Revisiting Review Depth in Search for Helpful Online Reviews

Shardul Dorwat
The University of Queensland
uqsdorwa@uq.edu.au

Morteza Namvar
The University of Queensland
m.namvar@uq.edu.au

Saeed Akhlaghpour
The University of Queensland
s.akhlaghpour@uq.edu.au

Abstract

This study investigates online review features that constitute review depth and assess their impacts on review helpfulness. It develops a model capturing the moderating effects of heuristic and systematic cues of an online review on the relationship between review length and its helpfulness. This study examines the moderating effects of price, product type, review readability and the presence of two-sided arguments. For testing the model, a dataset of 568,454 reviews from 256,059 different reviewers on Amazon.com were analyzed. The variables were operationalized using test processing techniques and relationships were empirically tested using regression and machine learning models. The results highlight significant moderating effects of review readability and the presence of two-sided arguments on the relationship between review length and its helpfulness. However, the results did not confirm the moderating effects of price and product type. This article discusses the significant implications for a better understanding of review depth and helpfulness in e-commerce platforms.

Keywords: online review, helpfulness, depth, length, readability, two-sided argument

1. Introduction

E-commerce websites have become an essential distribution channel for most businesses in the 21st century. These revenue streams are often larger than traditional brick and mortar stores that consumers interact with (Fresneda & Gefen, 2019; Namvar & Chua, 2022). In traditional brick and mortar stores, consumers can rely on product experts and sales assistants to help them understand the appropriateness of a product. However, this physical interaction or advice is not available in the e-commerce space. Instead, consumers rely on online reviews published by fellow patrons of the e-commerce platforms who have purchased the products (Banerjee & Chua, 2021; Mudambi & Schuff, 2010). These reviews act as e-word of mouth (eWoM), as consumers share their anecdotal experience when they use the products or services. Consumers will evaluate multiple reviews available in the online review platform to form their comprehensive opinion before acting on their purchasing intensions (Wu et al., 2021), which involves high level of interaction between the consumers and e-commerce platforms. This is referred to as the online consumers' stickiness (Chen, 2016). Since the introduction of online reviews in 2009, Amazon alone, estimates that the online reviews have added $2.7 billion to the organization's revenue annually (Fresneda & Gefen, 2019). Although consumers rely on online reviews for information about the products, the reviews often sway consumers away from their purchasing intention (Chen, 2016). The accessibility and ease of writing online reviews mean that consumers are often overloaded with information as they try to form a holistic impression of the product (Namvar & Chua, 2022). The higher cognitive effort required to analyze the overpopulated reviews means consumers might not even act on their prior purchasing intentions (Gholizadeh et al., 2021).

To counter this problem of information overload, e-commerce platforms have developed a helpfulness rating for a review (Godes & Silva, 2011). The helpfulness rating of a product is evaluated by finding the ratio of the total helpfulness votes (e.g., thumbs-up) and the number of votes (e.g., sum of thumbs-up and thumbs-down) the review receives (Mudambi & Schuff, 2010). The reviews are then sorted by the e-commerce platforms to present reviews that are voted helpful before showing other reviews. This also helps the overall quality of the reviews: authors of the reviews present more in-depth reviews as they try to get higher helpfulness ratings on their review. However, the helpfulness rating systems come with some drawbacks. First, reviews that are published closer the launch of the product listing might receive a higher helpfulness rating regardless of the information provided (Namvar et al., 2021). Older reviews have a higher interaction rate with the patrons on the online platforms as they have been displayed higher in the review corpus (Zhou & Guo, 2017). A more recent review, despite being more elaborate and providing a more comprehensive evaluation of the product is listed lower in the corpus (Namvar et al., 2022).

Generally, there is a consensus that a more in-depth review is more helpful to the online consumer (Wu et al., 2021, 2021). In the current literature, review depth is usually determined as a function of the number of words (Wu et al., 2021). A longer review reflects a higher cognitive effort from the reviewer as they provide their opinions when interacting with the product. However, the current definition of review depth as a function of the number of words, does not consider other factors of a review that can contribute to the helpfulness rating of an online review.

Subsequently, the aim of this research is to go beyond a simple measure of review length and investigating the depth of online reviews by considering other interacting elements to the relationship between review length and review helpfulness. The extant literature on online review helpfulness elaborates on the review sentiment to discover the presence of a two-sided arguments and its impact on the overall review helpfulness as it provided an unbiased opinion of the products of services. This was particularly preferred by consumers over a review that only presented extreme one-sided arguments (Chen, 2016).

This study, therefore, conceptualizes product review depth as an interaction among review length, review features (readability and two-sided arguments), and product features (product type and price). It then investigates the relationship between review length its and helpfulness. The review sentiment and its rating were used as control variables. To test the conceptual model, we explored the relationships on a data set of 568,454 reviews from 256,059 different reviewers.

