FACULTY OF ECONOMICS AND MANAGEMENT

Chair of Marketing

The Effect of Rating Mechanism Design in Consumers' Online Rating Behavior

Supervisors: Prof. Dr. Sören Köcher
Dr. Frauke Kühn

Seminar: Recent Issues in Marketing Research

Winter term 2021/2022
Date of submission: 23.01.2022

Name:	E-Mail Address:	Student ID
Ihsan Lashari	ihsan.lashari@st.ovgu.de	236813
Lingrui Sun	LIngrui.sun@st.ovgu.de	234124
Saifur Rahman	saifur.rahman@st.ovgu.de	230218

Abstract

The study attempts to answer a question: what effect do today's rating system designs have on the consumer's online buying behavior. Customers can utilize information from online product reviews and can evaluate products during online shopping. The review platform usually features an open social component that allows customers to see how other members of the community have evaluated particular product reviews as well as the social standing of reviewers. In this research, we evaluate the relationship between online rating and review. We employ a fresh dataset derived from consumer ratings and reviews of 3000 products which were collected from Amazon.de. This dataset has been analyzed by using Excel and SPSS. Therefore, We have found that rating without reviews is still one of the popular sources of consumer feedback.

Keywords: Consumer Behavior, Online product reviews, Ratings, E-commerce, Electronic Word-of-Mouth

Table of content

List of Figures

List of Tables

1. Introduction

Ratings, including reviews, help customers become more aware of selecting the products and decide whether it is the desired products to purchase (Amazon.com). Ratings and reviews generally mean permitting customers to share their views and experiences. The consumers give it an overall star rating after using a product or service (power reviews, 2022). It has been observed that rating and reviews have significantly influenced the customer when taking a pre-purchase decision (Ho, Tan, Wu, 2014, P.1). Because consumer-generated ratings and reviews are the most reliable source for consumers, they can reduce product ambiguity and help consumers assess product quality (Koh, Hu, Clemons, 2010, P.1). There are mainly two types of online rating systems. First, consumers gather information about the products that have already been stored in various formats, such as rating, textual content, etc. Secondly, customers can share their experiences by giving only ratings and posting reviews (Ho, Tan, Wu, 2014 P.1). According to a survey from Power Reviews, about 97% of consumers typically read product ratings and reviews before making an online purchase, and about 81% of online purchasing occurs when the product page has ratings and reviews.

The internet revolution has changed our lifestyle and human behavior by increasing internet access and adoption worldwide (Statista, 2021). Since the past two decades, we have noticed that fast-growing advanced technology and the availability of the internet have led to a significant change in our shopping behavior (Köcher, 2018). Consumer purchasing behavior has changed from physical to online, and it has also become an indispensable part of the global business framework (Statista, 2021). In the past, we have seen that customers were made purchase decisions based on recommendations from their friends and family. However, they can now take decisions independently from a variety of reviews platforms (Köcher &

Köcher, 2021, P.582). While since the covid pandemic has been started, the immense amount of online purchase volume has been growing consistently. Consequently, the number and impact of rating and reviews have also increased during the pandemic, where it consistently converted at 25% higher rate than a year ago. (Power reviews, 2022). Meanwhile, during the pandemic, about 31% of online shoppers read more ratings and reviews, while 34% of consumers take fewer recommendations before making any online purchase (findstack, 2021).

The principal purpose of this seminar paper is to provide an overview of prior research on the effects of rating mechanisms on consumers' rating behavior and to identify gaps in the literature. In addition, differences in rating behavior across different rating mechanisms are explored by collecting and analyzing secondary data from Amazon.de. In this study, however, we will get an overall understanding of average rating and average reviews, the difference between rating and reviews, standard deviation (SD) for rating and Standard deviation (SD) for reviews, and skewness (Rating) and skewness (reviews) from nearly 3000 secondary data which has been collected from amazon.de. This finding gives us a statistical explanation of online ratings and reviews where it shows a deep insight between ratings and posting reviews.

In this seminar paper, we divided it into five main chapters. The second chapter is included theoretical background on existing and previous research about ratings and reviews, and was described to effects of rating mechanisms on consumers' rating behavior. The following chapter, chapter 3, explains research methodology, where the first part of it is analyzed pair sample t-test among three main variables, average ratings, standard deviation and skewness. Then, the bivariate correlation between rating and reviews shows a graphical frequency representation based on the output of P-value, T score, the difference between average rating and reviews, etc. In chapter 4 is discussed findings and makes a statement about online

shopping. In addition to chapter 4.2, the limitation is be described and mentioned further research direction. Finally, chapter 5 ends with a conclusion.

2.Theoretical Background

2.1 Literature Review about Previous Research

This research is most closely related to a stream of literature investigating the impact of online reviews and ratings on consumer rating behavior. The principle question of this research is: what is the effect of rating mechanism design on consumers' online rating behavior?. the aim of this research is to understand the impact of a rating mechanism design on a consumer's online purchase behavior. Customer reviews and ratings are available in a broad range of products and services, and they have become an important aspect of many consumers' decision-making processes (Mudambi & Schuff, 2010).

When consumers share their past experiences of particular products, it reduces the uncertainty associated with the quality of the product; consequently, it can impact consumers" decision-making (Chen et al., 2004). If consumer feedback is an important driver of sales, authors, publishers, and retailers may want to devote more resources to managing it. Because it can provide an insights into the development and marketing of new products that would otherwise be available only through costly customer interviews. Furthermore, understanding the effectiveness of rating systems can aid in the development of organizations' future technology initiatives (Chen et al., 2004).

