A Meta-Analysis on the Determinants of Online Review Helpfulness

Hong Hong
School of Management, Xiamen University, Xiamen, 361005, China, dxu@xmu.edu.cn

Di Xu
School of Management, Xiamen University, Xiamen, 361005, China, dxu@xmu.edu.cn

Dapeng Xu
School of Management, Xiamen University, Xiamen, 361005, China, dxu@xmu.edu.cn

G. Alan Wang
Pamplin College of Business, Virginia Tech, Blacksburg, 24061, USA

Weiguuo Fan
Pamplin College of Business, Virginia Tech, Blacksburg, 24061, USA

Follow this and additional works at: http://aisel.aisnet.org/whiceb2017

Recommended Citation
Hong, Hong; Xu, Di; Xu, Dapeng; Wang, G. Alan; and Fan, Weiguuo, "A Meta-Analysis on the Determinants of Online Review Helpfulness" (2017). WHICEB 2017 Proceedings. 34.
http://aisel.aisnet.org/whiceb2017/34

This material is brought to you by the Wuhan International Conference on e-Business at AIS Electronic Library (AISeL). It has been accepted for inclusion in WHICEB 2017 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
A Meta-Analysis on the Determinants of Online Review Helpfulness

Hong Hong\(^1\), Di Xu\(^*\), Dapeng Xu\(^1\), G. Alan Wang\(^2\), Weiguo Fan\(^2\)

\(^1\)School of Management, Xiamen University, Xiamen, 361005, China
\(^2\)Pamplin College of Business, Virginia Tech, Blacksburg, 24061, USA

Abstract: Online consumer reviews can help customers decrease uncertainty and risk faced in online shopping. However, information overload and conflicting comments in online reviews can get consumers confused. Therefore, it is important for both researchers and practitioners to understand the characteristics of helpful reviews. But studies examining the determinants of perceived review helpfulness produce mixed findings. We review extant research about the determinant factors of perceived helpfulness. Conflicting findings exist for six review related factors, namely review extremity, review readability, review total votes, linear review rating, quadratic review rating, and review sentiment. We conduct a meta-analysis to reconcile the contradictory findings on the influence of review related factors over perceived review helpfulness. The meta-analysis results confirm that review extremity, readability, total votes, and positive sentiment have a negative influence on helpfulness, but review rating is positively related to helpfulness. We also examine those studies whose findings are contradictory with the meta-analysis results. Measure discrepancy and reviewed product type are the two main reasons why mixed findings exist in extant research.

Keywords: online reviews, helpfulness, meta-analysis, readability, sentiment

1. INTRODUCTION

Nowadays, online shopping plays an important role in our daily lives because of its low cost and convenience. Compared to traditional shopping, online shopping is unique in its temporal and spatial separation of buyers and sellers \(^1\). Given the unique characteristic of online shopping, it is impossible for consumers to experience products or services before buying. Hence, consumers face more uncertainty and risk while shopping online. However, user-generated content such as online customer reviews can help consumers decrease the uncertainty and risk. Online customer reviews are defined as peer-generated evaluations about products or services \(^2\). Typically, an online review includes a star rating and written comments about the experience of using a product or service and critique about product features \(^2\). It is no doubt that online reviews are helpful to potential online shoppers, but information overload and conflicting comments in reviews can also get consumers confused. Therefore, it is important for both researchers and practitioners to understand the characteristics of helpful reviews \(^3\).

Both practitioners and researchers have examined ways to identify helpful reviews. Many websites, such as Amazon and Yahoo! Movie, provide a helpfulness feedback mechanism for online reviews. The mechanism has been found effective in promoting sales. However, this indicator needs long time accumulation, and it cannot provide usefulness information about latest reviews. Lu et al. (2010) find that a large proportion of reviews obtain few or no helpfulness feedback, particularly the more recent ones \(^4\). In order to help sellers use online reviews to promote products and consumers improve decision efficiency, a great deal of research has been carried out to investigate the helpfulness of online reviews, but there is no consensus on the determinants of review helpfulness\(^5\)\(^6\). The mixed findings on the determinants of helpfulness create confusion to both researchers and practitioners.

