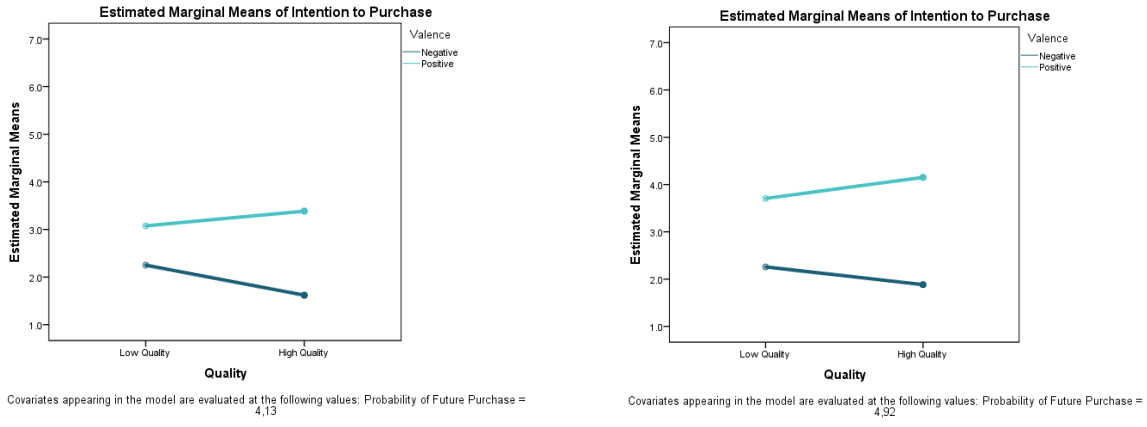


Figure 10: The effects of valence and quality on intention to purchase for males (left) and females (right)

6. Discussion

In this section, we will first go through the findings from the analyses, before we give a discussion on how the results of this study contributes to previous literature on eWOM and then proceed with an elaboration on managerial implications.

6.1 Summary of Findings

Our data collection and analyses have been part of trying to answer our research question, which is as follows:

***RQ**: What effect does online reviews have on purchase intention, and what is the impact of valence and argument quality of the online reviews?*

We developed three hypotheses to answer the research question and in the following the results of the analyses conducted will be presented. The hypotheses were developed based on previous literature and theory and indicated the predicted outcomes of the study. With this background, the expected findings were that valence would have a direct effect on purchase intention, with an enhanced effect of negative online consumer reviews compared to positive. In addition, it was anticipated that argument quality of online consumer reviews would have a moderating effect on the relationship between valence and intention to purchase.

The results show that valence does have a direct effect on purchase intention, and that the negative effect is the strongest. However, the anticipated moderating effect of quality was not found significant. Nevertheless, the results from the analysis indicate that the effect is stronger for both types of negative reviews (NHQ and NLQ) compared to the positives (PHQ and PLQ). A summary of the findings are shown in Table 10.

Furthermore, we ran an additional analysis to investigate whether the effect of negative reviews would surpass the positive, regardless of level of quality. We found that this holds true, further proving the presence of the negativity effect.

Hypotheses	Accepted (yes or no)
H1: Valence of online reviews has a direct effect on consumers' purchase intention	Yes
H2: Negative online consumer reviews have a stronger effect on consumers' purchase intention than positive online consumer reviews.	Yes
H3: The effect that valence of online consumer reviews has on consumers' purchase intention, is stronger when it is moderated by argument quality.	No

Table 10: Summary of findings

6.2 Implications

6.2.1 Theoretical Implications

Based on our previously stated hypotheses, we expected to find that consumer reviews would have a statistically significant impact on purchase intention. Furthermore, we indicated that the relationship between valence and argument quality would be an important factor in influencing consumers' intention to purchase a product. Based on prior findings, we expected to accept H1 and H2, indicating that the valence of online consumer reviews would impact purchase intentions and in addition, that negative reviews would have a larger impact than positive reviews. The acceptance of both H1 and H2 is thus in correspondence with previous literature (Park and Lee, 2009; Lee and Lee, 2009; Lee and Youn, 2009; Sen and Lerman, 2007). This finding enhances the belief that the negativity effect is applicable for Norwegian students as well. As a conclusion of the first and second hypothesis, valence does have an impact on intention to purchase a DAB radio, and the negative reviews have a larger impact than the positive reviews. This makes sense on both an intuitive and theoretical level. However, a surprising aspect of the analysis is that the means regarding intention to purchase were rather low regardless of valence. The scale ranged from 1 to 7, and the highest mean

was just below 4, which is neutral. The purchase intention in general is thus lower than expected.

In the third hypothesis, argument quality of online reviews was proposed to moderate the effect of valence on intention to purchase. Both an ANOVA and an ANCOVA with *Probability of Future Purchase* as a covariate was conducted in order to investigate if there was a significant moderating effect of argument quality. The covariate did improve the significance level, making the effect weakly significant ($\rho<.10$). However, we set the accepted significance level for this study to $\rho<.05$. To summarize, none of the analyses performed found this relationship significant, resulting in rejection of H3. This outcome contradicts with previous research done on argument quality (Cheung et al., 2008; Cheung et al., 2009). Park et al. (2007) found that involvement played an important role regarding the influence of argument quality. They portrayed that in order for argument quality to have an effect on purchase intention, the consumers need to be highly involved in the process. Considering this, the results reported in section 5.6.3 *H3* might not be as unexpected as initially suspected. This thesis does not measure involvement, but given that the negativity effect was present, this could be interpreted as a sign of the respondents not being highly involved when participating in this study. A reason for involvement being low rather than high, could be that the respondents did not recognize a need for the product and as a result stopped the consumer decision-making process at need recognition. As a result, the information search stage will not be optimal and most likely conducted with low involvement.

Lastly, we will discuss the implications of the control variables. Former research have used the variables *Review Attitude* and *Prior Knowledge* as covariates in their analysis (Lee et al., 2008). One of the requirements for safely including a covariate in the analysis is correlation between this variable and the dependent variable (Pallant, 2007). Thus, neither *Review Attitudes* nor *Prior Knowledge* could be included in the analysis, which is in contrast with previous research. As discussed in section 5.2 *Control Variables*, the only control variable with sufficient significance level towards *Intention to Purchase* was *Probability of Future Purchase*. This might be an indication that future research on eWOM also should consider controlling for this variable in their analysis.

We also did an additional analysis on whether the valence would trump the effect of argument quality on purchase intention. By sorting the significance levels of each treatment

group from the post hoc test, the order of the effects was as expected, that the negativity effect was apparent. In light of the previous discussion regarding H3, this could be another sign of low involvement amongst the respondents. Based on this, we encourage future researchers on effects of eWOM and argument quality to include a measure of involvement.

6.2.2 Managerial Implications

The findings of this study could have several implications for marketers and managers. Firstly, we found that over 80 % of the respondents state that in a purchase decision process, they regularly use the Internet to search for information. In addition, through measuring the respondents' use of online consumer reviews, we discovered that approximately 80 % use consumer reviews in their decision-making process and that online reviews are useful to make a purchase decision. We believe that this emphasizes the importance of eWOM and this should be a large focus for companies.

Through testing our hypotheses, we found that online consumer reviews have an impact on consumers' intention to purchase, where negative reviews influence purchase intention negatively and positive reviews influence positively. Furthermore, we show that the negativity effect is present. For this it is possible to draw the conclusion that when consumers read online reviews, they are more likely to avoid a purchase when reading a negative review, than to purchase a product based on a positive review. Therefore, companies should carefully address negative reviews about their products or brands because they can damage sales or reputation. A surprising result is that argument quality has no significant effect on consumers purchase intention. Even though we found that the respondents were able to distinguish between high- and low-quality reviews, this factor did not influence their purchase intention. This indicates that when consumers come across negative online consumer reviews, it is likely that consumers will form negative associations and attitudes towards the brand or product, regardless of the relevance of the review. The fact that the negativity effect is shown significant, increases the potential damage consumer reviews can cause.

Companies and marketing managers should become more aware of the implied effect that online consumer reviews can have on their product or company's reputation. As suggested in section 6.2.1 *Theoretical Implications*, a possible explanation for why

consumers do not distinguish bad quality from good when processing information, is the excessive use of system 1 processing instead of system 2. This enhances the possibility of the consumers being affected by limited attention. In addition, even if system 2 is being used, the likelihood that the consumer is affected by limited attention is still high, as discussed in section 2.1 *Purchasing Process* (Kahneman, 2013). It could thus be beneficial to make consumer reviews of good quality more salient. As an example, Amazon provide the opportunity of rating consumer reviews as useful, where the top-rated reviews are displayed on top of the list of reviews. On popular products with an overload of reviews this appears to be a good way to improve the problem and the website shows that consumers are in fact using this feature. Several other websites use this type of consumer reviews model (elkjop.no, yelp.com) as well, and the recommendation from our side is to implement this as a standard option.

A popular nudge is to take advantage of the peer effect and articulate that most consumers write online reviews after purchasing a product (Thaler and Sunstein, 2008). Furthermore, it has been shown that standardization and simplification influences consumer behavior. We therefore believe it could be beneficial to make the process of writing online consumer reviews more standardized. This could be done by for instance introducing questions for the consumers to answer or rate instead of open fields where they have to compose their own review. Less required effort makes it more likely that consumers will perform an action.