This study provides the theoretical contribution by elaborating on the definition and operationalization of review depth. The use of moderating factors will help to understand how the relationship between length and helpfulness is altered. As a managerial implication, the insights from this research can be used by e-commerce platforms to increase consumers' stickiness to their websites (Chen, 2016; Wu et. al., 2021). In an already saturated e-commerce space, insights from this research can be used to filter reviews to increase interactions with potential buyers. It will also ensure that patrons of the platforms act on their purchasing intentions instead of getting overloaded by information from less helpful reviews.

The rest of this paper structured as follows. Next section reviews the literature of online review helpfulness. We elaborate on the different review features and their effects on the review helpfulness. Subsequently, we discuss the hypothesis development and our conceptual model. Results of the empirical findings are presented next. Finally, we discuss the results and outline the theoretical and practical contributions.

2. Literature review

This section outlines the elements of an online review that contribute to review helpfulness in an online shopping scenario. Additionally, it will describe the quantitative methods used by well-known e-commerce platforms to tackle the consequences of information overload. It will first explore the current definitions of online review depth n. Then, it elaborates on the use of heuristic-systematic model (HSM) of information processing (Chen & Chaiken, 1999) to derive the pillars of our proposed conceptual model. Finally, we synthesize a holistic method to define the depth of an online review.

Literature in the online e-commerce space draws parallels from the literature in the consumer-decision making theories. These theories elaborate that consumer use online reviews to assist their information search and evaluation stage in the decision-making model (Namvar, 2020). The inputs from the review elements act as cues to inform consumers about the products listed online. These factors are branched into two categories based on HSM. This model posits that information from a review can be categorized as 1) heuristic cues that are easily interpreted by an online consumer 2) systematic cues that are embedded in the message and understanding them requires higher levels of cognitive effort by the consumers. Heuristic cues such as the star rating of the review or expert rating of the reviewers are easily interpreted by the online consumers. More relevant to this research, systematic cues such as the length of the review (often considered synonymous with depth), readability of the message and two-sided arguments. Unlike heuristic cues, which are easily identifiable from the review, systematic cues require some aptitude and effort from the users to be processed.

Mudambi and Schuff (2010) first introduced the depth of an online review as a function of the total number of words. The simple rationale is that a longer review will explore the product themes more extensively compared to a shorter more direct review (Mudambi & Schuff, 2010). The current literature suggest that more in-depth reviews are easily identified and noticed compared to shorter reviews as they take up more screen space. They contain rich information about the desired product. They also establish credibility and expertise of the reviewer due to their cognitive involvement when publishing the review (Namvar & Chua, 2022; Wu et al., 2021). For potential buyers, a longer review will make it easier to fulfil their information search about the product, evaluate the product features, confirm the appropriateness of the use scenario, and compare the various products available in that product category (Choi & Leon, 2020; Chua & Banerjee, 2016). Also, it is likely that a longer review is more balanced (Chua & Banerjee, 2015) . Furthermore, an in-depth online review helps establish credibility and impose the expertise of the reviewer (Lutz & PrÃ, 2019). A longer review is harder to fabricate, readers of the review acknowledge that the author of the review possesses a certain level of credibility, interest in the product category and experience to comment in such detail. Source credibility theory suggests buyers (especially, first-time consumers in a product category) tend to prioritize the opinions of product experts when considering the purchase of a product (Banerjee et al., 2017). Leaving consistent long reviews in a product category grants the author a level of authority and expertise. Often, a regular pattern of such reviews helps them receive “expert” badges from the e-commerce platform. Therefore, review length considered as a measure of depth affect the helpfulness on an online review.

Adding to the discussion on the systematic features, extant literature identifies readability and two-sided argument of review text contribute to the helpfulness rating of the review as well (Wu et al., 2021). A review with easier readability (one that can be easily understood in the first reading attempt) increases the helpfulness rating received by the online review. On the other hand, a message that uses complex language often requires higher cognitive effort from the reader to understand, which means it may not be read in completion. Therefore, it may receive a lower helpfulness rating (Hu et al., 2012).