Our research found out that the reviews (textual description) are more effective than overall ratings (star ratings), which is also evident from some recent research. this research proposes that the reviews that include text opinion (reviews) in a rating system make the attributes of a

product/service more appealing than the overall rating (Schneider et al., 2020). Furthermore, compared to numerical ratings, textual comments (reviews) are more efficient in providing fine-grained details about a retailer's previous transactions, which can not be known with raw data such as numerical ratings (Sun et al., 2020b). Nevertheless, another study based on psychology theory, "Accessibility Diagnosticity framework," proposed that multidimensional rating systems (detailed ratings about specific product's attributes) are more appealing to consumers than the overall rating for a product or service (Schneider et al., 2020b). The attribution theory also proves that. According to this, people try to collect as much information as possible, and then they make a casual judgment out of it (Chakraborty & Bhat, 2017). According to (Schneider et al., 2020), the impact of electronic word of mouth (ratings, reviews, tweets, etc) depends on platforms, product characteristics, matrix, and their overall interplay. Moreover, keeping the effects of the product price and rating as a focus, during purchase decisions with the use of "EEG Technique" it was proved by a study that, the behavioral results of their study reflected that lower product price and positive online feedback had a positive impact on customer's purchase decision (Sun et al., 2020). Understanding how online recommendations and user feedback impact retailers' strategy and market dynamics is critical. Therefore, the study suggests that previous research has put that the presence of a customer review system can improve the consumer perception of the website. For instance, sites such as, amazon.com obtain customer reviews for several reasons to work as a mechanism to enhance the site's "stickiness" (Mudambi & Schuff, 2010).

(Weathers et al., 2015) revealed that Customers must overlook pages having text-based information (reviews) to acquire a good qualitative review of the product. That way, retailers can also know the customer desires the type of data and as a result they can lower the costs and become more efficient communicators. Considering the importance of textual reviews, another study suggested a relation between incidence and evaluation. Their work identified

four types of review posters Individuals with Low-involvement, community builders, Bandwagoners and differentiators. The overall findings suggested that online opinions on a review platform are dominated by the "community builder individuals" the participation by this group enhances the overtime, unlike the low-involvement individuals whose participation results in the decrease of the overtime (Moe & Schweidel, 2011). Search cost might be a factor that may discourage the customers from making a purchase decision and affect the efficiency of allocation of resources. However, with the advancement in information and technology, the search cost and product-related information have been reduced significantly (Chen et al., 2004). Another very recent study's results suggested that the length of a review impacts the overall product star ratings; longer reviews are packed with negative comments, resulting in a negative perception compared to the shorter reviews (Ramachandran et al., 2021).

Now talking about the gaps, the community's impression of the validity of the opinions represented in the review is impacted by the effectiveness of individual reviews and the reputation of individual reviewers. Because according to the theory of social psychology, credibility has an impact on the persuasiveness of the message, and here the credibility may depend on the author or the content itself (Chen et al., 2008). Likewise, another research has indicated the same gap recently and has suggested that in future research, more and more focus should be devoted to analyzing the content of eWom, which is currently limited. Many studies use a narrow set of matrices, namely, numerical ratings or volume, while neglecting the reviews' information full of consumer expressions (Babić Rosario et al., 2016). However, there has been much research conducted on the helpfulness of reviews. However, importance should also be given to research mechanisms to enhance the effectiveness of reviews to ease the decision-making process (Lackermair et al., 2013).

Another important gap was discussed in a study. There is very little information in the previous research about how a community's perception is altered by the quality of individual reviews and the reviewer's reputation (Dhanasobhon & Smith, 2007). Likewise, another study emphasizes that it is significantly more critical to reveal how reviews can reflect a product's actual quality (Köcher & Köcher, 2021). In addition, the misalignment of the star rating and the review text may also result in a systematic distortion of product quality to consumers. Reading all of the reviews in detail is not a practical option because many items have thousands of reviews (Mudambi et al., 2014).

Throughout the literature, there is steady proof that the rating mechanism effectively reduces customer uncertainty and helps in customer decision-making. It is evident that from the above discussion that the rating and reviews system work as a driving force for sales. It reduces customers' costs in terms of time and efforts and facilitates the decision-making process. Most importantly, it has been observed that reviews compared to numerical ratings are more efficient in providing detailed information regarding the product and service. limited research is conducted on the content of reviews (Babić Rosario et al., 2016). Although star ratings and text features have been utilized to explain aspects of online evaluation and customer behavior, the relationship between stars and language needs further analysis (Mudambi et al., 2014).

2.2 Introduction to “The Effect of Rating Mechanism Design in Consumers’ Online Rating Behavior”

Research has shown that in today's digital world, consumers' decisions, especially online purchase-related decisions, are affected by the available customer evaluations (rating and reviews) on the rating platforms. In the present day, many online platforms give customers

access to online ratings and reviews. Hence the product review is a text-based review of a customer which describes all the characteristics (Pros and Cons) of a product or service. Unlike reviews, product rating reflects the customers' views on a specific scale (Lackermair et al., 2013). Furthermore, following the same path, "The theory of Buying decision" states that consumers collect the information of a product during the pre-purchase stage, once they recognize the need for it (Ho et al., 2014). Because of the rating platforms, businesses can engage directly with their customers or accurately distribute current information about their products and services to their customers (Tantrabundit & Jamrozy, 2018). Many studies have also enhanced the understanding of the effect of eWOM in the previous 15 years. The study's overall research revealed that the use of eWOM decreases the uncertainty and aids them in choosing a suitable offering (Babić Rosario et al., 2016b). Information conveyed via ratings, reviews, tweets, etc., is regarded as an electronic word of mouth (eWom) (Babić Rosario et al., 2016). Consumers use eWOM because it decreases their uncertainties and helps them choose the best product, which impacts the bottom line (Babić Rosario et al., 2016). The rating mechanism's objective is that the rating system providers tend to be seen as trusted providers of the product or services, consequently attracting the customers and monetizing the website (Schneider et al., 2020b). Based on how they interact with the system, users of an online rating system can be divided into two types. The first type of user is the review reader (information receiver), who collects product information in various formats, including numeric ratings, written material, and multimedia files. Then the second type includes the people who share their opinions mostly after a post-purchase decision (Ho et al., 2014).

3.Methodology

3.1 Method of collecting data

In this research, secondary data is the main object of analysis. The advantages of secondary data analysis are cost-effectiveness and time savings (Johnston, M., 2014, p.648). The data were collected from publicly available information on Amazon's Ranking. Compared to other e-commerce websites, such as eBay, Douglas, Amazon's rating system has a long history, is more sophisticated, and is more suitable for research. Amazon has been opening up the rating function regarding books since 2004, and as technology has developed, this rating system has been used for various products. The extensive experience of star ratings and the long-term optimisation of the rating system has provided a wealth of actual data for the research. Amazon is one of the world's largest online providers, with around $950 million daily turnovers. People often look at products and product reviews before placing an order. This means that a large number of real users can take part in the survey. The data collected is authentic and reliable.

