



What makes a helpful online review? A meta-analysis of review characteristics

Yani Wang¹ · Jun Wang¹ · Tang Yao^{1,2}

Published online: 28 June 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

In this study, we aim to clarify the determinants of online review helpfulness concerning review depth, extremity and timeliness. Based on a meta-analysis, we examine the effects of important characteristics of reviews employing 53 empirical studies yielding 191 effect sizes. Findings reveal that review depth has a greater impact on helpfulness than review extremity and timeliness with the exception of its sub-metric of review volume, which exerts the negative influence on review helpfulness. Specifically, readability is the most important factor in evaluating review helpfulness. Furthermore, we discuss important moderators of the relationships and find interesting insights regarding website and culture background. In accordance with the results, we propose several implications for researchers and E-business firms. Our study provides a much needed quantitative synthesis of this burgeoning stream of research.

Keywords Customer reviews · Online review helpfulness · Review depth · Review extremity · Timeliness · Meta-analysis

✉ Jun Wang
king.wang@buaa.edu.cn

Yani Wang
yn_wang@buaa.edu.cn

Tang Yao
yaot@buaa.edu.cn

¹ School of Economics and Management, Beihang University, Xueyuan Road, 37, Haidian District, Beijing 100191, People's Republic of China

² Key Laboratory of Complex System Analysis, Management and Decision, Beihang University, Ministry of Education, Beijing 100191, People's Republic of China

1 Introduction

Customer reviews are increasingly available online with the development of electronic commerce. As consumers search online for product information and evaluate product alternatives, they often have access to hundreds of product reviews from other consumers. Reviews work as a form of “sales assistance” that inform prospective consumers and have the potential to add value for them [1]. Among the many characteristics associated with online customer reviews, review helpfulness is the most important [2]. “Helpfulness” has been commonly used as the primary way of measuring how consumers evaluate a review and consumers can sort reviews by their level of helpfulness. The wide adoption and perceived helpfulness of online user reviews on consumers’ decision making has energized academic research on the assessment of what constitutes a helpful review. For instance, Mudambi and Schuff [3] examined the role of word count and rating on the perceived helpfulness of reviews. Ghose and Ipeirotis [4] studied the role of review volume on helpfulness. Recently, Yin et al. [5] studied the relationship between review systematic cues (e.g. reviews length and readability) and helpfulness.

While there are a large amount of studies expatiating what makes a review helpful, findings regarding determinants of review helpfulness have not get unanimous conclusion. The mixed findings on the determinants of review helpfulness lead to confusions to both academic researchers and e-commerce practitioners. Accordingly, we perform a meta-analysis to clarify factors related to review characteristics contributing to review helpfulness and to what extent their influence is. The Meta-analytic method provides a quantitative integration of the main constructs that are associated with a certain topic and a summary of the relationships after adjusting the measurement and sampling error based on the previous studies. It enables researchers to (1) obtain empirical generalizations about the specific effect sizes across varying substantive and methodological conditions and (2) examine whether and how these conditions affect the focal effect size [6]. First insights along these lines come from the recent work of Floyd et al. [7], You et al. [8] and Rosario [9]. Floyd et al. [7] examined 26 studies in order to explain how review valence and volume influence the elasticity of retailer sales. You et al. [8] extended the study to 51 papers and the moderating effect of product characteristics, industry characteristics, and platform characteristics. Purnawirawan et al. [10] clarified the influence of online review valence on various dependent variables and Rosario et al. [9] investigated the influence of platform characteristics, product characteristics, and online WOM metrics on the relationship between online WOM and sales by 96 studies. Although the literature has probed online review and its impact on sales, seldom meta-analysis has been utilized to explain how online customers perceive the helpfulness of online consumer reviews. In this study, our meta-analysis aims to make contributions to the understanding of what kind of reviews can be classified as helpful ones and can create value for e-commerce firms.

We identify important metrics that can represent different dimensions of reviews when evaluating a review and give a systematic review of their

performance. While helpfulness works as the primary way of measuring how consumers evaluate a review [3], customer reviews serve as informants in the information seeking process [10]. Review extremity helps consumers learn about a product at the first glance and gives consumers an overall product assessment. However, it may not convey huge amount of information to potential buyers who try to understand the importance of aspects of the product [3]. When consumers are willing to read and compare open-ended comments from peers, review depth can matter. Meanwhile, timeliness is a factor that could not be ignored when consumers consider a review. Timeliness of the review content could influence information diagnosticity in that recent information differs from the older one on its influence [11]. Consequently, when determining factors accounting for the helpfulness of a review, consumers usually take review extremity, review depth and timeliness into account which can capture the different aspects of review characteristics.

Our findings contribute to the literature as follows. First, we provide an integrative review of the empirical literature and develop a conceptual understanding of components of helpfulness. To do so, we compile a unique data set by collecting variables that reflects the relationship between review information and helpfulness. Second, by calculating mean effect sizes, we show whether and how these metrics affect helpfulness, thus resolving the inconsistency in findings across primary research studies. The meta-analytic results indicate a pattern of weak and strong effects of review helpfulness that can guide future research and practice. Specifically, we show that one sub-metric often overlooked—readability (ease of comprehension by consumers about a certain product)—can increase helpfulness on condition that consumers provide comments that is easier to understand. Third, by means of moderator analysis, we deliberate website and culture background as two key moderators that can explain the mixed findings of previous studies. In particular, we find that fewer reviews on Amazon.com will be conducive for its review helpfulness and American consumers prefer earlier ones when scanning the product reviews. We give a holistic explanation on what constitutes a helpful review that provides critical insights for scholars, managers, and website owners.

The study is structured as follows: First, we present literature review of previous studies and a conceptual framework and hypotheses are following to guide the meta-analysis. Second, we describe the method, including literature retrieval process, coding and computation of effect sizes. Third, we present a quantitative summary of the adjusted mean effect sizes for the pair-wise relationships between review helpfulness and its determinants. Fourth, we propose hierarchical linear model using the maximum likelihood estimation method for the moderating analysis. Finally, we discuss the main findings and propose research and managerial implications.

2 Literature review

Recent studies regarding online customer reviews have drawn their attention on the two aspects: (1) product sale effects and (2) psychological outcomes, in particular, online review helpfulness. From the marketing perspective, a number of researchers

have indicated that online reviews can be a valuable tool for promoting products, collecting consumer feedback and boosting sales [12–14]. Duan et al. [15] examined the persuasive effect and awareness effect of online reviews on movies' daily box office performance by considering reviews both influencing and influenced by movie sales. Hu et al. [16] concluded that online consumer reviews inferred product quality and reduced product uncertainty, in turn aiding the final purchase decision. Ghose and Ipeirotis [12] identified that the valence (average numerical rating) and number of online consumer reviews were vital predictors of product sales employing a panel data set of electronic products. Cui et al. [17] demonstrated that the percentage of negative reviews had a greater effect than that of positive reviews, confirming the negativity bias. They also showed that review volume had a significant effect on new product sales in the early period while such effect decreased over time. Meanwhile, researchers have utilized meta-analysis to quantitatively synthesize this developing literature stream and explore the consequences of online product reviews. For instance, Floyd et al. [7] examined 26 studies investigating how review valence and volume influenced the elasticity of retailer sales. You et al. [8] focused on the effects of consumer-generated information on firm performance, extending the analysis to 51 papers and other online platforms (e.g., blogs, discussion forums, and Twitter). More recently, Rosario et al. [9] extended the meta-analysis employing more effect sizes to give a more comprehensive explanation of the effect of electronic word of mouth on sales.

A growing body of research has paid attention to review helpfulness and what makes a helpful review. Based on Amazon.com data, Mudambi and Schuff [3] concluded that review extremity, review depth, and product type affected the perceived helpfulness of reviews. Cao [18] empirically examined the impact of the various features (including rating, length, posted time) of online user reviews on the number of helpfulness votes by employing dataset from CNET Download. With data collected from a real online retailer, Pan and Zhang [11] provided empirical evidence to support their conceptual predictions and found that both review valence and length had positive effects on review helpfulness. Korfiatis et al. [19] explored the determinants of online review helpfulness and concluded that review readability had a greater effect on the helpfulness ratio of a review than its length; In addition, extremely helpful reviews received a higher score than those considered less helpful. Using Amazon.com data, Baek et al.'s [20] study supported that both review rating and word count influenced the helpfulness of reviews. Lee and Choeh [21] predicted the level of review helpfulness using HPNN and suggested that factors such as the average number of words in a sentence and the length of a review in words had positive effects on the degree of helpfulness. Huang et al. [22] suggested that word count had a threshold in its effects on review helpfulness. Beyond this threshold, its effect diminished significantly or became near non-existent. They also found that reviewer experience and their impact were not statistically significant predictors of helpfulness. More recently, Fu et al. [23] tested mediation effects of review rating, depth and delay between membership tiers and review helpfulness and indicated that review rating was a reflection of customer attitudes and affective tendencies toward sellers and review depth had a positive effect on review helpfulness. Singh et al. [24] developed models based on machine learning to predict the helpfulness of

the consumer reviews using several textual features, where the parameters could be ranked and mapped to helpfulness.