Finally, the extant research on online review helpfulness elaborates on the presence of two-sided arguments and its effect on the helpfulness of an online review. Long reviews with a two-sided argument are deemed more helpful (M.-Y. Chen, 2016; Crowley & Hoyer, 1994). Consumers prefer to read a review that presents both pros and cons of the product feature (in contrast to a one-sided argument that might suggest a biased review) (M.-Y. Chen, 2016). A review that presents a two-sided argument signals the reviewer’s truthfulness. Acknowledging both the positive and the negative aspects of a product, shows an absence of self-interest and reviewer’s attempt market the product (Chen et. al., 2016). Consumers find the two-sided argument a fair evaluation of the product. This is more helpful than the one-sided argument that often strongly compliment or condemn it. The presence of both perspectives helps the readers to gain an unbiased evaluation of the product features or the use scenario. Additionally, a two-sided argument increases the review’s attractiveness and increases the chances of the review getting a helpfulness rating when compared to a one-sided review. In the e-commerce environment, consumers prefer to align their attitudes after evaluating a two-sided argument that explains both pros and cons of the product rather than a biased argument.

The existing research also shows that product type plays a moderating role when consumers evaluate the depth of an online review (Chua & Banerjee, 2016; Lutz & PrÃ, 2019). Mudambi and Schuff (2010) introduced the idea that consumers rely on different stimuli to consider review helpfulness for experience goods and search good. Search goods can be classified as products where the consumers can attain information on the quality of the product before its purchase (Kim et al., 2019). This may be a laptop or a camera or a mobile phone where it is easier to purchase a product based on use context: for example, a laptop's storage or operating software. Experience goods, on the other hand, are defined as products that require purchase before the quality of the product can be established: for example, an album or a dish at a restaurant. Studies have identified the elements that constitute the helpfulness of a review change based on the type of product (Chua & Banerjee, 2016).

A heuristic cue, such as the price of goods, impact the features considered important to a helpful review (Aghakhani et al., 2020). Consumers spend a lot more time analyzing reviews for more expensive products. Higher-priced products are associated with quality. When looking to purchase more expensive products, a higher cognitive effort is applied on average to confirm the product attributes publicized by the seller of the product. Consumers will prefer more in-depth reviews that provide a complete discussion to confirm the product attributes, understand the sentiment of the reviewer when interacting with the product and compare more reviews before they confirm the purchase (Choi & Leon, 2020). Subsequently, when purchasing a cheaper product, fewer information cues are considered, and consumers may not even compare it with other reviews before confirming the purchase. Indeed, limited search and evaluation tactics are applied when consumers want to purchase lower-priced goods online. Therefore, shorter more objective reviews that give direct cues about the product are deemed more helpful. An in-depth review, in this case, would be a much shorter review that provides a broad evaluation of the product.

The current literature on online review helpfulness considers reviews depth as a product of review length as a singular review feature. Our study highlights the importance of adopting a holistic approach in conceptualizing review depth by considering 1) product type, 2) price, 3) readability and 4) two-sided argument.

3. Research model and hypotheses

Figure 1 shows the conceptual model of this study. It demonstrates the variables that moderate the relationship between review length and review helpfulness. Along with review length, these variables combine to form the concept of review depth in our paper. These variables are components of two broad categories. Firstly, product related features cover price and product types. Secondly, the review related variables that explain readability, and two-sided argument.

In line with research from Mudambi and Schuff (2010), the first relationship explored in this research looks at the effects of review length on helpfulness of a review. A longer review gains more attention as it occupies a larger section of the screen, provides a greater number of informative and contextual cues about the product, and finally might provide a more balanced argument of the product features (Chua & Banerjee, 2015; Ghasemaghaei et al., 2018; Lutz & PrÃ, 2019). Hence, the first hypothesis of this research explores depth as a measure of the length of the review.

H1: Review length positively affects review helpfulness.

We expect readability to play a moderating role on the relationship between review length and review helpfulness. Prior research confirm that an easily comprehensible review is considered a lot more helpful by a consumer (Choi & Leon, 2020; Wang et al., 2018; Wu et al., 2021). It requires less cognitive effort from the consumer to understand the information cues in the review. The readability score indicates the level of cognitive effort required by the readers to understand the message in the first reading. Hence, a lower readability score makes a message easily understood and analyzed by consumers. Therefore, if a review is long but easy to comprehend (requires a low readability score) it will be more helpful than a long review that is written using a complex language (requires a high readability score).

H2: Readability score required for understanding a review negatively moderates the relationship between review length and review helpfulness.

Two-sided arguments moderate the relationship between review length and helpfulness. A two-sided argument highlights both the positive and the negative elements of a product in a review (Chen, 2016; Chen et. al., 2016). Extant research highlights that consumers prefer to be presented with both pros and cons of the product attributes. This allows the consumers to align their attitudes after a complete evaluation of a two-sided and holistic review. Moreover, a two-sided review may suggest no self-interest from the reviewer to sell the product to the consumer, instead a fair evaluation of the actual product (Lutz et al., 2018). We expect that a longer review that considers both the pros and cons of a product, it will be considered more helpful than a long review that presents one side in an extensive manner.