Based on the findings in this study showing the effect of negative online reviews on consumers' purchase intention, companies should initiate actions to minimize the consequences. As online shopping is consistently increasing in popularity, it will be more and more important for companies to be present at the channels their customers use. Many customers with negative experiences leave a review on the public site. Their negative experiences are thus available for all possible future customers to see, possibly affecting their purchase intention negatively. The strategy of responding to eWOM is increasing, and is commonly called webcare (Willemsen, 2013). The aim of webcare is to limit the effects of negative eWOM, and has by Willemsen (2013, p. 13) been described as "the act of engaging in online interactions with consumers, by actively searching the web to address consumer feedback". With a focus on webcare, consumers can see that their responses are taken seriously by the company and limit their negative experience. Thus, good customer service is key also in an online environment.

7. Limitations and future research

This section will elaborate on the quality of this study regarding research design and procedures. Firstly, we understand that a research study is not without its limitations, and will discuss this in the first section. Subsequently, we will assess validity and reliability, followed by a discussion of future research and lastly, we will provide a conclusion of our thesis.

7.1 Limitations

In the following, we will discuss and reflect upon what we believe to be the weaknesses of this study. Firstly, the decision of the product, DAB radio, was based on the criteria that it should be a product that people use and would want other's opinions and experiences of. We ran a pretest to get an overview of DAB radio as a potential product and the results from this showed that all respondents knew of the product and there were no extreme answers on the expertise of this type of product. However, we did not assess the probability of future purchase of a DAB radio, which could be a valuable aspect to consider. In addition, the respondents gathered for the pretest were not all NHH students, as in our actual experiment. This could lead to the answers from the pretest not being fully representative for the chosen population. As discovered from the experiment, fewer than expected answered that it was likely that they would buy a DAB radio in the future, and there were also respondents that had no prior knowledge of this product. One potential interpretation could be that the respondents in this study, consisting of students at NHH, might not be interested in the product in general. Most students live on a tight budget and thus have less money to spend. Moreover, many live in shared flats or a rented apartment and might have a landlord responsible for purchasing products. The product choice could therefore be a limitation. Nevertheless, we assessed both prior knowledge of the product and probability of future purchase and controlled for the variables in the analysis as an attempt to control for this limitation.

Regarding the treatment set-up, there are several potential limitations that we need to assess. We modified the DAB radio used for the treatments by adding a fictitious brand name to avoid brand loyalty or prior attitude affecting the results. The fact that this would be an unknown brand for the respondents could in itself have a negative effect on the intention

to purchase. We discussed briefly in section 4.1.3 *Manipulations* that using a fabricated brand name instead of an existing brand could lead to lower influence on the purchase intention. Additionally, we used a picture of a DAB radio from an existing page, which implies that there is a risk that some respondents might have recognized it and thus have prior established attitudes. It is also possible that the DAB radio depicted was not desirable due to price, design or specifications. Other reasons for the general low mean of intention to purchase could be that the respondents have strong brand loyalty to another radio, for instance Tivoli. In addition, the arrangement of the reviews could have an impact on the respondents' purchase intention. Statista (2017) show through statistics that 67 % of consumers need 4 reviews or more to be able to trust a business, which indicates that having only three reviews in the experiment, where one was neutral, could be insufficient for the respondents to acquire a strong intention to purchase.

Another possible limitation could be that it is difficult to get the respondents in a mindset of an actual purchasing process. As described in section 2.1 *Purchasing Process*, the first step in the consumer decision-making process is need recognition, which is necessary to proceed to the following steps of information search, evaluation and decision. If the respondents did not recognize a need, they might not be as aware of the product information and they might not be willing to decide on whether they would want to buy the product or not. The low probability of future purchase can be an indication of low motivation to process the information that followed. Respondents who already have a need for a radio and wish to purchase in the foreseeable future, would more easily familiarize with the situation.

We did not find the moderating effect of argument quality significant and a discussion of possible limitations regarding this measurement is therefore beneficial. By comparing the gap between the means of both levels of valence and argument quality, we see that the respondents could distinguish between negative and positive reviews to a much higher degree than between low and high argument quality (gap was respectively 4.49 and 1.15). One interpretation of this is that compared to valence, people can have different perceptions of quality. The interpretation of the reviews computed, could thus be different amongst the respondents. This could be some of the reason why we did not find the effect of argument quality on the purchase intention significant. A drawback from the pretest regarding argument quality is that all subjects evaluated all items instead of dividing the questionnaire into two groups, e. g. high quality and low quality, which is the procedure for a between-subjects design. The design chosen was based on limitations regarding time and

gathering sufficient responses, seeing that a between subject design with two groups requires twice as many responses. It is thus possible that the answers made regarding argument quality for each review were influencing each other, indicating that the respondents "learned" how to answer the questions after assessing reviews of both types of quality. We believe that the study could have benefited from fine-tuning the argument quality with an additional pretest, but the previously mentioned limitations made this difficult.

We conducted the experiment as an online questionnaire. The online questionnaire was open for participation at any time and any place, making it impossible for us to control for the surroundings or other external factors. We consider this a weakness compared to for instance a laboratory experiment where the probability of distraction is lower. The box plot (Figure 7) illustrating the timing variable shows seven extremes who used over 95 seconds on the treatment. We consider it unlikely that it is necessary to spend that much time to process the treatment, and it is therefore likely that these respondents were distracted at some point. We addressed this issue in the analysis under section 5.2 *Manipulation checks*.

The final data sample of 155 respondents could be a weakness to this study. As discussed earlier, two treatment groups had less than 30 responses, which Pallant (2007) considers as a minimum.

7.2 Validity

A discussion of validity and reliability is crucial to ensure the quality of the research. Validity considers aspects such as the appropriateness of the measures used, accuracy of the results from the analysis and generalizability of the findings (Saunders et al., 2015). The three main categories of validity are internal validity, external validity and measurement validity.

Internal validity in an experiment is achieved if the manipulations of the independent variables can be shown to statistically lead to an outcome (Saunders et al., 2015). Some factors that can threaten the internal validity is important to have in mind. Change in instruments or measurements is an example of this. We have used the same measurements throughout the study ensuring that the data are comparable and possible to analyze. Mortality

is when participants chose to withdraw from the study which can affect the validity of the research. However, as the time frame in this study is cross-sectional, non-completed questionnaires will not affect the results or findings as only the completed answers will be analyzed. Nevertheless, this will affect the response rate which in turn may affect the accuracy of the results. The last threat to internal validity is ambiguity about causal direction. This may be present if there are other factors that can explain causal relationships found in the study. To minimize this threat, we included a control group in the experiment and assigned the respondents randomly to the five different treatment groups. In addition, we have included and accounted for several control variables that have been well documented in previous research on eWOM. A past or recent event can change the perceptions of the participants. It is possible that the participants of this study have acquired strong opinions about the chosen product, DAB radio. There have been several discussions of this in the media in Norway, and many people have been affected by the closing of the FM radio broadcasting. The threat of the participants not responding accurately due to fear of consequences is limited in this study, as it is anonymous and participation is voluntary.

The external validity regards the generalizability of the findings to other situations or populations (Saunders et al., 2015). External validity can be established by selecting a sample that represents the population. The sample in this study consist only of students at NHH, which can be a threat to the external validity. This sample is likely to be quite homogenous as they belong to the same age group, live in the same area and study within the same field of study. An experiment requires creating fictive situations where the respondents are asked to adapt to a given thought situation, which might further reduce the external validity. In addition, we measured the effect on a specific product and we can therefore not generalize our findings to other products or services, or the effect of eWOM on purchase intention in general.

Measurement validity includes face validity, construct validity, content validity and predictive validity (Saunders et al., 2015). In the context of questionnaires, content validity, predictive validity and construct validity are most commonly discussed. It is important to ensure that the questionnaire measure what you intend to measure. Content validity concerns whether the questions provide a sufficient data basis to answer the research questions. By doing a thorough literature review and discussions on the topic, one can ensure that the questions provide the necessary coverage. We have made sure that all measurements regarding independent and dependent variables have been collected from previous research

done on eWOM. The predictive validity, also known as criterion-related validity, refers to the accuracy of predictions of the questions. In this study, *Intention to Purchase* is used as a measure of predicting future buying behavior. This study does not include actual purchase or behavior, and the accuracy of the predictive questions are thus difficult to measure. Construct validity is concerned with whether a set of questions, or scale items, measure the intended concept. As mentioned above, we have used well established scales for our measurements and in addition, we have made sure that when merged together, they are internally consistent (provided by Cronbach's alpha). To limit the convergent and discriminant validity, we used factor analysis to measure the multi-item scales against each other, ensuring that the items did not overlap each other in terms of concepts (Breivik, 2017b).

7.3 Reliability

Reliability regards elements concerning replication (external reliability) and consistency (internal reliability) (Saunders et al., 2015). Internal reliability can be achieved by using more than one researcher and being consistent when coding data, analyzing it and interpreting the results. The reliability analysis from section 4.6 *Multi-Item Scales* shows that for the multi-item scales, we accept all factors. Thus, the internal reliability is acceptable. In addition, the research is considered reliable if it is possible to replicate an earlier research design and obtain the same findings. The external reliability is thus harder to achieve. It is therefore important to be transparent when reporting the findings and making it possible to replicate. To secure transparency, we provide the entire questionnaire and report all the results from our analyses in an appendix.

