Marketing Science Institute Working Paper Series 2018
Report No. 18-119

Making Words Speak: Leveraging Consumer Insights from Online Review Text to Improve Service Quality

Andrea Ordanini, Raji Srinivasan, and Anastasia Nanni

Report Summary

How do managers make sense of online reviews? Although numeric online ratings provide summary consumer feedback, they are not highly diagnostic (e.g., there is often a j-shaped distribution). Review text, which includes details of consumers' experiences, may offer deeper insights but such text is unstructured, with contextually-driven meanings, and thus is challenging to exploit.

In this report, Andrea Ordanini, Raji Srinivasan, and Anastasia Nanni examine whether managers can leverage consumer insights from text mining analysis of online review text to improve their firm's performance.

They examine whether managerial use of online review text analytics, derived from extraction and visual representation of insights in review text, affects the service quality of their firm's offering. They focus on service quality as online consumer reviews are critical in consumers' purchase decisions in many service sectors (e.g., hotels, airlines, restaurants, and retailing).

They also examine the heterogeneous effects of managers' use of online review text analytics on service quality based on two contextual characteristics of the managers' environment: managerial accountability and prior firm performance.

They use a longitudinal randomized control trial (RCT) in a field setting in 135 Italian hotels to provide causal evidence for effects. They conducted the RCT over a period of eight months, using monthly average online ratings from TripAdvisor for June-August 2015 and June-August 2016 as measures of pre- and post-treatment service quality of the hotel respectively.

Results indicate that the managerial use of online review text analytics improved service quality by 5.4% during the treatment window. The effect size was not trivial and was robust to several alternative assumptions. Additional analysis indicated that managers who used online review text analytics were more likely to identify points of weakness in their offering and take actions to address those.

Importantly, the positive effect of review text analytics disappears when managers do not feel accountable for their actions or when managers are satisfied with the existing level of performance. Thus, decisional stimuli and organizational characteristics are important boundary conditions for the positive effect.

Andrea Ordanini is BNP Paribas Professor of Marketing and Service Analytics, Bocconi University. Raji Srinivasan is Sam Barshop Centennial Professor of Marketing Administration, Red McCombs School of Business, University of Texas at Austin. Anastasia Nanni is a doctoral student in Marketing, Bocconi University.

Online reviews which are central to consumers' decision-making secure businesses' visibility in organic online search rankings, drive traffic to websites, and play a key role on firm performance. For example, the demand and prices of hotels are positively correlated with their average online ratings (Lewis and Zervas 2016). As Hu, Liu, and Zhang (2008, p. 201) note "online product reviews provided by consumers who previously purchased products have become a major information source for consumers and *marketers* regarding product quality".

While there are several insights in the literature on the effects of online reviews on consumers' decision-making (see You, Vadakkepatt, and Joshi 2015 for a comprehensive review of the literature on electronic word-of-mouth), we know little about whether managers can use the feedback in online reviews to improve their firm's performance, the issue that we focus on in this research.

Most online (numeric) ratings, typically on a scale of 1-5 stars, display a ceiling effect and have limited variance (Hu, Pavlou, and Zhang 2009) with a disproportionate number of 4 or 5 star ratings. Further, the text of online consumer reviews of some 4-star (online) ratings may read like reviews of 3-star ratings whereas others may read like those of 5-star ratings, which makes online numeric ratings only weakly diagnostic. In contrast, the text of online consumer reviews which provide nuanced descriptions of consumers' positive and negative product experiences can be more diagnostic for consumers and managers alike.

From a managerial perspective, while managers are skilled with numerical data, such as online ratings, which can be summarized and analyzed, unfortunately, they have little experience with unstructured review text which have contextually-driven meanings. Further, although online review text can be read and also, manually coded, managers may not make reliable inferences because of their subjective biases in working with qualitative data.

One approach to analyzing textual data is text mining, extracting and representing insights obtained from the unstructured text. Text mining analyzes the words in the corpus of text to identify key concepts and the inter-relationships among the concepts (Blei 2012). Text mining applications of online reviews and chats in the marketing literature have been used to identify underlying dimensions of product quality (Tirunillai and Tellis 2014); and predict online ratings (Büschken and Allenby 2016), product sales, and stock prices (Xiong and Bharadwaj 2014). Overlooked in the literature is whether managers can leverage consumer insights from text mining analysis of online review text to improve their firm's performance. It is not, a priori, clear, whether online review text analytics will improve managerial decision making. As Lurie and Mason (2007, p. 160) note "there has been little systematic analysis of the implications of these tools [visualization of information with text mining of textual data] for decision making."

We examine the effects of managerial use of online review text analytics on the service quality of their firm's offering. We focus on service quality, as online consumer reviews are critical in consumers' purchase decisions in many service sectors (e.g., hotels, airlines, restaurants, and retailing). We also examine the heterogeneous effects of managers' use of online review text analytics on service quality based on two contextual characteristics of the managers' environment, managerial accountability for their actions and prior firm performance. The Italian hotel industry serves as the context for the empirical testing.

We integrate two distinctive features of text mining analysis with developments in information processing theory to motivate the online review text analytics that we use in this research. First, two distinctive features of text mining analysis are: knowledge extraction, i.e., the discovery of key concepts which give sense to unstructured text (Kodratoff 1999) and knowledge visualization, i.e., the visualization of these concepts which facilitate comprehension and learning (Seifert et al. 2014). Second, information processing theory suggests that individuals can more effectively process unstructured information, as those in online review text, if there is visualization of the information (Berlyne 1960; Fiske and Taylor 1984), redundancy in its presentation (Feustal, Shiffrin, and Salasso 1983), and dual coding of the information as both visuals and text (Paivio 1991). Thus, we conceptualize online review text analytics to include 1) a visual representation of key concepts (i.e. service attributes) in the online review text as tag clouds of words i.e. green and red tag clouds with related positive and negative sentiment respectively and 2) a summary table of the ratings of key attributes extracted from the online review text.

Applying the Motivation-Opportunity-Ability (MOA) framework (e.g., MacInnis, Moorman, and Jaworski 1991), we hypothesize that online review text analytics provide an opportunity to managers to make sense of extracted consumer insights and motivate and enable them to translate the consumer insights to improve the service quality of their firm's offering. Further, as managers' actions are affected by the contextual characteristics of their decision-making environment (Heath, Larrick, and Klayman 1998; Powell, Lovallo, and Fox 2011), we hypothesize two heterogeneous effects of the managerial use of online review text analytics, managerial accountability for their actions (Tetlock 1983) and prior firm performance (Greve 1998) on the service quality of firm's offering.