H3: Two-sided argument positively moderates the relationship between review length and helpfulness.

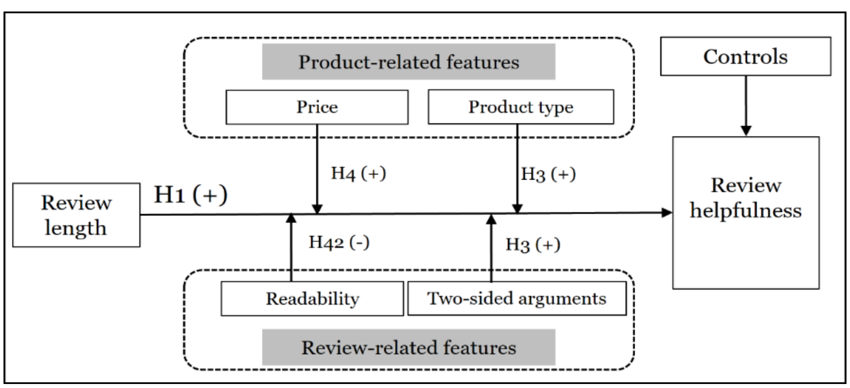

Figure 1. Research conceptual model

The price also plays a moderating role on the relationship between review length and helpfulness. Customers apply higher cognitive effort when shopping for more expensive items online. As stated earlier, the higher price is often associated with quality. To confirm the quality of the product, consumers will require a more in-depth evaluation in the review (Choi & Leon, 2020). Longer time will also be spent exploring the reviews and evaluating an individual review even if it is long. However, a shorter review that highlights the key product attributes can be considered sufficiently helpful for lower-priced items.

H4: Price positively moderates the relationship between review length and review helpfulness.

The type of product plays a pivotal role in the relationship between a review length and its helpfulness rating. Experience goods can be thoroughly evaluated only after they have been purchased and experienced (Brunner et al., 2019; Chua & Banerjee, 2016). The perception of the product quality can also be very subjective and anecdotal when compared to search goods, i.e., those with attributes that can be evaluated prior to purchase or consumption. Hence, we expect that a longer and more in-depth review that explores a detailed account of the reviewer's interaction with an experience good will be considered more helpful than a shorter review with few details. However, for search goods, a short review and outlining product descriptions provides sufficient information on the product attributes and confirm the appropriateness of the use context for the consumer (Wu et. al., 2021). Hence, we hypothesize that lengthier reviews become even more helpful for experience goods, while for search goods, a shorter review can be considered sufficiently helpful.

H5: The relationship between review length and review helpfulness is stronger for experience goods (vs. search goods).

4. Data collection and operationalization

Data was collected for this research via a secondary resource to find the Amazon online reviews. Review data is readily available from Amazon.com or through several repositories of datasets that outline data based on product categories, price, and user information. Amazon reviews are regularly studied and analyzed to study online consumer patterns and behavior online as they are robust and detailed (Mudambi & Schuff, 2010; Namvar & Chua, 2022). The dataset used in this research consists of users with verified purchases. It was collected over 10 years (Oct 1999 - Oct 2012) and consists of a broad range of product categories.

The following explains the methods used to operationalize the variables that were tested from the hypothesis. We discuss how independent variable, and the moderating variables were formed, followed by the target variable.

Review length is the main independent variable in this research. Measuring depth will help to operationalize how dispersed or focused the opinions of the review authors are. The length of the review is measured in by counting the number or words used in the review.

Readability scores: The Gunning fog index is applied to compare the readability scores of the reviews. This index integrates the formal level of education required to understand a given text in English in the first reading. The test has since been applied in various scenarios to confirm the appropriateness of the text for the targeted audience (Wu et. al., 2021). The readability score is given by:

$$Readability\ score = 0.4\left[\left(\frac{words}{sentences}\right) + 100\left(\frac{complex\ words}{total\ words}\right)\right]$$

Two-sided argument is measured using the sentiment analysis at the sentence level for each review. The sentiment is operationalized using the VaderSentiement package in Python. We first split each review by the number of sentences. Then we assigned a sentiment score for each sentence. Then we checked if in a review both negative and positive sentiments are present. Then we assigned values of -1 to reviews that are missing two-sided argument and 1 for reviews that consist of two-sided arguments.

Price of products is listed under the product images for the consumers to compare. Both the cost and the sale price were available in for the dataset attained. Considering the sale price will reflect value consumers will use to compare the price of items in their products in their product consideration.

Product type: Reviews were collected over an extensive range of products. We labeled experienced goods as "1" and search goods as "-1".

Review helpfulness: An online review has two constructs present to measure helpfulness, the number of helpfulness votes and the total number of votes received by the review. We operationalize review helpfulness as the number of helpfulness votes divided by the total number of votes.