The data was collected by Webharvy, a javascript to parse data from the relevant HTML pages in Windows System. Webharvy can grab data from Amazon's Rating automatically. There are rankings for every category of product in Amazon, ranking the 100 most popular items, with the ranking updated every hour. In this rating system, the act of evaluating a product is divided into two types: star ratings and reviews. Before writing a review, the author has to light up the stars before commenting, while some consumers will only light up the stars. The number of ratings and reviews is the leading data.

Figure 1: Percentage of rating and review.

However, according to our dataset, 75% of consumers only gave ratings, and 25% wrote a text about the product and their experiences.

Different price ranges and different purchase frequencies are conditions for selecting objects for analysis. The data was divided into three catalogs (Stationary, Electronics and Sports). Each catalog contains ten different product items. Each product item is collected with basic information on the top 100 sellers. In this research, information from a total of 3,000 items was collected. Table 1 shows 30 Product Categories about this research.

Table 1: Summary of product categories

Main Categories	Product Categories
Electronics	Action Camera
	Digital Camera
	Earphone

	SD card
	Laptop
	Mouse
	Smartphone
	Tablet
	Monitor
	Keyboard
Office & Stationary	Calculator
	Copy Paper
	Desk Pad
	Lever
	Liquid
	Permanent Maker
	Scissor
	Shredder
	Stamp Pad
	White Board
	Billiards Cloths
	Gloves

Sports	Match Ball
	Official Football
	Sport Sock
	Sport T Shirt
	Track Suit
	Meta Towel
	Track Botton
	Trouser

The software consists of 10 attributes, including Products names, Average ratings, Number of ratings, Price of product, URL, global 5-star ratings, global 4-star ratings, global 3-star ratings, global 2-star ratings, and global 1-star ratings.

Price was the primary condition for the selection of the research subjects. Figure 2 summarizes the average prices of the 100 items under each product category. As Figure 2 shows, the prices of the three selected items vary significantly. The highest prices were for electronic devices, including Desk Pads Mouse Micro SD Earphone Action Cam Monitor Smartphone Tablet Digital Cameras Laptop. The prices of these products fluctuate the most (from €29.96 to €707.95) and are the least frequently purchased.

Figure 2: The Average Price of Products.

The next is Sporting Goods (Billiards Cloth, Official football, Sport sock, Meta Towels, Gloves, Tracksuit, Track bottom, T-shirt, Match ball), which has a very stable price fluctuation (from 19.11 to 50.66€) and is purchased more frequently.

The last category is Office Stationery (Calculator, Whiteboard, Desk Pads, Stam Pad, Permanent Markers, Copy paper, Shredders, Liquid, Scissors, and Lever), which has the lowest average price of all the data. With prices ranging from €7.66 to €92.19, the median stays at €11.77 and is a high purchase frequency for some consumable items.

After collecting this data, it was divided into two parts, 5-star ratings to 1-star ratings and 5-star reviews to 1-star reviews. As Amazon's website does not give the data which only gives star ratings, this is obtained by the total number of comments minus the number of reviews. Then Excel calculated Standard Deviation, Variance, and Skewness for original data. The data needs to be cleaned to get a proper text review on which analysis in SPSS can be performed. When cleaning up the scraped data, the next step is to remove some calculation errors in Excel. Missing data has existed when Webharvy collects data. For example, zero reviews due to being out of stock or to removing from Amazon will be recorded as 0 by

webharvy, and excel may make errors when using formulas, such as #DIV/0! It is essential to clear this missing data.

All processed output is stored in an Excel file for further use. And the file is then loaded into the SPSS for use in analysis and summarization. The final database consisted of nearly 3,000 products across three categories.

3.2 Methodology of analyzing data

3.2.1 Study1: Analyzing the relationship between total rating and total review

The purpose of this section of the study is to initially examine whether there are differences between the rating system and review system.

3.2.1.1 Material and Hypothesis

With the rapid development of the internet, people's lives are becoming more dependent on the internet. Online shopping platforms can provide consumers with a large number of similar items in a short time. Although many useful online objects are helpful, it is not easy to accurately rank their quality. Therefore, Amazon has introduced an online rating system(Liao et al., 2014). As feedback from consumers, these reviews largely influence potential consumers' willingness to buy or purchase behavior. Therefore, online reviews are an essential influencing factor for online sales(Zhang et al., 2020). Amazon offers star ratings, text reviews, or a combination of both. This platform provides well-established research data (Lackermair et al., 2013). In the following study, the relationship between ratings and reviews is the purpose of the study. Therefore, this paper will take the relationship between

rating and review as the object of observation. In order to reduce the correlation between ratings and reviews, in the following hypotheses and experiments, the data used are the number of only ratings and the number of only reviews.

Table 2: Statistical Features of the Number

Statistics

		Number of Rating	Number of Review
N	Valid	2896	2896
	Missing	101	101
Mean		2761.387463	631.8649427
Median		135.0000000	51.00000000
Skewness		22.026	18.528
Std. Error of Skewness		.045	.045
Kurtosis		616.241	523.110
Std. Error of Kurtosis		.091	.091
Minimum		-1.00000000	.0000000000
Maximum		666349.0000	118283.0000

In November 2021, 3,000 products on Amazon were collected. As table 2 shows, there were 2896 reliable messages about the number of ratings and 2,896 reliable messages about the number of reviews. The missing data is because the products are out of stock or there are no consumers involved in the rating. However, the overall reliability of the database is very high. The mean about the number of ratings is 2,761, and the mean number of reviews is 632. The median is less influenced by limit values than the mean, and the median is 135 ratings and 51 reviews. The maximum value of ratings is 666,349, and the maximum value of reviews is 118,283. The overall number of ratings is greater than the number of reviews from the data obtained. Consumers prefer star ratings to text reviews of products. Regarding the predictions of the two review mechanisms, we believe that there is a difference and that the number of ratings is higher than the number of reviews.

H0 of study1 is no difference between rating system and review system for Average score, Standard deviation and skewness. This Hypothesis implies no relationship between two different rating mechanisms on consumers' online rating behavior. H1 is a significant difference between these two systems for three variables. Different review mechanisms affect consumers' online rating behavior. Among them, the rating system means no text review, just involvement in star rating. Review system means only text comments.