It is a common problem to have mixed research findings in social and behavioral sciences \(^7\). Meta-analysis is an appropriate research methodology to solve this problem \(^8\). To the best of our knowledge, no meta-analysis
has been conducted to study the complex relationships between online review characteristics and review helpfulness. We therefore attempt to fill this gap. Aggregating existing literature allows us to validate their findings and clarify the inconsistency amongst existing studies on review helpfulness[8]. In this study, we first review extant research about the determinants of online review helpfulness and identify those determinants with mixed findings. We then conduct a meta-analysis to reconcile the contradictory findings on review extremity, readability, total votes, rating, and sentiment.

In next Section, we extensively review existing studies related to perceived review helpfulness. In Section 3 we present our research methodology and data collection process. In Section 4 we report the meta-analysis results and discuss the reasons why mixed findings exist on the relationships between review helpfulness and its determinants. In the final section, we conclude our paper by discussing the contributions, limitations, and future directions for this study.

## 2. LITERATURE REVIEW

<table>
<thead>
<tr>
<th>Category</th>
<th>Factors</th>
<th>Definitions or other names</th>
<th>Relationships</th>
<th>Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review related factors</td>
<td>Review length</td>
<td>Review words number; word count; review depth; review elaborateness.</td>
<td>Positive</td>
<td>[9], [10], [11], [5], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [6], [26], [2], [27], [28], [29], [30], [31]</td>
</tr>
<tr>
<td></td>
<td>Review extremity</td>
<td>Difference between a review rating and the average rating; review deviation; rating difference; rating inconsistency.</td>
<td>Positive</td>
<td>[12], [32], [30], [6]</td>
</tr>
<tr>
<td></td>
<td>Review readability</td>
<td>Ease of understanding of reviews, use Gunning’s fog index, Automated readability index and the Coleman-Liau index to measure, the lower the grade, the more readable the text.</td>
<td>Positive</td>
<td>[24], [33], [15], [21]</td>
</tr>
<tr>
<td></td>
<td>Review total votes</td>
<td>Total number of votes received for a review.</td>
<td>Positive</td>
<td>[9], [19], [23], [33], [34]</td>
</tr>
<tr>
<td></td>
<td>Linear review rating</td>
<td>Review rating usually ranges from one star to five stars.</td>
<td>Positive</td>
<td>[14], [11], [20], [23], [24], [26], [37], [38], [28], [36], [34]</td>
</tr>
<tr>
<td></td>
<td>Quadratic review rating</td>
<td>Review rating*Review rating; quadratic term of review rating.</td>
<td>Positive</td>
<td>[5], [13], [18], [22], [27], [29], [39], [31]</td>
</tr>
<tr>
<td></td>
<td>Review positive sentiment</td>
<td>Review positive valence; positive degree of the review.</td>
<td>Positive</td>
<td>[17], [19], [3], [11]</td>
</tr>
<tr>
<td></td>
<td>Review age</td>
<td>Review timeliness; review elapsed days; days elapsed after the review being posted, the post day minus the first review post day or the product release day.</td>
<td>Positive</td>
<td>[13], [16], [22], [27], [6], [28], [36]</td>
</tr>
<tr>
<td></td>
<td>Total review number</td>
<td>Total number of reviews for the product.</td>
<td>Negative</td>
<td>[19], [22], [38], [28]</td>
</tr>
<tr>
<td>Reviewer related factors</td>
<td>Information disclosure</td>
<td>Disclosure of self-information, e.g., real name, self-photo, location, reviewer identity.</td>
<td>Positive</td>
<td>[14], [15], [21], [23], [26], [38], [33]</td>
</tr>
<tr>
<td></td>
<td>Reviewer experience</td>
<td>Number of reviews on the platform written by the reviewer.</td>
<td>Positive</td>
<td>[13], [15], [6]</td>
</tr>
<tr>
<td></td>
<td>Reviewer expert label</td>
<td>Dummy variable of whether the reviewer has expert/elite badge, rank 10,000 label; credibility.</td>
<td>Positive</td>
<td>[17], [19], [22], [36], [12]</td>
</tr>
<tr>
<td></td>
<td>Reviewer friend number</td>
<td>Reviewers’ friends; reviewer out-degree centrality.</td>
<td>Positive</td>
<td>[14], [13], [16], [22], [6]</td>
</tr>
</tbody>
</table>

Table 1 summarizes the determinants of review helpfulness identified from existing 35 papers. Factors related to review helpfulness can be divided into two categories: (1) Review related factors that are related to
review ratings or contents. (2) Reviewer related factors that are derived from review authors' background and self-described labels. As Table 1 shows, the findings about the influence of reviewer related factors, namely information disclosure, experience, self-described expert label, and friend number, over perceived helpfulness are consistent across different studies. They are all positively related to helpfulness. Only 3 out of 9 review related factors, namely review length (positive), age (positive), and total review number (negative), have consistent findings over their influence on review helpfulness. The other 6 review related factors, namely review extremity, readability, total votes, linear and quadratic review ratings, positive sentiment, are found to have mixed findings over their influence on perceived review helpfulness. Therefore, the literature review suggests that a meta-analysis is necessary to understand and reconcile the contradictory findings on those review related factors.

