What makes information in online consumer reviews diagnostic over time? The role of review relevancy, factuality, currency, source credibility and ranking score

Abstract

Online consumer reviews (OCRs) have become one of the most helpful and influential information in consumers purchase decisions. However, the proliferation of OCRs has made it difficult for consumers to orientate themselves with the wealth of reviews available. Therefore, it is paramount for online organizations to understand the determinants of perceived information diagnosticity in OCRs. In this study, we investigate consumer perceptions and we adopt the Elaboration Likelihood Model to analyze the influence of central (long, relevant, current, and factual OCRs) and peripheral cues (source credibility, overall ranking scores) on perceived information diagnosticity (PID). We consider the potential moderating effect of consumer involvement, and tested the robustness of the theoretical framework across time.

Based on two surveys carried out in 2011 and in 2016, this study demonstrates the dynamic nature of the antecedents of PID in e-WOM. We found that long reviews are not perceived as helpful, while relevant and current reviews as well as overall ranking scores are perceived as diagnostic information in both samples. The significance of the predicting power of review factuality and source credibility has evolved over time. Both central (review quality dimensions) and peripheral cues (ranking score) were found to influence PID in high-involvement decisions.

Keywords electronic word of mouth; online consumer reviews; information quality dimensions; source credibility; overall ranking score; information diagnosticity.

1. Introduction

Various online platforms (e.g. social commerce, ecommerce, online communities) are increasingly facilitating consumers in sharing their experiences, opinions, and feedback regarding people, products, services and brands in the form of online reviews, ratings, and ranking scores. Online consumer reviews, a form of electronic word of mouth (e-WOM), can be defined as any positive, neutral, or negative evaluation of a product, a service, a person, or a brand presumably posted by former customers on websites that host consumer reviews. According to a report from Mintel (2015) 81 percent of consumers aged 18-34 in the United States seek out opinions from others before purchasing a product or service. Research has established the power of online consumer reviews in predicting product sales and revenues in different product categories, such as books, beers, restaurants, movies, and hotels (e.g. Liu, 2006; Duan et al., 2005; Clemons et al., 2006; Chevalier & Mayzlin, 2006; Ye et al., 2011; Cui et al., 2012).

Different organizations are increasingly enabling consumers to leave a helpful vote to each review in an attempt to signal to consumers the most helpful reviews for assessing products and services' quality and performance. Scholars have recently started to examine what makes online review helpful by importing the data from these e-retailers (e.g. Amazon) and using the voting mechanism to measure the characteristics of the reviews that receive more helpful votes (e.g. Mudambi & Schuff, 2010; Pan & Zhang, 2011; Racherla & Friske, 2012; Baek, Ahn, & Choi, 2012; Yin, Bond, & Zhang, 2014; Jabr & Zheng, 2014; Ahmad & Laroche, 2015; Huang et al., 2015; Chua & Banerjee, 2016).

However, there is a dearth of studies on the determinants of information diagnosticity from a consumer perspective in e-WOM research (Filieri, 2015). Although existing studies are useful, they have mainly investigated the 'visible' aspects of review helpfulness focusing on textual elements such as review extremity, review sentiment, review valence, review length or

profile information of the reviewer (e.g. Mudambi & Schuff, 2010; Chua & Banerjee, 2016). Researching consumer's perception of information diagnosticity is important for several reasons: the voting mechanism can be easily manipulated (Lim et al., 2010; Pan & Zhang, 2011; Filieri, 2016); for example it is plausible to expect that given the importance that consumer reviews have on sales, managers may vote as more helpful those reviews that provide a positive rather than a negative evaluation of their business. Moreover, some important 'qualitative' information dimensions cannot be measured through quantitative textual analysis. For example, the perceived credibility of a source (i.e. the reviewer), the capacity of a review message to satisfy a consumer's specific information needs (i.e. review relevancy) or to provide plausible, fact-based information (i.e. review factuality) or up to date information (review currency). However, these factors can still be important to consumers to assess information diagnosticity.

This study applies the definition of information diagnosticity to consumer reviews and assesses consumers' perception regarding the ability of the information contained in OCRs to enable consumers to learn and to evaluate the quality and performance of services (information diagnosticity) before purchasing them. Understanding information diagnosticity is paramount for social commerce organizations because the higher the perceived diagnosticity of the information they host the better will be consumer's attitude towards shopping online (Jiang & Benbasat, 2007) and the higher will be the influence on purchase intentions (Filieri, 2015).

Additionally, most studies on e-WOM are cross-sectional and no study has measured the variations of the determinants of information diagnosticity in e-WOM over time. For instance, previous scholars have looked at the temporal evolution of e-WOM, investigating the evolution of different marketing variables and consumer posting behavior (Chen et al., 2011), while other scholars emphasized the temporal dynamics in the evolution of ratings (Godes &

3

Silva, 2012) or on how e-WOM volume evolves in movies' releases (pre-release and post-release) (Liu, 2006). In this study instead, we focus on the determinants of consumer perception of information helpfulness and how it evolves over time. We conjecture that due to the increasing importance of OCRs, the global echo produced by mass media on the phenomenon of fake reviews (e.g. Tuttle, 2012; Gartner, 2012; Smith, 2013), consumers may have become more cautious and attentive when they scrutinize the recommendations contained in websites publishing OCRs (Filieri, 2016). Thus, what makes a consumer review diagnostic may be subject to changes due to external factors (i.e. negative publicity from mass media) that might have changed consumers' attitudes towards OCRs. An analysis at different points in time can provide us with some insights into how (and if) the influence of various antecedents of information diagnosticity has changed over time.

Elaboration Likelihood Model (ELM) has been adopted in e-WOM research to explain consumer cognitive processing of product reviews and evaluation of review messages (e.g. Park, Lee, & Ahn, 2007; Park & Lee, 2008; Zhang & Watts, 2008; Lee, Park, & Ahn, 2008; Lee & Lee, 2009; Cheung et al., 2012; Filieri & McLeay, 2014). In this study, ELM (Petty & Cacioppo, 1986) has been used to investigate whether central and peripheral cues of information processing affect perceived information diagnosticity considering the potential moderating effect of consumer's involvement with a purchase. In line with the ELM, we have developed and tested a model that measures the influence of some central cues, namely length, relevancy, currency, and factuality; and of some peripheral cues of information processing, namely source credibility, and overall ranking scores; on information diagnosticity (dependent variable) considering the moderating role of consumer involvement with a purchase. The model was tested using regression analysis respectively in 2011 with 334 respondents and in 2016 with 297 respondents.

2. Theoretical Background and Hypotheses

2.1 The Elaboration Likelihood Model

We have adopted Petty and Cacioppo's (1986) Elaboration Likelihood Model (ELM) to understand the determinants of information diagnosticity in different involvement conditions. ELM postulates that consumers may take a central or a peripheral route when they process information from advertising messages (Petty, Cacioppo, & Schumann, 1983). Consumers take the central route when they are capable, highly motivated or willing to process information, spending more time and providing a rational response to advertising messages using criteria such as information quality (Petty et al., 1983). On the other hand, consumers take a peripheral route of information processing when they are less motivated or not willing or capable of processing information. In this situation, they will use information shortcuts, such as source likeability or source credibility or volume of information to make a decision (Petty et al., 1983).