Although these studies have provided important insights into the impact of reviews on their helpfulness, little research has examined them jointly. Past research has not provided a theoretically grounded explanation of what constitutes a helpful review, intimating the need for a systematic integration of this body of work. Consequently, with the experienced researchers providing concise, helpful information about review helpfulness, this study aims to provide a quantitative integration of the main constructs that are associated with review helpfulness and a summary of the relationships between these constructs and helpfulness. We illustrate three metrics that consumers may take into account when determining the helpfulness of a review and refine them the most important antecedents of review helpfulness which have not been fully accounted within the context of online reviews.

3 Conceptual framework

In this study, we examine the determinant factors that may influence review helpfulness based on review characteristics. Our model illustrates the above metrics that consumers take into account when determining the helpfulness of a review. We complement Mudambi and Schuff's research by adding review volume and readability for review depth besides length. Length and readability are regarded as micro-level of depth because they are directly related to the extent to which a review text is understood [19]. Volume can be an indication of macro-level depth in that more reviews could attract more consumers participate in and increase the likelihood of review helpfulness intangibly. Retail websites allow consumers to post product evaluations in the form of numerical star ratings, and one star indicates an extremely negative view of the product and five stars reflects an extremely positive view; Quadratic review rating is utilized to represent the quadratic term with an extreme rating of one or five and a moderate rating of three. Besides, timeliness is quantified by review age of a review in our study, which is calculated by the time elapsed (in days) since the date on which a review was posted. Figure 1 provides an overview of our research model.

3.1 Dependent and independent variables

The dependent variable in this meta-analysis is review helpfulness. Review helpfulness is defined as consumers' perceived value of online reviews while shopping online [3, 25, 26]. Many websites have designed peer reviewing systems that allow people to vote on whether they find a review helpful in their decision making [27], which can reduce information overload by identifying which among the massive number of reviews is helpful [28]. Some websites use the ratio of helpful votes to total votes received as the measure of perceived review helpfulness (e.g. [29, 30]) and others choose to use the absolute total number of helpful votes received as the helpfulness measure (e.g. [31, 32]). As an indicator of psychological outcomes of

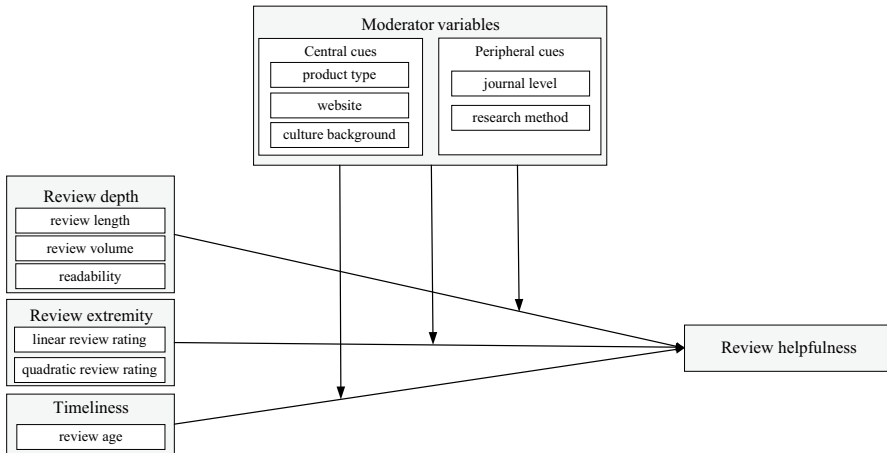


Fig. 1 Research model

reviews, helpful reviews enable consumers to make better decisions and experience greater satisfaction with the online channel [33].

3.1.1 Review length

In the context of online reviews, review length means the word count of a review. Word becomes a natural way to look at insights of online reviews because reviews are typically composed and delivered in text forms [22]. Reviews provide information to help in the decision making process of consumers but helpfulness varies when word count is taken into account. Short reviews are more likely to be shallow and lack the comprehensive evaluation of product features [34]. Longer reviews contain relatively more information, which help consumers to obtain indirect consumption experiences [27, 35]. Yin et al. [5] indicated that length had a stronger effect on helpfulness than other variables. There are also cases that the likelihood of a review being valuable gradually decreases as the number of words surpass a value [20, 22, 36]. Kuan et al. [37] argued that a longer review attracted attention and motivated reading but discouraged processing when the additional cognitive effort anticipated exceeds the incremental value expected from extra length. Therefore, we need to clarify the relationship between review length and helpfulness and propose:

H1 There is a significant positive effect of review length on review helpfulness.

3.1.2 Review volume

The volume of online reviews is an extrinsic, high-scope cue that increases the information diagnosticity and consumer awareness about a certain product [38, 39]. Opinion expressed by more people usually conveys more possibility of helpfulness [40]. However, when a product has a large number of available reviews, the

abundance of product information may dilute the helpfulness of any single review [21, 41]. Zhou and Guo [42] found review count exerting positive impact on helpfulness, but when the reviews are classified into short and long reviews, the influence is only significant in short reviews. Meanwhile, too large volume of reviews constantly being generated may be considered as a big data challenge for both online businesses and consumers and makes it difficult for buyers to go through all the reviews to make purchase decisions [37]. Consequently, we put forward the following hypothesis:

H2 There is a significant negative effect of review volume on review helpfulness.

3.1.3 Review readability

Another more sophisticated measurement of review depth is review readability—cognitive efforts required to read and comprehend reviews [43]. Online reviews are information resources that consumers utilize to gain knowledge about products and services, the extent to which an individual requires to comprehend the information can present the level of readability [27]. A review is considered more helpful on condition that it is easier to comprehend [4, 5, 19, 44]. Reviews that are too readable may appear unprofessional, which will be more likely to be disregarded or evaluated unfavorably [37]. Chua and Banerjee [45] classified reviews into favorable and unfavorable reviews and found that, for search products, unfavorable reviews ease of comprehension are most likely to attract helpful votes. In our analysis, we are interested to examine the extent to which a review easy to read will influence helpfulness and propose:

H3 There is a significant positive effect of review readability on review helpfulness.

3.1.4 Review extremity

Review extremity is one of the key features which can be used to determine the helpfulness of online reviews [46]. Review extremity in our study is measured by both linear review rating and quadratic review rating. Linear review rating refers to the star rating of a review; the more stars a review received, the more positive the review is. Quadratic review rating refers to the quadratic term of review rating and is included to account for the nonlinear relationship between rating and helpfulness [3]. In most review platforms, reviewers can rate their experiences of products or services by using a single indicator to reflect the overall valence of their reviews [47]. The extant literature has demonstrated a positive association between online review ratings and review helpfulness [19, 22]. However, it is possible that reviews with lower ratings and/or higher extremity are considered more helpful [48, 49]. Forman et al. [41] found that for books, moderate reviews were less helpful than extreme reviews. Park and Nicolau also indicated that the extreme ratings of Amazon products were more influential than moderate ratings on consumers' decision

to buy or not buy, and extreme bad reviews are more useful than extreme ones [50]. Hence, we pose the following hypotheses:

H4a There is a significant positive effect of linear review rating on review helpfulness.

H4b There is a significant positive effect of quadratic review rating on review helpfulness.

3.1.5 Timeliness

Timeliness has different descriptions such as elapsed days, days since posting and age of review. Research has demonstrated the significant relationship between timeliness and review helpfulness. Otterbacher [25], Lee and Choeh [21], Cao [18] all testified that reviews posted more recently tend to be more helpful than those posted earlier, while Pan and Zhang [11] hold the view that older reviews might be more helpful since they might exhaust the aspects worth reviewing, leaving less perspective for newer reviewer to add on. Wan [51] thought early reviews not only have a higher chance of being identified as helpful reviews but have advantage in retaining that position because of Matthew effect. Salehan and Kim [52] pointed out older reviews were perceived to be more helpful because older reviews started receiving votes early and has a better chance of appearing on the first page which would lead to more votes and perhaps better perceptions regarding their helpfulness. Therefore, we expect that there exists correlation between timeliness and review helpfulness and propose:

H5 There is a significant positive effect of review age on review helpfulness.

Table 1 provides the main study variables and their description and the expected relationships would be tested by meta-analysis results.

3.2 Moderator effects

To explain the inconsistent findings in previous research, we select product type, websites and culture background as central cues, and journal quality and research methods as peripheral cues in the effect of customer reviews on helpfulness. Similar to other meta-analysis studies on online reviews, such as Rosario et al. [9] and Purnawirawan et al. [10], we include these potential moderators for the two reasons: First, they can be coded from the primary studies included in the meta-analysis; Secondly, they do exert impacts on the relationship between helpfulness and its determinants.