Table 1 descriptive statistics of the generated features

Variable	Price	Rating	Length	Label	Readability	Sentiment	Target	Two Sided
Mean	26.61	4.03	110.54	0.51	7.95	0.64	89%	-0.18
Standard Error	0.12	0.01	0.60	0.00	0.01	0.00	9%	0.01
Median	22.50	5.00	77.00	1.00	7.72	0.87	100%	-1.00
Mode	29.99	5.00	32.00	1.00	6.00	0.00	100%	-1.00
SD	23.19	1.48	116.71	0.86	2.63	0.50	0.17	0.98
Sample Variance	537.94	2.18	13621.50	0.74	6.91	0.25	300.63	0.97
Range	639.79	4.00	3425.00	2.00	11.7	2.00	96.55	2.00
Minimum	0.19	1.00	7.00	-1.00	1.32	-1.00	3.45%	-1.00
Maximum	639.98	5.00	3432.00	1.00	17.02	1.00	100.00	1.00

The interactions were formed after finding the product between the dependent variable and the interaction terms. For example, Readability*Length indicated the moderation effect of readability on the relationship between length and readability. Table 1 shows the descriptive statistics of the generated features.

5. Results

Once we operationalized and cleansed the data, we had 37,652 reviews to be considered for the regression model. We conducted correlation to inspect the presence of multicollinearity between independent variables. The Bivariate Person's correlations coefficient helped to assess the magnitude of the relationship between the seven metric variables. The correlations for this dataset (n= 37,652) confirm little to no multicollinearity between the variables that would alter the regression results (see Table 2).

Table 2 Correlation coefficients for variables

Variable	1	2	3	4	5	6	7
1.Price	1.0						
2.Rating	- 0.0	1.0					
3.Length	0.0	- 0.0	1.0				
4.Label	0.00	0.0	-0.0	1.0			
5.Readability	0.0	- 0.0	0.2	0.0	1.0		
6.Sentiment	0.0	0.5	0.1	0.0	0.0	1.0	
7.Two-sided	- 0.0	0.2	- 0.2	0.0	- 0.0	0.1	1.0

We developed three models to confirm the regression model. Model 0 represents the base model with the control variables. Results shows that a neutral rating and the presence of the sentiment were found to be more useful. The significant results from model confirm the effects of rating and sentiment on the helpfulness, these are consistent with prior studies (Mudambi & Schuff, 2010; Namvar & Chua, 2022; Siering et al., 2018).

Model 1 represents the direct effect of review length on the review helpfulness rating, with the prior variables as controls. Model 2 then studies the moderating relationships (see Table 3). Hypothesis 1 predicts that a longer review will relate to a higher helpfulness rating. Our analysis revealed that the result for length is positively significant (coefficient = 0.011, $p < 0.000$). Therefore, hypothesis 1 is supported. Hypothesis 2 predicts that a higher readability will positively affect the relationship between review length and helpfulness rating of a review. Model 2 revealed that the interaction effect of readability is statistically significant (coefficient = - 0.001, $p < 0.000$).

Hypothesis 3 predicts that the presence of a two-sided argument will positively affect the relationship between review length and review helpfulness. Our analysis revealed that the interaction effect of two-sided argument is statistically significant. The presence of a two-sided argument in the review is represented by "1" and the absence of the two-sided argument is represented by "-1". The presence of a two-sided argument has a positive effect on review helpfulness. Hence, the third hypothesis is confirmed.

Table 3 Linear-Regression estimates explaining review helpfulness

	Variable	Model 0	Model 1	Model 2
	Intercept	65.096 *** (0.226)	64.906***(0.241)	63.232***(0.429)
	Rating	5.678*** (0.062)	5.697***(0.062)	5.751***(0.062)
	Sentiment	1.130*** (0.182)	1.049*** (0.186)	1.189***(0.018)
H1	**Length**		0.001*(0.000)	0.010***(0.003)
H2	**Readability*Length**			-0.001***(0.000)
H3	**Two-sided argument*Length**			0.002***(0.000)
H4	**Price*Length**			0.0003(0.000)
H5	**Type*Length**			-0.009(0.0007)
	F stat	6206	4139	1141
	df	2, 37648	3,37647	11,37639
	P value	0.000	0.000	

Note: Robust standard errors reported in parentheses for coefficients.

***p < 0.001, **p < 0.01, and *p < 0.05

We utilize spotlight analysis to depict the interaction effect of readability of a review (Aiken and Stephen, 1991). For this purpose, we used “Interactions” package in RStudio. We plotted three lines for readability, the first one at the mean level of readability, second at one deviation above, and the third at one standard deviation lower than the mean of the readability score. As shown in figure 2, a lower readability has a positive effect on the review helpfulness. Therefore, H2 is confirmed that a lower readability score positively affects the relationship between review depth and helpfulness of an online review.