To ensure causal identification of the effect of managerial use of online review text analytics on the hotel's service quality, we use a randomized control trial (RCT) in a field setting in hotels in Italy. In January 2016, we enlisted the assistance of Federalberghi, the Italian hotel industry association, to conduct a RCT in 135 Italian hotels who agreed to participate. Managers in the control condition received a report with their hotel's unprocessed online consumer reviews for the previous year from Tripadvisor, Italy. Managers in the treatment condition received a report that included the online review text analytics (i.e., our stimulus) and unprocessed online consumer reviews. Managers in both the treatment and control conditions were instructed to pay attention to the online reviews and take actions needed to improve their hotel's service quality.

Since we are not in the condition to ensure that all managers in the treatment condition effectively see and read the report we provide, our analysis is more correctly defined as 'intention-to-treat' (Newel 1992), although we then employ a mechanism to identify potential drop-outs (see below). For the sake of simplicity, however, we use the term 'treatment' in the rest of the paper, but each time it should be intended as 'intention-to-treat'.

We conducted the RCT over a period of eight months. The monthly average online ratings from TripAdvisor for June-August 2015 and June-August 2016 are the measures of pre- and post-treatment service quality of the hotel respectively.

Difference-in-difference and mixed model analyses indicate that, as expected, the managerial use of online review text analytics improves the hotel's service quality (+5.4%) in the treatment condition (vs. control condition). Additional analyses indicate no threats to randomization of hotels to the two conditions. The findings are robust to alternative measures, model specifications, and samples.

Consistent with our hypotheses, this effect is heterogeneous; when managers are not accountable for their actions or when the hotel's prior performance is satisfactory, managerial use of online review text analytics has no effect on its service quality. Additional analysis indicates that managers who use online review text analytics were more likely to identify the main weakness of their offering and take actions to address it.

The paper's findings make several theoretical contributions. First, to the best of our knowledge, this is the first study to provide evidence of the managerial use of consumer insights from online consumer review text in improving firm outcomes. Extending current work on online reviews, we show yet another way, managerial use of online review text analytics, by which online reviews can improve firm performance. In doing so, the findings extend the literature on electronic word-of-mouth, which has focused on consumer-level effects of numeric online ratings, to the managerial-level and to review text. Second, we update the extant literature on managerial use of market intelligence developed in the 1980s and 1990s to the present time, when vast consumer insights are being generated in online review text. Third, the findings on the heterogeneous effects of the managerial use of review text analytics, based on managerial accountability and prior firm performance, identify two boundary conditions for benefits from the managerial use of online review text analytics.

For managerial practice, the findings indicate that the use of simple online review text analytics can help improve product quality. The findings of the heterogeneous effects suggest two actionable guidelines to motivate and enable managers to leverage online review text analytics to improve service quality: 1) managers who are entrusted with managing their online consumer reviews must be held accountable for their actions and 2) if prior firm performance is satisfactory, then additional inducements (e.g., bonuses) may be useful to motivate them.

The rest of the paper is organized as follows. We first develop theory and hypotheses of the main and heterogeneous effects of managerial use of online review text analytics on service quality. Following that, we describe the RCT that we use to test the hypotheses and results. We conclude with a discussion of the findings' contributions to marketing theory, managerial implications, and limitations and opportunities for further research.

Hypotheses

We apply two distinctive features of text mining analysis with key developments in the information processing literature on the processing of complex, unstructured information to develop the hypotheses. The two distinctive features of text mining analysis are 1) knowledge extraction, the discovery of key concepts in the unstructured text (i.e. online review text) (Kodratoff 1999) and 2) knowledge visualization (Seifert et al. 2014), the representation of the key concepts in the text to facilitate comprehension and learning. These two features affects the "depth of field" of the decision maker (Lurie and Mason 2007), that is her/his capability to simultaneously consider both contextual overview and detailed information when accessing and evaluating new information.

Developments in information processing theory also suggests that individuals can process unstructured information (as in online review text) and improve their comprehension and learning of such information when there is visualization of the information (Berlyne 1960; Fiske and Taylor 1984), redundancy in its presentation (Feustal, Shiffrin, and Salasso 1983) and dual coding of the information as both visuals and text (Paivio 1991).

Accordingly, we conceptualize text-mined online review text analytics to include a visual representation in the form of tag clouds of the key concepts in the online review text (i.e. green and red tag clouds with related positive and negative sentiment respectively) and a summary table of the ratings of the service attributes extracted from the online review text.

Main Effect

Applying the Motivation-Opportunity-Ability (MOA) theory (e.g., MacInnis, Moorman, and Jaworski 1991), we propose that the managerial use of online review text analytics will provide an opportunity to managers make sense of the consumer insights extracted from the online review text and motivate and enable them to translate the consumer insights into improved service quality.

Because of the knowledge extraction feature of text mining analysis (Blei 2012), online review text analytics includes aggregate information on key concepts extracted from consumers' online review text, across multiple consumers. The availability of hitherto unknown insights provides managers accounts of consumers' experiences that enable corrective actions that may be needed to improve consumers' experiences (Maitlis and Christianson 2014). An aggregate summary of consumer insights implies that multiple consumers have similar concerns and that they are not idiosyncratic based on the experiences of a few consumers. In contrast, when managers process disaggregated online reviews of multiple consumers, it may be effortful, impeding their motivation and ability. Moreover, as each online consumer review is a disaggregate data point, managers may discount it, further reducing their motivation to process raw disaggregate online consumer reviews.

Second, knowledge extraction includes sentiment analysis of online review text which summarizes the valence (positive and negative) of consumer feedback in the online reviews. This vivid simultaneous presentation of positive and negative consumer feedback may be considered to be analogous to the presentation of two-sided (negative and positive) information in advertising, which has been shown to be both credible and persuasive (Crowley and Hoyer 1994; Pechmann 1992). Indeed, there is some evidence at the consumer-level that sentiment analysis of online reviews with positive and negative sentiment is persuasive and diagnostic (Archak, Ghose, and Ipeirotis 2011). Hence, managers using online review text analytics may find them to be persuasive, increasing their motivation and ability to process them.

Third, visual representation of complex information enhances attention to it (Berlyne 1960; Fiske and Taylor 1984). Visually prominent attribute information, for example, in an advertisement's headline and copy, is persuasive and increasing its retention (Gardner 1983). Moreover, dual coding theory (Paivio 1991) proposes that unlike verbal information which is sequentially processed, visual information is simultaneously processed and encoded in memory, as both images and verbal traces, again improving both learning and retention.

Hence, visual representation of key concepts with associated positive and negative sentiment as red and green tag clouds respectively (Seifert et al. 2014) and the summary table of key attributes in online review text analytics should enable managers to effectively process and retain consumer insights.

Taken together, these ideas suggest that online reviews text analytics can simultaneously provide both overviews and detailed views of customer feedback to managers, who can in turn improve their overall understanding of content and attenuate the trade-off between accuracy and effort that occurs in decision making involving complex information (Lurie and Mason 2007). This should result in a more precise diagnosis of consumers' concerns and the remedial actions necessary to improve the quality of the firm's offering.