Products are commonly classified as search or experience and consumers process information differently based on their categories [3]. Quality evaluation of search products is objective in that a search product can be evaluated based on its information [45]. As consumers are lacking product experience before purchase

Table 1 The main study variables

| Variable type | Category | Main variables | Description | Expected relationship |
|-----------------------|--------------------|--|---|-----------------------|
| Dependent variables | | Review helpfulness | The ratio of helpful votes to total votes received or the absolute total number of helpful votes received | |
| Independent variables | Review depth | Review length | The total number of words of a review | H1 |
| | | Review volume | The total number of reviews of a product | H2 |
| | Review readability | Ease of understanding of reviews, measured by the Coleman-Liau Index | H3 | |
| | Review extremity | Linear review rating | Review rating ranging from one star to five stars | H4a H4b |
| | | Quadratic review rating | Measured by review rating*review rating | |
| Timeliness | Review age | The number of days since the review was posted. | H5 | |

of experience products, they might seek additional information as an effort to reduce perceived risk and uncertainty [22]. In this case, the difference in product type may change how review helpfulness determinants influence perceived review helpfulness. Consumers care about not only review sender pertaining to the information but also where the information comes from. Given that American consumers have stronger individualistic values and can express their opinions more freely without reproach, which leads to more variation in presenting their review and increases the value of reviews [53], we examine whether the cultural background of American individuals may take an effect in our study. Besides, reviews are more effective when consumers trust them and are confident about the website where they are displayed [54–56]. In the study, we identify the potential moderator of website on the relationship between online reviews and helpfulness; In particular, we examine the performance of Amazon.com which is perceived as one of the largest B2C online forums for user-generated product reviews. Therefore, we propose the following hypotheses:

H6a The relationship between online review helpfulness and its determinants is stronger (weaker) for experience (search) products.

H6b The relationship between online review helpfulness and its determinants is stronger (weaker) for American (non-American) culture.

H6c The relationship between online review helpfulness and its determinants is stronger (weaker) for Amazon (other) website.

In addition, the papers comprising our database come from different levels, which may disturb the mean effect sizes that we calculate. As a result, similar to other meta-analysis studies on online reviews, such as Floyd et al. [7] and

Johnson et al. [57], we code whether each study published in top journals (whose article is more rigorous and persuasive) or some other outlet in an attempt to capture variance in the rigor of each paper's research as well as the review process it withstood. Measurement methods might also contribute to identifying systematic patterns in a meta-analysis [6, 58]. The studies comprising our database differ in the models applied to estimate review helpfulness. Thus, we capture differences in estimation method by differentiating between econometric model like ordinary least squares (OLS) regression and non-econometric model like experiments. Following the moderators in other meta-analysis studies (e.g. [7–10]), each of the moderator variables is coded as binary variable. Hence, we pose the following hypotheses:

H7a The relationship between online review helpfulness and its determinants is stronger (weaker) in top (non-top) journals.

H7b The relationship between online review helpfulness and its determinants is stronger (weaker) by econometric (non-econometric) model.

Table 2 provides the moderator variables, their description and coding scheme and the expected relationship would be tested by meta-analysis results.

4 Methods

4.1 Selection of studies

To identify the empirical studies concerning factors of review helpfulness, we conduct a rigorous and thorough literature search. We use the following keywords in the search process: “consumer-generated content”, “electronic word of mouth”, “online review”, “review helpfulness”, “review usefulness”, “review perceived helpfulness”, “credibility of online consumer review”, and “perceived value of reviews”. First, we do an elaborate computerized bibliographic search the commonly used digital library database such as EBSCO, Science Direct, Springer, and Web of science and check the database using relevant subjects of online reviews and identify 829 articles related to our study initially. Second, we conduct a manual search of relevant journals where research related to perceived review helpfulness is mostly likely to be published and focus on the following categories: (1) marketing, (2) economics and management, and (3) information systems and computer science. Third, to ensure the comprehensiveness of the literature, we scan the Internet using Google Scholar, Research Gate and so on to request working papers. Fourth, we apply a snowballing procedure, in which we examine the references in the publications obtained to find additional studies.

After completing the search process, we exclude theoretical papers, qualitative investigations and quantitative studies that do not report factors we need. The studies could not be included for several reasons: (1) results are not originated from empirical models and no correlations are presented (e.g. [2]); (2) review helpfulness is

Table 2 The moderator variables

| Moderator variables | Description | Coding scheme | Expected relationship |
|---------------------|---|--------------------------------------|-----------------------|
| Product type | Whether a product is experience or search product | 1 = Experience, 0 = search | H6a |
| Website | Data is collected from Amazon.com or not | 1 = Amazon.com, 0 = others | H6b |
| Culture | Reviews are posted by Americans or not | 1 = American consumers, 0 = others | H6c |
| Journal level | Whether the article is on top journal or not | 1 = Top journals, 0 = others | H7a |
| Method | Analysis is econometric model or not | 1 = Econometric, 0 = non-econometric | H7b |

proposed in the context of consumer complaining behavior, but it is not treated as dependent variables of the proposed model (e.g. [59]); and (3) review helpfulness is measured but it is not related to any of the variables of the proposed model in this study (e.g. [28]). Finally, we confirm 53 manuscripts that provide sufficient information to calculate effect sizes. The “Appendix A” provides a list of these manuscripts.

4.2 Coding and effect size integration

To ensure our analysis includes maximum number of effect sizes so that the results are more generalizable, the effect size metric selected is the r-family of effect size indicators (including correlation coefficient, the standardized regression beta coefficient and t value) between review helpfulness and its determinants.¹ The coefficient is chosen because it is an easily interpretable, scale-free measure and provides a metric to which the reported non-correlations could be converted [60], and it has been widely used in recent meta-analysis of marketing literature (e.g. Purnawirawan [10]; Rosario [9]). Following standard procedures employed in other meta-analysis (e.g. Rosenthal [61]; Peterson and Brown [62]), we convert other effect size indicators (the standardized regression beta coefficient and t statistics) into correlation coefficients to compute the mean effect sizes. After completing the coding process, we obtain a total of 191 effects from the 53 publications. The frequency distribution of observed effects is presented in Fig. 2.

We integrate the effect sizes according to guidelines in the meta-analysis literature [63]. Correlations are first adjusted by reliability for measurement error. We then convert the adjusted correlations to Fisher’s Z statistic, and weight them applying the

¹ We convert t value to the correlation coefficient effect size by using the formula suggested by Rosenthal [46]: $r = t / \sqrt{(t^2 + d.f.)}$ where t is the t-value associated with the regression parameter that

captures the effect and d.f. is the degree of freedom of the reported regression model; We convert β to coefficient effect size by using the formula suggested by Peterson and Brown [47]: $r = 0.98\beta + 0.05\lambda$, where λ is a variable that equals 1 when β is non-negative and 0 when β is negative.

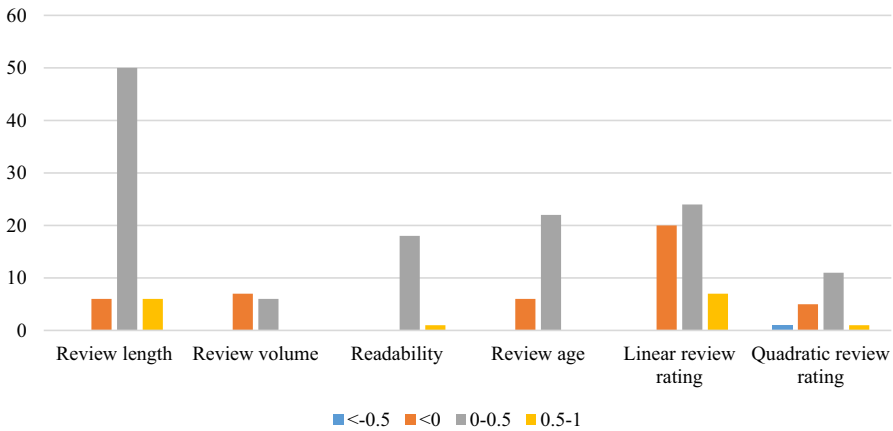


Fig. 2 Frequency distribution of effect sizes

square of sampling standard error to compute the mean values and confidence intervals. We convert them back to \bar{r} using appropriate formulas. In this approach, the effect size is corrected both for sampling and measurement error.² Homogeneity of the effect size distribution is tested by Q-statistics [64]. If the null hypothesis of homogeneity is rejected, difference in effect sizes may be attributed to other factors other than sampling error alone. Confidence interval is presented for each effect size to test if it is significant. When the mean effect size is significant, a fail safe N is an indication of the robustness of the results. It is calculated to estimate the number of unavailable studies necessary to bring the cumulative effect size to a non-significant value [65].³

4.3 Estimation model

We conduct hierarchical linear model analysis (HLMA) using the moderators including product type (experience vs. search) website (Amazon vs. others), culture background (American culture vs. others), and journal level (top journal vs. others),

² We use the following formulae for Fisher’s Z: (1) transformation: $z_r = 0.5 \ln\left(\frac{1+r}{1-r}\right)$, (2)

Weighted average: $\bar{z}_r = \frac{\sum (n_i - 3) * z_r}{\sum (n_i - 3)}$ and (3) back-transformation to correlation units:

$$\bar{r} = \frac{(e^{2\bar{z}_r} - 1)}{(e^{2\bar{z}_r} + 1)} \text{ [49].}$$

³ We calculate the 95% confidence interval as: lower $CI = \bar{z}_r - 1.96 / \sqrt{\sum (n_i - 3)}$, upper

$CI = \bar{z}_r + 1.96 / \sqrt{\sum (n_i - 3)}$; the variance of effect size as: $S_r^2 = \frac{\sum n_i (r_i - \bar{r})^2}{\sum n_i}$ and variation

caused by sampling error as: $S_e^2 = \frac{\sum n_i (1 - \bar{r})^2}{\sum n_i}$; Q-value is calculated as:

$Q = \sum (n_i - 3) * (z_r - \bar{z}_r)^2$ and the fail safe N is as: $N = k * \left(\frac{\bar{r}}{r_c} - 1\right)$, where r_c is the “just significant” level or critical effect size which usually use 0.01.

research method (econometrics vs. non-econometrics) to explain the variation in the correlation between reviews and helpfulness. The HLMA is preferable to more conventional subgroup moderator analysis for its use of a regression-based format and explicating the multi-level nested or hierarchical structure, which overcome the statistical error brought by all level factors regarded as the same [66]. The model is estimated by using the maximum likelihood estimation method because it produces robust, efficient, and consistent estimates [67].