3. RESEARCH METHODOLOGY

Originated from Fisher’s “combining P value”, meta-analysis was developed to “combining statistics” by Glass (1976) [40]. It is a popular method to combine and analyze the quantitative results of empirical results [40] and can offer directions for future studies [41]. Meta-analysis was used in medical and psychological fields initially. King and He (2005) discuss the application of meta-analysis in the field of information systems and consider it as a formal and systematic literature review method [8].

3.1 Study selection and coding

In order to avoid publication bias, we used multi-channel literature search. For English studies, we searched literature from commonly used digital databases such as ScienceDirect, EBSCO, SAGE, and Taylor & Francis. In addition, we manually searched related papers from four prestigious information systems journals where research related to perceived review helpfulness are most likely to be published, including Decision Support Systems, Information Systems Research, Journal of Management Information Systems, and MIS Quarterly. Online consumer reviews have also been extensively studied in the field of marketing because of its impact on product sales. Therefore, we also search papers from three prestigious marketing journals, namely Journal of Consumer Research, Journal of Marketing, and Journal of Marketing Research. additionally, we also downloaded working papers from the Social Science Research Network database. For Chinese literature, we searched papers from the Chinese National Knowledge Infrastructure database, which is the most popular literature database in China.

We analyzed 35 papers on online review helpfulness in literature review section and found conflicting influences of review extremity, readability, total votes, linear review rating, quadratic review rating, and positive sentiment on online review helpfulness. We will conduct a meta-analysis to find out the relationships between these online review antecedent factors and helpfulness. Kirca et al. (2005) argue that meta-analysis could be conducted with at least three studies [42], hence we could not do meta-analysis on quadratic review rating after deleting those studies with correlation coefficients greater than the critical value of 1. At last, 31 studies are included in our meta-analysis conducted on review extremity, review readability, review total votes, review rating, and review positive sentiment.

3.2 Statistical analysis

We used the Comprehensive Meta-analysis (CMA) 2.0 software to conduct our analysis. CMA 2.0 generates either a fixed-effect model or a random-effect model. Based on the result of Q-statistics which rejects the homogeneity assumption across studies [43], we adopt the fixed-effect model for our analysis. To conduct the meta-analysis, we extracted effect sizes from extant research first. In this study, we adopted the correlation coefficient $r$ and sample size as the effect size. There are three main steps to do the meta-analysis.

Step 1: Calculate the Fisher's $Z$ and combined effect size (i.e., the combination of correlation coefficients).
Fisher’s Z can be calculated using Equation 1 \([44]\).  
\[
Fisher’s \ Z_i = 0.5 \log \frac{1 + r_i}{1 - r_i} 
\]
Where \( r_i \) is the correlation coefficient extracted from study \( i \).

The weighted-average Fisher’s Z was calculated using Equation 2.
\[
\text{Fisher’s } Z = \frac{\sum_{i=1}^{n} w_i \times Fisher’s \ Z_i}{\sum_{i=1}^{n} w_i} 
\]
Where \( w_i \) is the weight of study \( i \), which equals to the ratio of sample size of study \( i \) to the overall sample size of all the studies considered in the meta-analysis.

The weighted-average Fisher’s Z was converted to a combined effect size \( \tilde{r} \) using Equation 3.
\[
\tilde{r} = (e^{2 \text{Fisher’s } Z_i} - 1)(e^{2 \text{Fisher’s } Z_i} + 1) 
\]

Step 2: Test the significance of the combined effect size. CMA 2.0 reports the \( P \)-value and confidence interval of the combined effect sizes in order to test their significance.

Step 3: Test the validity of the meta-analysis results. A fail-safe number is used to deal with the concern of publication bias. Rosenthal (1991) suggests that the critical value of fail-safe number is five times as large as the number of studies and then plus 10 \([45]\).