Paired t-analysis is used. The paired sample t-test procedure compares the means of two variables in a separate group. This procedure calculates the difference between the values of the two variables for each case and tests whether the mean difference is not 0. Many researchers use paired t-tests to assess the mean difference between matched data points(Hedberg & Ayers, 2015). We begin our experiment by exploring differences between paired data, so paired t-tests will be applied in the following experiments.

3.2.1.2 results and discussion

The data under study were the number of only ratings and only reviews and their mean, variance, standard deviation, and skewness. These data, i.e., the differences in pairings, follow a normal probability distribution. The paired samples were simple random samples from Amazon with the same probability of being selected into the data set. In data analysis, we did descriptive statistics and Paired t-tests. All these analyses were implemented by using SPSS. Descriptive Statistics allows us to observe whether the object under study is normally distributed or not. The more normally distributed the data are, the more suitable they are for paired t-testing. Descriptive Statistics can visually show the skewness of the data and filter the data for paired t-testing. These data are then used later for paired t-testing. The significance value of the t-test is set to 0.05 and the results assess the difference between two

independent factors. If the resultant value is higher than the typical value of 0.05, no significant difference between these two factors indicates that H0 holds, and vice versa, H1 holds (Hedberg & Ayers, 2015).

In general, the skewness calculated by SPSS can tell whether the data are normally distributed or not. When skewness approximately equals to 0, the distribution can be considered symmetric and obeys normal distribution, and the data are suitable for t-test. The skewness of the normal distribution affects the t-test, but not as much as the confidence interval (Lim & Lim, 2016).

Table 3: Skewness of data

Statistics							
		Average Rating	SD (Rating)	Skewness (Rating)	Average Rating (Review)	SD (Review)	Skewness (Review)
N	Valid	2996.000	2996.000	2996.000	2996.000	2996.000	2996.000
	Missing	1.000	1.000	1.000	1.000	1.000	1.000
Skewness		-2.256	-0.224	51.710	-1.798	15.343	1.334
Std. Error of Skewness		0.045	0.045	0.045	0.045	0.045	0.045
Kurtosis		3.846	0.166	2780.439	2.000	407.017	13.221
Std. Error of Kurtosis		0.089	0.089	0.089	0.089	0.089	0.089

From Table 3, compared with other data, the skewness of average rating and review, SD of ratings, and skewness of review are closer to 0, which are eligible for the paired t-test, and conform to the normal distribution in this case. The normal distribution dose not significantly characterize the other data, but they still could be examined by paired t-test. This has little impact on the next test.

In the following paired t-test, the p-value and t-value are emphasized. The p-value is <0.05, which means that the difference between without text review and text review is very significant. And t-value reflects how significant the difference is.

Table 4: Parameter Estimates for Paired Sample Statistics

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Average Rating	3.97052489	2996	1.35519451	.024758865
	Average Rating (Review)	3.61516272	2996	1.41959236	.025935388
Pair 2	SD(Rating)	.841103113	2996	.423005940	.007728150
	SD(Review)	1.13728423	2996	1.02302027	.018690174
Pair 3	Skewness(Rating)	-1.5566889	2996	12.1877979	.222666222
	Skewness (Review)	-1.1819210	2996	1.14145530	.020853935

Table 5: Parameter Estimates for Paired Sample Test

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	Average Rating - Average Rating (Review)	.355362168	.936483295	.017109177	.321815239	.388909096	20.770	2995	<.001
Pair 2	SD(Rating) - SD(Review)	-.29618112	1.04637796	.019116909	-.33366472	-.25869752	-15.493	2995	<.001
Pair 3	Skewness(Rating) - Skewness (Review)	-.37476790	12.1227824	.221478414	-.80903311	.059497317	-1.692	2995	.091

The results are shown in table 4 and table 5. In the comparison between average rating point for without text review M=3.97, SD=1.35) and with text review (M=3.61, SD=1.42); t=20.77, p =0.00, the difference is significant. The difference about standard deviation between without text review (M=0.84, SD=0.42) and with text review (M=1.14, SD=1.02) ; t=-15.49, p = 0.00 is significant. There is not a significant difference about Skewness in the scores for without text review (M=-1.56, SD=12.19) and with text review (M=-1.18, SD=1.14) ; t=-1.69, p = .091.

In order to obtain more detailed results, a more granular examination was performed in study 2, where the database was sorted by product type.

3.2.2 Study 2: Analyzing the relationship between rating and review in different product categories

3.2.2.1 Material and Hypothesis

The aim of this study 2 was to provide a descriptive overview and illustrate a framework about the average rating system and average reviews system with text among N=30 product categories, which contains nearly 3000 individual product lists from amazon.de. After that, we conducted a pair sample t-test for every 30 categories where we got a precise result about the mean value of variables, t-test score, and p-value. We randomly assigned and compared the results among three main dependent variables: average rating and average rating (reviews), which was described the overall mean between ratings and reviews. In addition, the mean differences between Standard Deviation (Rating) and Standard Deviation (Reviews) were described how far the observed value from the mean. Moreover, Skewness (Rating) and Skewness (Reviews) were better described to evaluate the likelihood of events falling in the tails of a probability distribution.

3.2.2.2 results and discussion

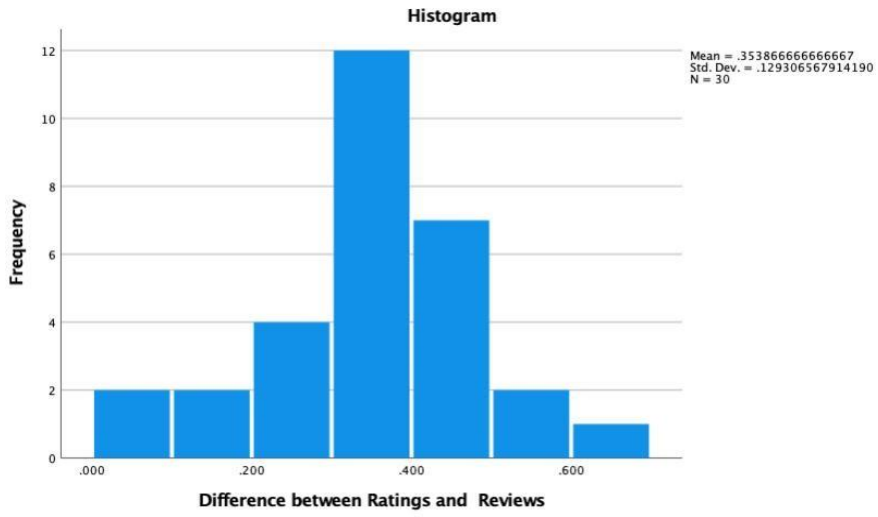
The following table 6 compares average ratings and average rating with review, which have very close mean value where the difference mean score was at M=0.35. The mean difference of standard deviation is at 0.12. In addition to, the t-test score indicates the difference between two variables where the more significant the t score means, the more difference between the variables. And more miniature t score means similarity between the variables (Wan Nor, 2014). In this Table 6, the t score was noticed a positive value with the highest value t=9.11 and lowest t=0.99. Furthermore, almost in all categories (N=30), the p-value was p=<.001, which is statistically significant, therefore, the null hypothesis can be possibly rejected at p<.001.