Before estimating the HLMA, we conduct several checks to ensure the robustness of this meta-analysis. We examine the bivariate correlations among the moderators to avoid the potential collinearity problems and perform sensitivity analysis by omitting each of the factors with at least one correlation greater than .5, one at a time, as proposed in previous meta-analysis [68]. The coefficient estimates are stable in all cases in the models. In addition, we perform a residual analysis of errors to determine whether the assumptions of HLMA are satisfied. The extensive robustness checks can rule out multicollinearity and ensure the stability of our model and results. Based on the HLMA analysis, we specially depict the different trends of Amazon.com and all websites, American culture and all cultures to explain the variation related to website and culture background in the correlation between review helpfulness and its determinants.

5 Results

Table 3 provides a summary of mean effect sizes and homogeneity analysis of the influence of review characteristics on helpfulness. #Effect size, sample sizes, mean effect sizes, 95% confidence intervals, the variance of correlation coefficients, variance caused by sampling error, the residual variance, heterogeneity test Q and fail safe N are reported. Regarding review depth, the mean effect size is strongest for readability ($r = .110$, $p < .1$), followed by review length ($r = .094$, $p < .001$) and volume ($r = -.023$, $p < .05$), indicating that review readability exerts greater impact on review helpfulness than review length and volume in terms of review depth. Besides, the negative coefficient between volume and helpfulness indicates that review helpfulness decreases as the number of reviews increase. We also find that the relationship between review rating and review helpfulness is $.056$ ($p < .1$) and between rating² and review helpfulness is $.086$ ($p < .1$). Meanwhile, timeliness exerts positive influence on review helpfulness with mean effect sizes of $.037$ ($p < .05$). All the proposed relationships are found to be significant, implying that on average these variables do have significant influence on review helpfulness. Thus, hypotheses H1, H2, H3, H4a, H4b, and H5 are supported.

Examining the confidence intervals, we notice that all correlation coefficients do not include zero and have lower dispersion around the mean effects, indicating the mean effect sizes can significantly reflect the relationships. The greater the file safe N, the stronger stability of the mean effect size and the less likely it is to overthrow it. We can find that file safe N is higher for the links length-helpfulness and rating-helpfulness, which is an indication that the two mean effect sizes are more unlikely to be null. For example, to bring length-helpfulness down to the level of .05 (the

Table 3 Meta-analytic effect sizes of the determinants on helpfulness

| Variables | #Effect size | Sample size | Mean effect size | Confidence interval | Sr ² | Se ² | Q statistics | Fail safe N |
|---------------------|--------------|-------------|------------------|---------------------|-----------------|-----------------|--------------|-------------|
| Depth | 94 | | | | | | | |
| Length | 62 | 2,083,056 | .094*** | (.093, .096) | .018 | .000 | 55,459 | 521 |
| Volume | 13 | 811,236 | -.023** | (-.025, -.021) | .001 | .000 | 6258 | 43 |
| Readability | 19 | 493,268 | .110* | (.107, .113) | .001 | .000 | 2458 | 189 |
| Extremity | 69 | | | | | | | |
| Rating | 51 | 363,379 | .056* | (.055, .057) | .0161 | .000 | 71,153 | 237 |
| Rating ² | 18 | 179,616 | .086* | (.081, .090) | 2.59 | .000 | 61,717 | 136 |
| Timeliness | 28 | | | | | | | |
| Review age | 28 | 832,844 | .037** | (.035, .040) | .016 | .000 | 15,563 | 76 |

*** $p < .01$, ** $p < .05$, * $p < .1$

“just significant” level), there would be 529 studies with null results to be included in our analysis. Even the lower fail safe N of 43 studies for volume suggests that it is unlikely that there is this number of studies (compared with 13) finding null results for the effects of volume on helpfulness.

The interpretation of sampling variance to variation of the effect sizes are all very small (see columns of Sr² and Se²). According to the 75% rule of Hunter and Schmidt [63], there may be some moderating variables that might contribute to explaining the variance in these effect sizes, Q test also supports this conclusion. Q-statistics are significant for all variables, indicating the presence of highly heterogeneity. This means that the variability in the effect sizes exceeds what would be expected from sampling error. The variability can be explained by substantial differences between studies and effect sizes. For that purpose, we apply moderator variables that describe differences between studies and effect sizes.

Table 4 shows the moderator effects of product type, website, culture background, journal level and method on the relationships between helpfulness and its determinants. We can see from the table that the maximum variance inflation factors (VIF) are all below the recommended threshold of 10 [69], indicating that the regression models are relatively free of collinearity. Besides, the models present different proportions of variance explained by the predictors (see Pseudo-R²) and the significant results (e.g. $\chi^2 = 802.7$) of Chi square test indicate that there may be still other moderators affecting the relationship between review helpfulness and its determinants besides the selected moderators in this study.

For the moderating effect of product type, the results show that product type significantly moderates the effects of review extremity and depth (with the exception of review readability). Specially, review extremity exerts relatively strong

Table 4 The HLMA results for moderator analysis

| Moderators | Depth | | | Extremity | | Timeliness |
|---------------------------|------------|------------|---------------|------------|---------------------|---------------|
| | Length | Volume | Readability | Rating | Rating ² | Review age |
| Product type | | | | | | |
| Beta | .12* | .18** | .01 | .18* | .41** | .11 |
| #Effect sizes | 28/34 | 5/8 | 7/12 | 19/32 | 8/10 | 13/15 |
| Mean effect size | .04/.11 | -.01/-.03 | .03/.07 | .03/.16 | .01/.12 | -.04/.06 |
| Website | | | | | | |
| Beta | .09 | -.28*** | -.03 | .03 | -.21 | .05 |
| #Effect sizes | 26/36 | 4/9 | 8/11 | 23/28 | 6/12 | 13/15 |
| Mean effect size | .036/.058 | .06/-.13 | .02/.06 | .022/.043 | .06/.076 | .05/-.01 |
| Culture background | | | | | | |
| Beta | .10 | -.13 | -.06** | -.11 | -.18 | .14** |
| #Effect sizes | 20/42 | 5/8 | 10/9 | 24/27 | 7/11 | 8/20 |
| Mean effect size | .08/.034 | -.02/-.01 | .06/.04 | .068/.015 | .08/.06 | -.01/.05 |
| Journal level | | | | | | |
| Beta | -.15** | -.05* | -.08 | -.18** | .13 | -.07 |
| #Effect sizes | 27/35 | 8/5 | 12/7 | 25/26 | 8/10 | 22/6 |
| Mean effect size | .04/.03 | .00/-.05 | .05/.06 | .04/.03 | .12/-.01 | .02/.01 |
| Method | | | | | | |
| Beta | .19** | .2*** | -.13 | -.05 | -.105 | -.09 |
| #Effect sizes | 14/48 | 2/11 | 3/16 | 11/40 | 2/16 | 9/19 |
| Mean effect size | -.05/.06 | -.04/-.01 | .06/.03 | .02/.04 | .30/.06 | -.11/.04 |
| Max VIF | 1.63 | 2.72 | 2.93 | 1.53 | 2.50 | 3.38 |
| Pseudo-R ² | .26 | .88 | .14 | .35 | .32 | .16 |
| Chi square test | 2.12 (>.1) | 1.93 (>.1) | 802.72 (<.01) | 4.87 (>.1) | 192.05 (<.1) | 235.39 (<.05) |

Standardized coefficients are presented; #Effect sizes and mean effect sizes are displayed for each category of the significant moderators. The code first mentioned is 0 and latter mentioned is 1. *** $p < .01$, ** $p < .05$, * $p < .1$

influence on review helpfulness for experience products than for search products with mean effect sizes of .16 versus .03 on rating and of .12 versus .01 on rating². The negative coefficient ($\beta = -.28, p < .001$) reveals that the effect of volume on review helpfulness is stronger for other websites (.06) than for Amazon.com (-.13). On Amazon.com, review volume exerts a negative impact on the helpfulness; While on other websites, the impact of volume on review helpfulness is positive. Culture background is significant for the links readability-helpfulness ($\beta = -.06, p < .05$) and review age-helpfulness ($\beta = .14, p < .05$), with the negative coefficient indicating that the association between readability and helpfulness is stronger (.06) for the reviews wrote by others while weaker (.04) for those wrote by Americans and the positive coefficient indicating that the association between review age and helpfulness is stronger (.05) for the reviews wrote by Americans

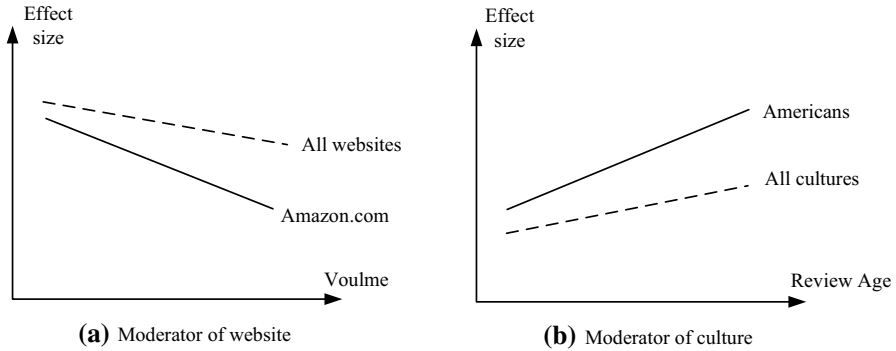


Fig. 3 Trends of effect sizes for the two moderators

while weaker ($-.01$) for those wrote by others. The findings are partially in support of our hypotheses H6a, H6b, H6c.