Thus, the ELM has been adopted to predict consumer information adoption and purchase intentions in e-WOM research (Park et al., 2007; Zhang & Watts, 2008; Cheung, Lee, & Rabjhon, 2008; Filieri & McLeay, 2014). Argument quality, which implies the adoption of a central route of information processing, has often been conceptualized as the quality of the arguments available in a consumer review as judged by a consumer (Park et al., 2007), while source credibility (i.e. perceived expertise and trustworthiness of a reviewer) and ranking scores are information shortcuts and thus refer to a peripheral route of information processing. By adopting ELM, we expect that central cues, namely information quality dimensions, will be adopted by users in the evaluation of the quality of a service in high-involvement conditions; while the use of peripheral cues, such as source credibility and overall ranking scores, will be minimal or absent in such a context.

2.2 Central cues of information processing: Information quality dimensions

Information quality is defined as 'the quality of the content of a consumer review from the perspective of information characteristics' (Park et al. 2007, p. 128). Information quality has proved to be an important influencer of consumer attitude towards products (Lee et al. 2008); consumer purchasing intentions (Park et al., 2007; Lee & Shin, 2014), review credibility (e.g. Cheung, Sia, & Kuan, 2012), and information diagnosticity (Filieri, 2015) in e-WOM research.

Previous studies on review helpfulness have used datasets of customer reviews from e-retailers and have consistently proven that *review length* - an information quality dimension - affects the helpfulness of OCRs (Mudambi & Schuff, 2010; Pan & Zhang, 2011; Baek et al., 2012). Review length - measured in terms of words count – refers to the quantitative aspects of information in a review (e.g. Huang et al., 2015). In this study we focus on the qualitative dimensions of information quality that are likely to contribute to consumers' perceived information diagnosticity. Thus, based on information systems literature (Wang & Strong, 1996) we included the following dimensions: long review, relevant review, current review (or timeliness), and factual review. We explain each of them below.

2.2.1 Long review

In previous studies, scholars have used the number of words per review to measure review length (e.g. Mudambi & Schuff, 2010; Pan & Zhang, 2011; Huang et al., 2015) and found that the review word count has a positive impact on review helpfulness (Mudambi & Schuff, 2010; Pan & Zhang, 2011; Baek et al., 2012). Long reviews contain more words and therefore are expected to provide sufficient amount of information about a service and its features/components in order for a user to be able to evaluate it. Long reviews are more likely

to contain more details about a service being reviewed; hence they may be perceived as useful to assess service quality and performance. According to Pan and Zhang (2011) the length of a review signifies how involved the reviewer is in writing a review. It is plausible to expect that users will find more credible a reviewer who has spent more time in writing a long review about a product/service rather than a reviewer who has spent less time in writing a couple of lines about his/her experience. Following this literature we argue that long reviews can be perceived as more diagnostic information by consumer because they are more likely to contain more information of the service that a consumer is considering buying. Therefore, we hypothesize:

H1. *Long review has a positive effect on perceived information diagnosticity.*

2.2.2 Review relevancy

Review relevancy refers to the extent to which a review message is appropriate to and helpful for understanding the quality and performance of a product/service and depends on a specific customer need in a specific situation (Filieri & McLeay, 2014). Information from OCRs is perceived to be diagnostic if it contains information that is relevant to a consumer, namely it matches the product/service information a consumer is looking for. Information in OCRs will be perceived as diagnostic if it satisfies the needs of a specific customer's segment. For example, a young couple with kids will be looking for reviews from people travelling with their family members because the latter are more relevant and diagnostic to them to evaluate the quality of accommodation they are planning to book than reviews from single travelers, drifters, or adventurers. Each consumer group searches for reviews that are more likely to satisfy their information needs regarding the specific aspects/features of a service that are more important to them. Thus, we hypothesize that the higher the capacity of a review to

satisfy consumers' information needs the higher the perceived diagnosticity of the information contained in a review will be.

H2. *Review relevancy has a positive effect on perceived information diagnosticity.*

2.2.3 Review factuality

A factual review is a review that contains fact-based information, objective and logical discussion around the consumer's experience with a product or service. Conversely, an emotional review is a review that contains a subjective content, which reduces the perceived objective value of a description. Ahmad and Laroche (2015) using a dataset of reviews from Amazon found that reviews containing strong emotions (happiness, anxiety, disgust) are positively related to the number of helpful votes received.

Conversely, in this study we argue that consumers may find more diagnostic the reviews that appear to be more logical, objective, and fact-based. Consumers may find factual reviews to be diagnostic information because they provide information on specific facts or events related to experiencing a product, which increases the perceived trustworthiness of the review (Filieri, 2016). Generally, facts-based reviews objectively and rationally assess the quality and performance of a service. Thus, we believe that factual, objective, and logical reviews can be perceived as particularly diagnostic because they differ from information coming from commercial or potentially biased sources, which are often less fact-based and contain promotional language (Filieri, 2016). Thus, we hypothesize:

H3. *Review factuality has a positive effect on perceived information diagnosticity.*

2.2.4 Review currency

Information currency refers to information that is up to date, current, and represents the state of the art of a product or service (Wang & Strong, 1996). A review is rapidly available after

the contributor has submitted it to a website publishing consumer reviews. For example, TripAdvisor takes an average of two days to check review that do not break guidelines before publishing them online (TripAdvisor, 2015). Consumers with an internet connection can access the reviews of a recent experience a customer has had with a service. A recent account about a consumption experience can be more diagnostic for a consumer to assess the quality and performance of a service available in the marketplace because it provides an overview of its current state of the art or actual performance. Services may be subject to frequent changes, for example accommodation can undergo through renovation or managerial changes, which can have a huge impact on the perceived quality of the service offered over time. Therefore, we hypothesize that the more current and updated the information in OCRs is the more consumers will find those reviews to be diagnostic to assess service quality and performance. Thus, we can hypothesize:

H4. *Review currency has a positive effect on perceived information diagnosticity.*

2.3 Peripheral cues of information processing

2.3.1 Overall ranking score

The ranking score of a service is a summary statistics, which is information about average consumer evaluation, which takes into account individual positive, negative, and neutral ratings of a service (Filieri, 2015). Overall ranking score is a common feature in social commerce websites and refers to the average of all individual consumer ratings of a product/service in a specific product category (e.g. hotels in a tourism destination); thus, it does not refer to the quality of arguments rather it is an information shortcut with respect to how all reviewers have evaluated a specific organization (e.g. Hotel 'X'). Overall ranking scores are often graphically represented as star ratings.