Several threats to the reliability may exist (Saunders et al., 2015). Participant errors is a common problem with online questionnaires. The mood or situational factors of the respondent can affect the answers and the optimal time to send out the questionnaire should be discussed. Right before lunch-break, Monday (start of new week could be stressful) and Friday (weekend-mood) should be avoided. We had in total three rounds of distributions to the student emails, and made sure to send the invitation email at different days and time of day. Furthermore, the questionnaire was in Norwegian and thus easily understandable for all respondents. The pilot test also ensured that the questionnaire would not be misunderstood. Another threat to reliability is participant bias, which refers to factors inducing fake

responses during the questionnaire. Fake responses could either be the participants trying to answer what they think we would want them to answer or that the motivation to complete the questionnaire is solely based on the possibility of winning a price. Questions addressing online reviews directly, were placed after the treatments to avoid the respondents guessing the purpose of the study and answer the questions thereafter. We made sure that we did not explicitly mention the purpose of the study in either the invitation email, introduction or debrief of the experiment. Furthermore, we clearly stated their anonymity and encouraged honest answers in the introduction of the questionnaire in order to reduce participant bias. In addition, the price we provided as incentive to participate in our questionnaire is considered fairly small, and we believe that the possibility of participant bias occurring because of this is limited.

Researcher error may occur if the interpretation is affected by external factors, such as stress, tiredness or mood (Saunders et al., 2015). With the use of an online questionnaire providing quantitative data that can easily be analyzed in a statistical software, we reduce the errors that might occur when the interpretation needs to be done manually as in a qualitative study. In addition, researcher bias could be a threat to reliability of the questionnaire. With this in mind, we have made sure that the questions are not leading or judgmental in any way, the names of the variables are not included in any of the questions and all questions are closed-ended. This limits the possibility of interpreting the answers subjectively. The only responses we have removed from the dataset are the incomplete responses, all other responses have been included in the analysis.

7.4 Research Ethics

It is important to consider ethical concerns throughout the research, from topic choice to discussion of results (Saunders et al., 2015). Saunders et al. (2015 p. 239) define research ethics as “the standards of behavior that guide your conduct in relation to the rights of those who become the subject of your work, or are affected by it”. Several codes of ethics within different research areas has been established to avoid ethical dilemmas based on different social norms or philosophical positions. The codes of ethics address how to avoid poor practice, malpractice and harm and promote ethical practices. The Norwegian National Research Ethics Committees (2014) has come up with some general guidelines for research

ethics in Norway. We have done our best to follow these guidelines and corresponding principles throughout our research, including behaving according to norms, respecting the people involved and avoiding negative consequences of our research.

Within the limits of the topic of our Major, Marketing and Brand Management, we were free to choose a topic of our own interest for this thesis, ensuring our academic freedom (NNREC, 2014). To the best of our ability, we have made choices regarding methodology that was suitable to answer our research question. We also designed the questionnaire with the ethical principles in mind. Before starting the questionnaire, the respondents received information about the purpose of the study, that participation was voluntary, and that we would keep their identity anonymous. To emphasize that the participation was voluntary, we clearly stated that by proceeding the questionnaire the respondents gave consent of participation. We used a setting in Qualtrics to avoid logging the IP addresses of respondents to keep all the responses anonymous. Furthermore, we did not collect any personal information to ensure the confidentiality of the respondents. We also constructed a separate questionnaire to receive the email addresses of respondents who wanted to take part in the drawing of gift cards. It was thus not possible to connect their email address to their responses on the prior questions.

A dilemma when designing the questionnaire was whether to force responses on the questions or not. From an ethical point of view, respondents should have the opportunity to not answer a question (Saunders et al., 2015). However, since the questionnaire was self-completed online, the consumers had the possibility to withdraw whenever they wanted, and thus had an opportunity to not answer a question. We included the forced responses to avoid respondents forgetting to answer a question by mistake, and in this circumstance, we considered this aspect more important. Moreover, we ensured that there were no questions in the questionnaire that would cause embarrassment or harm to any of the respondents, which we got a confirmation of from the pilot test. As a closure of the questionnaire, we included a short debrief, explaining to the respondents the purpose of the study as well as stating that we constructed the product information for research purposes only, to avoid for respondents having negative or unfavorable associations towards the actual product depicted.

We have done our best effort to report our findings with honesty, openness, systematicness and documentation, by including outputs from our analyses and provide thorough explanations of our procedures and the purpose of our study (NNREC, 2014). We

declare that neither of us has conflicts of interest and all of the sources we have used, have been reported and we have followed reference practices. To the best of our knowledge, we have not crossed national laws or regulations.

7.5 Future research

As we have elaborated on earlier, there have been several studies on eWOM and online reviews. The purpose of this thesis was to contribute to this literature by exploring valence and argument quality combined. However, given the time constraints and limitations of a master thesis, there are several additional aspects that we did not cover and could be of interest for further research within the topic. In this section we will therefore provide suggestions for further research.

Firstly, our findings did not support the hypothesis that the effect that valence of consumer reviews has on intention to purchase was moderated by argument quality. We believe that to be able to see the differences between the two levels of argument quality, involvement should be included. Involvement might make high quality reviews more visible because argument quality requires more effort and elaboration than valence.

In this thesis, we chose DAB radio as a product to measure the effects of consumer reviews. An aspect that we did not account for is brand effects. We encourage future researchers to test the effect of reviews on brands with strong brand loyalty, and compare them with less familiar brands. Furthermore, investigating the effects of online reviews on different types of products or services could provide useful information. We have looked at an electronic product. The eWOM effect might be different for experience or search goods, and hedonic or utilitarian products. Moreover, eWOM can appear in different forms and on different platforms. We have used a product page with corresponding reviews. Research on eWOM on other platforms, such as review sites (Tripadvisor) or social media (Facebook) might give different results.

The sample used in this thesis is quite small, and consists of a relatively homogenous group of people. Students, or young adults, might have different behavior online than other

age groups. We therefore believe that replicating our study to a larger sample including different age groups would provide a different result.

Lastly, we believe that our findings confirm the importance of eWOM today and encourage researchers to investigate the different aspects of this mentioned above.

7.6 Conclusion

In this thesis we have studied the effects of online consumer reviews on Norwegian students' intention to purchase. Previous research on eWOM has investigated several types of stimuli, including valence and argument quality, but never combined these two factors as far as we know. For this reason, we wanted to focus on the interaction effect of valence and quality on consumers' purchase intention, and developed the following research question:

RQ: *What effect does online reviews have on purchase intention, and what is the impact of valence and argument quality of the online reviews?*

In order to investigate our research question, we developed three hypotheses. Firstly, we expected that valence would have a direct effect on consumers' purchase intention (H1). Furthermore, we anticipated that negative online consumer reviews would have a stronger impact than positive online consumer reviews on intention to purchase (H2). Lastly, we proposed that argument quality would moderate the effect that valence has on consumers' intention to purchase (H3). We further developed an experiment, where the respondents were presented a product with corresponding consumer reviews. The reviews were manipulated to include two levels of both valence and quality, resulting in a total of five treatment groups when including the control group.

From our analysis and results we can draw two main conclusions. Firstly, that valence does have a direct effect on purchase intention, where the effect of negative online consumer reviews exceeds the effect of positive online consumer reviews. This is an indication that the negativity effect is present, which is also a prominent finding in previous studies. Secondly, we rather surprisingly found that argument quality did not moderate the effect of valence on consumers' intention to purchase. This suggests that consumers focus

more on the valence of an online review rather than the relevance of the content. However, we believe that higher involvement could lead to this interaction effect being significant.

Our results shed light on several important aspects regarding eWOM. Online reviews are proven to be an important part of consumers' decision-making process, and should thus be an important focus for companies in the future. In addition, we encourage future researchers to further investigate the role of argument quality and involvement.

References

Arndt, J. 1967. Role of Product-Related Conversations in the Diffusion of a New Product. *Journal of Marketing Research,* 4(3)**,** pp 291-295.

Bartholomew, D. J., Steele, F., Moustaki, I. & Galbraith, J. 2008. *Analysis of Multivariate Social Science Data*, Second Edition: Taylor & Francis Group.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C. & Vohs, K. D. 2001. Bad is stronger than good. *Review of General Psychology,* 5(4)**,** pp 323-370.

Berger, J. 2014. Word of mouth and interpersonal communication: A review and directions for future research. *Journal of Consumer Psychology,* 24(4)**,** pp 586-607.

Bickart, B. & Schindler, R. M. 2001. Internet forums as influential sources of consumer information. *Journal of Interactive Marketing,* 15(3)**,** pp 31-40.

Bone, P. F. 1995. Word-of-mouth effects on short-term and long-term product judgments. *Journal of Business Research,* 32(3)**,** pp 213-223.

Breivik, E. 2017a. Experiments. Forelesningsmateriell MBM400A: NHH.

Breivik, E. 2017b. Measurement. Forelesningsmateriell MBM400A: NHH.

BrightLocal. 2017. *Local Consumer Review Survey* [Online]. Available: https://www.brightlocal.com/learn/local-consumer-review-survey/.

Chatterjee, P. 2001. Online Reviews: Do Consumers Use Them? *Advances in Consumer Research,* 28(129-133.

Chen, Y. & Xie, J. 2008. Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix. *Management Science,* 54(3)**,** pp 477-491.

Cheung, C. M. K., Lee, M. K. O. & Rabjohn, N. 2008. The impact of electronic word-of-mouth: The adoption of online opinions in online customer communities. *Internet Research,* 18(3)**,** pp 229-247.

Cheung, C. M. K. & Thadani, D. R. 2012. The impact of electronic word-of-mouth communication: A literature analysis and integrative model. *Decision Support Systems,* 54(1)**,** pp 461-470.