Overall, journal level presents a similar pattern when considering the relationships between review helpfulness and its determinants. The effect of length, volume and rating on review helpfulness is stronger for other journals than for top journals. For example, the negative coefficient ($\beta = -.15, p < .05$) reveals that the relationship length-helpfulness on other journals (.04) is stronger than that on top journals (.03). Furthermore, econometric method is significant for length-helpfulness ($\beta = .19, p < .05$) and volume-helpfulness ($\beta = .2, p < .001$), with the positive coefficients indicating that the link length-helpfulness is stronger (.06) for the studies measured by econometrics and weaker ($-.05$) for those measured by other methods and that the association volume-helpfulness is stronger ($-.01$) for the studies measured by econometrics and weaker ($-.04$) for those measured by other methods. The findings are partially in support of our hypotheses H7a, H7b.

Given the contradictory effects that the influence of volume on helpfulness on other websites is positive (.06) while negative ($-.13$) on Amazon.com and that the influence of review age on helpfulness moderated by American culture is positive (.05) while negative ($-.01$) by others, we further check the moderator effects of website and culture on the links volume-helpfulness and timelines-helpfulness, respectively. As depicted in Fig. 3a, when volume of reviews increases, the probability of reviews found helpful decreases on Amazon.com while increases on other websites, but this downward trend is more pronounced on Amazon.com. Consequently, the helpfulness of whole reviews decreases as the volume increases. This implies that consumers perceive too many reviews on Amazon.com as less helpful than that on the others. Excessive opinion may well bring only marginal information, which is not particularly important to consumers [11]. As a whole, the influence of review volume on helpfulness depends on websites in the sense that volume has a weaker and negative influence on helpfulness for reviews that are posted on Amazon.com compared with those appearing on other websites.

With regard to culture, the upward trend of effect sizes for all cultures suggests that review age has a positive effect on helpfulness. For non-Americans, they are

inclined to hold that later posted reviews would have more opportunities to be noticed by consumers, thus receiving more helpful votes, which make them easier to get the chance being ranked in the most helpful positions [18]. The influence, however, is distinctive and positive for Americans. As depicted in Fig. 3b, the probability of reviews being helpful for Americans greatly increases as the posted time increases. American reviewers pursue freedom of expression and they can show their opinions more openly; besides, they tend to be more specific in their reviews and express their own opinions without reproach [70]. All these lead to consequences that older reviews may capture the aspects worthy of reviewing, leaving less perspective for later reviewer to add on.

6 Discussion

Based on prior empirical studies about review helpfulness, we illustrate three metrics that consumers may take into account when determining the helpfulness of a review and refine them the most important antecedents of review helpfulness which has not been fully accounted within the context of online reviews. Table 5 summarizes our propositions and findings.

The most impactful influencing variable in our meta-analytic model is review readability. Online product reviews exert significantly greater influence on their helpfulness when they are delivered by containing judgments related to the product being evaluated in a level of understandability and comprehension [19]. Consumers have propensity to search for product reviews that are easy to read, which facilitates them obtaining the specific information necessary within the overwhelming amount of reviews posted online [27]. They value opinions from reviewers who provide easily understood usage experience with the product, and require supporting evidence for arguments. Park and Kim [71] once theorized it cognitive effort and, in particular, review text's cognitive fit to an average consumer with a normal level of expertise regarding the product evaluated. Theoretically, when the information expressed in the text matches the consumer's own information-processing strategy, a cognitive fit occurs [72]. This finding provides a proxy for investigating the impact of readability as an independent measure for helpfulness forecasting.

Our analysis also provides more solid evidence for the discussed interplay between the length and helpfulness attributed to an online review. The attributes of text-based information make customer reviews can be measured using the number of words or pages. Therefore, word becomes a natural way to look at insights of online reviews because reviews are typically composed and delivered in text forms [22]. Although some studies have documented the adverse effect of too much information [73, 74], our study still supports that a review filled with depth is conducive for people's online shopping decision. Longer reviews can draw more attention because consumers can have a better chance to find the review content they are interested from longer ones [35].

Regarding the relationship between review volume and helpfulness, there is a weaker and inverse association between them. When a product has a small number of reviews, the scarcity of information elevates the importance of each available

Table 5 Results summary

| Hypothesis | Result | Depth | | | Extremity | | Timeliness review age |
|------------|---------------------|--------|--------|-------------|-----------|---------------------|--------------------------|
| | | Length | Volume | Readability | Rating | Rating ² | |
| H1 | Confirmed | + | | | | | |
| H2 | Confirmed | | - | | | | |
| H3 | Confirmed | | | + | | | |
| H4a | Confirmed | | | | + | | |
| H4b | Confirmed | | | | | + | |
| H5 | Confirmed | | | | | | + |
| H6a | Partially confirmed | + | + | | + | + | |
| H6b | Partially confirmed | | - | | | | |
| H6c | Partially confirmed | | | - | | | + |
| H7a | Partially confirmed | + | + | | + | | |
| H7b | Partially confirmed | + | + | | | | |

review; With reviews increasing, an additional opinion may well bring only marginal information, which is not particularly important to consumers [11]. Alternatively, when there are only a handful of product opinions, each opinion may bring new information, and therefore matters to consumers; While the abundance of product information appears to dilute the helpfulness of any single review. The moderator analysis demonstrates that website significantly moderates the effect of review volume on helpfulness within Amazon.com exerting stronger negative impact. We notice that the probability of a whole review being found helpful on Amazon.com and the whole websites all decrease with the increase of volume, but this downward trend is more pronounced on Amazon.com. As one of the earliest retail websites, Amazon.com enables consumers to submit product reviews about their opinion without censorship. Pan and Zhang [11] once pointed out that Amazon.com appeared to interfere with publishing consumer reviews at a low level. With so many reviews, people may feel disgusted and doubt about its motives. In contrast, for other websites, more reviews may relate more to personal experiences, and contain more content to express feelings and opinion, which can increase the consumer's viscosity and the visibility of the sites [75]. To sum up, from the perspective of review readers, the volume-helpfulness relationship differs based on the websites and the negative effect of review volume on helpfulness is expected to be more manifest on Amazon.com than on others.

Our results show that the reviews with the older generated opinions have a higher probability of getting more votes, suggesting that people tend to pay more attention to those earlier posted opinions. One of the possible explanations for the fact is that the reviewers of the earlier reviews could provide and integrate more contents and opinions and tend to be more accurate, leaving less information could be written by later reviewers [18]. It is also interesting to find that the cultural background of individuals significantly moderates the relationship timeliness-helpfulness with the evidence that American consumers differ from the others towards the posted reviews. As for □ Americans, earlier reviews have significant advantage over later reviews

in terms of getting attention from online shoppers and garnering helpfulness votes. American consumers are probably more receptive to e-commerce activities and rely more on online recommendations than consumers elsewhere do because of the developed Internet [10]. Additionally, as opposed to many other cultures, the American culture is the most prudential one in the world and American consumers might scan the reviews more seriously with full passion. As a result, American consumers might react stronger to earlier reviews than consumers in other countries.

Our empirical analysis also confirms the positive relationship between review extremity and helpfulness. The positive coefficient for rating means that consumers tend to rate positive reviews to be more helpful than negative ones. This is consistent with the general online shopping experience because consumers are more likely to believe that review content are objective posted by reviewers without personal emotion bias. Besides, review rating has a “U” shape relationship with review helpfulness, indicating that online consumers perceive extreme reviews (positive or negative) as more helpful than moderate reviews. Meanwhile, we also provide insights about important moderators relating to journal level and research method, which may augment or diminish the influence that product reviews exert on helpfulness. Findings from moderator analysis support the moderating role of journal level on length, review volume and rating and results of top journals have more reference value when compared with mean effect sizes of review characteristics. Results of moderator analysis also show that method moderators the influence of length and volume on review helpfulness.