Developments in the accounting literature on analysts' and investors' use of information (Elliott, Hobson, and Jackson 2011; Hirst, Koonce, and Venkataraman 2007) suggest that disaggregate (vs. aggregate) earnings forecasts (i.e. revenues, research and development expenses, etc.) are more credible and diagnostic. Applying this idea to our setting suggests reduced motivation and ability of managers to use aggregate online review text analytics, resulting in no improvements in service quality. However, we do not anticipate this to occur in our context as managers using online review text analytics can supplement online review text analytics with disaggregate online consumer reviews, which are also available readily to managers. On net, we anticipate the positive effects of managerial use of online review text analytics on service quality to dominate and propose H_1:

> *H_1: Managerial use (vs non-use) of online review text analytics will improve the service quality of the firm's offering.*

Heterogeneous Effects

Managerial information processing and decision-making is affected by the architecture of their environments including their motivational features (Heath, Larrick, and Klayman 1998; Powell, Lovallo, and Fox 2011) which may either constrain or enable managerial actions. Applying this idea, we develop hypotheses of two heterogeneous effects of contextual characteristics of the managers' environment—managerial accountability for their actions and prior firm performance.

Managerial accountability. We define managerial accountability for his/her actions as the explicit expectation that one's decisions or actions will evaluated by some salient audience(s) with the belief that there is potential to receive either rewards or sanctions based on the evaluation (Hall et al. 2003). Managerial accountability improves individuals' awareness of their task responsibilities guiding their effort toward the achievement of a specified set of objectives (Hall et al. 2003). Hence, managerial accountability is positively related to the manager's motivation, actions (Lerner and Tetlock 1999) and job performance (Wallace et al. 2011).

When managers are held accountable for their actions, they may be motivated and able to change their firm's service offering based on the diagnostic consumer insights in the online review text analytics. Accordingly, we propose that the positive effect of managerial use of online review text analytics on service quality (H_1 above) will be strengthened by managerial accountability. Thus, we propose H_2:

H_2: The positive effect of managerial use of review text analytics on the service quality of the firm's offering will be stronger when managerial accountability is higher.

Prior firm performance. Individuals' information search behaviors and responses are driven by their problems (Cyert and March 1963). Failure and/or poor performance increases individuals' information search, usage, and their ability to change in response to this information (Kiesler and Sproull 1982). Hence, we propose a second contingency, prior firm performance.

When prior firm performance is satisfactory, managers may not perceive a need to change anything, including the firm's offering (Greve 1998). In contrast, when prior firm performance is unsatisfactory, managers may be both more motivated and willing to take the actions needed to improve their firm's unsatisfactory performance (Merchant and Manzoni 1989). In our setting, this may include improving the firm's service offering based on the consumer insights in online review text analytics which, as we noted above, provide diagnostic information from their consumers on the problems with the firm's offering. Accordingly, we offer H_3:

H_3: The positive effect of managerial use of review text analytics on the service quality of the firm's offering will be weaker when prior firm performance is satisfactory.

Method

We use a RCT, traditionally considered to be the gold standard to detect causality, in a field setting for the empirical testing of the hypotheses of the causal effect of managerial use of online review text analytics on service quality. In an RCT, subjects in their real-world setting are randomly assigned to one of two conditions, one in which they receive the stimulus, online review text analytics (treatment condition) and the other in which they do not (control condition). Following random assignment to either one of the two conditions, subjects pursue their 'regular' activity (in our case, managers go about the business of managing hotels). At the conclusion of the RCT, the difference in the outcome (i.e. the hotel's service quality) between the treatment and control conditions represents the average treatment effect of the online review text analytics.

The first advantage of RCTs in field settings, which also applies to lab experiments, is their ability identify the causal effect, ensuring their internal validity. The randomization process eliminates potential bias (e.g., endogeneity) from observed and unobserved sources so that the control condition may be considered as an average counterfactual for the treatment condition. The second advantage of RCTs in field settings is their ability to investigate the causal effect of interest in a realistic setting, ensuring a greater external validity of the findings. The third advantage of RCTs in field settings is their ability to assess heterogeneous effect i.e. whether the treatment effect varies with the characteristics of the environment. The absence of heterogeneity in the effect highlights the robustness of the (main) treatment effect, while its presence identifies its boundary conditions. See Banerjee and Duflo (2017) for a complete discussion of the use of RCTs in field settings.

Empirical Context and Sampling

We test the hypotheses in the hotel industry where online reviews are central to consumers' decision-making and therefore to hotels' performance (Litvin et al. 2008; PWC 2015). We enlisted the cooperation of Federalberghi, the Italian hotel industry association, to conduct a RCT on a sample of Italian hotels, whose consumers had posted online reviews on TripAdvisor, Italy. We sought managers of Italian hotels willing to participate in a longitudinal study on the role of online reviews. In Figure 1, we provide the time frames of the RCT and the different steps in data collection.

Federalberghi fields a quarterly survey of the Italian economic situation among member hotels. We included a section introducing our study in this survey in January 2016, with questions 1) pertaining to the use of online consumer reviews by the hotel's manager and his/her attitudes toward online consumer reviews and 2) describing our proposed study with a request for participation. Appendix A contains the measures collected in the different stages of the RCT. To attenuate potential demand effects, we informed managers that we sought their participation in a study investigating hotels' responses to online consumer feedback. We indicated that we would send them consumers' online reviews of their hotels at the beginning of the period (March 1, 2016) related to the one-year period before, followed by another similar report after three months (June 1, 2016) as a stimulus reinforcement. As an incentive for participation, at the end of the study, participating hotels would receive a summary report of their hotel's service quality synthesized from their online consumer reviews.

Sample profile. Overall, 598 hotel managers responded to this initial survey and 215 hotel managers (34%) expressed willingness to participate in the study. As the study focuses on the role of online review text analytics, we excluded from the RCT, 12 hotels already using consultants and/or software to process online consumer reviews. This resulted in an initial sample size of 203 hotels. An ex-ante power analysis for the repeated measures design (i.e. we have pre-treatment and post-treatment measures of the hotel's service quality) revealed that this sample size is adequate for the anticipated size of the effect (details in Appendix B).

On average, the 203 hotels had 42 rooms (range: 3-389) and 19 employees (range: 2-140). Most (58%) and were moderately-priced (i.e., class three hotels) with 20% being budget hotels (i.e., class one or class two) and 21% being boutique hotels (i.e., class four or class five). All the hotels participating in our study were independent hotels (vs. chain). Further, the hotels were equally split between city (48%) and tourist locations (52%). The managers who participated in our RCT spent an average of 3.3 hours per week on online consumer reviews and considered them to be both credible and correct (μ=3.97; σ=1.57).