Cheung, M. Y., Luo, C., Sia, C. L. & Chen, H. 2009. Credibility of Electronic Word-of-Mouth: Informational and Normative Determinants of On-line Consumer Recommendations. *International Journal of Electronic Commerce,* 13(4)**,** pp 9-38.

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

Chu, S.-C. & Kamal, S. 2008. The Effect of Perceived Blogger Credibility and Argument Quality on Message Elaboration and Brand Attitudes. *Journal of Interactive Advertising,* 8(2)**,** pp 26-37.

Dellarocas, C. 2003. The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. *Management Science,* 49(10)**,** pp 1407-1424.

Dellarocas, C. a. Z. X. a. A. N. F. 2007. Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing,* 21(4)**,** pp 23--45.

DMR. 2017. *63 Interesting Yelp Statistics and Facts* [Online]. Available: https://expandedramblings.com/index.php/yelp-statistics/3/ [Accessed November 12th.

DN. 2017. *Londons beste restaurant var «fake news»* [Online]. Dagens Næringsliv. Available: https://www.dn.no/nyheter/2017/12/08/0640/Utenriks/londons-beste-restaurant-var-fake-news [Accessed December 12th.

Fan&Fuel. 2016. *No online customer reviews means BIG problems in 2017* [Online]. Available: https://fanandfuel.com/no-online-customer-reviews-means-big-problems-2017/ [Accessed December.

Godes, D., Mayzlin, D., Chen, Y., Das, S., Dellarocas, C., Pfeiffer, B., Libai, B., Sen, S., Shi, M. & Verlegh, P. 2005. The Firm's Management of Social Interactions. *Marketing Letters,* 16(3)**,** pp 415-428.

Hayes, A. F. 2013. *Introduction to Mediation, Moderation, and Conditional Process Analysis, First Edition: A Regression-Based Approach,* First Edition: The Guilford Press.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G. & Gremler, D. D. 2004. Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing,* 18(1)**,** pp 38-52.

Hennig-Thurau, T. & Walsh, G. 2003. Electronic Word-of-Mouth: Motives for and Consequences of Reading Customer Articulations on the Internet. *International Journal of Electronic Commerce,* 8(2)**,** pp 51-74.

Herr, P. M., Kardes, F. R. & Kim, J. 1991. Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective. *Journal of Consumer Research,* 17(4)**,** pp 454-462.

Hoyer, W. D., MacInnis, D. J. & Pieters, R. 2012. *Consumer Behavior*, Sixth Edition, Ohio, USA: South-Western.

Ismagilova, E., Dwivedi, Y. K., Slade, E. & Williams, M. D. 2017. Electronic Word of Mouth (eWOM) in the Marketing Context: A State of the Art Analysis and Future Directions 1ed. Wales, UK: Springer.

Jiménez, F. R. & Mendoza, N. A. 2013. Too Popular to Ignore: The Influence of Online Reviews on Purchase Intentions of Search and Experience Products. *Journal of Interactive Marketing,* 27(3)**,** pp 226-235.

Kahneman, D. 2013. *Thinking, Fast and Slow*: Farrar, Straus and Giroux.

Kanouse, D. E. 1984. Explaining Negativity Biases in Evaluation and Choice Behavior: Theory and Research. *NA - Advances in Consumer Research*, 11(703-708.

Keller, K. L. 2013. *Strategic Brand Management: Building, Measuring, and Managing Brand Equity*, Fourth Edition, Harlow, UK: Pearson Education Limited.

KPMG. 2017. *The truth about online consumers: 2017 Global Online Consumer Report* [Online]. Available: https://assets.kpmg.com/content/dam/kpmg/xx/pdf/2017/01/the-truth-about-online-consumers.pdf.

Kulturdepartementet. 2017. *Dette betyr DAB for deg* [Online]. Regjeringen.no. Available: https://www.regjeringen.no/no/tema/kultur-idrett-og-frivillighet/film-og-medier/innsiktsartikler/dette-betyr-dab-for-deg/id2507614/ [Accessed 20.09.2017.

Lee, J. & Lee, J.-N. 2009. Understanding the product information inference process in electronic word-of-mouth: An objectivity–subjectivity dichotomy perspective. 46.

Lee, J., Park, D.-H. & Han, I. 2008. The effect of negative online consumer reviews on product attitude: An information processing view. *Electronic Commerce Research and Applications,* 7(3)**,** pp 341-352.

Lee, M., Rodgers, S. & Kim, M. 2009. Effects of Valence and Extremity of eWOM on Attitude toward the Brand and Website. *Journal of Current Issues & Research in Advertising,* 31(2)**,** pp 1-11.

Lee, M. & Youn, S. 2009. Electronic word of mouth (eWOM). *International Journal of Advertising,* 28(3)**,** pp 473-499.

Mayzlin, D. 2006. Promotional Chat on the Internet. *Marketing Science,* 25(2)**,** pp 155-163.

Metzger, M. J. 2007. Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology,* 58(13)**,** pp 2078-2091.

Mizerski, R. W. 1982. An Attribution Explanation of the Disproportionate Influence of Unfavorable Information. *Journal of Consumer Research,* 9(3)**,** pp 301-310.

NNREC. 2014. *General guidelines for research ethics* [Online]. The Norwegian National Research Ethics Committees. Available: https://www.etikkom.no/globalassets/general-guidelines.pdf.

OECD. 2007. Participative Web and User-Created Content, OECD (Organisation for Economic Co-operation and Development).

Pallant, J. 2007. SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS for Windows. Third Edition ed. New York, USA: Open University Press.

Park, C. & Lee, T. M. 2009. Information direction, website reputation and eWOM effect: A moderating role of product type. *Journal of Business Research,* 62(1)**,** pp 61-67.

Park, D.-H. & Kim, S. 2008. The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews. *Electronic Commerce Research and Applications,* 7(4)**,** pp 399-410.

Park, D.-H. & Lee, J. 2008. eWOM overload and its effect on consumer behavioral intention depending on consumer involvement. *Electronic Commerce Research and Applications,* 7(4)**,** pp 386-398.

Park, D.-H., Lee, J. & Han, I. 2007. The Effect of On-Line Consumer Reviews on Consumer Purchasing Intention: The Moderating Role of Involvement. *International Journal of Electronic Commerce,* 11(4)**,** pp 125-148.

Rozin, P. & Royzman, E. B. 2001. Negativity Bias, Negativity Dominance, and Contagion. *Personality and Social Psychology Review,* 5(4)**,** pp 296-320.

Saunders, M., Lewis, P. & Thornhill, A. 2015. *Research Methods for Business Students*, Seventh Edition, Harlow, UK: Pearson Education Limited.

Schindler, R. M. & Bickart, B. 2005. Published "Word of Mouth:" Referable, Consumer-Generated Information on the Internet. *Online Consumer Psychology: Understanding and Influencing Consumer Behavior in the Virtual World (Advertising and Consumer Psychology).* 1 ed. New Jersey: Lawrence Erlbaum Associates Inc.

Sen, S. & Lerman, D. 2007. Why are you telling me this? An examination into negative consumer reviews on the Web. *Journal of Interactive Marketing,* 21(4)**,** pp 76-94.

Skowronski, J. J. & Carlston, D. E. 1989. Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin,* 105(1)**,** pp 131-142.

Solomon, M. R., Marshall, G. W. & Stuart, E. W. 2011. *Marketing: Real People, Real Choices*, Seventh Edition, New Jersey, USA: Pearson.

Spears, N. & Singh, S. N. 2004. Measuring Attitude Toward the Brand

and Purchase Intentions. 26.

Statista. 2017. *Trust in businesses based on amount of online reviews read 2014-2017* [Online]. The Statistics Portal. Available: https://www.statista.com/statistics/315738/business-trust-number-online-reviews/ [Accessed 10.12.2017.

Tabachnick, B. G. & Fidell, L. S. 2012. *Using Multivariate Statistics*, Sixth Edition: Pearson.

Tabbane, R. S. & Debabi, M. 2015. Electronic Word of Mouth: Definitions and Concepts. *Advances in Marketing, Customer Relationship Management, and E-Services.* California, USA: IGI Global.

Thaler, R. H. & Sunstein, C. R. 2008. Nudge: Improving Decisions about Health, Wealth and Happiness. New Haven & London: Yale University Press.

Weinberger, M. G. & Dillon, W. R. 1980. The Effects of Unfavorable Product Rating Information. *NA - Advances in Consumer Research,* 07(528-532.

Willemsen, L. M. 2013. Electronic word of mouth: Challenges for consumers and companies,

Wright, P. 1974. The harassed decision maker: Time pressures, distractions, and the use of evidence. *Journal of Applied Psychology,* 59(5)**,** pp 555-561.

Xie, H., Miao, L., Kuo, P.-J. & Lee, B.-Y. 2011. Consumers' responses to ambivalent online hotel reviews: The role of perceived source credibility and pre-decisional disposition. *International Journal of Hospitality Management,* 30(1)**,** pp 178-183.

Zhang, W. & Watts, S. A. 2008. Capitalizing on Content: Information Adoption in

Two Online communities. *Journal of the Association for Information Systems,* 9(2)**,** pp 73-94.

Appendix

Content

A. Pretest

A.1 Descriptive statistics

Firstly, the respondents of the pretest were asked about their knowledge of DAB radio, given by two questions presented below.