7 Implications

By including collected variables used in prior studies influencing online review helpfulness with the moderator variables used in meta-analysis and modeling them on review depth, extremity and timeliness, we not only identify important factors driving helpfulness but also arrive at a rich set of academic and managerial implications.

7.1 Theoretical implications

First, our research helps summarize the literature on helpfulness within the context of review characteristics. Different with previous meta-analysis which has explored the influence of customer reviews on economic behavior such as how online reviews influence sales, we focus on how and to what extent a review affects consumers' psychological outcomes. By contrasting and combining results from existing studies concerning factors influencing review helpfulness, we provide conclusive take-away with respect to some important aspects of online product reviews and their helpfulness. We give a combination of all the metrics in terms of review depth, extremity and timeliness and provide further researchers a comprehensive model so as to estimate the relative importance of reviews on predicting review helpfulness, which will be conducive to set the agenda for future research efforts.

Second, we ground the measure of review depth in Mudambi and Schuff's study by adding readability and volume. The present study constitutes a complement of earlier work of Mudambi and Schuff [3], where a similar mode was examined with a theoretical model involving review depth and extremity. We propose a more appropriate theoretical framework to explain the interplay between review depth and its helpfulness. Readability is an intrinsic and qualitative cue of depth. To some extent, easily understood reviews indicate a lack of depth in content but save the time consumers look through them. In contrast with review length, volume is an extrinsic and quantitative cue of depth, more reviews mean information for a product. The results of this study enhance prior knowledge by showing that not only length of reviews plays a role, but also the ease of understanding and volume of reviews do.

Third, we test for a wide variety of commonly raised concerns for the relationships, including product type, website, culture background, journal level and research model. Because the focus of our meta-analysis is on the characteristics of the reviews, the effect sizes we obtain differ concerning its source influencing the impact of online product reviews. We find that website significantly moderates the effect of review volume on helpfulness with Amazon.com exerting a negative impact as volume increases. Given that more than 60% of observations in our sample are from Amazon, the huge proportion ultimately make negative volume-helpfulness link. With the rapid growth of internet, China has seen the biggest e-commerce market such as Alibaba and JD.com [23] in the world. More research should be encouraged to enhance the diversity of the websites and should not be limited on Amazon.com.

7.2 Practical implications

Our findings also provide better analytic tools for E-business firms, who can benefit from practical guidance based on the rigorous analysis of specific design elements of online feedback mechanisms and contextual variables that could improve their marketing initiatives.

First, the different influence of review characteristics on helpfulness suggests that they should be separated as different goals in online activities. For instance, if the goal is to maximize review helpfulness, it is perhaps that review depth (such as length and readability) should be taken into account, because it may have great chance to induce consumers to browse through the products. Online retailers therefore could consider different incentive mechanism for customers to provide as much depth, or detail, as possible. The remarkably strong effect of readability suggests that online review readability is an important tool to influence review helpfulness. In practice, encouraging easily understood customer reviews does appear to be an important component of the strategy of many online retailers. E-business firms may attempt to encourage understandable reviews from existing customers as part of their marketing strategy, they need to think about mechanisms to encourage not only more information-rich but also more understood reviews that are helpful to future customers. For example, websites like Amazon could include a readability assessment tool showing the readability scores while a customer is writing his or her review.

Second, it is more of a concern for Amazon.com to restrict the volume of product reviews. Reviews have the potential to attract consumer visits, increase the time spent on the site (“stickiness”), and create a sense of community among frequent shoppers. However, as the availability of customer reviews becomes widespread, the strategic focus shifts from the mere presence of customer reviews to the customer evaluation and use of the reviews [3]. Making a better decision more easily is the main reason consumers use a website, so it is important for Amazon.com to draw consumers’ attention by providing detailed guidelines for writing reviews valuable and censor them and remove irrelevant information so that consumers have an easier access to find the review content they are interested from the dense ones.

Third, our study supports that American consumers are more likely to recognize the older reviews to be helpful. Given that timeliness is a signal influencing online consumers when they do product research, E-business firms should devote efforts to strategically highlight the earlier reviews posted by Americans and bring them to the attention of prospective consumers, even when they are not rated by users as most helpful. In practice, we suggest that online retailers should develop measures regarding how prospective consumers scan the user-generated product information. For example, the review system can distinguish American reviews and strategically highlight them and ordering likely helpful recent reviews. Such practice could help consumers make more informed buying decisions. This implication may be particularly important to those retailers whose target consumers are across the whole world. Offering accurate location of the culture of a reviewer can enhance consumers’ shopping experiences, which, in turn, may help retailers achieve their long-term financial goals.

8 Conclusion and limitations

Previous studies on the impact of review characteristics on consumers’ evaluations in terms of review helpfulness have shown controversial conclusions. Our meta-analysis partly resolves these inconsistencies in the literature. Based on prior empirical studies about review helpfulness, we provide a theoretically grounded explanation of what constitutes a helpful review employing 191 effect sizes of 53 empirical studies. Overall, our study shows that review depth, extremity and timeliness have varying effects on review helpfulness. Review depth exerts greater impact on review helpfulness except for its sub-metric of review volume, which exerts the least impact among the whole variables. This study also confirms the significant positive relationship between review readability and helpfulness. Besides, our findings concerning the moderating effects of websites and culture background on the relationships imply that fewer reviews on Amazon.com will be conducive for its review helpfulness and that American consumers prefer earlier ones when scanning the product reviews. We believe this work could provide multiple contributions to this important field.

Several limitations that are common to meta-analytic reviews are also presented here. First, the quantitative synthesis is constrained by the nature and scope of the original studies on which it is based. Despite the fact that we make every effort to search the original studies, it is inevitable that we still ignore some articles, for

example those published not in English or working papers undisclosed. Second, our analysis is constrained to examining variables that could be coded from the extant literature. While the independent variables studied here provide scholars and practitioners useful information, more factors are still needed to be modeled and reported in future studies of online product reviews. For instance, it would be interesting to explore the effectiveness of review valence, which is used to evaluate the content of individual reviews. Third, this meta-analysis is limited by the quality of the original studies and the availability of moderators. Following Cohen [76] suggestion for analyzing the magnitude of effect sizes ($r > .5$ as large; $r = .3$ as medium; and $r < .1$ as small effect size), given that all the effect sizes are small, there may be quality-difference existing among the selected studies. Fourth, because some moderators such as study characteristics are not readily available in all studies that we consider for this meta-analysis, further research is needed to understand how other moderators may influence the impact of reviews on helpfulness.

Acknowledgements The work described in this paper is supported by National Natural Science Foundation of China (Grant Nos. 71531001 and 71572006).

Appendix A

See Table 6.

Table 6 Studies included in the meta-analysis

| References | Website | Study provides effect sizes of valence influence | | | | | Review age on |
|--------------------------|---------|--|--------|-------------|--------|---------------------|---------------|
| | | Length | Volume | Readability | Rating | Rating ² | |
| Otterbacher [25] | Amazon | | ✓ | ✓ | ✓ | | ✓ |
| Huang et al. [22] | Amazon | ✓ | | | | | |
| Zhang [100] | Amazon | ✓ | | | ✓ | | |
| Pan and Zhang [11] | Amazon | ✓ | ✓ | | ✓ | | ✓ |
| Ghose and Ipeiritis [99] | Amazon | | | ✓ | | | ✓ |
| Ghose and Ipeiritis [4] | Amazon | | ✓ | ✓ | | | |
| Baek et al. [20] | Amazon | ✓ | | | ✓ | ✓ | |
| Lee and Choeh [21] | Amazon | ✓ | ✓ | | ✓ | | ✓ |
| Huang and Yen [77] | Amazon | ✓ | | | ✓ | ✓ | |
| Wan [51] | Amazon | ✓ | | | | | ✓ |
| Chua and Banerjee [49] | Amazon | ✓ | | | ✓ | | |
| Zhang [78] | Amazon | ✓ | | | ✓ | | |
| Mudambi and Schuff [3] | Amazon | ✓ | | | ✓ | ✓ | |
| Li and Zhan [44] | Amazon | ✓ | | | | | |
| Forman et al. [41] | Amazon | | ✓ | | | | |
| Salehan and Dan [52] | Amazon | ✓ | | | | | |