Sample selection. Participant hotels did not differ from non-participant hotels on type (segment, size, location) and the characteristics of managers (education, age, work experience, and their attitudes toward online consumer reviews) (details of this analysis in Appendix C). The only difference was that managers in participant hotels (vs. non-participant hotels) perceived online consumer reviews to be marginally more useful (p =.053). Thus, there is no evidence of a selection bias in the sample of participant hotels.

Randomization and Manipulation

In March 2016, in Step 2 of the RCT, we randomly assigned each of the 203 hotels willing and qualified to participate in the RCT to either the control or treatment condition. We stratified the random assignment of hotels to the two conditions to achieve an unbiased assignment of hotels based on size, type, and location.

In the control condition (n =102), we emailed a report that included their hotel's online consumer reviews and ratings from TripAdvisor for the previous twelve months (March 2015-February 2016) (details in Appendix D, panel 1). The information for managers of hotels in the control condition was identical to their hotel's TripAdvisor, Italy's consumer reviews webpages and represents the “normal” condition of how hotel managers may process online consumer reviews.

In the treatment condition (n=101), we emailed a similar report to hotel managers, but, in addition, this report included the online review text analytics (our manipulation). To test the hypotheses, as the stimulus for the treatment condition, we used a basic form of text mining analyses’ outputs from the hotels’ online review text of its consumers for the previous twelve months (March 2015-February 2016), as we next describe.

We first extracted key concepts from the text data and represent them in tag clouds of the key concepts. The system performs a sentiment analysis of the review text and identifies the key concepts as having either net positive (green) or negative sentiment (red), based on which they are represented as a green or red tag cloud respectively of the word/term denoting the concept (e.g., bathroom, location). The size of the tag cloud is based on the overall frequency of occurrence and strength of the sentiment associated with the concept in the textual data.

We present the sentiment analysis of the key attributes which are common across all hotels (location, rooms, service, food, welcome, cleaning, convenience, and Internet connection) and represent their net positive sentiment score (ranging from 0 to 100) in a summary table including the number of responses, and a bar chart denoting the positive and negative sentiment of the attribute. This summary table is parsimonious and standardized across hotels and over time, providing managers diagnostic insights on their hotel’s performance. The tag clouds and the summary table of key attributes are analogous to the unsupervised and supervised learning approaches in data mining, respectively. We provide the stimulus for the treatment condition in Appendix D, Panel 2. The average number of online consumer reviews included in our reports is 46 per hotel (standard deviation = 52). To ensure that the reports were read by the hotel’s managers during the RCT, we communicated with the manager in the hotel responsible for managing its online consumer reviews identified in Step 1 above.

To reinforce the stimuli, at the end of May 2016, we sent an email to managers in both the conditions, i.e., a second report with the same information as in the first report, updated with online consumer reviews between March 2016 and May 2016. We obtained responses from 146 hotels. This represents Step 3 of the RCT (see Figure 1) and is a reinforcement of the stimulus to reduce potential non-compliance effects (Glennerster 2017).

At the conclusion of the RCT, at the end of August 2016, we fielded a brief survey of the hotel managers to ensure that they had been active in the RCT. As expected, there was some attrition. We obtained responses from 65 and 70 hotels in the control and treatment conditions respectively, which constitute the final sample for analysis. This is Step 4 of the RCT.

In December 2016, three months after the conclusion of the RCT, managers of hotels in both conditions responded to a brief online survey ($n = 99$). The purpose of collecting this primary data from a post-hoc survey was to generate insights on the underlying mechanism (Simester 2017) and further examine the robustness of the results. This is Step 5 of the RCT.

Sample Attrition

We estimated a Probit model to assess whether sample attrition in the RCT may be explained by observable characteristics of the hotels. We provide the estimation results in Appendix E. Only the hotel's size was significantly associated with likelihood of attrition ($b = .57$; $p < .02$); large hotels (more than 30 rooms) were 19% more likely to drop out from the RCT. To correct for this attrition, we employ the inverse probability weighting approach assigning greater weights to larger hotels (Seaman and White 2011). Sample attrition was not related to the dependent variable, the hotel's service quality, and the attitude of managers toward online consumer reviews (details in Appendix E).

Dependent Variable

To measure the service quality of the hotel's offering, we use its online consumer ratings on TripAdvisor, Italy. The pre-treatment service quality is the monthly weighted (by the number of reviews) average online ratings of the hotel on TripAdvisor between June and August 2015. We chose the period between June and August as the period for the RCT as it is the peak season for Italian hotels.

A t-test reveal no difference in the pre-treatment service quality of hotels in the control and treatment conditions ($\mu_{treat} = 4.07$; $\mu_{ctrl} = 4.12$; $t = 0.56$; $p = .58$). As the stimulus for each hotel's treatment is customized, we note that the outcome of a hotel cannot be affected by the treatments received by other hotels, thus ensuring stability of the unit treatment values (Athey and Imbens 2017). The monthly weighted average online ratings of the hotels between June and August 2016 is the measure of post-treatment service quality.

Results

To empirically test the hypotheses, we report the results using multiple empirical approaches.

Difference-in-Difference Analysis

We first estimated a difference-in-difference (DID) model, comparing the average change (pre-/post-treatment) in the hotel's service quality in the treatment condition (with online review text analytics) to the average change in the control condition (no online review text analytics). This model included the factorial interaction between the treatment variable and the time dummy (pre/post) as a predictor and incorporated the inverse probability weights based on the hotel's size (as discussed above) to account for sample attrition. We cluster standard errors at the hotel-level to account for non-independence of observations (Bertrand, Duflo, and Mullainathan 2004).

The results provide evidence of a positive treatment effect (b = .23; p < .03) indicating that following the RCT, hotels in the treatment condition whose managers were provided with online review text analytics improved the service quality of their offering significantly more than managers of hotels in the control condition (Table 2). Marginal effects analysis reveals that average service quality in the control condition remains unchanged (μ_{pre} = 4.19; μ_{post} = 4.21; F=0.10; p = .75) while, in the treatment condition, average service quality increased (μ_{pre} = 4.08; μ_{post}=4.34; F=10.72; p = .00) during the treatment period.

Mixed Model Analysis

While the DID analysis provides evidence of the treatment effect, by clustering the standard errors at the hotel level, it only partially addresses the non-independent nature of the data, as we have pre- and post-treatment service quality measures of the same hotel. Hence, we re-estimated the model using mixed regression which accounts for the nested structure of the data incorporating a random effect at the hotel-level.

The results of a mixed model substantially replicate the treatment effect (b = .22; p = .02) identified in the DID analysis and support H_1 that hotels in the treatment condition (online review text analytics) improved their service quality significantly more than hotels in the control condition (no online review text analytics). The estimated random effect intercept is positive and statistically significant (mean = .29; confidence interval: .21-.42) suggesting the need to account for the nested structure of the data.