Previous studies on the role of ranking and rating score in e-WOM have investigated: the ranking behavior of reviewers (Moe & Schweidel, 2012); how ratings change over time and sequence (Godes & Silva, 2012), the importance of ratings on retailers' perceived trustworthiness (Aiken & Boush, 2006; Benedicktus et al., 2011), the role of extremely positive and extremely negative individual rating on sales (Chevalier & Mayzlin, 2006) and on review helpfulness (Mudambi & Schuff, 2010). However, these studies have focused their attention on individual review ratings, namely the rating (or squared rating) provided by each review. The current study instead looks at the role of overall ranking score of a product in a specific category and investigates its influence on information diagnosticity in e-WOM. Following previous studies' findings (Filieri, 2015), we argue that consumers benefit from the aggregation of single-review ratings into summary statistics (i.e. overall average ranking score) because classifications of products in a category provide information about the quality and performance of service compared to main competitors. Thus, such classifications may ultimately ease consumers in their decisions making because only the best performing services will be considered. Accordingly, we hypothesize:

H5. *Overall ranking score has a positive effect on perceived information diagnosticity.*

2.3.2 Source credibility

Source credibility and trustworthiness are considered as fundamental predictors of a consumer's acceptance of a message in WOM (McGinnies & Ward, 1980). In an online environment it is often difficult to infer the credibility and the trustworthiness of a source (Chatterjee, 2001) and this is even harder in the context of OCRs, who are often generated by anonymous users that have no prior relationship with the receiver (Sen & Lerman, 2007).

Existing studies on the influence of source credibility have produced contrasting results in e-WOM research. For instance, in a study of an online food community in Hong Kong, Cheung et al. (2008) found that source expertise and trustworthiness did not influence perceived information usefulness, while Racherla & Friske (2012) using a dataset of 3.000 reviews from Yelp of products (furniture stores) and services (restaurants and beauty spas) found that reviewer's expertise is negatively correlated with information usefulness for search, experience and credence products. Similarly, Huang et al (2015) found that reviewers who write more reviews do not necessarily write more helpful reviews.

Reviewer experience is often measured in relation to the total number of reviews contributed by a reviewer (e.g. Huang et al., 2015). In this study we focus instead on the perceived credibility of the reviewer in relation to the reviewed service, which can be inferred by various ways: by simply reading his/her review or by looking at his/her profile picture (Filieri, 2016). We argue that the more a source is perceived as experienced and trustworthy, the more a consumer will be likely to find the information provided by that source to be more diagnostic to assess service quality. Therefore we hypothesize:

H6. *Source credibility has a positive effect on perceived information diagnosticity.*

2.4 Moderator variable

### 2.4.1 Involvement

Based on the ELM (Petty & Cacioppo, 1986), involvement in the processing of information is considered as a moderating variable. According to the ELM, once a consumer receives a message, he or she begins to process it and depending on the consumer's degree of involvement in a purchase, two routes can be adopted: a central route or a peripheral route (Petty & Cacioppo 1986).

In high-involvement situations consumers will spend more time and exert greater effort to carefully scrutinize the arguments contained in OCRs and they will primarily focus on the quality of the arguments. Conversely, in low-involvement conditions consumers are less motivated or less capable of thinking about a message and thus exert less cognitive effort and focus on things like brand image and source attractiveness or credibility to assess a product's quality.

Research has provided evidence that in high-involvement situations, the quality of arguments used in OCRs influences consumers' decisions, while in low involvement conditions consumers prefer to use peripheral cues or information shortcuts to evaluate a message rather than analyzing its content (Petty & Cacioppo, 1986; Park et al., 2007; Lee et al., 2008). In accordance with the ELM and previous studies' findings, we hypothesize that the higher the involvement with a purchase the higher will be the perceived diagnosticity of an online review that is long, current, factual, and relevant to consumers' information needs. Thus, we hypothesize:

H7. *The higher consumer's involvement with a purchase, the higher the influence that long, current, relevant, and factual reviews will have on perceived information diagnosticity.*
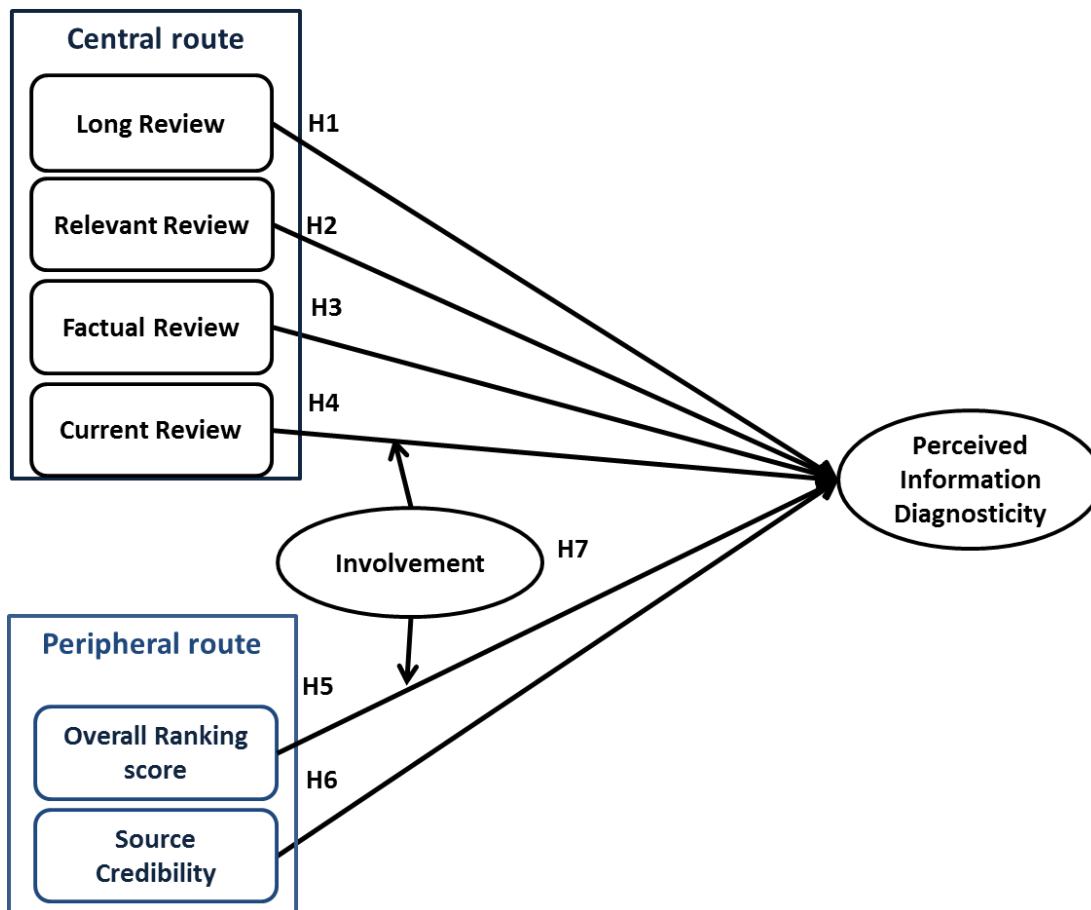
Figure 1. Framework and hypothesized relationships in this study.