Table 6: Pair Sample T Test result Average Rating and Average Rating (reviews)

Product Categories	Average Rating	Average Ratings(Reviews)	T	P
Action Camera	3,4	3,2	1,18	0,241
Digital Camera	4,1	4,0	0,99	0,326
Earphone	3,9	3,6	5,19	<.001
SD card	3,6	3,3	6,53	<,001
Laptop	4,1	3,8	4,78	<,001
Mouse	4,5	4,1	9,11	<,001
Smartphone	4,2	3,9	4,60	<.001
Tablet	4,6	4,3	6,11	<,001
Monitor	4,4	4,1	5,32	<,001
Keyboard	4,4	4,0	8,51	<,001
Calculator	3,4	3,1	3,06	0,003
Copy Paper	4,5	4,1	7,13	<.001
Desk Pad	4,3	4,0	3,89	<,001
Lever	4,7	4,4	4,42	<,001
Liquid	4,3	3,9	7,96	<,001
Permanent Maker	4,4	4,0	6,67	<,001
Scissor	4,2	3,8	5,02	<,001
Shredder	4,1	3,8	4,87	<,001

Stamp Pad	3,5	3,0	3,43	<,001
white Board	4,2	3,7	6,00	<.001
Billiards Cloths	2,1	1,4	4,03	<,001
Gloves	3,8	3,6	3,18	0,002
Match Ball	3,8	3,5	3,56	<,001
Official Football	3,7	3,2	4,55	<,001
Sport Sock	4,3	4,1	4,45	<,001
Sport T Shirt	4,2	3,9	3,37	0,001
Track Suit	3,9	3,8	1,11	0,268
Meta Towel	2,8	2,5	2,17	0,032
Track Botton	3,8	3,3	3,32	0,001
Trouser	3,8	3,3	3,56	<,001

Figure 3: Difference between Average Rating and Average Ratings (reviews).



In figure 3, the histogram reveals the differences between rating and review in the horizontal axis and the frequency level with the vertical axis. In general, the histogram (figure: 03) measures the difference between variables, average ratings and average rating(reviews) of N=30 product categories. Using SPSS to make differences between rating and reviews, figure 3 shows that the mean difference between rating and reviews was at M=0.35 with frequency 12 and SD=0.13, and is distributed the difference value between the two variables contains between 0.088 and 0.681. So, there is a slight difference between ratings and reviews. Therefore, when the difference between the variables is equal to 0 or higher, that means the value of the average ratings is higher than the average rating (reviews).

Meanwhile, standard deviation generally represents an indication of how far the observed value is from the mean. The following Table 7 shows the standard deviation results for average ratings and reviews along with t score and p-value score from each product category (N=30), which was analyzed based on a pair sample t-test. In particular, the analysis was conducted based on three main categories, electronics, office & stationery, and sports. The product category, Trouser, has the highest standard deviation score (1.16 for rating and 3.20 for reviews) and lowest score for SD card at 0,527 for rating and billiards cloths at 0.501 for rating reviews. Billiards clothes and Track bottom have the most t score at t=2.93 and t=2.7 respectively, while monitor has the lowest score at t=-16.278. It does mean that the difference between average ratings (SD) and reviews (SD) is higher in case of Billiards cloths and Track bottom, and lower in terms of monitor. However, the p-value of Stamp Pad, Gloves, and Track Suit did not appear significant at p >.05. In contrast, the p-value from the rest of the attributes is less than 0.05. Therefore, H0 can be rejected, meaning that there is a significant difference between SD (ratings) and SD (reviews).

Table 7: Pair Sample T Test result of SD (Rating) and SD (reviews)

Product Categories	SD (Rating)	SD (Reviews)	T	P
Action Camera	0,859	1,169	-4,061	<.001
Digital Camera	0,870	1,235	-9,116	<.001
Earphone	0,866	1,176	-7,905	<.001
SD card	0,527	0,914	-10,73	<.001
Laptop	0,814	1,208	-7,529	<.001
Mouse	0,786	1,247	-18,12	<.001
Smartphone	0,912	1,187	-5,547	<.001
Tablet	0,796	1,120	-8,844	<.001
Monitor	0,743	1,209	-16,278	<.001
Keyboard	0,823	1,291	-13,498	<.001
Calculator	0,672	0,922	-3,548	<.001
Copy Paper	0,758	1,140	-7,697	<.001
Desk Pad	0,868	1,000	-1,942	0,055
Lever	0,722	0,953	-4,098	<.001
Liquid	0,756	1,122	-8,604	<.001
Permanent Maker	0,839	1,185	-8,909	<.001
Scissor	0,817	1,037	-3,683	<.001
Shredder	0,906	1,234	-5,754	<.001

Stamp Pad	0,867	0,887	-0,251	0,803
White Board	0,910	1,166	-4,099	<.001
Billiards Cloths	0,785	0,501	2,93	0,004
Gloves	0,935	0,993	-0,869	0,387
Match Ball	0,851	1,783	-3,626	<.001
Official Football	0,823	0,968	-1,568	0,12
Sport Sock	0,807	1,127	-10,643	<.001
Sport T Shirt	0,879	1,078	-4,218	<.001
Track Suit	1,160	1,167	-0,632	0,529
Meta Towel	0,641	0,724	-1,051	0,296
Track Bottom	1,075	0,841	2,7	0,0008
Trouser	1,164	3,205	-4,672	<.001

Figure 4: Mean difference between SD (Rating) and SD (Reviews).