4. META-ANALYSIS RESULTS

4.1 Calculation of effect sizes

The meta-analysis results are summarized in Table 2. As the \( P \)-values indicate, the influences of five antecedent factors on perceived review helpfulness are all statistically significant. All fail-safe numbers exceed their corresponding critical values, indicating a high level of validity in our meta-analysis results.

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>Study number</th>
<th>Sample size</th>
<th>Combined effect size</th>
<th>( P )-value</th>
<th>( Q )-value</th>
<th>Confidence interval (CI)</th>
<th>Fail-safe number</th>
<th>Fail-safe number critical value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review extremity</td>
<td>12</td>
<td>304546</td>
<td>-0.247</td>
<td>0.000*</td>
<td>61074.704</td>
<td>[-0.251, -0.244]</td>
<td>16610</td>
<td>70*</td>
</tr>
<tr>
<td>Review readability</td>
<td>11</td>
<td>196197</td>
<td>-0.014</td>
<td>0.000*</td>
<td>303.462</td>
<td>[-0.018, -0.009]</td>
<td>83</td>
<td>65*</td>
</tr>
<tr>
<td>Review total votes</td>
<td>9</td>
<td>196634</td>
<td>-0.029</td>
<td>0.000*</td>
<td>654.031</td>
<td>[-0.034, -0.025]</td>
<td>342</td>
<td>55*</td>
</tr>
<tr>
<td>Linear review rating</td>
<td>16</td>
<td>467754</td>
<td>0.003</td>
<td>0.000*</td>
<td>51022.173</td>
<td>(0.000, 0.006)</td>
<td>1324</td>
<td>90*</td>
</tr>
<tr>
<td>Review positive sentiment</td>
<td>3</td>
<td>34886</td>
<td>-0.162</td>
<td>0.000*</td>
<td>512.811</td>
<td>[-0.173, -0.152]</td>
<td>203</td>
<td>25*</td>
</tr>
</tbody>
</table>

Twelve extant studies reported the correlation coefficient of review extremity. Seven of them were reported to have a negative impact on review helpfulness while five were reported to have a positive influence. The meta-analysis result of review extremity showed a negative impact on review helpfulness, confirming that moderate reviews are perceived to be more helpful than extreme ones.

Eleven studies reported the coefficient of review readability. Five of them showed a positive impact on perceived review helpfulness while others showed a negative impact. The meta-analysis result confirmed that review readability grade negatively influences review helpfulness. The result implies that the more readable a review is, the more likely the customers will perceive it as being helpful. We extracted the correlation coefficient of review total votes from nine extant studies. Six of them showed a positive influence over perceived helpfulness while the other three showed a negative influence. The meta-analysis result confirmed the negative impact of review total votes on perceived helpfulness. This result is reasonable as many studies use ratio of helpful vote number and total vote number to measure perceived helpfulness.

Sixteen studies reported mixed results on the relation between the linear review rating and perceived...
helpfulness. Five of them showed a negative impact on helpfulness while others showed a positive impact. The meta-analysis result confirmed the positive influence of the linear review rating on perceived review helpfulness. We obtained the coefficient of review positive sentiment on review helpfulness from three prior studies. The result confirmed a negative influence of review positive sentiment on helpfulness, indicating that consumers are more likely to perceive negative reviews as being helpful.

4.2 Further analysis of the results

In this section, we compare the findings confirmed by the meta-analysis to those in extant. Our goal is to provide possible explanations for the mixed findings in those studies. It is our hope that the discussion could provide useful insights for researchers and practitioners when they interpret the findings in research related to perceived review helpfulness. Table 3 lists the studies that are consistent and inconsistent with the meta-analysis results, respectively. Possible explanations are also provided.

<table>
<thead>
<tr>
<th>Consistent or not</th>
<th>IVs</th>
<th>Review extremity</th>
<th>Review readability</th>
<th>Review total votes</th>
<th>Linear review rating</th>
<th>Review positive sentiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confirmed influence on perceived review helpfulness</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Positive</td>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>Consistent studies</td>
<td>[32], [9], [5], [17], [19]</td>
<td>[32], [12], [16], [22]</td>
<td>[35], [16], [13]</td>
<td>[14], [11], [22], [24], [26], [37], [27], [38], [28]</td>
<td>[17]</td>
<td></td>
</tr>
<tr>
<td>Inconsistent studies</td>
<td>[32], [30], [6], [12]</td>
<td>[33], [14], [15]</td>
<td>[9], [19], [23], [34], [33]</td>
<td>[5], [13], [18], [27], [39], [31]</td>
<td>[3], [9]</td>
<td></td>
</tr>
<tr>
<td>Possible reasons</td>
<td>Different measurements for review helpfulness and extremity.</td>
<td>Different measurements for readability, different online review contexts.</td>
<td>Different online review contexts</td>
<td>Different product types</td>
<td>Different measurements for positive sentiment.</td>
<td></td>
</tr>
</tbody>
</table>