Marginal effects analysis (see Figure 2) indicates that for hotels in the control condition, service quality remained unchanged (μ_{pre} = 4.19; μ_{post} = 4.19; χ^2=0.01; p = .91), while in the treatment condition, service quality increased (μ_{pre} = 4.08; μ_{post} = 4.31; $\chi^2_{(1)}$ = 9.28; p = .00) in the treatment period. In terms of elasticity, the treatment increased the service quality of hotels by 5.4% in the six-month period.

We measure the effect size of the treatment using the f'^2 statistic which measures the difference between variance explained by a model with the treatment over a model without it (Selya et al. 2012). The f'^2 statistic (f^2 = .09) indicates that the treatment effect may be considered as a small to medium effect (Cohen 1988). From a practical perspective, given the low standard deviation of online ratings (σ = 0.56), we consider this effect size is substantive. Moreover, as Tripadvisor (and other online review platforms) provide rounded, not continuous ratings (e.g., 3.5, 4, and 4.5), a small change can move the displayed, rounded rating by a large amount.

Robustness Checks

We next report analyses that assess the finding's robustness.

Measures of dependent variable. We examine the finding's robustness to alternative measures of the hotel's service quality. In robustness check #1, we account for the heterogeneity in the number of online reviews of the hotel, providing greater weight to online ratings of hotels with more online reviews. As the distribution of online reviews is left-skewed (median = 26; mean = 41; skewness = 2.39), we weight the online ratings of hotels by the natural logarithm of the number of their online reviews. Results show that the two groups do not differ in the average number of online reviews in the treatment period (μ_{CTRL} = 44.6; μ_{TREAT} = 37.1; p = 0.32) and, importantly, the results of the weighted analysis still support a significant treatment effect (b = .17; p = .00).

In robustness check #2, we extended the period for the dependent variable from June to September 2016 instead of from June to August 2016. We do this as some consumers who visited the hotel in August 2016 may have posted online reviews on Tripadvisor only in September 2016. We replicated the analysis of the weighted average online ratings (see check #1) between June and September 2015 and 2016, for the “pre” and “post” dependent variable measures respectively. The results indicate that while the effect size is smaller, the treatment effect is still positive and significant (b = .12; p = .00).

Influential outliers. In robustness check #3, we examine the role of influential outliers in the sample on the finding. We identified influential cases (n = 3) as those with a DFBETA diagnostic greater than the threshold value of $2/\sqrt{n}$ (Belsley, Kuh, and Welsch 1980). We re-estimated the mixed model excluding these three cases and find that our treatment effect is again positive and significant (b = .24; p = .01).

Potential confounders of treatment. The randomization procedure in the RCT eliminates confounding from variables used for sample stratification. Nonetheless, we examine whether these variables result in heterogeneous treatment effects. In robustness check #4, we re-estimated the mixed model by interacting sample characteristics with the treatment. The results indicated no heterogeneous treatment effect with the hotel’s size ($\chi^2_{(1)} = 0.25$; p = .61), class ($\chi^2_{(2)} = 2.46$; p = .29), and location ($\chi^2_{(1)} = 0.01$; p = .92).

Moreover, exogenous events that affect hotels in one condition but not in the other condition may create heterogeneous treatment effects. To rule out this potential confound, in robustness check #5, at the conclusion of the RCT, we asked hotel managers if their hotel had experienced any events (e.g., flooding, earthquake, closing for refurbishing) during the RCT. Twelve (9.8%) of 135 hotels reported such events; this proportion was equal across the two conditions ($\chi^2_{(1)} = 0.41$; $p = 0.52$). When we interacted the treatment with the external event variable, the heterogeneous treatment effect was not significant ($b = -0.16$; $p = 0.62$).

Assumption about sample attrition. Sample attrition is an inherent feature of longitudinal RCTs (Glennerster 2017). As described above, we excluded from our analysis, hotel managers who did not respond to our intermediate and final surveys, assuming that they were "not compliant" with the RCT. To further test this assumption, in robustness check #6, we obtained the pre- and post-treatment measures of service quality for these 67 hotels that dropped out from the RCT.

The simple effects analysis in this sub-sample of hotels reveals no treatment effect ($b = -0.03$; $p = .842$) while the treatment effect was confirmed for the original sample of 135 participants in the RCT ($b = .23$; $p = .019$). Overall, our results appear to be robust to various threats.

Heterogeneous Effects: Managerial Accountability and Prior Firm Performance

We next discuss analyses that examine the hypothesized heterogeneous effects of two features of the manager's environment, managerial accountability for their actions and prior firm performance (H_2 and H_3 respectively) on the effects of managerial use of online review text analytics on service quality.

We measured managerial accountability using a three-item scale ($\alpha = .85$) on a scale from 1 to 7 at the end of the RCT in August 2016. Although managerial accountability may be considered as a fixed factor with respect to the treatment (so collecting the measure post-treatment should not matter), we nonetheless, found no difference across the two conditions (t = 0.29; p = 0.77).

To test this heterogeneous treatment effect, we re-estimate the mixed model for hypotheses testing (for H_1) with the inclusion of a three-way interaction term (treatment × time × managerial accountability). This three-way interaction term was positive and significant (b = .16; p < .03) indicating that, in support of H_2, the positive effect of managerial use of online review text analytics on service quality is stronger as managerial accountability increases. In Figure 3, we plot the pre/post difference in the hotel's service quality (Y axis) across the control and the treatment conditions at different levels of managerial accountability (X axis).

Figure 3 reveals that hotels in the treatment condition perform better than hotels in the control condition only when managerial accountability is high, i.e., greater than 5. While this situation reflects most cases in our sample (78%), when contextual factors are not favorable, managers may not be motivated to use online review text analytics to improve service quality i.e. the average treatment effect when the moderator is below 4 is not statistically significant (despite the apparent visual difference).

We next examine the heterogeneous effect of prior firm performance on the average treatment effect (H_3). We measure managers' perceptions of their hotel's prior performance (hotel occupancy rate, a widely-used performance metric in Italian hotels, as per Federalberghi) on a scale from 1 to 5 for each month of the pre-treatment period (June-August 2015). There is no difference in the level of pre-treatment prior performance across treatment and control conditions (t = 0.62; p = 0.54).

We again re-estimate the mixed model with the inclusion of a three-way interaction term (treatment × time × prior hotel performance). This interaction term is negative and statistically significant (b = -.17; p < .05), indicating that, in support of H_3, when prior hotel performance is not satisfactory, managerial use of online review text analytics in the treatment condition improves service quality, but not when prior hotel performance is satisfactory.

In Figure 4, we plot the pre/post difference in service quality (Y axis) across the control and treatment conditions across different levels of prior hotel performance (X axis). Additional marginal analysis reveals that the treatment effect occurs when prior hotel performance is low (i.e., value of the moderator < 4; 69% of the cases).