3. Methodology

3.1 Data collection and sample

We collected the data for this study at two different points in time: the first survey was carried out in September 2011, while the second one took place in January 2016 with users of OCRs of accommodation. Accommodation is a classic example of service (e.g. Lovelock, 1983; Lovelock & Wirtz, 2011) and they constitute one of the most important items for which consumers search information online. We have decided to focus on services because the intangibility, variability, perishability, inseparability, and non-standardized nature of services (Zeithaml, Parasuraman, & Berry, 1985) make services more difficult to evaluate prior to purchase than goods. Thus, understanding information diagnosticity becomes even more important and influential in a service, rather than in a goods, context. Following previous

studies, we decided to focus on users of a specific website. We chose Tripadvisor.com, which represent the largest and most popular social commerce website in the travel industry.

We decided to collect data at two separate points in time with an attempt to explore whether the nature and strength of the relationship between the antecedents of information diagnosticity changes over time. In both data collections, an online questionnaire was developed using Survey Monkey. A 7-point Likert scale requiring an answer from strongly disagree (1) to strongly agree (7) was used to measure the items in the questionnaire.

Before starting to collect the data for the first study, the questionnaire was pilot-tested with a total of 3 academics experienced with survey designs and 45 users of OCRs among acquaintances and friends of one of the researchers. In both the first and second data collection, an email with a link to the online survey explaining the research and conditions for participation, was sent to a convenience sample of staff and students of a university located in Northern Europe.

With regard to ensuring that only the right participants would participate to this study, the email message clearly stated that only people with recent experience with using OCRs of accommodation could take part to the survey. In order to ensure that only individuals who had recently used OCRs would have participated to the study, some compulsory questions asked respondents to provide information about the website where they read OCRs.

In a period of two months in 2011 a total of 398 responses were received. However, 64 questionnaires were removed because they were filled too quickly or not properly, which gives a total of 334 usable questionnaires. The same process was implemented between February and March 2016, where a total of 322 responses were received. Also in this case, some questionnaires were removed (N = 25) because they were not filled properly or not complete, which gives a total of 297 usable answers.

In relation to the first and second data collection the two samples were rather homogeneous, which facilitate the comparison of findings. The sample of respondents (see Table 1) was almost equally distributed between males and females (45% males versus 55% females in the 2011 sample and 43% males versus 57% females in the 2016 sample), was primarily composed of individuals aged 18-35 (95% of participants in the 2011 sample versus 93% of participants in the 2016 sample) and who originated from European countries (86% of participants in the 2011 sample versus 83% of participants in the 2016 sample).

3.2 Measures

Source credibility was measured by a scale developed by Ohanian (1990), and used in e-WOM research by Senecal and Nantel (2004). Information diagnosticity was measured using three items derived from Jiang and Benbasat (2007). The scale for measuring overall ranking score and the items used to measure long review were developed based on the conceptualizations in previous studies (e.g. Mudambi & Schuff, 2010; Filieri, 2015). Information systems literature was useful in identifying some of the information quality items used in this research such as review currency, factuality, and relevancy (Wang & Strong, 1996).

Moderators in the model included consumer involvement with message processing, which was measured using a scale that was developed and tested by Wheeler, Petty, and Bizer (2005).

-----------------------ADD TABLE 1 HERE------------------

4. Findings

Confirmatory factor analysis (CFA) was conducted using AMOS 22 to examine the measurement validity of the constructs used in our study. The data of both the first and the second dataset show a good model fit: $\chi^2 = 385.31$; $p < 0.001$; CFI = 0.94; TLI = 0.92; NFI = 0.91; RMSEA = 0.05; and $\chi^2 = 494.78$; $p < 0.001$; CFI = 0.92; TLI = 0.91; NFI = 0.90; and RMSEA = 0.05 (Hair et al., 2010).

Additionally, both the convergent and discriminant validity in the 2011 and 2016 samples were assessed (see Appendix 1). Convergent Validity was assessed through average variance extracted (AVE). All of the constructs' AVE values were above the recommended level of 0.5 (Fornell & Larcker, 1981), thus demonstrating that the scales measure the concepts that they were designed to measure. Scale reliability was assessed for each construct with Cronbach's α, and coefficients ranged from 0.730 to 0.900 (see Appendix 1).

For discriminant validity to be supported, the AVE of each latent variable included in the model should be greater than the squared correlation estimate (Fornell & Larcker, 1981). The results presented in Appendix 1 demonstrate that these requirements have been satisfied for both samples.

The regression models were tested using hierarchical regression analysis and the statistical software SPSS 22.0. The results are presented in Tables 2, 3, and 4 and in Figure 2. Table 2 shows the descriptive statistics, including number of respondents, mean, and standard deviation in the two samples. Table 3 shows the results of the regression analysis for the two samples: information relevancy ($\beta = 0.387$; $p < 0.000$ in 2011 and $\beta = 0.247$; $p < 0.000$ in 2016) and overall ranking scores ($\beta = 0.344$; $p < 0.000$ in 2011 and $\beta = 0.190$ in 2016) are consistently the most important predictors of perceived information diagnosticity in the two samples. Review currency also resulted to be a significant predictor in both samples ($\beta = 0.215$; $p < 0.000$), but its predicting power is reduced in the last data collection ($\beta = 0.115$; $p < 0.05$).

Thus, hypotheses 2, 4 and 5 are supported in the two samples. We also found that long review do not significantly predict information diagnosticity in both samples, thus H1 is rejected. We could not find the same consistency over time in the results regarding the other predictors in our model. For example, source credibility is a significant predictor of information diagnosticity in the 2016 sample ($\beta = 0.210$; $p < 0.000$) while its predicting power was not significant in 2011 ($\beta = 0.021$; p = non-significant). Similarly, information factuality is not a significant antecedent of information diagnosticity in the 2011 sample ($\beta = -0.062$; p = non-significant) but its predicting power is significant in the 2016 sample ($\beta = 0.163$; $p < 0.05$). Thus, hypotheses 3 and 6 are supported only in the 2016 sample.

Table 4 shows the results regarding the moderation effect of involvement in the relationship between the independent variables and perceived information diagnosticity. Drawing on ELM, hypothesis 7 assumed that the higher the involvement of a user in the purchase of a product the higher the influence that a long, current, relevant, and factual review (central cues) would have on perceived information diagnosticity. Table 4 shows the results of the moderation analysis. The strength of the relationship between central cues of information processing and the dependent variable does not increase when involvement is added to the equation. Additionally, overall ranking score, considered as a peripheral cue of information processing, remains a positive and significant predictor of information diagnosticity in high involvement conditions, which contrasts with our expectations. On the other side, the influence of source credibility becomes negative when involvement is added to the equation and this result is similar across the two samples. Therefore, hypothesis 7 is not fully supported because overall ranking score, a peripheral cue of information processing, plays a significant influence in high involvement purchase decisions.

-------------------ADD TABLE 2, 3, 4 HERE -------------------

17

5. Discussion

This study has adopted the ELM model and has tested a model to investigate the influence of central and peripheral cues of information processing on perceived information diagnosticity in OCRs of accommodation. This study has tested the proposed framework using data collected at different points in time (2011-2016).