Do you know what a DAB radio is?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	19	100.0	100.0	100.0

I have expert knowledge of DAB radio

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	1	5.3	5.3	5.3
	Partly disagree	4	21.1	21.1	26.3
	Neither agree nor disagree	5	26.3	26.3	52.6
	Partly agree	9	47.4	47.4	100.0
	Total	19	100.0	100.0	

Report

Expert Knowledge		
Mean	N	Std. Deviation
3.16	19	.958

A.2 Online Reviews

The respondents were presented four reviews from all four treatments, i.e. a total of 16 reviews. Negative reviews of high quality are given by NHQ1-4, Negative reviews with low quality are NLQ1-4, Positive reviews of high quality are PLQ1-4 and Positive reviews of high quality are PHQ1-4. NHQ1, NHQ3, NLQ1, NLQ4, PLQ1, PLQ4, PHQ1 and PHQ3 were kept for the experiment.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
NHQ1	19	2.60	5.00	4.0000	.67330
NHQ2	19	1.00	5.00	3.5895	.84715
NHQ3	19	1.00	5.00	3.7684	.90003
NHQ4	19	2.40	4.80	3.4737	.65730
NLQ1	19	1.00	3.20	1.8105	.71639
NLQ2	19	1.00	3.60	2.1158	.74927
NLQ3	19	1.00	4.80	2.4526	.92097
NLQ4	19	1.00	3.80	1.9895	.83924
PLQ1	19	1.00	3.40	2.0105	.75269
PLQ2	19	1.00	4.00	2.3895	.80684
PLQ3	19	1.40	4.80	2.9158	.99402
PLQ4	19	1.00	3.80	2.3684	.77535
PHQ1	19	3.40	5.00	4.2105	.55966
PHQ2	19	2.80	5.00	4.1053	.71295
PHQ3	19	3.40	5.00	4.4000	.62539
PHQ4	19	2.80	4.60	3.8632	.47165
Valid N (listwise)	19				

Table 1: Descriptive statistics (pretest)

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	HQ	4.0123	19	.44141	.10127
	LQ	2.1140	19	.55124	.12646

Table 2: Descriptive statistics argument quality

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	HQ - LQ	1.89825	.81077	.18600	1.50747	2.28902	10.205	18	.000

Table 3: Paired t-test on argument quality

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	P	4.5351	19	.77694	.17824
	N	1.2719	19	.61667	.14147

Table 4: Descriptive statistics valence

Paired Samples Test

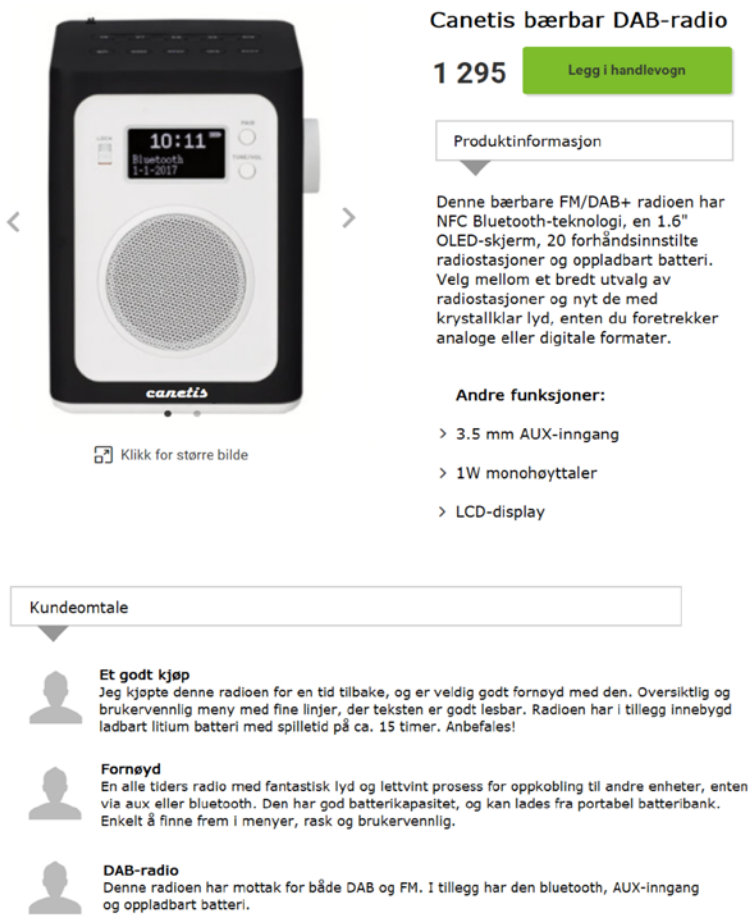
		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	P - N	3.26316	1.09194	.25051	2.73686	3.78946	13.026	18	.000

Table 5: Paired t-test valence

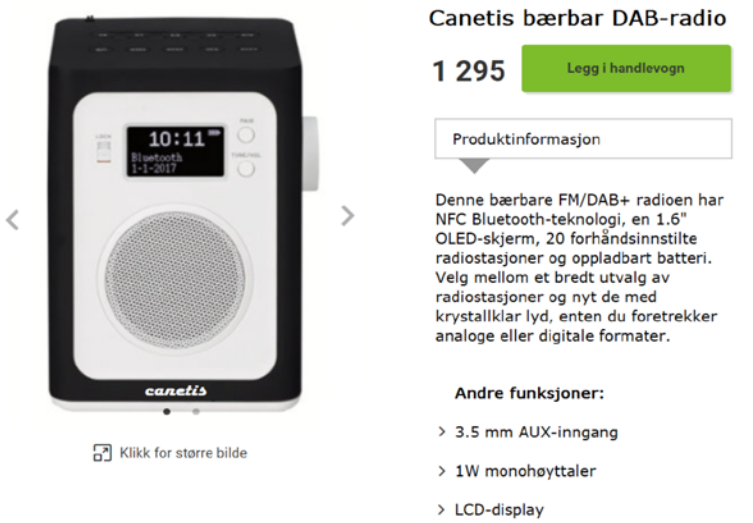
B. Experiment

B.1 Manipulations

B.1.1 Treatment 1: Positive High Quality



B.1.2 Treatment 2: Negative High Quality



B.1.3 Treatment 3: Negative Low Quality

Klikk for større bilde

Canetis bærbar DAB-radio

1 295 Legg i handlevogn

Produktinformasjon

Denne bærbare FM/DAB+ radioen har NFC Bluetooth-teknologi, en 1.6" OLED-skjerm, 20 forhåndsinnstilte radiostasjoner og oppladbart batteri. Velg mellom et bredt utvalg av radiostasjoner og nyt de med krystallklar lyd, enten du foretrekker analoge eller digitale formater.

Andre funksjoner:

- 3.5 mm AUX-inngang
- 1W monohøyttaler
- LCD-display

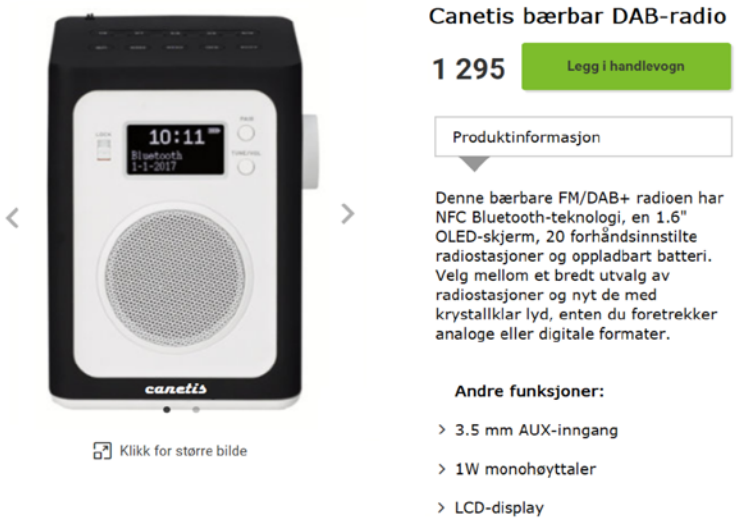
Kundeomtale

Et dårlig kjøp
Denne radioen var utrolig dårlig:(Skjønner ikke hvorfor jeg kjøpte den, dette var et mega bom kjøp. Nesten ikke brukt den å anbefalt alle mine venner å familie og holde seg unna Å dem har gjort det:) Ikke kjøp den!

Misfornøyd
Denne radioen var et impulskjøp fordi jentungen ville ha en egen på rommet sitt. Angrer veldig på kjøpet. Det værste med dette produktet er at dattera mi på 13 ble misfornøyd. Ikke en radio jeg vil ambefale!

DAB-radio
Denne radioen har mottak for både DAB og FM. I tillegg har den bluetooth, AUX-inngang og oppladbart batteri.

B.1.4 Treatment 4: Positive Low Quality



B.1.5 Treatment 5: Control group

B.2 The Questionnaire

B.2.1 Invitation

Kjære student,

I forbindelse med vår masteroppgave gjennomfører vi en spørreundersøkelse om forbrukeratferd på nett.

Undersøkelsen tar ikke mer enn **5 minutter** og ved å delta har du muligheten til å vinne kinogavekort.

Klikk her for å delta!

På forhånd tusen takk for hjelpen!

Mvh
Vilde Larsen og Ingrid Rabben

Follow this link to the Survey:
Take the Survey

Or copy and paste the URL below into your internet browser:
https://nhh.eu.qualtrics.com/jfe/form/SV_0ujE8RmY8lVyl1z?Q_DL=eX4iCI1P3lzCYkJ_0ujE8RmY8lVyl1z_MLRP_9WxRAAxOUKFNkVL&Q_CHL=email

Follow the link to opt out of future emails:
Click here to unsubscribe

B.2.2 Introduction

NHH

Kjære respondent,

Dette er en frivillig undersøkelse. Ved å delta bidrar du med verdifull informasjon i arbeidet med en masteroppgave ved Norges Handelshøyskole (NHH).