To summarize, the findings indicate that the treatment effect of managerial use of online review text analytics on the hotel's service quality is 1) positive, 2) robust to various threats from measures, confounding variables, and sample attrition, and 3) heterogeneous with respect to managerial accountability and hotel's prior performance.

Additional Analysis: Mechanism Underlying Service Quality

RCTs are not generally well-suited to investigating the mechanism underlying the causal effect as on the one hand, it is difficult to maintain complete control of the conditions in the field context and on the other hand, collecting process measures during the RCT may interfere with the treatment effect. Nonetheless, we decided to investigate a process explanation for the main effect, that is, whether the treatment effect caused an identification and follow-through of the actions that the manager needs to take to address the problems affecting service quality.

To do this, we obtained the sentiment analysis data of the key attributes in the summary table in our treatment condition for the hotels in the control condition. We anticipate that managers of the hotels in the treatment (vs. control) condition would be able to detect and address their hotel's problem attributes.

Therefore, we identify, for each hotel, the attribute with the lowest sentiment, i.e. the most problematic attribute. We excluded from this analysis, location and rooms, which cannot be modified during the short time period of the RCT. The worst hotel attribute is equally distributed across the different attributes (exceptions, cleaning and welcome are less frequent) and the distribution is equivalent across treatment and control groups ($\chi^2_{(5)} = 1.53$; $p = .91$).

We build the criterion variable for each hotel as the difference between the level of sentiment on the worst attribute after and before the treatment ($\mu = 18.4$; $\sigma = 32.8$) and predict this difference (i.e. improvement) using the treatment variable. A regression analysis reveals preliminary evidence that the average increase in the treatment condition (22.7) was higher than in the control condition (13.3). Given the high variance of the dependent variable, the average effect is noisy and does not reach statistical significance ($b = 9.4$; $p = .12$).

Hence, we tested for this effect using inter-quantile regression, i.e. we estimate the treatment effect at 25 and 75 quantiles of the dependent variable. The results indicate a significant difference in the treatment effect of the sentiment score for the worst attribute ($b = 18.9$; $t = 1.70$; $p = .09$) between the 25 and the 75 quantiles. These results suggest that managers in hotels in the treatment (vs. control) condition were able to identify and remedy their weak attribute of their hotels.

We next follow the suggestion of Simester (2017) to complement the findings of a RCT with a post-hoc survey of participants to generate inferences about the underlying mechanism explored above. Hence, we fielded a post-treatment survey in December 2016 (Step 5 in Figure 1) collect data on the underlying mechanism. In this post-treatment survey (n = 99), managers were asked about the extent to which their hotel had improved its service offering in the last six months and whether they had reduced price, as service quality and price are two factors which influence perceived customer value (Bolton and Lemon 1999) and can be modified during the short period of the RCT. If our intuition that managers in the treatment condition are able to act on the weak attribute is correct, we expect that managers in the treatment condition should have changed the service attribute more than those in the control condition.

Given the small sample size (n = 99) and the non-normal distribution of price and service offering improvement measures (e.g., kurtosis = 4.7 and skewness = 4.4 respectively) we tested the difference between treatment and control conditions by using a two-sample Wilcoxon rank-sum test. In support of our intuition, managers of hotels in the treatment condition improved their service offering more than managers of hotels in the control condition ($z = 1.65$; $p < .10$) but they did not change their prices ($z = 0.49$; p = .62). These results of the additional analyses suggest that managers in the treatment (vs. control) condition were more likely to identify the weakest attribute of their offering and take remedial actions to improve it.

Discussion

A recent industry report estimated that the market for text analytics solutions for user-generated content, including online consumer reviews, is expected to grow dramatically from USD 3.97 Billion in 2017 to USD 8.79 Billion by 2022, at a compound annual growth rate of 17.2% (Marketsandmarkets 2017). Business experts (McKinsey 2016) and policy makers (EU Commission 2016) predict that text analytic solutions will enable managers exploit consumer insights embedded in textual data. Yet, there are few insights in the marketing literature on whether managerial usage of online review text analytics will improve their firm's performance, which we address in this research.

We examine the effect of the managerial use of online review text analytics on service quality in hotels. Using a RCT of Italian hotels, we find that managerial use of online review text analytics does, indeed, improve the service quality of the hotel during the treatment period. Further, this effect is heterogeneous based on managerial accountability and prior firm performance. We conclude with a discussion of the findings' theoretical contributions, managerial implications, and limitations and opportunities for further research.

Theoretical Contributions

First, the extant marketing literature on the effects of online word-of-mouth (You, Vadakkepatt, and Joshi 2015) has, hitherto, focused on the effects of online ratings, primarily at the consumer-level, with three notable exceptions (Büschken and Allenby 2016; Tirunillai and Tellis 2014; Xiong and Bharadwaj 2014). We extend the literature on online reviews by focusing on managers instead of consumers and on the effects of the usage of consumer insights in online review text instead of data in online numeric ratings. In doing so, we show yet another way, managerial use of online review text analytics, by which online reviews can improve firm performance.

Moreover, we test the effects of managerial use of online review text analytics on service quality, using a RCT in a field setting, which allows us to establish causality ensuring external validity. Indeed, our study is the first to show that the managerial use of online review text analytics leads to an increase in service quality (i.e., +5.4% on average). The post-hoc analysis provides preliminary evidence that improvements in service quality are achieved through improvements in service attributes and not through price decreases, which augurs well for the hotel's profitability.

Second, we extend the literature on managerial use of market intelligence (Menon and Varadarajan 1992; Moorman, Zaltman, and Deshpandé 1992) to contemporary approaches i.e., online feedback on social media, in general, and to consumer insights in online review text, in particular. Given the anticipated growth in online feedback, we hope that our study's findings spur additional research on how managers can effectively learn from online consumer-generated text, (e.g., relationships or co-occurrence of compliments and complaints).

Third, our study's focus on the benefits of managerial use of consumer insights extracted from online review text responds to the call for further research in the under-represented area of managerial actions (Wierenga 2011). We extend this literature, which has focused on the use of traditional dashboards using numerical information, where managers have considerable proficiency to the relatively unexplored area of textual information, where they have limited expertise, if at all.

Finally, the study's findings on the heterogeneous effects of contextual characteristics of the managers' environment, managerial accountability and prior firm performance, on the effects of managerial use of online review text analytics on service quality, contribute to the marketing literature on accountability (e.g., Verhoef and Leeflang 2009), which has focused at the level of the marketing function. Our study's finding on the role of managerial accountability, extends this work at the functional level, to the individual level of managers.

Managerial Implications

The research's findings generate some actionable guidance for managers, who may be interested in using online consumer review text as a source of consumer insights, and yet, have no formal training and/or experience on how to do so. First, the findings indicate that the use of even simple online review text analytics, of the type that we use in our study, is beneficial to managers in improving the service quality of their firm's offering. As noted, displayed online ratings are rounded to a half rating (e.g., 3, 3.5, and 4). Hence, even a small increase in the raw online rating, when it crosses the threshold for rounding, is 'in the eye of consumers', a large increase in the displayed online rating. Additional analysis indicates that 37% of the hotels in the treatment condition increased the online ratings of their service quality by more than 0.5, an increase at least greater than 20%. We consider this to be a good improvement in the hotel's short-term service quality, especially without any structural changes.