Previous studies have focused on review helpfulness and mainly used a database of consumer reviews from Amazon and 'helpful votes' left by readers to shed light on the review that are voted as most helpful (e.g. Mudambi & Schuff, 2010; Pan & Zhang, 2011; Baek et al., 2012; Huang et al., 2015). The current research instead has analyzed consumers' perception of information diagnosticity of services and has used two samples at different points in time (2011 and 2016). The findings of this study advance the literature on e-WOM and reveal important patterns and dynamics in consumer information processing in this context. Central cues of information processing adopted in this study included the following information quality dimensions: long review, factual review, current review, and relevant review; while peripheral cues included variables such as overall ranking score and source credibility. Below we discuss the main findings.

Contrary to previous studies' findings (Mudambi & Schuff, 2010; Pan & Zhang, 2011; Baek et al., 2012), we found that *a long review* is not perceived as diagnostic and this result was consistent in both our samples (2011 and 2016). This finding can be explained by the fact that millennials and centennials are attention and time-poor (PowerReviews, 2015) and as such they do not perceive lengthy, narrative reviews as diagnostic. Consumers probably pick only the information that is relevant to them instead of reading the full review. Thus, the length of a review *per se* does not make a review more diagnostic to assess the quality and performance

of accommodation. Long and narrative reviews are not perceived as helpful even in high involvement conditions. From this result it is evident that what is important to readers is not how extensive the discussion in a review is but rather how relevant, factual, and current is the content in the review. This study supports the finding of a previous study that revealed that there might be a threshold in a review's word count over which a review becomes less or not helpful at all (Huang et al., 2015).

Information relevancy was found to be the strongest predictor of perceived information diagnosticity in both samples (2011 and 2016). This finding may imply that consumers want to find relevant information when they search for consumer reviews, namely information that matches their specific needs. This result supports the argument that different reviews have a varying degree of impact on perceived information diagnosticity; such impact will depend on how relevant the information contained in a review is in providing an answer to the consumer's specific information needs.

The increasing significance of *review factuality* on information diagnosticity can be interpreted by the fact that consumers are increasingly looking for reviews that report accounts of facts and events related to their experience, so consumer reviews that display objective, logical and fact-based information tend to be perceived as diagnostic to assess the quality and performance of services by consumers. This study confirms consumers prefer fact-based, objective, and logical reviews especially in high involvement purchase decisions.

Findings show that *review currency* emerged as a predictor of information diagnosticity in both samples, however its predicting power and significance is reduced in the most recent sample (2016). This result can be explained by the fact that review websites nowadays contain many more reviews than in the past and accommodation receive several reviews every week, which implies that consumers are more likely to see and read only recent reviews than

old reviews. Therefore, consumers are provided with an increasing amount of current reviews per service, which may explain why review timeliness is less important than in the past.

Overall ranking score is summary statistics that subsumes the overall (average) evaluation of all of those who have purchased and subsequently reviewed/rated a product. This study reveals that overall ranking scores are perceived as diagnostic information by consumers who use them to assess service quality, which finding does not agree with assumptions of social cognition theorists about the fact that consumers do not sufficiently use base-rate information (i.e. summary statistics) while making judgments (Bar-Hillel, 1980; Borgida & Nisbett, 1977; Qiu et al., 2012); rather, this study found that they use aggregate ratings to learn about a service's performance and quality. Thus, the "base rate fallacy" occurring in previous studies on social cognition does not seem to apply to e-WOM contexts (Bar-Hillel, 1980). In e-WOM settings both base-rate and individuating information are available, however people use both (e.g., overall ranking scores) in making judgments even in conditions of high involvement. Although we found that the role of overall ranking score is consistently significant in both samples, its predicting power of perceived information diagnosticity has decreased over time. This finding may be attributed to the reduced reliability of overall ranking scores, which, according to many consumers, are becoming increasingly biased due to more promotional and fake reviews being posted on consumer review websites. In fact, many consumers are now aware that service providers attempt to inflate ranking scores through fake reviews so that they can appear in the top positions when consumer search for an accommodation in a specific location (Filieri, 2016). Accordingly, we found that consumers are less likely than in the past to consider the ranking score as diagnostic information to inform their product evaluations.

Interestingly, we found that the predicting power of *source credibility* has changed over time, from not being a significant predictor of information diagnosticity in the 2011 sample to

becoming a significant predictor of information diagnosticity in the last sample. This result can be explained by the fact that consumers in the past were not much motivated in assessing the credibility of a reviewer, because they assumed reviewers were credible information providers (Bickart & Schindler, 2001). However, the increasing number of scandals reported by mass media around the world about the most important review websites may have affected consumer's attitude towards them (e.g. Tuttle, 2012; Gartner, 2012; Smith, 2013). Thus, consumers are probably becoming increasingly suspicious about some reviewers and their reviews. Thus, some degree of expertise or evidence of having experienced a product/service must be shown to be considered as a 'diagnostic' information provider. Therefore, reviewers' expertise and reliability are increasingly important to assess information diagnosticity. This finding contrasts with Racherla and Friske (2012), who found that source credibility is negatively correlated with usefulness using a dataset of reviews from Yelp.

Overall, the reduced influence of overall ranking scores and the significance of the influence of review factuality and source credibility can lead us to speculate that contrarily to 10-15 years ago when consumers viewed forum opinions as highly trustworthy information source (Bickart & Schindler, 2001), today consumers seem to have a more cautious attitude towards them. This may be due to echo given by mass media on the scandals surrounding popular websites that publish consumer reviews.

This study also advances consumer behavior theory. Theoretically speaking, previous findings suggest that consumers take a central route when they process information in high-involvement situations (i.e. the quality of information) (Park et al., 2007; Lee et al., 2008). This study's findings did not show any particular influence of involvement as moderating factor between central cues, peripheral cues, and perceived information diagnosticity. In fact, the only information quality dimension that increased its predicting power in high-involvement conditions is review factuality, whereas overall ranking score, considered as a

peripheral cue, was also a significant predictor of perceived information diagnosticity. We can speculate that when involvement increases, consumers may not only rely on the quality of arguments of few reviews and ranking score from a single source but rather they will seek for additional information sources (e.g. friends, comparing reviews and rating scores with other websites) to evaluate service quality and performance. The ELM presents these two routes (central and peripheral) as alternatives (Bitner & Obermiller, 1985); however, these findings show that the two routes can occur together and both central and peripheral cues can be suitable to influence consumers' evaluations in high involvement conditions. Someone could argue that consumer reviews are different than information shared in inter-personal and advertising communications (Petty & Cacioppo, 1986). In the e-WOM context, consumers actively search for information they need about the products and services they want on their own volition. It is plausible to expect that in this context consumers are *per se* willing and motivated to source and read reviews and other types of visual information shortcuts like overall ranking score to assess service quality. This result implies the possibility that consumers will be adopting and considering both central and peripheral information cues when evaluating the quality and performance of a service because they are both willing and motivated to do so. Our results may also be interpreted in the following way: when involvement increases, consumers are willing to spend more time and use several information sources (or channels) to make up their mind about the quality of different alternatives. Thus, we can conclude by saying that in e-WOM communications there is support for the central and the peripheral routes to be considered as complementary ways of information processing rather than as alternatives.

6. Managerial implications

Online organizations still struggle to understand what types of information are more important to consumers when they retrieve information from their website. This study provides some insights into the nature of diagnostic information to consumers' eyes.