Table 6 (continued)

| References | Website | Study provides effect sizes of valence influence on | | | | | Review age |
|-----------------------------|--------------|---|--------|-------------|--------|---------------------|------------|
| | | Length | Volume | Readability | Rating | Rating ² | |
| Willemsen et al. [79] | Amazon | ✓ | | | ✓ | | |
| Chua and Banerjee [45] | Amazon | ✓ | | ✓ | | | |
| Siering et al. [80] | Amazon | ✓ | | ✓ | | | ✓ |
| Zhang et al. [81] | Amazon | ✓ | | | ✓ | | |
| Einar et al. [82] | Amazon | ✓ | | | | | |
| Siering and Muntermann [83] | Amazon | ✓ | | | | | |
| Wang et al. [84] | Amazon | ✓ | | ✓ | | | |
| Wu [85] | Amazon | ✓ | | ✓ | | | |
| Wu et al. [47] | Amazon | | | | ✓ | ✓ | |
| Kuan and Hui [37] | Amazon | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Wu [30] | Amazon | ✓ | | | ✓ | ✓ | |
| Chen et al. [86] | Amazon | ✓ | | | | | |
| Lee and Choeh [87] | Amazon | ✓ | | | | | |
| Korfiatis et al. [19] | Amazon UK | ✓ | | ✓ | ✓ | ✓ | |
| Karimi and Wang [88] | Applications | ✓ | ✓ | | ✓ | | |
| Yu et al. [89] | Blog, IMDB | ✓ | | | ✓ | | |
| Cao et al. [18] | CNET | ✓ | | | ✓ | | ✓ |
| Hong and Xu [90] | Douban.com | ✓ | | | ✓ | | |
| Schindler and Bickart [36] | Experiment | ✓ | | | | | |
| Filieri [91] | Experiment | | | ✓ | ✓ | | |
| Ullah et al. [92] | IMDB | | | | ✓ | ✓ | |
| Cheng and Ho [93] | Restaurant | ✓ | | | | | |
| Hu and Chen [94] | TripAdvisor | | | | ✓ | | |
| Qazi et al. [34] | TripAdvisor | ✓ | | | | | |
| Fang et al. [95] | TripAdvisor | ✓ | | | ✓ | ✓ | ✓ |
| Kwok and Xie [96] | TripAdvisor | | | | ✓ | | |
| Xiang et al. [29] | TripAdvisor | ✓ | | | | | |
| Yin et al. [97] | Yahoo | ✓ | | ✓ | ✓ | ✓ | |
| Zhou and Guo [42] | Yelp | ✓ | ✓ | | ✓ | | ✓ |
| Liu and Park [27] | Yelp | ✓ | | ✓ | ✓ | ✓ | |
| Racherla and Friske [35] | Yelp | ✓ | | | ✓ | ✓ | ✓ |
| Yin et al. [97] | Yelp | ✓ | | ✓ | ✓ | ✓ | |
| Zhu et al. [43] | Yelp | ✓ | | ✓ | ✓ | ✓ | |
| Guo and Zhou [31] | Yelp | ✓ | ✓ | | | | ✓ |
| Li et al. [98] | Yelp | ✓ | | ✓ | | | |
| Chen and Lurie [32] | Yelp | ✓ | | | ✓ | | ✓ |
| Yin et al. [97] | Yelp. | ✓ | | ✓ | ✓ | | ✓ |

References

1. Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, *49*(10), 1407–1424.
2. Malik, M., & Hussain, A. (2018). An analysis of review content and reviewer variables that contribute to review helpfulness. *Information Processing and Management*, *54*(1), 88–104.
3. Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on amazon.com. *MIS Quarterly*, *34*(1), 185–200.
4. Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, *23*(10), 1498–1512.
5. Yin, G., Wei, L., Xu, W., & Chen, M. (2014). Exploring heuristic cues for consumer perceptions of online reviews helpfulness: The case of yelp.com. In *Pacific Asia conference on information systems*.
6. Farley, J. U., Lehmann, D. R., & Sawyer, A. (1995). Empirical marketing generalizations using meta-analysis. *Marketing Science*, *14*(3), G36–G46.
7. Floyd, K., Freling, R., Alhoqail, S., Cho, H. Y., & Freling, T. (2014). How online product reviews affect retail sales: A meta-analysis. *Journal of Retailing*, *90*(2), 217–232.
8. You, Y., Vadakkepatt, G., & Joshi, A. M. (2015). A meta-analysis of electronic word-of-mouth elasticity. *Journal of Marketing*, *79*(2), 19–39.
9. Rosario, A. B., Sotgiu, F., Valck, K. D., & Bijmolt, T. H. A. (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.
10. Purnawirawan, N., Eisend, M., Pelsmacker, P. D., & Dens, N. (2015). A meta-analytic investigation of the role of valence in online reviews. *Journal of Interactive Marketing*, *31*, 17–27.
11. Pan, Y., & Zhang, J. Q. (2011). Born unequal: A study of the helpfulness of user-generated product reviews. *Journal of Retailing*, *87*(4), 598–612.
12. Ghose, A., & Ipeirotis, P. G. (2010). Estimating the helpfulness and economic impact of product reviews: mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, *23*(10), 1498–1512.
13. Ogut, H., & Tas, B. (2012). The influence of internet customer reviews on the online sales and prices in hotel industry. *The Service Industry Journal*, *32*(2), 197–214.
14. Chu, W., & Roh, M. (2014). Exploring the role of preference heterogeneity and causal attribution in online ratings dynamics. *Asia Marketing Journal*, *15*(4), 61–101.
15. Duan, W., Gu, B., & Whinston, A. (2008). Do online reviews matter?—An empirical investigation of panel data. *Decision Support Systems*, *45*(4), 1007–1016.
16. Hu, N., Liu, L., & Zhang, J. (2008). Do online reviews affect product sales? The role of reviewer characteristics and temporal effects. *Information Technology Management*, *9*(3), 201–214.
17. Cui, G., Lui, H. K., & Guo, X. (2012). The effect of online consumer reviews on new product sales[M]. *International Journal of Electronic Commerce*, *17*(1), 39–57.
18. Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the “helpfulness” of online user reviews: A text mining approach. *Decision Support Systems*, *50*(2), 511–521.
19. Korfiatis, N., García, B. E., & Sánchez, A. S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness versus review content. *Electronic Commerce Research and Applications*, *11*(3), 205–217.
20. 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.
21. Lee, S., & Choeh, J. Y. (2014). Predicting the helpfulness of online reviews using multilayer perceptron neural networks. *Expert Systems with Applications*, *41*(6), 3041–3046.
22. Huang, A. H., Chen, K., Yen, D. C., & Tran, T. P. (2015). A study of factors that contribute to online review helpfulness. *Computers in Human Behavior*, *48*(C), 17–27.
23. Fu, D., Hong, Y., Wang, K., & Fan, W. (2017). Effects of membership tier on user content generation behaviors: Evidence from online reviews. *Electronic Commerce Research*, *7*, 1–27.
24. Singh, J., Irani, S., & Rana, N. (2016). Predicting the “helpfulness” of online consumer reviews. *Journal of Business Research*, *70*, 346–355.
25. Otterbacher, J. (2009). ‘Helpfulness’ in online communities: a measure of message quality. In *Sigchi conference on human factors in computing systems* (pp. 955–964).

26. Baek, H., Lee, S., Oh, S., & Ahn, J. (2015). Normative social influence and online review helpfulness: Polynomial modeling and response surface analysis. *Journal of Electronic Commerce Research*, 16(4), 290–306.
27. Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management*, 47(47), 140–151.
28. Ngoye, T. L., & Sinha, A. P. (2014). The influence of reviewer engagement characteristics on online review helpfulness: A text regression model. *Decision Support Systems*, 61(4), 47–58.
29. Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, 58, 51–65.
30. Wu, J. (2017). Review popularity and review helpfulness: A model for user review effectiveness. *Decision Support Systems*, 97, 92–103.
31. Guo, B., & Zhou, S. (2017). What makes population perception of review helpfulness: An information processing perspective. *Electronic Commerce Research*, 17(4), 1–24.
32. Chen, Z., & Lurie, N. (2013). Temporal contiguity and negativity bias in the impact of online word of mouth. *Journal of Marketing Research*, 50(4), 463–476.
33. Kohli, R., Devaraj, S., & Mahmood, M. A. (2004). Understanding determinants of online consumer satisfaction: A decision process perspective. *Journal of Management Information Systems*, 21(1), 115–135.
34. Qazi, A., Syed, K., Raj, R. G., & Cambria, E. (2016). A concept-level approach to the analysis of online review helpfulness. *Computers in Human Behavior*, 58(C), 75–81.
35. Racherla, P., & Friske, W. (2012). Perceived usefulness of online consumer reviews: An exploratory investigation across three services categories. *Electronic Commerce Research and Applications*, 11(6), 548–559.
36. Schindler, R. M., & Bickart, B. (2012). Perceived helpfulness of online consumer reviews: The role of message content and style. *Journal of Consumer Behavior*, 11(3), 234–342.
37. Kuan, K., & Hui, K. (2015). What makes a review voted? An empirical investigation of review voting in online review systems. *Journal of the Association for Information Systems*, 16(1), 48–71.
38. Anderson, W. E., & Salisbury, C. L. (2003). The formation of market-level expectations and its covariates. *Journal of Consumer Research*, 30(1), 115–124.
39. Khare, A., Lauren, I. L., & Anthony, K. A. (2011). The assimilative and contrastive effects of word-of-mouth volume: An experimental examination of online consumer ratings. *Journal of Retailing*, 87(1), 111–126.
40. Salganik, M. J., Peter, S. D., & Duncan, J. W. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762), 854–856.
41. Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, 19(3), 291–313.
42. Zhou, S., & Guo, B. (2015). The interactive effect of review rating and text sentiment on review helpfulness. In *E-Commerce and web technologies* (pp. 100–111).
43. Zhu, L., Yin, G., & He, W. (2014). Is this opinion leader's review useful? Peripheral cues for online review helpfulness. *Journal of Electronic Commerce Research*, 15(4), 267–280.
44. Li, J., & Zhan, L. (2011). Online persuasion: How the written word drives WOM—Evidence from consumer-generated product reviews. *Journal of Advertising Research*, 51(1), 239–257.
45. 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(C), 547–554.
46. Kim, S. M., Pantel, P., Chklovski, T., & Pennacchiotti, M. (2006). Automatically assessing review helpfulness. In *Conference on empirical methods in natural language processing* (pp. 423–430). Association for Computational Linguistics.
47. Wu, P. F., Hans, V. D. H., & Korfiatis, N. (2011). The influences of negativity and review quality on the helpfulness of online reviews. In *International conference on information systems*. Shanghai, China.
48. Yin, D., Bond, S. D., & Zhang, H. (2014). Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quarterly*, 38(2), 539–560.
49. 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.