Second, the findings indicate heterogeneous effects of the managerial use of online review text analytics on the hotel's service quality. The benefits of online review text analytics can be leveraged only when managers are accountable for their actions and when the firm has had poor prior performance. These findings suggest two actionable guidelines to motivate and enable managers to leverage and act on the consumer insights in online review text analytics to improve service quality: 1) managers who are entrusted with managing online consumer reviews must be held accountable for their actions and 2) if prior firm performance is satisfactory, then additional inducements (e.g., bonuses) may be useful to motivate them.

Third, our research's findings generate guidelines for online reputation management firms in identifying prospects to sell their software solutions. These firms should target customer firms where managers, responsible for managing online consumer reviews, are held accountable for their actions and where performance is not satisfactory, as these managers are likely to be motivated to use online review text analytics and able to leverage their benefits.

Limitations and Opportunities for Further Research

First, in this initial empirical study on the causal effect of the managerial use of online review text analytics, we used a RCT in a field setting over six months, which some may consider to be a short time period. As a result, we focused on the short-term benefits of managerial use of online review text analytics on service quality. A longer study using a qualitative approach that examines the long-term benefits of using online review text analytics, can focus on the structural changes, including additional capacity and improved facilities, will be a useful extension to this work.

Second, in the interest of parsimony and to ensure a clean test of the hypotheses, we used a rather simple online review text analytics (tag cloud of words denoting key concepts and summary table of key attributes). Further, we focused on the benefits of using online review text analytics on service quality. Thus, our findings may be considered as a conservative test of the benefits of managerial use of online review text analytics. A question for further research is whether the findings hold for more complex online review text analytics and other firm outcomes including, innovations, profitability, and stock returns.

Third, as we conducted a RCT in a field setting to test the hypotheses which provides external validity to the findings, we are only able to generate preliminary insights on the mechanism underlying the effect of managerial use of online review text analytics on service quality. Lab studies using simulation exercises with business students which can generate additional insights on the underlying mechanism, will be a useful extension to this work.

In sum, we view our study as a first step in exploring the role of managerial use of online review text analytics in improving firm performance. Given the growing importance of online review text in business practice, we hope that this research stimulates further work on how managerial use of unstructured online reviews can improve other firm outcomes.

References

Archak, Nikolay, Anindya Ghose, and Panagiotis G. Ipeirotis (2011), "Deriving the Pricing Power of Product Features by Mining Consumer Reviews," *Management Science*, 57(8), 1485-1509.

Athey, Susan, and Guido W. Imbens (2017), "The Econometrics of Randomized Experiments", in Banerjee Abhijit Vinayak and Esther Duflo (eds.) *Handbook of Economic Field Experiments (Volume 1)*, North-Holland, Elsevier.

Banerjee, Abhijit Vinayak, and Esther Duflo (2017), *Handbook of Field Experiments. Vol. 1 and 2*, North-Holland, Elsevier.

Belsley, David A., Edwin Kuh, and Roy E. Welsch (2005), *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York. John Wiley & Sons.

Berlyne, Daniel E. (1960), *Conflict, Arousal, and Curiosity*. New York: McGraw-Hill.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004), "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics,* 119(1), 249-275.

Blei, David M (2012), "Probabilistic Topic Models," *Communications of the ACM*, 55(4), 77-84.

Bolton, Ruth N., and Katherine N. Lemon (1999), "A Dynamic Model of Customers' Usage of Services: Usage as an Antecedent and Consequence of Satisfaction," *Journal of Marketing Research,* 36(2), 171-186.

Büschken, Joachim, and Greg M. Allenby (2016), "Sentence-Based Text Analysis for Customer Reviews," *Marketing Science,* 35(6), 953-975.

Chenhall, Robert H., and Deigan Morris (1986), "The Impact of Structure, Environment, and Interdependence on the Perceived Usefulness of Management Accounting Systems," *The Accounting Review,* 61(1), 16-35.

Cohen, Jacob (1988), *Statistical Power Analysis for the Behavioral Sciences*. Hilsdale, NJ: Lawrence Earlbaum Associates.

Crowley, Ayn E., and Wayne D. Hoyer (1994), "An Integrative Framework for Understanding Two-Sided Persuasion," *Journal of Consumer Research,* 20(4), 561-574.

Englewood Cliffs, New Jersey: Prentice-Hall.

Elliott, W. Brooke, Jessen L. Hobson, and Kevin E. Jackson (2011), "Disaggregating Management Forecasts to Reduce Investors' Susceptibility to Earnings Fixation," *The Accounting Review,* 86(1), 185-208.

EU Commission (2016), "Competence Centre Text Mining and Analysis", Workshop on Text Mining in Policy making, Brussels, Belgium, December 13 2016.

Feustel, Timothy C., Richard M. Shiffrin, and Aita Salasso (1983), "Episodic and Lexical Contributions to the Repetition Effect in Word Identification," *Journal of Experimental Psychology: General,* 112, 309–346.

Fiske, Susan T., and Shelley E. Taylor (1984), *Social Cognition Reading*. MA: Addison-Wesley.

Fitzgerald, John M., Peter Gottschalk, and Robert Moffitt (1998), "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics," *The Journal of Human Resources*, 33(2), 251-299.

Gardner, Howard (1983), *Frames of Mind: The Theory of Multiple Intelligences*, Basic Books.

Glennerster, Rachel. (2017), "The Practicalities of Running Randomized Evaluations: Partnerships, Measurement, Ethics, and Transparency", in Banerjee Abhijit Vinayak and Esther Duflo (eds.) *Handbook of Economic Field Experiments (Volume 1)*, North-Holland, Elsevier.

Greve, Henrich R. (1998), "Performance, Aspirations, and Risky Organizational Change," *Administrative Science Quarterly*, 43 (March): 58-86.

Hall, Angela T., Dwight D. Frink, Gerald R. Ferris, Wayne A. Hochwarter , Ceasar J. Kacmar, and Michael G. Bowen, (2003), "Accountability in Human Resources Management," in: C. A. Schriesheim and L. Neider (Eds.), *New Directions in Human Resource Management*, 29-63, Greenwich, CT: Information Age.

Heath, Chip, Richard P. Larrick, and Joshua Klayman (1998), "Cognitive repairs: How organizational practices can compensate for individual shortcomings," *Review of Organizational Behavior*, 20, 1–37.

Hirst, Eric D., Lisa Koonce, and Shankar Venkataraman (2007), "How Disaggregation Enhances the Credibility of Management Earnings Forecasts," *Journal of Accounting Research* 45(4), 811-837.