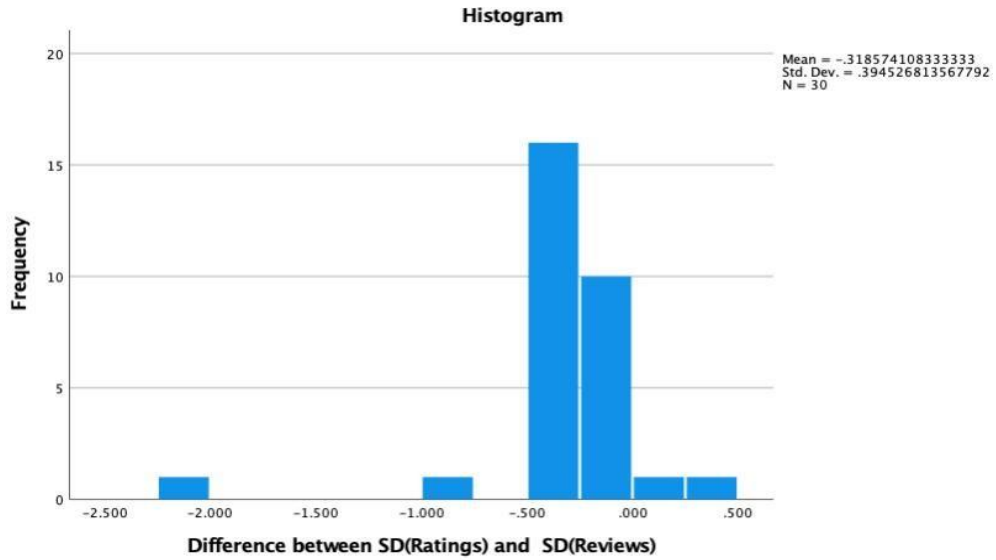


Figure 4 is compared the mean difference of SD (Rating) and SD(Reviews) at the level from 0.5 to -2.5. In general, the histogram (Figure:4) measures the difference between variables, SD (ratings) and SD (reviews) from N=30 product categories. The bar chart (figure 4) shows that the mean difference between SD (rating) and SD (reviews) is at M=-0.32 with a frequency about 17 and SD=0.39. So, The result from the difference between the variables is negative, which means the value of the SD(Reviews) is higher than SD (Rating).

Skewness generally means the measure of how much the probability distribution of a random variable deviates from the normal distribution (Analytics Vidhya, 2020). From Table 8, it is noticed that regarding the skewness, most of the values of skewness, both in ratings and reviews, have negative, whereas trousers have only positive value at 4.97 for ratings and 0.94 for reviews. Moving on to the attribute t, where most product categories have negative value apart from two categories, namely copy paper and Trouser, have positive results at t=0.82 and t=16.98 respectively. On the other hand, the p-value of SD card, Billiards Cloths, Sport T-Shirt is not considered statistically significant at p >.05. At the same time, the remaining categories are observed statistically significant at p <.05.

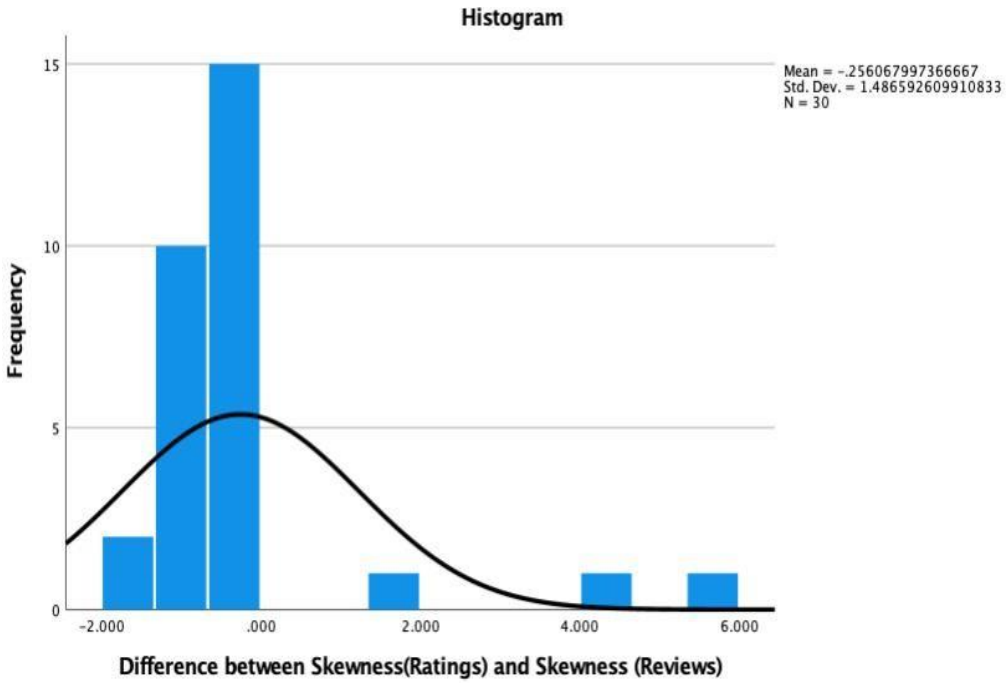
Table 8: Pair Sample T Test result of Skeness (Rating) and Skewness(reviews)

Product Categories	Skewness (Rating)	Skewness (Reviews)	T	P
Action Camera	-1	-0,807826	-4,127	<.001
Digital Camera	-2,208975	-1,406959	-7,055	<.001
Earphone	-2,482474	-1,063165	-3,031	<.001
SD card	-1,759858	-1,532366	-0,88	0,381
Laptop	-2,220212	-1,237818	-9,358	<.001
Mouse	-2,992503	-1,448629	-8,343	<.001
Smartphone	-2,184033	-1,285752	-11,178	<.001