The importance of information relevancy emerged in this study, which suggests that social commerce organizations should refine their information search criteria (e.g. product's style, price, availability, friends' recommendations, and the like) to facilitate consumers' retrieval of the information that they want to find. For example, different consumer segments may require or seek information about some specific features of a service instead of others. Thus, social media commerce organizations could help them by providing more information about the target customer of a specific service by adapting their review submission forms (e.g., asking reviewers to indicate the potential target group of the reviewed service). By doing so, social commerce organizations can increase the perceived diagnosticity of the information they provide.

This study's findings also imply that social commerce organizations should consider adopting a wider range of peripheral cues (e.g. crowd opinions) in order to ease consumer's products and services evaluation and ultimately their decision making. The study has found that consumers appreciate summary statistics regarding all other customers' evaluations to learn about service quality and performance, even in high involvement conditions. Therefore, e-retailers selling high and low-involvement products should provide more of summary statistics on their website.

Considering that the expertise of a reviewer was found to be a significant predictor of perceived information diagnosticity, we suggest social commerce organizations to provide more of profile information to facilitate consumers in the assessment of a reviewer's credibility. Thus, a reviewer's webpage profile should provide information about a reviewer's expertise of a specific product or service. By doing so, a user can rapidly find out if and how

expert a reviewer is in a specific product category. We recommend placing profile information as close as possible to the posted review so that consumers can easily and rapidly make inferences on a reviewer's credibility.

Finally, consumers have become more vigilant and suspicious towards consumer reviews than in the past (Filieri, 2016). Therefore, we recommend websites that publish consumer reviews and that want to improve the diagnosticity of information by investing in fraud management software or in engaging the users of the website to flag out suspicious reviews.

7. Limitations and future research

The present study has some limitations. First, the sample was composed mainly by European respondents, which may hinder the generalizability of this study's findings to other geographical contexts. Further research in other countries with a more diverse sample is advised.

Another limitation of this study is that it focuses on travel services only (accommodation). However studies focusing on goods or on different service types (e.g. financial service providers) may obtain different results. Therefore, to extend the comparability and generalizability of this study's findings, scholars should use different services or goods.

Existing research on e-WOM helpfulness is mainly based on quantitative studies often relying on databases of reviews downloaded from a specific website (e.g. Amazon, Yelp). These studies are valuable in that they provide significant findings, however consumer perceptions can reveal the underside of the iceberg, telling how consumers perceive and evaluate information diagnosticity in OCRs. We therefore believe that additional studies using experimental methods on consumer perceptions are advisable, as the measurable textual

aspect of a review may not enable to fully grasp consumer perceptions of the information shared through OCRs.

Moreover, this study has found that long reviews are not perceived as diagnostic information, while relevant and factual reviews are particularly diagnostic to consumers to assess service quality. Based on this result, future research could investigate more information quality dimensions and/or test the moderation effect of information quality dimensions in the relationship between review length and review diagnosticity. For example, researchers could test whether long reviews containing specific types of information (e.g. relevant information) will be perceived as more helpful than long reviews alone. Moreover, research on information quality dimensions in OCRs is still scant.

Furthermore, research in e-WOM is still lacking qualitative and longitudinal studies. This study's findings provide evidence of the dynamic nature of information cues that consumers adopt to diagnose the quality and performance of services in e-WOM communications. We believe that mixed-method studies can provide an in-depth understanding of consumer perceptions of information and review diagnosticity. Preliminary qualitative research could also be useful to identify new predictors of information diagnosticity, which can inform subsequent data acquisition and measurements.

Finally, we believe that demographic and cultural differences in consumers' information processing of OCRs should be explored. For instance, it is plausible to expect that consumers from different cultures and countries seek, process, and evaluate information from online reviews differently. At present, the literature lacks of a cross-cultural study examining how users of reviews from different countries evaluate the diagnosticity of online reviews. However, understanding these differences could help system developers to identify and display diagnostic information in the best manner possible for organizations operating in different countries.

References

Aiken, D.K., & Boush, D.M. (2006). Trustmarks, objective-source ratings, and implied investments in advertising: investigating online trust and the context-specific nature of internet signals. *Journal of the Academy of Marketing Science*, *34*(3), 308-323.

Ahmad, S. N., & Laroche, M. (2015). How do Expressed Emotions Affect the Helpfulness of a Product Review? Evidence from Reviews Using Latent Semantic Analysis. *International Journal of Electronic Commerce*, *20*(1), 76-111.

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.

Bar-Hillel, M. (1980). The base rate-fallacy in probability judgements. *Acta Psychologica*, *44*(3), 211-233.

Benedicktus, R.L. (2011). The effects of 3rd party consensus information on service expectations and online trust. *Journal of Business Research*, *64*(8), 846-853.

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

Bitner, M.J., & Obermiller, C. (1985). The elaboration likelihood model: Limitations and extensions in marketing. *Advances in Consumer Research, 12*(1), 420-425.

Borgida, E., & Nisbett, R.E. (1977). The differential impact of abstract vs. concrete information on decisions. *Journal of Applied Social Psychology*, *7*(3), 258–271.

Chatterjee, P. (2001). Online reviews: Do consumers use them? *Advances in Consumer Research*, *28*(1), 129-133.

Chen, Y., Fay, S., & Wang, Q. (2011). The Role of Marketing in Social Media: How Online Consumer Reviews Evolve. *Journal of Interactive Marketing*, *25*(2), 85-94.

Chevalier, J., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, *43*(August), 345-54.

Cheung, C.M.K., Lee, M.K.O., & Rabjhon, N. (2008). The impact of e-WOM – the adoption of online opinions in online customer communities. *Internet Research*, *18*(3), 229-247.

Cheung, M. Y., Sia, C. L., & Kuan, K. K. (2012). Is this review believable? A study of factors affecting the credibility of online consumer reviews from an ELM perspective. *Journal of the Association for Information Systems*, *13*(8), 618-635.

Chua, A. Y., & Banerjee, S. (2016). Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality. *Computers in Human Behavior*, *54*, 547-554.

Churchill, G.A. Jr. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, *16*(1), 64-73.

Clemons, E., Gao, G., & Hitt, L. (2006). When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of Management Information Systems*, *23*(2), 149-171.

Cui, G., Lui, H. K., & Guo, X. (2012). The Effect of Online Consumer Reviews on New Product Sales. *International Journal of Electronic Commerce*, *17*(1), 39-58.

Duan, W., Bin, G., & Whinston, A.B. (2005). Do online reviews matter? – An empirical investigation of panel data. *Decision Support Systems*, *45*(4), 1007-1016.

Filieri, R., & McLeay, F. (2014). E-WOM and accommodation: An analysis of the factors that influence travelers' adoption of information from online reviews. *Journal of Travel Research*, *53*(1), 44-57.

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.

Filieri, R. (2016). What makes an online consumer review trustworthy? *Annals of Tourism Research*, *58*, 46-64.

Fornell, C., & Larcker, D.F. (1981). Evaluating structural equation models with unobserved variables and measurement error. *Journal of Marketing Research*, *18*(1), 39-50.