Undersøkelsen er anonym og tar ca 5 minutter å fullføre. På slutten av undersøkelsen vil du få valg om å delta i en trekning av kinogavekort.

Vi anbefaler at du bruker PC eller nettbrett, og undersøkelsen skal besvares indviduelt uten å kommunisere med andre. Les spørsmålene nøye og svar ærlig.

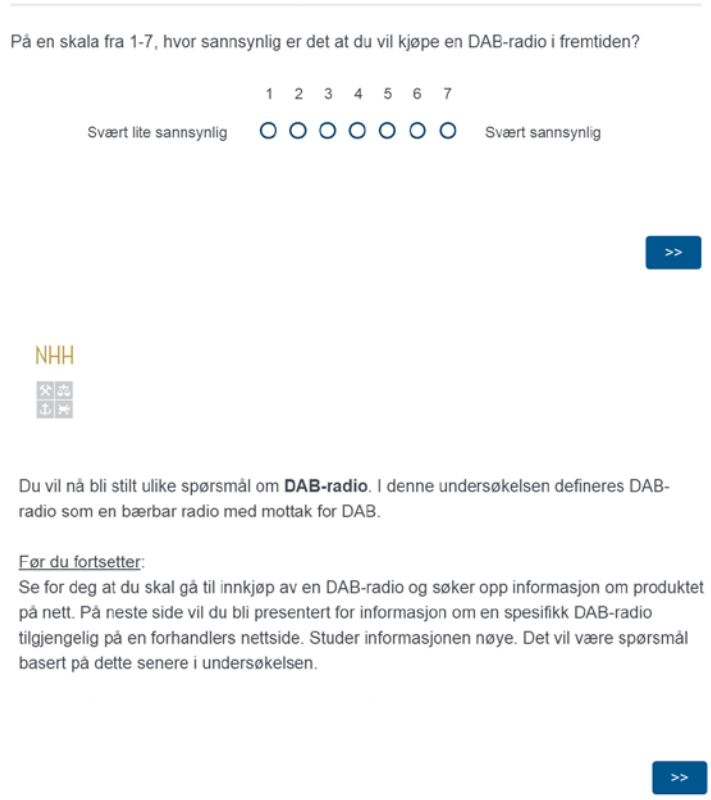
Ved å trykke deg videre til neste side, gir du samtykke til å delta i denne undersøkelsen.

Takk for at du tar deg tid til å delta!

>>

B.2.3 Questions

The online questionnaire used in the thesis is presented below. All four treatment groups received the same questions, while the control group did not receive questions regarding the manipulation checks of the reviews.



[Treatment here]

Hvor sannsynlig er det at du ville anbefalt DAB-radioen du nettopp ble presentert til venner?

	1	2	3	4	5	6	7	
Svært lite sannsynlig	○	○	○	○	○	○	○	Svært sannsynlig

Hvis jeg skulle tatt en kjøpsavgjørelse nå, ville informasjonen som ble gitt om DAB-radioen vært nyttig for meg

	1	2	3	4	5	6	7	
Svært uenig	○	○	○	○	○	○	○	Svært enig

>>

NHH

Du skal nå vurdere kundeomtalene som ble gitt om DAB-radioen du ble presentert tidligere. Ta utgangspunkt i den sammenlagte evalueringen og svar på en skala fra 1-7.

Kundeomtalene er...

	Svært uenig 1	2	3	4	5	6	Svært enig 7
...objektive	○	○	○	○	○	○	○
...forståelige	○	○	○	○	○	○	○
...troverdige	○	○	○	○	○	○	○
...tydelige	○	○	○	○	○	○	○
...av høy kvalitet	○	○	○	○	○	○	○
...positive	○	○	○	○	○	○	○

>>

NHH

Ta stilling til følgende påstander og svar på en skala fra 1-7.

Når jeg planlegger kjøp av et produkt, leser jeg alltid kundeomtaler om produktet på nett på forhånd

	1	2	3	4	5	6	7	
Svært uenig	○	○	○	○	○	○	○	Svært enig

Når jeg skal kjøpe et produkt, er kundeomtaler på nett hjelpsomme i beslutningsprosessen

	1	2	3	4	5	6	7	
Svært uenig	○	○	○	○	○	○	○	Svært enig

Kundeomtaler om produktet på nett gjør meg trygg på å kjøpe produktet

	1	2	3	4	5	6	7	
Svært uenig	○	○	○	○	○	○	○	Svært enig

Kundeomtaler om produktet på nett er irriterende

	1	2	3	4	5	6	7	
Svært uenig	○	○	○	○	○	○	○	Svært enig

Hvis jeg ikke leser kundomtaler om produktet på forhånd, blir jeg usikker på om jeg har gjort et riktig kjøp

	1	2	3	4	5	6	7	
Svært uenig	○	○	○	○	○	○	○	Svært enig

>>

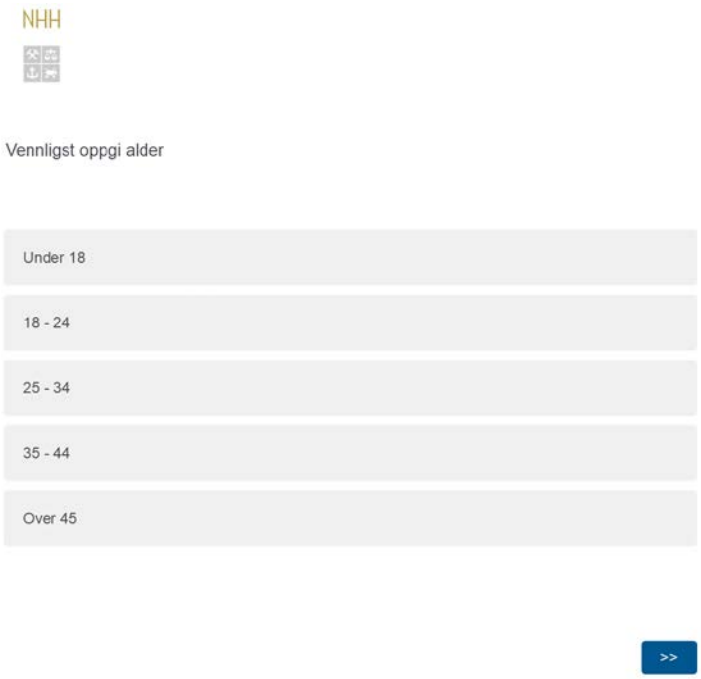
NHH

Vennligst oppgi kjønn

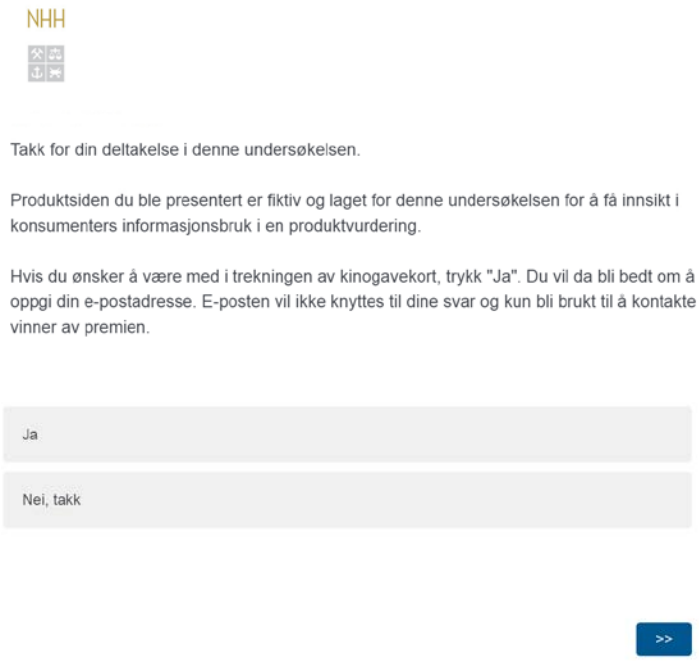
Mann

Kvinne

>>


NHH
Vennligst oppgi alder
Under 18
18 - 24
25 - 34
35 - 44
Over 45
>>

B.2.4 Debrief


NHH
Takk for din deltakelse i denne undersøkelsen.
Produktsiden du ble presentert er fiktiv og laget for denne undersøkelsen for å få innsikt i konsumenters informasjonsbruk i en produktvurdering.
Hvis du ønsker å være med i trekningen av kinogavekort, trykk "Ja". Du vil da bli bedt om å oppgi din e-postadresse. E-posten vil ikke knyttes til dine svar og kun bli brukt til å kontakte vinner av premien.
Ja
Nei, takk
>>

NHH

Takk for at du tok deg tid til å ta denne spørreundersøkelsen.
Svaret ditt er registrert.

Levert av Qualtrics

C. Measurement

C.1 Factor Analysis

Factor analysis for the variables *Review Attitudes, Argument Quality* and *Intention to Purchase.*

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.704
Bartlett's Test of Sphericity	Approx. Chi-Square	562.417
	df	66
	Sig.	.000

Table 6: KMO and Bartlett's test

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	3.839	31.992	31.992	3.839	31.992	31.992	3.486
2	2.036	16.966	48.958	2.036	16.966	48.958	1.954
3	1.615	13.456	62.414	1.615	13.456	62.414	2.505
4	.846	7.047	69.461				
5	.778	6.483	75.944				
6	.713	5.942	81.886				
7	.565	4.706	86.592				
8	.494	4.116	90.708				
9	.365	3.040	93.748				
10	.334	2.784	96.532				
11	.257	2.143	98.675				
12	.159	1.325	100.000				

Extraction Method: Principal Component Analysis.

a. When components are correlated, sums of squared loadings cannot be added to obtain a total variance.