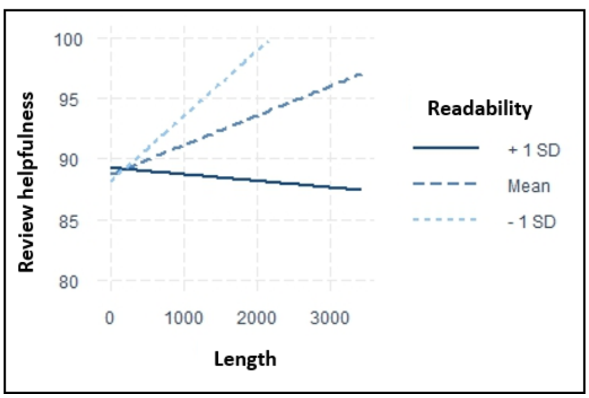

Figure 2 Moderating effect of readability

We also plotted two lines for two-sided argument, the first one showing the presence of two-sided argument (1), and the showing the absence of this argument (-1). As shown in figure 3, the presence of two-sided argument has positive impact on the relationship between review length and its helpfulness.

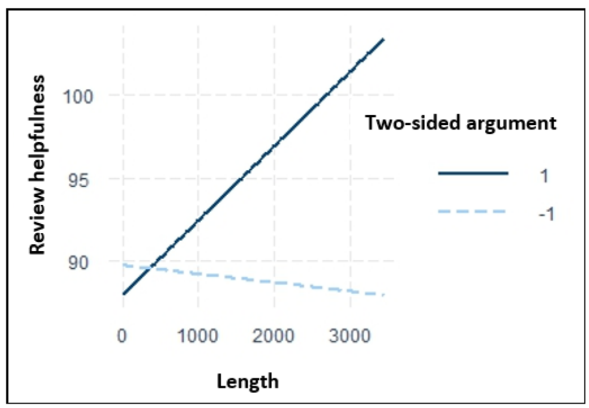

Figure 3 Moderating effect of two-sided argument

Robustness check: To check for the robustness of the model we applied a classic machine learning technique for classification, Artificial Neural Network (ANN). Our descriptive analysis of the dataset revealed that 75% of the data had a helpfulness rating of 80% or higher. We labelled helpfulness rating higher than 80% as “1”, indicating the review is helpful. We labelled them as “0” if reviews had a helpfulness rating lower than 80% indicating the review is not helpfulness. Rating, Sentiment, Length, Length*Readability, and Two-sided argument*Length were used in the feature matrix to classify reviews as helpful or not helpful. We evaluated the model by running a 5-fold cross-validation. The model showed an accuracy of above 80% (n= 37652).

Next, we ran the Tobit regression model. It is widely used in the online review literature to confirm the relationships between various review elements and their helpfulness rating (Mudambi & Schuff, 2010). The Tobit regression model had very similar results to the original regression model (see Table 4). The model significantly explained 24% of the total regression (Log likelihood = -155641.1, p = 0.000). Therefore, both the classification model and the Tobit regression confirmed the robustness of the original regression model for this study. It worth to note that, even though we do not claim that our proposed model provides a comprehensive understanding of the factors that have impact on review helpfulness, Model 2 in Tables 3 and 4 report a high R Squared of 2.440 and 0.238.

6. Discussion and conclusion

Online reviews highlight the prevalence of e-WOM in the e-commerce space. In this online space, people rely on the anecdotal accounts of fellow shoppers to gain information and act on their purchasing intention. Past research informs us that an in-depth review, measured as a function of the total number of words is one of the most critical factors to determine the helpfulness of the review (Baek et al., 2012; Mudambi & Schuff, 2010; Wu et al., 2021). We also found a significant relationship between the length of a review and the review helpfulness. A longer review gives the indication of a higher effort taken by then reviewer to carefully evaluate the product listed online. Readers can gain ample insight about the product in a scenario where physical interaction to judge the appropriateness of the product is not possible.

The aim of this research was to further the knowledge on the factors that are considered essential to determine the helpfulness of an online review. This research theorized variables that have an interaction effect to the relationship between review length and its helpfulness. We confirmed that systematic cues such as readability and the presence of a two-sided argument have an interaction effect on this relationship.

6.1 Theoretical contribution

This study makes two significant theoretical contributions to the online review helpfulness literature. We were able to prove the positive moderating effect of readability and the presence of a two-sided argument on the relationship between review length and helpfulness. Past research (e.g., Wang et al., 2018) have confirmed the effect of the readability score on online review helpfulness. This research further extends the scope of the prior research. We showed a positive effect on the relationship between review length and its helpfulness as the readability score of a review decreased. Next, this study also furthers the research by proving a positive effect of the presence of a two-sided argument on the relationship between review length and its helpfulness.