Gartner (2012). *Gartner Says By 2014, 10-15 Percent of Social Media Reviews to Be Promotional, Paid for By Companies* [Press release]. , September 17, 2012. Retrieved from http://www.gartner.com/newsroom/id/2161315.

Godes, D., & Silva, J. C. (2012). Sequential and temporal dynamics of online opinion, *Marketing Science*, *31*(3), 448-473.

Hair, J. F., Anderson, R. E., Babin, B. J., & Black, W. C. (2010). *Multivariate data analysis: A global perspective* (Vol. 7). Upper Saddle River, NJ: Pearson.

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*, 17-27.

Jabr, W. & Zheng, Z. (2014). Know Yourself and Know Your Enemy: An Analysis of Firm Recommendations and Consumer Reviews in a Competitive Environment. *MIS Quarterly*, *38*(3), 635-654.

Jiang, Z., & Benbasat, I. (2007). The effects of presentation formats and task complexity on online consumers' product understanding. *MIS Quarterly*, *31*(3), 475-500.

Lee, J., Park, D.H., & Han, I. (2008). The effect of negative online consumer reviews on product attitude: An information processing view. *Electronic Commerce Research and Applications*, *7*(3), 341-52.

Lee, K., & Lee, J.M. (2009). Understanding the product information inference process in electronic word-of-mouth: an objectivity-subjectivity dichotomy perspective, *Information & Management, 46*(5), 302-11.

Lee, E. J., & Shin, S. Y. (2014). When do consumers buy online product reviews? Effects of review quality, product type, and reviewer's photo. *Computers in Human Behavior*, *31*, 356-366.

Lim, E. P., Nguyen, V.-A., Jindal, N. Liu, B. & Lauw, H. W. (2010). Detecting Product Review Spammers Using Rating Behaviors. *Proceedings of the 19th ACM International Conference on Information and Knowledge Management,* 939-48.

Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing*, *70*(3), 74-89.

Lovelock, C. H. (1983). Classifying services to gain strategic marketing insights. *Journal of Marketing*, *47*(3), 9-20.

Lovelock, C. H., & Wirtz, J. (2011). *Services Marketing. People, Technology, Strategy*, Seventh edition, Pearson Education.

McGinnies, E. & Ward, C. D. (1980). Better liked than right: Trustworthiness and expertise as factors in credibility. *Personality and Social Psychology Bulletin*, *6*(3), 467-472.

Moe, W. W., & Schweidel, D. A. (2012). Online product opinions: Incidence, evaluation, and evolution. *Marketing Science*, *31*(3), 372-386.

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.

Nunnally, J.C. (1978). *Psychometric theory*. 2nd ed., New York: McGraw-Hill.

Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising*, *19*(3), 39-52.

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.

Park, D.-H., Lee, J., & Han, I. (2007). The effect of online consumer reviews on consumer purchasing intention: The moderating role of involvement. *International Journal of*

Electronic Commerce, *11*(4), 125-148.

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

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

Petty, R.E., & Cacioppo, J.T. (1986). *Communication and Persuasion: Central and Peripheral Routes to Persuasion*. New York: Springer-Verlag.

Petty, R. E., Cacioppo, J. T., & Schumann, D. (1983). Central and Peripheral Routes to Advertising Effectiveness: The Moderating Role of Involvement. *Journal of Consumer Research*, *10*(2), 135-46.

PowerReviews (2015). *2015 Centennial Shopper Study*. Available at: https://www.powerreviews.com/wp-content/uploads/2015/12/Centennials_US.pdf.

Qiu, L., Pang, J., & Lim, K. H. (2012). Effects of Conflicting Aggregated Rating on eWOM Review Credibility and Diagnosticity: The Moderating Role of Review Valence. *Decision Support Systems*, *54*(1), 631-43.

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), 548-59.

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

Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of retailing*, *80*(2), 159-169.

Smith, O. (2013). TripAdvisor fails to spot fake restaurant, The Telegraph, 25 Jul 2013, available at http://www.telegraph.co.uk/travel/travelnews/10201754/TripAdvisor-fails-to-spot-fake-restaurant.html.

Tuttle, B. (2012). Why you shouldn't trust positive online reviews—or negative ones, for that matter, *Time*, Aug 28 2012, accessible at: http://business.time.com/2012/08/28/why-you-shouldnt-trust-positive-online-reviews-or-negative-ones-for-that-matter/

Wang, R.Y., & Strong, D.M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5-34.

Wheeler, S. C., Petty, R.E., & Bizer, G.Y. (2005). Self-Schema Matching and Attitude Change: Situational and Dispositional Determinants of Message Elaboration, *Journal of Consumer Research*, *31*(4), 787-797.

Ye, Q., Law, R., Gu, B., & Chen, W. (2011). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Computers in Human Behavior*, *27*(2), 634-639.

Yin, D., Bond, S. & Zhang, H. (2014). Anxious or Angry? Effects of Discrete Emotions on the Perceived Helpfulness of Online Reviews. *MIS Quarterly, 38*(2), 539-60.

Zhang, W. & Watts, S. A. (2008). Capitalizing on Content: Information Adoption in Two Online Communities. *Journal of the Association for Information Systems*, *9*(2), 73-94.

Zeithaml, V.A., Parasuraman, & Berry, L. L. (1985). Problems and Strategies in Services Marketing, *Journal of Marketing*, *49*(Spring), 33–46.

Table 1.

Sociodemographic characteristics of the samples.

		2011		2016	
Dimensions		**%**	**N**	**%**	**N**
Gender	Female	55	185	57	169
	Male	45	149	43	128
Age	18 – 25	80	268	77	229
	26 – 35	15	50	16	47
	36 – 45	3	10	5	15
	>46	2	6	2	6
Economic Status	50.000 € and above	3	11	3	9
	30.000 - 49.000 €	5	18	7	21
	10.000 - 29.000 €	16	56	15	44
	Under 9.999 €	76	249	75	223
Nationality	European	86	288	83	247
	Asian	5	16	14	41
	American	4	12	1	3
	Others	5	18	2	6

Table 2.

Descriptive Statistics

Variable	**2011** **N = 334** **Mean** (*Std.*)	**2016** **N = 297** **Mean** (*Std.*)
Long Review	4.8 (1.120)	4.3 (1.3231)
Review Relevancy	5.6 (1.112)	6.0 (0.9371)
Review Factuality	5.1 (1.054)	6.1 (0.9123)
Review Currency	5.5 (1.143)	6.2 (0.9079)
Source Credibility	5.0 (1.091)	5.3 (1.196)
Overall Ranking	5.7	5.5

Score	(1.097)	(1.010)
Information diagnosticity	5.7 (0.979)	5.9 (0.7550)
Involvement	5.1 (1.339)	5.6 (1.089)

Table 3.