50. Sangwon, P., & Nicolau, J. L. (2015). Asymmetric effects of online consumer reviews. *Annals of Tourism Research*, 50, 67–83.
51. Wan, Y. (2015). The Matthew effect in social commerce: The case of online review helpfulness. *Electronic Markets*, 25(4), 313–324.
52. Salehan, M., & Dan, J. K. (2016). Predicting the performance of online consumer reviews. *Decision Support Systems*, 81(C), 30–40.
53. Koh, N. S., Hu, N., & Clemons, E. K. (2010). Do online reviews reflect a product's true perceived quality? An investigation of online movie reviews across cultures. *Electronic Commerce Research and Applications*, 9(5), 374–385.
54. McGinnies, E., & Ward, D. C. (1980). Better liked than right: Trustworthiness and expertise as factors in credibility. *Personality and Social Psychology Bulletin*, 6(3), 467–472.
55. Grewal, D., Gotlieb, J., & Marmorstein, H. (1994). The moderating effects of message framing and source credibility on the price-perceived risk relationship. *Journal of Consumer Research*, 21(1), 145–153.
56. Schweidel, D., & Moe, W. (2014). Listening in on social media: A joint model of sentiment and venue format choice. *Journal of Marketing Research*, 51(4), 387–402.
57. Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). Effects of cooperative, competitive and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin*, 89(1), 47–62.
58. Assmus, G., Farley, J. U., & Lehmann, D. R. (1984). How advertising affects sales: Meta-analysis of econometric results. *Journal of Marketing Research*, 21(1), 65–74.
59. Lee, J. (2013). What makes people read an online review? The relative effects of posting time and helpfulness on review readership. *Cyberpsychology Behavior & Social Networking*, 16(7), 529.
60. Völckner, F., & Hofmann, J. (2007). The price-perceived quality relationship: A meta-analytic review and assessment of its determinants. *Marketing Letters*, 18(3), 181–196.
61. Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York: Russell Sage Foundation.
62. Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology*, 90(1), 175–181.
63. Hunter, J. E., & Schmidt, F. L. (2006). Methods of meta-analysis: Correcting error and bias in research findings. *Evaluation & Program Planning*, 29(3), 236–237.
64. Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Beverly Hills: Sage.
65. Rosenthal, R. (1979). The 'file drawer problem' and tolerance for null results. *Psychological Bulletin*, 86, 638–641.
66. Tellis, G. J. (1988). The price elasticity of selective demand: A meta-analysis of econometric models of sales. *Journal of Marketing Research*, 25(4), 331–341.
67. Singer, J. D., & Willet, J. B. (2003). *Applied longitudinal data analysis*. Berlin: Springer.
68. Bijmolt, T. H. A., & Pieters, R. G. M. (2001). Meta-analysis in marketing when studies contain multiple measurements. *Marketing Letters*, 12(2), 157–169.
69. Hair, J. F., Anderson, R. E., Tatham, R. L., & Black W. C. (1998). *Multivariate data analysis* (5th ed.). All Publications.
70. Huang, H. (2005). A cross-cultural test of the spiral of silence. *International Journal of Public Opinion Research*, 17(3), 324–345.
71. Park, D. H., & Kim, S. (2009). The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews. *Electronic Commerce Research and Applications*, 7(4), 399–410.
72. Vessey, I., & Galletta, D. (1991). Cognitive fit: An empirical study of information acquisition. *Information Systems Research*, 2(1), 63–84.
73. Jackson, T. W., & Farzaneh, P. (2012). Theory-based model of factors affecting information overload. *International Journal of Information Management*, 32(6), 523–532.
74. Schultz, C., Schreyoegg, J., & Von, R. C. (2013). The moderating role of internal and external resources on the performance effect of multitasking: Evidence from the R&D performance of surgeons. *Research Policy*, 42(8), 1356–1365.
75. Jiang, Z., & Benbasat, I. (2007). Investigating the influence of the functional mechanisms of online product presentations. *Information Systems Research*, 18(4), 221–244.
76. Cohen, J. (1988). Statistical power analysis for the behavioral science. *Journal of the American Statistical Association*, 84(363), 19–74.

77. Huang, H. A., & Yen, C. D. (2013). Predicting the helpfulness of online reviews—A replication. *International Journal of Human-computer Interaction*, 29(2), 129–138.
78. Zhang, Z. (2008). Weighing stars: Aggregating online product reviews for intelligent E-commerce applications. *IEEE Intelligent Systems*, 23(5), 42–49.
79. Willemsen, L. M., Neijens, P. C., Bronner, F., & Ridder, J. (2011). “Highly Recommended!” The content characteristics and perceived usefulness of online consumer reviews. *Journal of Computer-Mediated Communication*, 17(1), 19–38.
80. 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.
81. Zhang, J. Q., Craciun, G., & Shin, D. (2010). When does electronic word-of-mouth matter? A study of consumer product reviews. *Journal of Business Research*, 63(12), 1336–1341.
82. Einar, B., Havro, L. J., & Moen, O. (2015). An empirical investigation of self-selection bias and factors influencing review helpfulness. *International Journal of Business & Management*, 10(7), 16–30.
83. Siering, M., & Muntermann, J. (2013). What drives the helpfulness of online product reviews? From stars to facts and emotions. In *11th international conference on Wirtschaft sinformatik*.
84. Wang, C. C., Li, M. Z., & Yang, Y. H. (2015). *Perceived usefulness of word-of-mouth: An analysis of sentimentality in product reviews* (pp. 448–459). Berlin: Springer.
85. Wu, P. F. (2013). In search of negativity bias: An empirical study of perceived helpfulness of online reviews. *Psychology & Marketing*, 30(11), 971–984.
86. Chen, X., Sheng, J., Wang, X., & Deng, J. S. (2016). Exploring determinants of attraction and helpfulness of online product review: A consumer behavior perspective. *Discrete Dynamics in Nature and Society*, 1, 1–19.
87. Lee, S., & Choeh, J. Y. (2017). Exploring the determinants of and predicting the helpfulness of online user reviews. *Management Decision*, 40(3), 316–332.
88. Karimi, S., & Wang, F. (2017). Online review helpfulness: Impact of reviewer profile image. *Decision Support Systems*, 96, 39–48.
89. Yu, X., Liu, Y., Huang, X., & An, A. (2010). Mining online reviews for predicting sales performance: A case study in the movie domain. *IEEE Transactions on Knowledge and Data Engineering*, 24(4), 720–734.
90. Hong, H., & Xu, D. (2015). Research of online review helpfulness based on negative binary regress model the mediator role of reader participation. In *International conference on service systems and service management* (pp. 1–5). IEEE.
91. Filieri, R. (2015). What makes online reviews helpful? A diagnosticity-adoption framework to explain informational and normative influences in e-WOM. *Journal of Business Research*, 68(6), 1261–1270.
92. Ullah, R., Zeb, A., & Kim, W. (2015). The impact of emotions on the helpfulness of movie reviews. *Journal of Applied Research & Technology*, 13(3), 359–363.
93. Cheng, Y. H., & Ho, H. Y. (2015). Social influence’s impact on reader perceptions of online reviews. *Journal of Business Research*, 68(4), 883–887.
94. Hu, Y. H., & Chen, K. (2016). Predicting hotel review helpfulness: The impact of review visibility, and interaction between hotel stars and review ratings. *International Journal of Information Management*, 36(6), 929–944.
95. Fang, B., Ye, Q., Kucukusta, D., & Law, R. (2016). Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management*, 52, 498–506.
96. Kwok, L., & Xie, K. L. (2016). Factors contributing to the helpfulness of online hotel reviews: does manager response play a role? *International Journal of Contemporary Hospitality Management*, 28(10), 2156–2177.
97. Yin, G., Zhang, Q., & Li, Y. (2014). Effects of emotional valence and arousal on consumer perceptions of online review helpfulness. In *Twentieth Americas conference on information systems, Savannah*.
98. Li, H., Zhang, Z., Janakiraman, R., & Fang, M. (2016). How review sentiment and readability affect online peer evaluation votes?—An examination combining reviewer’s social identity and social network. In *Tourism travel and research association: Advancing tourism research globally* (p. 29).

99. Ghose, A., & Ipeirotis, P. G. (2007). Designing novel review ranking systems: Predicting the usefulness and impact of reviews. In *International conference on electronic commerce* (pp. 303–310).
100. Zhang, Y. (2014). Automatically predicting the helpfulness of online reviews. In *The fifteenth IEEE international conference on information reuse and integration* (pp. 662–668). IEEE.