Table 7: Total variance explained (eigenvalues)

C.2 Cronbach's Alpha

C.2.1 Intention to purchase

IP1 measures the probability of purchasing the product and IP2 measures the probability of recommending the product to friends.

Reliability Statistics

Cronbach's Alpha	N of Items
.860	2

Item Statistics

	Mean	Std. Deviation	N
IP1	2.9226	1.56892	155
IP2	2.7484	1.56916	155

C.2.2 Argument Quality

Reliability Statistics

Cronbach's Alpha	N of Items
.825	5

Item Statistics

	Mean	Std. Deviation	N
Objectivity	3.4634	1.73816	123
Understandability	5.0732	1.44954	123
Credibility	4.1382	1.46727	123
Clearness	4.7642	1.45468	123
Quality	3.4634	1.45023	123

C.2.3 Review Attitudes

The five variables are as follows: how often do the respondents read consumer reviews in a decision-making process (AR1), are consumer reviews considered helpful in a decision-making process (AR2), do consumer reviews make the respondents more confident in purchasing a product (AR3), do the respondents find consumer reviews irritating (AR4) and lastly, do not reading online reviews before a purchase make them worry about their decision (AR5).

Reliability Statistics

Cronbach's Alpha	N of Items
.714	5

Item Statistics

	Mean	Std. Deviation	N
AR1_1	5.17	1.325	155
AR2_1	5.65	1.194	155
AR3_1	5.14	1.276	155
AR4_1	5.85	1.223	155
AR5_1	3.25	1.613	155

Scale	AR1	AR2	AR3	AR4	AR5
1	4	1	0	52	30
2	2	1	5	62	29
3	13	8	11	19	24
4	13	11	29	13	28
5	55	41	46	6	34
6	48	52	40	2	9
7	20	41	24	1	1
Total	155	155	155	155	155

Table 8: Frequencies Review Attitudes

D. Analysis

D.1 Descriptive Statistics

Treatment

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	PHQ	4	11.4	28.6	28.6
	NHQ	6	17.1	42.9	71.4
	NLQ	2	5.7	14.3	85.7
	PLQ	1	2.9	7.1	92.9
	CON	1	2.9	7.1	100.0
	Total	14	40.0	100.0	
Missing	System	21	60.0		
Total		35	100.0		

Table 9: Distribution of incomplete responses

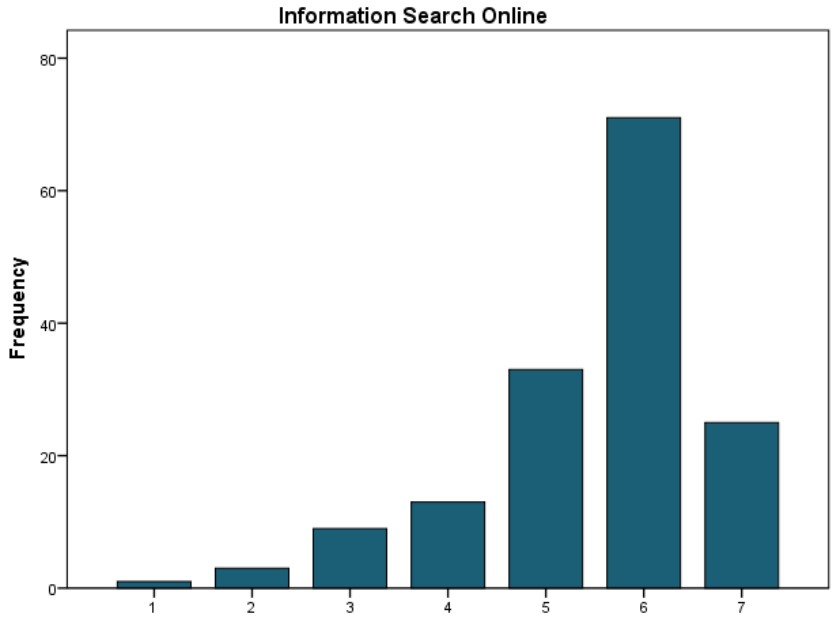
Treatment * Gender Crosstabulation

Count

		Gender		
		Male	Female	Total
Treatment	PHQ	20	9	29
	NHQ	9	19	28
	NLQ	18	15	33
	PLQ	16	17	33
	CON	14	18	32
Total		77	78	155

Table 10: Distribution of genders in treatment groups

The next chart shows the distribution on the question: when you are planning to buy a product, how often do you use the internet to find information about the product? The scale ranges from “Never” to “Always”.

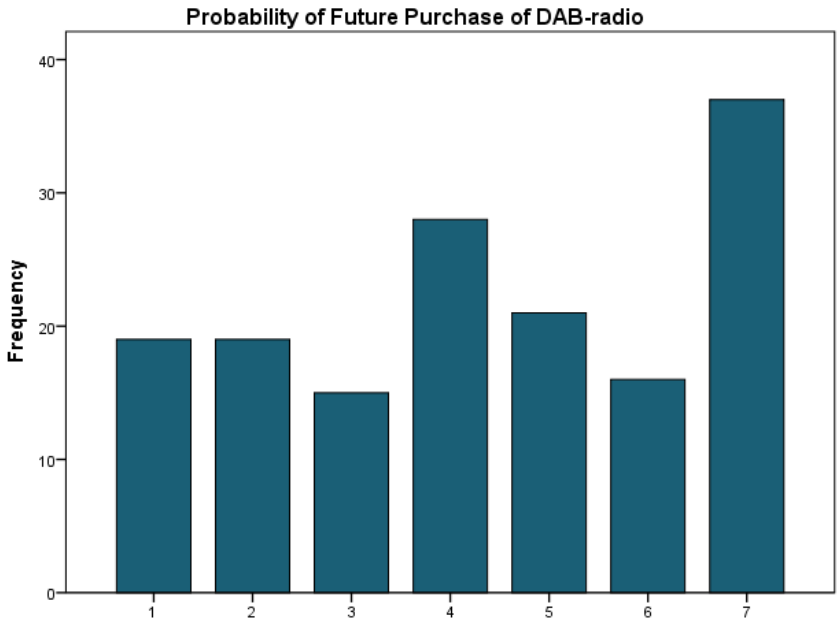


Prior Knowledge of DAB-radio

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	148	95.5	95.5	95.5
	No	7	4.5	4.5	100.0
	Total	155	100.0	100.0	

Prior Purchase of DAB-radio

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	98	63.2	63.2	63.2
	No	47	30.3	30.3	93.5
	Uncertain	10	6.5	6.5	100.0
	Total	155	100.0	100.0	



D.2 Manipulation checks

Manipulation check of Valence

Report

Positivity			
Valence	Mean	N	Std. Deviation
Negative	1.72	61	1.157
Positive	6.21	62	.871
Total	3.98	123	2.473

Table 11: Descriptive statistics of Valence

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
Positivity * Valence	Between Groups	(Combined)	619.431	1	619.431	592.328	.000
	Within Groups		126.536	121	1.046		
	Total		745.967	122			

Table 12: One-Way ANOVA of Valence

Manipulation check on argument quality

Report

Review Quality			
Quality	Mean	N	Std. Deviation
Low Quality	3.645	66	1.0993
High Quality	4.800	57	.9024
Total	4.180	123	1.1626

Table 13: Descriptive statistics of Argument Quality

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
Review Quality * Quality	Between Groups	(Combined)	40.770	1	40.770	39.737	.000
	Within Groups		124.144	121	1.026		
	Total		164.913	122			

Table 14: One-Way ANOVA of Argument Quality

Manipulation check on the timing variable

```
Timing Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     7,00        0 .  3333444
    15,00        0 .  555666667778889
     9,00        1 .  011122444
     9,00        1 .  555667788
    22,00        2 .  0001111122222233334444
    16,00        2 .  5666778888899999
    11,00        3 .  11112333444
    15,00        3 .  555666788889999
     5,00        4 .  01113
    13,00        4 .  5555577788999
     7,00        5 .  0012223
     6,00        5 .  567888
     3,00        6 .  011
     2,00        6 .  79
     4,00        7 .  0234
     2,00        7 .  89
     1,00        8 .  0
     1,00        8 .  6
     7,00 Extremes    (>=95)

 Stem width:     10,00
 Each leaf:       1 case(s)
```

Table 15: Stem-and-Leaf Plot of the time spent on the treatment

Descriptive statistics of the respondents' answers on *Intention to Purchase* divided into respondents spending more or less than 10 seconds on the treatment.

Descriptives

Intention to Purchase

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
>10	133	2.8459	1.46962	.12743	2.5938	3.0979	1.00	7.00
<10	22	2.7727	1.50180	.32019	2.1069	3.4386	1.00	5.50
Total	155	2.8355	1.46951	.11803	2.6023	3.0687	1.00	7.00

Table 16: Descriptive statistics

ANOVA

Intention to Purchase

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.101	1	.101	.046	.830
Within Groups	332.454	153	2.173		
Total	332.555	154			

Table 17: One-Way ANOVA of careless respondents (< 10 seconds)

Descriptive statistics of the respondents' answers on *Intention to Purchase* divided into respondents spending less or more than 95 seconds on the treatment. Respondents spending more than 95 seconds on the treatment were outliers.