Table 4 Tobit-Regression estimates explaining review helpfulness

	Variable	Model 0	Model 1	Model 2
	Intercept	65.096 *** (0.226)	64.906***(0.241)	63.232***(0.429)
	Intercept 2	2.716***(0.000)	2.710***(0.003)	2.715***(0.003)
	Rating	5.678*** (0.062)	5.697***(0.062)	5.751***(0.062)
	Sentiment	1.130*** (0.182)	1.049*** (0.185)	1.189***(0.018)
H1	**Length**		0.001*(0.000)	0.010***(0.003)
H2	**Readability*Length**			-0.001***(0.000)
H3	**Two-sided argument*Length**			0.002***(0.000)
H4	**Price*Length**			0.000(0.000)
H5	**Type*Length**			-0.000(0.0007)
	Log-likelihood:	-155694.6	-155691.8	-155641.1
	df	75298	75297	75292
	P value	0.000	0.000	0.000

Note: Robust standard errors reported in parentheses for coefficients.

***p < 0.001, **p < 0.01, and *p < 0.05

6.2 Managerial implications

The findings of this study can be applied to e-commerce platforms to retain consumers and increase the time spent by the consumers on their platforms. Information overload caused by the large number of reviews presented on these platforms may cause consumers sway away from their purchasing intention, as they are overloaded with information and unable to decide. E-commerce platforms can use the insights from this research to form a better way to arrange the reviews. This can be done by comparing the readability score and checking for the presence of a two-sided argument in reviews. The empirical approach highlighted in this study can be used to measure the in-depth reviews based on the review's readability score and the presence of two-sided arguments.

6.3 Limitations and future research

In this study, we analyzed online reviews that were listed helpfulness on Amazon to analyze the regression model tested in this study. While Amazon reviews provide easily available information on online reviews, only analyzing reviews that were voted helpful will not cover the perspectives of the patrons of these e-commerce platforms that did not vote a review helpful. While it is important to consider the reviews, all related online patrons of Amazon, it is not possible to consider the perspectives of patrons that did not participate in the voting of the helpfulness review. Furthermore, setting a lower limit of three votes of helpfulness rating provided a control for producing more accurate statistical relationships. However, this also neglected the perspectives of the patrons that might not vote due to the factors in this review. A more traditional survey collection method might be more insightful to consider the motivations of online shoppers who did not vote on the helpfulness of the review after reading it (Siering et. al., 2018). This will help to lower the selection bias that might develop as the perspectives of online patrons that did not vote on the helpfulness of the online reviews are also considered.

Retailers on Amazon can sell their products to a worldwide audience, therefore relying on more culture specific methods to operationalize these constructs can be explored in future research. With the rise of smartphone users and increased accessibility to internet in developing countries there is a large influx in the number of users on these e-commerce websites (Mayes et al., 2016). Understanding culture specific nuisances like slag used will help to expand the understanding of the factors considered more important to determine the helpfulness of an online review in these markets.

Future studies could also look to explore the moderating relationships of time-related and reviewer related constructs on the relationship between the depth and the helpfulness of a review. Comparing these relationships might offer new insights on the time-related factors such a review order and reviewer related factors such a reviewer expertise to offer moderating effect on the relationships between review depth and helpfulness. It will further enrich the comparison of in-depth reviews in the whole review corpus. Additionally, future studies can extend on the robustness of the relationships explored in this study, by applying it in varying context. Extending the analysis to data from websites such IMDB, clothing retailers and hotel and flight booking platforms might offer new insights to these single product related platforms.

7. References

Aghakhani, N., Oh, O., Gregg, D. G., & Karimi, J. (2020). Online Review Consistency Matters: An Elaboration Likelihood Model Perspective. *Information Systems Frontiers*.

Baek, H., Ahn, J., & Choi, Y. (2012). Helpfulness of online consumer reviews: Readers' objectives and review cues. *International Journal of Electronic Commerce*, *17*(2), 99–126.

Banerjee, S., Bhattacharyya, S., & Bose, I. (2017). Whose online reviews to trust? Understanding reviewer trustworthiness and its impact on business. *Decision Support Systems*, *96*, 17–26.

Banerjee, S., & Chua, A. Y. K. (2021). Calling out fake online reviews through robust epistemic belief. *Information & Management*, *58*(3), 103445.

Brunner, C. B., Ullrich, S., & De Oliveira, M. J. (2019). The most optimal way to deal with negative consumer review: Can positive brand and customer responses rebuild product purchase intentions? *Internet Research*, *29*(1), 104–122.