Tablet	-0,112646	-1,952965	-10,765	<.001
Monitor	-2,732266	-1,473162	-16,549	<.001
Keyboard	-2,299915	-1,371888	-10,09	<.001
Calculator	-1,630728	-1,01324	-4,279	<.001
Copy Paper	3,790233	-1,615675	0,826	<.001
Desk Pad	-2,053333	-1,437883	-5,516	<.001
Lever	-2,887942	-2,015932	-8,487	<.001
Liquid	-2,625592	-1,43544	-11,482	<.001
Permanent Maker	-2,315738	-1,419545	-4,273	<.001
Scissor	-2,044336	-1,408988	-5,93	<.001
Shredder	-1,234243	-1,109295	-9,018	<.001
Stamp Pad	-1,828531	-0,878428	-2,213	0,029
White Board	-1,791775	-1,203028	-5,439	<.001
Billiards Cloths	-0,225005	-0,155748	-1,535	0,128
Gloves	-1,539842	-1,058101	-6,24	<.001
Match Ball	-1,7829903	-1,167477	-7,344	<.001
Official Football	-1,451635	-0,846043	-5,411	<.001
Sport Sock	-2,248304	-1,810966	-6,219	<.001
Sport T Shirt	-2,3777	-1,544044	-1,841	0,069
Track Suit	-1,166583	-1,051032	-1,56	0,122

Meta Towel	-1,001209	-0,738304	-2,719	0,008
Track Botton	-1,392902	-0,903286	-5,356	<.001
Trouser	4,97889803	0,943706651	16,982	<.001

Figure 5: Mean difference between Skewness (Rating) and Skewness (Reviews).



It has been clearly seen that the above distribution is positively skewed with the difference value of two variables (6 to -2 on the horizontal axis) and frequency (0 to 15 on the vertical axis). This positive skew indicates the tail on the right side is longer than on the left (right-skewed). The mean value, however, is M=-0.26 with SD=1.48. This means that the difference between skewness is skewed positively when variables in a group score less than the average score for their group.

4.Summary of findings

4.1 General Discussion

The results showed that the average score of only star rating is higher than the average score of text review. When consumers are not very satisfied with the product being paid for, they prefer text reviews, leading to a lower mean score than star ratings (Ramachandran et al., 2021). The standard deviation implies the volatility of this data set, and the standard deviation of reviews is m larger than the standard deviation of ratings. This indicates that most of the values in the number of text reviews differ more from their mean values and the more extreme values. This result reflects the fact that there are many extremes in the number of reviews, and that consumers may concentrate on leaving text reviews for some products and none for others.

In prior studies, researchers have concluded that customers prefer to leave text reviews with negative emotions (Ramachandran et al., 2021). This is similar to the previous findings, where the average rating of reviews was lower than ratings, regardless of the product category. These reviews, in turn, influence potential customers to make judgments and thus affect product sales. In the future, researchers can continue to deepen their research on the extent to which rating and review systems interact with each other. Provide more reference information for online shopping platforms.

4.2 limitation of findings

Two research studies have both strengths and weaknesses, and the limitations in a paper are explicitly focused on methodology (Connelly, Lynne M., 2013). Limitations are very important, and a better discussion about limitations could prevent some losses (Ioannidis, 2007), and these shortcomings could be avoided in future studies.

First, the biggest obstacle in the preliminary investigation was the lack of prior research studies on the topic. The most discussed topic was the relationship between price, sale volume and eWOM. There are fewer studies about the direct relationship between rating and review, so it was difficult for the researcher to get opinions or past views as a guide and reference for this study. Therefore, we found many papers about the price-eWOM relationship, sales volume-eWOM relationship. In chapter 2, we summarized this literature, found the reasonable argument for this article, and did our research. However, due to the lack of sufficient literature support, our research results are novel, but some parts may not be mature.

Second, there are limitations in terms of sample size and collection method in data collection. The size of sample size can have some impact on qualitative research. A sample size that is too small does not adequately support the claim of drawing valid conclusions, while a sample size that is too large does not allow for the in-depth, naturalistic, inductive analysis that characterizes qualitative inquiry. Determining the appropriate sample size in qualitative research is ultimately a matter of judgment and experience (Huberman, A. M., & Miles, M. B., 1994). In this thesis, the samples were collected without prior market research and were randomly selected, thus leading to many missing data. In future experiments, the sample size can be expanded to reduce the impact of missing data.

With the development of online shopping platforms, consumers' choices, evaluations, and online shopping experiences are becoming more and more easily quantifiable. More and more variables can be analyzed, and more questions about marketing and consumers can be answered.

5.Conclusion

Digital innovations, particularly the rapid development of internet technology, have really changed consumer buying behavior (Jin, 2017). Consumers are giving more priority to ratings and reviews, which has significantly influenced the customer when making a final purchase decision. The early research on online product reviews has studied how user-generated ratings can be related to market performance, whereas the recent work focuses on the impact of existing ratings on subsequent ones from a social dynamics standpoint. This paper, however, examines rating behavior and attempts to explain the online rating mechanism where we have found that the majority of the consumer prefer to give only ratings than reviews, writing text. In addition to another finding, consumers are more willing to give more ratings with posting texts when they become dissatisfied with particular products. Therefore, although the review is recognized as the most effective source of information about the products, the product rating, without text, still becomes a more popular way of giving feedback about the experiences of the product from the customer side.

References

Babić Rosario, A., Sotgiu, F., de Valck, K., & Bijmolt, T. H. (2016). The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of Platform, Product, and Metric Factors. *Journal of Marketing Research*, *53*(3), 297–318. https://doi.org/10.1509/jmr.14.0380

Chakraborty, U., & Bhat, S. (2017). The Effects of Credible Online Reviews on Brand Equity Dimensions and Its Consequence on Consumer Behavior. *Journal of Promotion Management*, *24*(1), 57–82. https://doi.org/10.1080/10496491.2017.1346541

Chen, P.-Y., Dhanasobhon, S., & Smith, M. D. (2008). All Reviews are Not Created Equal: The Disaggregate Impact of Reviews and Reviewers at Amazon.Com. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.918083

Chen, Wu, & Yoon (Eds.). (2004). *The Impact of Online Recommendations and Consumer Feedback on Sales*. http://aisel.aisnet.org/icis2004

Deaton, A. (2021). COVID-19 and Global Income Inequality. *LSE Public Policy Review,1*(4). https://doi.org/10.31389/lseppr.26

Dhanasobhon, & Smith. (2007). An Analysis of the Differential Impact of Reviews and Reviewers at Amazon.com. *International Conference on Information Systems*. https://aisel.aisnet.org/icis2007/.