Descriptives

Intention to Purchase

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
<95	148	2.8074	1.44340	.11865	2.5730	3.0419	1.00	7.00
>95	7	3.4286	1.98806	.75142	1.5899	5.2672	1.00	5.50
Total	155	2.8355	1.46951	.11803	2.6023	3.0687	1.00	7.00

Table 18: Descriptive statistics (outliers)

ANOVA

Intention to Purchase

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.579	1	2.579	1.196	.276
Within Groups	329.976	153	2.157		
Total	332.555	154			

Table 19: One-Way ANOVA of careless respondents (outliers)

D.3 Test of Assumptions

D.3.1 ANOVA

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Intention to Purchase

F	df1	df2	Sig.
3.493	1	121	.064

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Valence

Table 20: Levene's test H1

Test of Homogeneity of Variances

Intention to Purchase

Levene Statistic	df1	df2	Sig.
1.424	4	143	.229

Table 21: Levene's test H2

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Intention to Purchase			
F	df1	df2	Sig.
1.452	3	115	.231

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + PBDAB_1 + Valence + Quality + Valence * Quality

Table 22: Levene's test H3

D.3.2 ANCOVA

Tests of Between-Subjects Effects

Dependent Variable: Intention to Purchase					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	90.594[a]	3	30.198	18.915	.000
Intercept	86.984	1	86.984	54.485	.000
Valence	1.691	1	1.691	1.059	.305
PBDAB_1	12.437	1	12.437	7.790	.006
Valence * PBDAB_1	5.103	1	5.103	3.196	.076
Error	189.980	119	1.596		
Total	1231.500	123			
Corrected Total	280.573	122			

a. R Squared = .323 (Adjusted R Squared = .306)

Table 23: Homogeneity of regression slopes

D.4 Hypothesis 1

Between-Subjects Factors

		Value Label	N
Valence	0	Negative	61
	1	Positive	62

Descriptive Statistics

Dependent Variable: Intention to Purchase			
Valence	Mean	Std. Deviation	N
Negative	2.0000	1.15470	61
Positive	3.5484	1.44208	62
Total	2.7805	1.51650	123

Table 24: Descripitve statistics Valence

D.5 H2

First analysis on H2 was a one-way ANOVA with planned contrasts. Two planned contrasts were run: The first showing positive reviews (PHQ and PLQ) against the control group (CON). The second was negative reviews (NHQ and NLQ) against the control group.

ANOVA

Intention to Purchase					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	78.258	4	19.565	11.540	.000
Within Groups	254.297	150	1.695		
Total	332.555	154			

Contrast Coefficients

	Treatment				
Contrast	PHQ	NHQ	NLQ	PLQ	CON
1	.5	0	0	.5	-1
2	0	.5	.5	0	-1

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
Intention to Purchase	Assume equal variances	1	.510	.2836	1.798	150	.074
		2	-1.060	.2845	-3.727	150	.000
	Does not assume equal variances	1	.510	.2912	1.752	69.982	.084
		2	-1.060	.2696	-3.934	57.913	.000

For the second analysis, we filtered out respondents answering "No" on *Prior Knowledge.* Thereafter, the same one-way ANOVA with planned contrasts was run.

ANOVA

Intention to Purchase

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	81.164	4	20.291	12.825	.000
Within Groups	226.255	143	1.582		
Total	307.419	147			

Contrast Coefficients

	Treatment				
Contrast	PHQ	NHQ	NLQ	PLQ	CON
1	.5	0	0	.5	-1
2	0	.5	.5	0	-1

D.6 H3

Firstly, we ran a two-way ANOVA with *Valence* and *Quality* on *Intention to Purchase*.

Between-Subjects Factors

		Value Label	N
Valence	0	Negative	61
	1	Positive	62
Quality	0	Low Quality	66
	1	High Quality	57

Descriptive Statistics

Dependent Variable: Intention to Purchase

Valence	Quality	Mean	Std. Deviation	N
Negative	Low Quality	2.152	1.1284	33
	High Quality	1.821	1.1802	28
	Total	2.000	1.1547	61
Positive	Low Quality	3.424	1.4094	33
	High Quality	3.690	1.4905	29
	Total	3.548	1.4421	62
Total	Low Quality	2.788	1.4198	66
	High Quality	2.772	1.6341	57
	Total	2.780	1.5165	123

Tests of Between-Subjects Effects

Dependent Variable: Intention to Purchase

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	76.456[a]	3	25.485	14.858	.000	.272
Intercept	939.718	1	939.718	547.855	.000	.822
Valence	75.423	1	75.423	43.972	.000	.270
Quality	.032	1	.032	.019	.892	.000
Valence * Quality	2.711	1	2.711	1.581	.211	.013
Error	204.117	119	1.715			
Total	1231.500	123				
Corrected Total	280.573	122				

a. R Squared = .272 (Adjusted R Squared = .254)

Next, we ran the same analysis, but this time include a filter for *Prior Knowledge*.

Between-Subjects Factors

		Value Label	N
Valence	0	Negative	59
	1	Positive	60
Quality	0	Low Quality	65
	1	High Quality	54

Tests of Between-Subjects Effects

Dependent Variable: Intention to Purchase

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	79.924[a]	3	26.641	16.695	.000	.303
Intercept	886.724	1	886.724	555.672	.000	.829
Valence	77.992	1	77.992	48.875	.000	.298
Quality	.376	1	.376	.235	.628	.002
Valence * Quality	4.121	1	4.121	2.583	.111	.022
Error	183.513	115	1.596			
Total	1164.750	119				
Corrected Total	263.437	118				

a. R Squared = .303 (Adjusted R Squared = .285)

Lastly, we ran a two-way ANCOVA using filter on *Prior Knowledge* and *Probability of Future Purchase* as a covariate.

Descriptive Statistics

Dependent Variable: Intention to Purchase

Valence	Quality	Mean	Std. Deviation	N
Negative	Low Quality	2.172	1.1402	32
	High Quality	1.685	.9522	27
	Total	1.949	1.0776	59
Positive	Low Quality	3.424	1.4094	33
	High Quality	3.685	1.4686	27
	Total	3.542	1.4300	60
Total	Low Quality	2.808	1.4216	65
	High Quality	2.685	1.5880	54
	Total	2.752	1.4942	119

In addition to the analyses above, we ran two planned contrasts: one comparing the positive reviews (PHQ and PLQ) with the control group (CON), and the second comparing the negative reviews (NHQ and NLQ) with the control group.

Descriptives

Intention to Purchase

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum	Between-Component Variance
						Lower Bound	Upper Bound			
PHQ		27	3.685	1.4686	.2826	3.104	4.266	1.0	7.0	
NHQ		27	1.685	.9522	.1833	1.309	2.062	1.0	4.0	
NLQ		32	2.172	1.1402	.2016	1.761	2.583	1.0	5.0	
PLQ		33	3.424	1.4094	.2453	2.925	3.924	1.0	5.5	
CON		29	2.983	1.2355	.2294	2.513	3.453	1.0	5.5	
Total		148	2.797	1.4461	.1189	2.562	3.032	1.0	7.0	
Model	Fixed Effects			1.2579	.1034	2.593	3.002			
	Random Effects				.3718	1.765	3.830			.6332

ANOVA

Intention to Purchase

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	81.164	4	20.291	12.825	.000
Within Groups	226.255	143	1.582		
Total	307.419	147			

Contrast Coefficients

Contrast	Treatment				
	PHQ	NHQ	NLQ	PLQ	CON
1	1	0	0	-1	0
2	0	1	-1	0	0

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
Intention to Purchase	Assume equal variances	1	.261	.3264	.799	143	.425
		2	-.487	.3287	-1.481	143	.141
	Does not assume equal variances	1	.261	.3743	.697	54.708	.489
		2	-.487	.2724	-1.787	56.997	.079

Table 25: Planned contrast H3

D.7 Additional Analysis

Multiple Comparisons

Dependent Variable: Intention to Purchase
Tukey HSD

(I) Treatment	(J) Treatment	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
PHQ	NHQ	1.8682*	.3450	.000	.916	2.821
	NLQ	1.5381*	.3314	.000	.623	2.453
	PLQ	.2654	.3314	.930	-.650	1.181
	CON	.6428	.3338	.308	-.279	1.565
NHQ	PHQ	-1.8682*	.3450	.000	-2.821	-.916
	NLQ	-.3301	.3345	.861	-1.254	.594
	PLQ	-1.6028*	.3345	.000	-2.527	-.679
	CON	-1.2254*	.3369	.003	-2.156	-.295
NLQ	PHQ	-1.5381*	.3314	.000	-2.453	-.623
	NHQ	.3301	.3345	.861	-.594	1.254
	PLQ	-1.2727*	.3205	.001	-2.158	-.388
	CON	-.8954*	.3230	.049	-1.787	-.003
PLQ	PHQ	-.2654	.3314	.930	-1.181	.650
	NHQ	1.6028*	.3345	.000	.679	2.527
	NLQ	1.2727*	.3205	.001	.388	2.158
	CON	.3774	.3230	.769	-.515	1.269
CON	PHQ	-.6428	.3338	.308	-1.565	.279
	NHQ	1.2254*	.3369	.003	.295	2.156
	NLQ	.8954*	.3230	.049	.003	1.787
	PLQ	-.3774	.3230	.769	-1.269	.515

*. The mean difference is significant at the 0.05 level.

ANOVA

Intention to Purchase

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	78.258	4	19.565	11.540	.000
Within Groups	254.297	150	1.695		
Total	332.555	154			