Discrepancy in the measures of review helpfulness and review extremity are the possible reasons why some studies found review extremity to be positively related to review helpfulness. For example, Fang et al. (2016) used total helpful votes to measure perceived review helpfulness [32], which is different from other studies using the ratio of helpful vote number and total vote number. User helpfulness feedback can be very sparse in user-generated content, so it is possible that some helpful reviews never receive any helpful vote due to lack of user motivation. Similarly, Yin et al. (2014) and Yin (2012) measured review extremity separately for positive ratings or negative ratings, which is also different from other studies where extremity is calculated for all ratings [6][12].

Conflicting findings for review readability can be found in those studies where discrepancy in the readability measure or product type exists. In those studies with findings consistent with the meta-analysis result, the measurement for readability is Gunning’s fog index [24][32] or the Coleman-Liau index [12][16][22]. Studies using the Automated Readability ease Index [14][15] reported findings inconsistent with the meta-analysis result. As for the product type, those studies focusing on experience good reviews reported the same relationship as the meta-analysis result, while those examining search goods reported different influence.

Online review context may contribute to the mixed findings about the influence of review total votes on perceived helpfulness. Those studies consistent with the meta-analysis result examined the reviews made on experience goods (e.g., hotel) from TripAdvisor or Yelp [16][35], while online reviews used by other studies are product reviews mainly from Amazon. Different review context may induce different results.

Product type may be the reason behind the inconsistent findings on review rating. Those studies examining reviews on search goods have findings consistent with the meta-analysis result. Those studying experience goods have findings inconsistent with the meta-analysis result.

The sentiment measure may cause the mixed findings on its influence on perceived review helpfulness.
Kuan et al. (2015) use the ratio of positive words in a review to measure positive sentiment, which achieved findings consistent with the meta-analysis results [17]. Other studies use different sentiment measures.

5. CONCLUSIONS AND FUTURE WORK

We reviewed extant research about the determinants of perceived online review helpfulness. Two types of factors were found to have influence on perceived helpfulness, reviewer related and review related factors. While reviewer related factors have consistent findings on their influence over helpfulness in extant research, conflicting findings exist for six review related factors, namely review extremity, readability, total votes, linear review rating, quadratic review rating, and review sentiment. We conducted a meta-analysis to reconcile the contradictory findings on the review related factors. The meta-analysis results confirmed that review extremity, readability, total votes, and positive sentiment have a negative influence on perceived helpfulness. Review rating was found to be positively related to helpfulness. We also examined those studies whose findings were contradictory with the meta-analysis results. Measure discrepancy and reviewed product type are two main reasons why mixed findings exist in extant research.

This study has both theoretical and practical implications. From a theoretical perspective, it enriches the study on online review helpfulness. Based on the mixed findings in regard to the determinants of review helpfulness and how they influence helpfulness, our study integrates existing research, reconciles their findings, and explores the reasons behind the inconsistencies in extant studies. From a practical perspective, our findings help both sellers and buyers better identify helpful reviews among an enormous amount of reviews and thus improve their decision efficiency.

However, our work still has several limitations. First, although the results of fail-safe number indicate that our analysis results are valid, this study is still not able to include all previous studies on online review helpfulness. Second, we just examined direct relations between the determinants and perceived helpfulness, while some research suggests moderating effects of product type [2][19], review type [35], and product price [16][19]. We will consider more complex models in our future research. Third, the weakness of meta-analysis method, i.e. losing contextual information cannot be completely avoided. The meta-analysis result cannot reveal all the differences of the research contexts in the studies considered. Therefore, more detailed analysis is needed to explain the causes of mixed findings in extant research in the next step.

ACKNOWLEDGEMENT

This research was supported by China Scholarship Council (Grant# 201506310121) and the Natural Science Foundation of China (Grant# 71531013, 71572122, and 71671153).

REFERENCES

Empirical Investigation of Confirmation Bias in Online Word of Mouth. Information Systems Research, 27(1): 131-144.