Regression analysis. Output for the 2011 and the 2016 samples

	Model 1 (2011 sample)	Model 1 (2016 sample)	Supported *vs.* Not supported
Long Review	-.061 (-.714) (.061)	-.001 (-.020) (.029)	H1: Rejected in the two samples
Review Relevancy	.387*** (6.057) (.062)	.247*** (3.939) (.051)	H2: Supported in the two samples
Review Factuality	-.062 (-.805) (.045)	.163* (2.581) (.052)	H3: Supported in the 2016 sample but not in the 2011 sample – increased significance
Review Currency	.215** (3.434) (.062)	.115* (1.124) (.053)	H4: Supported in the two samples - reduced significance
Overall Ranking Scores	.357*** (6.064) (.049)	.190*** (3.539) (.040)	H5: Supported in the two samples – reduced predicting power
Source Credibility	.021 (.369) (.053)	.210*** (4.164) (.052)	H6 Supported in the 2016 sample but not in the 2011 sample – increased significance
R^2	.619	.326	
Adjusted R^2	.626	.316	
F	86.941***	35.150***	

Note: Standardized Beta coefficient, Significance, t-value and standard error in parenthesis.
*** $p = .001$; ** $p = .005$; * $p = .05$

Table 4.

Regression analysis output. Moderating role of involvement

	Model 2011 sample x Involvement	**Model 2016 sample x Involvement**
Long Review	-.029	.004
	(-.546)	(.078)
	(.026)	(.029)
Review Relevancy	.301***	.202**
	(6.036)	(2.732)
	(.044)	(.052)
Review Factuality	-.033	.159*
	(-.638)	(2.322)
	(.023)	(.052)
Review Currency	.204***	.105*
	(4.177)	(1.326)
	(.043)	(.039)
Ranking Scores	.325***	.165**
	(6.588)	(3.038)
	(.046)	(.040)
Source Credibility	.066	.189**
	(1.411)	(3.704)
	(.047)	(.031)
Involvement	.106*	.112*
	(2.389)	(2.078)
	(.033)	(.037)
Long x involvement	.045	.054
	.959	.945
	(.058)	(.041)
Relevancy x involvement	.045	-.016
	.959	-.186
	(.058)	(.034)
Factuality x	-.033	.274***

involvement	-.705	(1.927)
	(-.043)	(.032)
Currency x	-.035	.137
involvement	-.778	.950
	(-.047)	(.055)
Ranking	.010	.172*
Scores x	(.199)	(1.913)
involvement	(.012)	(.48)
Source	-.090*	-.145**
Credibility x	(-2.177)	(-2.993)
involvement	(.033)	(.029)
R^2	.561	.370
Adjusted R^2	.555	.341
F	88.268***	29.007***

Note: Standardized Beta coefficient, Significance in *, t-value and standard error in parenthesis.

*** $p = .001$; ** $p = .005$; * $p = .05$

2011 sample

Items, Factor Loadings, Cronbach's alpha

Construct	Items	Factor Loadings	α
Long Review	Long	0.840	0.785
	Narrative	0.872	
Review Relevancy	Appropriate	0.896	0.752
	Relevant	0.896	
Review Factuality	Based on facts	0.833	0.797
	Logical	0.876	
	Objective	0.824	
Review Currency	Current	0.918	0.813
	Updated	0.918	
Overall Ranking Score	Has reduced the number of alternative services that I was considering buying	0.849	0.900
		0.876	
	Has helped me to rapidly identify the best services	0.873	
		0.843	
	Has facilitated my purchase decision		
	Has enabled me to identify the service that could satisfy my needs		
Source Credibility	The reviewers were credible	0.839	0.861
	The reviewers were experienced	0.749	
	The reviewers were trustworthy	0.898	
	The reviewers were reliable	0.877	
Information Diagnosticity	The information provided in online reviews was helpful for me to evaluate the service	0.855	0.847
		0.894	
	The information provided in online reviews was helpful in familiarizing me	0.875	

Construct	Items	Factor Loadings	α
	with the service		
	The information provided in online reviews was helpful for me to understand the performance of the service		
Involvement	How much effort did you put into evaluating the given information?	0.873	0.872
	Did you think deeply about the information contained in online reviews?	0.915	
	How personally involved did you feel with the issue you read about?	0.889	

2016 sample

Items, Factor Loadings, Cronbach's alpha

Construct	Items	Factor Loadings	α
Long review	Long	0.780	0.813
	Narrative	0.860	
Review Relevancy	Appropriate	0.805	0.807
	Relevant	0.717	
Review Factuality	Based on facts	0.890	0.730
	Logical	0.837	
	Objective	0.774	
Review Currency	Current	0.713	0.773
	Updated	0.716	

Overall Ranking Score	Reduced the number of alternative services that I was considering buying	0.800	0.769
		0.746	
	Has helped me to rapidly identify the best services	0.812	
		0.799	
	Has facilitated my purchase decision		
	Has enabled me to identify the service that could satisfy my needs		
Source Credibility	The reviewer was credible	0.803	0.789
	The reviewer was experienced	0.800	
	The reviewer was trustworthy	0.791	
	The reviewer was reliable	0.792	
Information Diagnosticity	The information provided in online reviews was helpful for me to evaluate the service	0.750	0.848
		0.751	
		0.770	
	The information provided in online reviews was helpful in familiarizing me with the service		
	The information provided in online reviews was helpful for me to understand the performance of the service		
Involvement	How much effort did you put into evaluating the given information?	0.911	0.881
		0.890	
	Did you think deeply about the information contained in online reviews?	0.900	
	How personally involved did you feel with the issue you read about?		

2011 Sample

Correlations and average variance extracted

Variable	1	2	3	4	5	6	7	8

Variable	1	2	3	4	5	6	7	8
1. Long	**0.699**	-	-	-	-	-	-	-
2. Factuality	0.436	**0.713**	-	-	-	-	-	-
3. Relevancy	0.526	0.609	**0.803**	-	-	-	-	-
4. Currency	0.579	0.451	0.518	**0.843**	-	-	-	-
5. Ranking	0.437	0.542	0.485	0.489	**0.719**	-	-	-
6. S. Credibility	0.334	0.511	0.383	0.382	0.416	**0.635**	-	-
7. Diagnosticity	0.581	0.650	0.669	0.562	0.628	0.467	**0.765**	-
8. Involvement	0.307	0.386	0.376	0.267	0.375	0.374	0.422	**0.797**

Note. All correlations were significant at $p = < 0.001$.

2016 Sample

Correlations and average variance extracted

Variable	1	2	3	4	5	6	7	8
1. Long	**0.674**	-	-	-	-	-	-	-
2. Factuality	0.256	**0.697**	-	-	-	-	-	-
3. Relevancy	0.144*	0.627	**0.581**	-	-	-	-	-
4. Currency	0.193	0.542	0.498	**0.510**	-	-	-	-
5. Ranking	0.194	0.354	0.338	0.369	**0.623**	-	-	-
6. S. Credibility	0.234	0.164*	0.191	0.191	0.285	**0.634**	-	-
7. Diagnosticity	0.162*	0.420	0.454	0.351	0.391	0.338	**0. 573**	-
8. Involvement	0.155*	0.387	0.465	0.395	0.224	0.125	0.326	**0.810**

Note. All correlations were significant at $p = < 0.001$ with the exception of those marked with * which are significant at $p = < 0.005$.