Chen, M.-Y. (2016). Can two-sided messages increase the helpfulness of online reviews? *Online Information Review*, *40*(3), 316–332.

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In *Dual-process theories in social psychology*. The Guilford Press.

Choi, H. S., & Leon, S. (2020). An empirical investigation of online review helpfulness: A big data perspective. *Decision Support Systems*, *139*, 113403.

Chua, A. Y. K., & Banerjee, S. (2015). Understanding review helpfulness as a function of reviewer

reputation, review rating, and review depth. *Journal of the Association for Information Science and Technology*, *66*(2), 354–362.
Chua, A. Y. K., & Banerjee, S. (2016). Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality. *Computers in Human Behavior*, *54*, 547–554.
Crowley, A. E., & Hoyer, W. D. (1994). An Integrative Framework for Understanding Two-sided Persuasion. *Journal of Consumer Research*, *20*(4), 561–574.
Fresneda, J. E., & Gefen, D. (2019). A semantic measure of online review helpfulness and the importance of message entropy. *Decision Support Systems*, *125*, 113117.
Ghasemaghaei, M., Eslami, S. P., Deal, K., & Hassanein, K. (2018). Reviews' length and sentiment as correlates of online reviews' ratings. *Internet Research*, *28*(3), 544–563.
Gholizadeh, M., Akhlaghpour, S., Isaias, P., & Namvar, M. (2021). Emergent affordances and potential challenges of mobile learning apps: Insights from online reviews. *Information Technology & People*, 0959–3845. https://doi.org/10.1108/ITP-05-2021-0412
Godes, D., & Silva, J. C. (2011). Sequential and Temporal Dynamics of Online Opinion. *Marketing Science*, *31*(3), 448–473.
Hu, N., Bose, I., Koh, N. S., & Liu, L. (2012). Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision Support Systems*, *52*(3), 674–684.
Kim, S. J., Maslowska, E., & Tamaddoni, A. (2019). The paradox of (dis)trust in sponsorship disclosure: The characteristics and effects of sponsored online consumer reviews. *Decision Support Systems*, *116*, 114–124.
Lutz, B., & PrÃ, N. (2019, December 13). The Longer the Better? The Interplay Between Review Length and Line of Argumentation in Online Consumer Reviews. *ICIS2019*. International Conference on Information Systems, Munich.
Lutz, B., PrÃllochs, N., & Neumann, D. (2018, December 13). *Understanding the Role of Two-Sided Argumentation in Online Consumer Reviews: A Language-Based Perspective*. International Conference on Information Systems, San Francisco.
Mayes, J., White, A., Byrne, M., & Mogg, J. (2016). How smartphone technology is changing healthcare in developing countries. *The Columbia University Journal of Global Health*, *6*(2), 36–38.
Mudambi, S. M., & Schuff, D. (2010). Research note: What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quarterly*, *34*(1), 185–200.
Namvar, M. (2020, January 7). *A Novel Approach to Predict the Helpfulness of Online Reviews*. 53rd Hawaii International Conference on System Sciences (HICSS), Hawaii.
Namvar, M., Akhlaghpour, S., Pool, J., & Priscilia, A. (2022, January 8). *Towards explaining user satisfaction with contact tracing mobile applications in a time of pandemic: A text analytics approach*. 55th Hawaii International Conference on System Sciences (HICSS), Hawaii.
Namvar, M., Boyce, J., Sarna, J., Zheng, Y., Chua Yeow Kuan, A., & Ameli, S. (2021, January 5). *Moderating Effects of Time-Related Factors in Predicting the Helpfulness of Online Reviews: A Deep Learning Approach*. 54th Hawaii International Conference on System Sciences (HICSS), Hawaii.
Namvar, M., & Chua, A. (2022). The impact of context clues on online review helpfulness. *Internet Research*, *ahead-of-print*. https://doi.org/10.1108/INTR-02-2021-0093
Siering, M., Muntermann, J., & Rajagopalan, B. (2018). Explaining and predicting online review helpfulness: The role of content and reviewer-related signals. *Decision Support Systems*, *108*, 1–12.
Wang, Y., Kung, L., Wang, W. Y. C., & Cegielski, C. G. (2018). An integrated big data analytics-enabled transformation model: Application to health care. *Information & Management*, *55*(1), 64–79.
Wu, C., Mai, F., & Li, X. (2021). The effect of content depth and deviation on online review helpfulness: Evidence from double-hurdle model. *Information & Management*, *58*(2), 103408.
Zhou, S., & Guo, B. (2017). The order effect on online review helpfulness: A social influence perspective. *Decision Support Systems*, *93*, 77–87.