Hedberg, E., & Ayers, S. (2015). The power of a paired t-test with a covariate. *Social Science Research*, *50*, 277–291. https://doi.org/10.1016/j.ssresearch.2014.12.004

Ho, Y.-C. C., Wu, J., & Tan, Y. (2014). Disconfirmation Effect on Online Rating Behavior: A Dynamic Analysis. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2516409

Ho, Y, Tan, Y., & Wu, J. (2014). *Effect of Disconfirmation on Online Rating Behavior: A Dynamic Analysis,* Information Systems Research 28(3):626-642. https://doi.org/10.1287/isre.2017.0694

Ioannidis, J. P. (2007). Limitations are not properly acknowledged in the scientific literature. *Journal of Clinical Epidemiology*, *60*(4), 324–329. https://doi.org/10.1016/j.jclinepi.2006.09.011

Kocher. S (2016), Reaching for stars: Consumers' *Interpretations Online Rating Distributions and Their Validity as an Indicator of Product Quality*, (thesis), TU Dortmund University

Kocher. S, Kocher. S, (2021), *The Mode Heuristic in Service Consumers' Interpretations of Online Rating Distributions*, (thesis), Journal of Service Research, Vol. 24(4) 582–600 https://www.analyticsvidhya.com/blog/2020/07/what-is-skewness-statistics/

Lackermair, G., Kailer, D., & Kanmaz, K. (2013). Importance of Online Product Reviews from a Consumer's Perspective. *Advances in Economics and Business*, *1*(1), 1–5. https://doi.org/10.13189/aeb.2013.010101

Liao, H., Zeng, A., Xiao, R., Ren, Z. M., Chen, D. B., & Zhang, Y. C. (2014). Ranking Reputation and Quality in Online Rating Systems. *PLoS ONE*, *9*(5), e97146. https://doi.org/10.1371/journal.pone.0097146

Lim, W. K., & Lim, A. W. (2016). A Comparison Of Usual t-Test Statistic and Modified t-Test Statistics on Skewed Distribution Functions. *Journal of Modern Applied Statistical Methods*, *15*(2), 67–89. https://doi.org/10.22237/jmasm/1478001960

Lin, G., Su, J., & Deng, Y. (2006). Improving the ease of use and efficiency of software tools. *Proceedings of the International Astronomical Union*, *2*(SPS5), 319–324. https://doi.org/10.1017/s1743921307007223

Moe, W. W., & Schweidel, D. A. (2011). Online Product Opinions: Incidence, Evaluation and Evolution. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1525205

Mudambi, & Schuff. (2010). Research Note: What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. *MIS Quarterly*, *34*(1), 185. https://doi.org/10.2307/20721420

Mudambi, S., Schuf, Zhang. (2014). Why Aren't the Stars Aligned?An Analysis of Online Review Content and Star Ratings. *Conference. 47th Hawaii International Conference on System Science.* https://ieeexplore.ieee.org/abstract/document/6758991.

Ramachandran, R., Sudhir, S., & Unnithan, A. B. (2021). Exploring the relationship between emotionality and product star ratings in online reviews. *IIMB Management Review.* https://doi.org/10.1016/j.iimb.2021.12.002

Schneider, C., Weinmann, M., Mohr, P. N. C., & vom Brocke, J. (2020a). When the Stars Shine Too Bright: The Influence of Multidimensional Ratings on Online Consumer Ratings. *Management Science*. https://doi.org/10.1287/mnsc.2020.3654

Schneider, C., Weinmann, M., Mohr, P. N., & vom Brocke, J. (2020b). When the Stars Shine Too Bright: The Influence of Multidimensional Ratings on Online Consumer Ratings. *Management Science*. https://doi.org/10.1287/mnsc.2020.3654

Sun, L., Zhao, Y., & Ling, B. (2020a). The Joint Influence of Online Rating and Product Price on Purchase Decision: An EEG Study. *Psychology Research and Behavior Management*, Volume 13, 291–301. https://doi.org/10.2147/prbm.s238063

Sun, L., Zhao, Y., & Ling, B. (2020b). The Joint Influence of Online Rating and Product Price on Purchase Decision: An EEG Study. *Psychology Research and Behavior Management*, Volume 13, 291–301. https://doi.org/10.2147/prbm.s238063

Tantrabundit, & Jamrozy. (2018). Influence of Online Review and Rating System towards Consumer Preferences in Hospitality Sector. *Journal of Business Diversity*, *18*(4). https://doi.org/10.33423/jbd.v18i4.248

Wan Nor, A. (2014). Calculating standard deviation of difference for determination of sample size for planned paired t-test analysis. *Education in Medicine Journal*, *6*(2). https://doi.org/10.5959/eimj.v6i2.252

Weathers, D., Swain, S. D., & Grover, V. (2015). Can online product reviews be more helpful? Examining characteristics of information content by product type. *Decision Support Systems*, *79*, 12–23. https://doi.org/10.1016/j.dss.2015.07.009

Zhang, J., Zheng, W., & Wang, S. (2020). The study of the effect of online review on purchase behavior. *International Journal of Crowd Science*, *4*(1), 73–86. https://doi.org/10.1108/ijcs-10-2019-0027

Amazon.com, Customer review, retrieve from. https://www.amazon.com/gp/help/customer/display.html?nodeId=G3UA5WC5S5UUKB5G

Findstack ,2021, Retrieve from https://findstack.com/online-review-statistics/

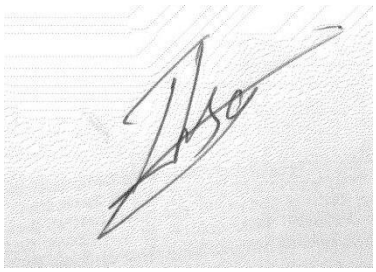
Power Review, Complete Guide for Ratings and Reviews, 2022, Retrieve from https://www.powerreviews.com/insights/ complete guide ratings-reviews-2022/

Statista , 2021, Retrieve from https://www.statista.com/topics/871/online-shopping/#dossierKeyfigures

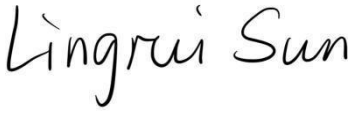
Declaration under oath

We hereby solemnly declare that we have created this seminar paper independently without any other aid but the one indicated in the paper. Every part of the seminar paper where we adapted ideas word by word or analogously from published or unpublished sources are labeled as such.
In addition, we declare that this seminar paper has not been (not even partly) submitted to any other examination board.

Ihsan Lashari
23.01.2022



Lingrui Sun
23.01.2022



Saifur Rahman
23.01.2022