Hu, Nan, Ling Liu, and Jie Jennifer Zhang (2008), "Do Online Reviews Affect Product Sales? The role of Reviewer Characteristics and Temporal Effects," *Information Technology and Management,* 9(3), 201-214.

Hu, Nan, Paul A. Pavlou, and Jie Jennifer Zhang (2009), "Why Do Product Reviews Have a J-Shaped Distribution?" *Communications of the ACM*, 52(10), 144-147.

Kodratoff, Yves (1999), "Knowledge Discovery in Texts: A Definition, and Applications," in: Raś Z.W., Skowron A. (eds) *Foundations of Intelligent Systems*, 16-29, Springer, Berlin.

Kiesler, Sara, and Lee Sproull (1982), "Managerial response to changing environments: Perspectives on problem sensing from social cognition," *Administrative Science Quarterly*: 548-570.

Lerner, Jennifer S., and Philip E. Tetlock (1999), "Accounting for the Effects of Accountability," *Psychological Bulletin*, 125(2), 255.

Lewis, Greg, and Georgios Zervas (2016), "The Welfare Impact of Consumer Reviews: A Case Study of the Hotel Industry," Working Paper (retrieved at http://faculty.chicagobooth.edu/workshops/marketing/pdf/tawelfare.pdf).

Litvin, Stephen W., Ronald E. Goldsmith, and Bing Pan (2008), "Electronic Word-of-Mouth in Hospitality and Tourism Management," *Tourism Management,* 29(3), 458-468.

Lurie, Nicholas H., and Charlotte H. Mason (2007), "Visual representation: Implications for decision making", *Journal of Marketing*, 71(1), 160–177.

MacInnis, Deborah J., Christine Moorman, and Bernard J. Jaworski (1991), “Enhancing and Measuring Consumers' Motivation, Opportunity, and Ability to Process Brand Information from Ads,” *Journal of Marketing*, 55, 32-53.

Maitlis, Sally, and Marlys Christianson (2014), “Sensemaking in Organizations: Taking Stock and Moving Forward,” *The Academy of Management Annals*, 8(1), 57-125.

Marketsandmarkets (2017), “Text Analytics Market - Global Forecast to 2022”, Business Report.

McKinsey (2016), “The Age of Analytics: Competing in a Data-Driven World”, Global Institute Report.

Menon, Anil, and P. Rajan Varadarajan (1992), “A Model of Marketing Knowledge Use within Firms,” *Journal of Marketing*, 56(4), 53-71.

Merchant, Kenneth A., and Jean-Francois Manzoni (1989), “The Achievability of Budget Targets in Profit Centers: a Field Study,” *The Accounting Review*, 64, 539-558.

Moorman, Christine, Gerald Zaltman, and Rohit Deshpandé (1992), “Relationships between Providers and Users of Market Research: The Dynamics of Trust within and between Organizations,” *Journal of Marketing Research*, 29(3), 314–328.

Muller, Keith E., Lisa M. Lavange, Sharon Landesman Ramey, and Craig T. Ramey (1992), “Power Calculations for General Linear Multivariate Models Including Repeated Measures Applications,” *Journal of the American Statistical Association*, 87, 1209-1226.

Newel, D.J. (1992), “Intention-to-treat analysis: Implications for quantitative and qualitative research,” *International Journal of Epidemiology*, 21, 837–841

Paivio, Allan (1991), *Images in Mind. The Evolution of a Theory*, Harvester Wheatsheaf, New York.

Pechmann, Cornelia (1992), “Predicting When Two-Sided Ads Will Be More Effective Than One-Sided Ads: The Role of Correlational and Correspondent Inferences,” *Journal of Marketing Research*, 29(November), 441-453.

Powell, Thomas C., Dan Lovallo, and Craig R. Fox (2011), “Behavioral Strategy,” *Strategic Management Journal*, 32 (13): 1369-1386.

PWC (2015), “Online Reputations Why Hotel Reviews Matter and how Hotels Respond”, retrieved at: https://www.hotelleriesuisse.ch/files/pdf10/pwc_global_hotels_report_2015_e2.pdf.

Royle, M. Todd., Angela T. Hall, Wayne A. Hochwarter, Pamela L. Perrewe, and Gerald R. Ferris (2005), “The Interactive Effects of Accountability and Job Self-Efficacy on Organizational Citizenship Behavior and Political Behavior,” *Organizational Analysis*, 13, 53-71.

Seaman, Shaun R., and Ian R. White (2011), “Review of Inverse Probability Weighting for Dealing with Missing Data,” *Statistical Methods Medical Research*, 22(3): 278–295.

Seifert, Christin, Sabol Vedran, Kienreich Wolfgang, Lex Elisabeth, and Granitzer, Michael (2014), “Visual Analysis and Knowledge Discovery for Text,” in: Gkoulalas-Divanis Aris, Labbi Abderrahim (eds.) *Large-Scale Data Analytics,* 189-218, Springer New York.

Selya, Arielle S., Jennifer S. Rose, Lisa C. Dierker, Donald Hedeker, and Robin J. Mermelstein, (2012), "A Practical Guide to Calculating Cohen's f^2, a Measure of Local Effect Size, from PROC MIXED," *Frontiers in Psychology*, 3, 1-6.

Simester, Duncan (2017), "Field Experiments in Marketing," in Banerjee Abhijit Vinayak and Esther Duflo (eds) *Handbook of Economic Field Experiments (Volume 1)*, North Holland: Elsevier.

Tetlock, Philip E. (1983), "Accountability and Complexity of Thought," *Journal of Personality and Social Psychology*, 45, 74-83.

Tirunillai, Seshadri, and Gerard J. Tellis (2014), "Mining Meaning from Online Chatter: Strategic Brand Analysis of Big Data with Latent Dirichlet Allocation," *Journal of Marketing Research*, 51, 463-479.

Verhoef, Peter C., and Peter S. H. Leeflang (2009), "Understanding Marketing Department's Influence within the Firm," *Journal of Marketing*, 73(2), 14–37.

Wallace, Craig, Paul D. Johnson, Kimberly Mathe, and Jeff Paul (2011), "Structural and Psychological Empowerment Climates Performance and the Moderating role of Shared Felt Accountability: A Managerial Perspective", *Journal of Applied Psychology*, 96(4), 840-850.

Wierenga, Berend (2011), "Managerial Decision Making in Marketing: The Next Research Frontier," *International Journal of Research in Marketing*, 28(2), 89-101.

Xiong, Guiyang, and Sundar Bharadwaj (2014), "Prerelease Buzz Evolution Patterns and New Product Performance," *Marketing Science*, 33(3), 401-421.

You, Ya, Gautham G. Vadakkepatt, and Amit M. Joshi (2015), "A Meta-Analysis of Electronic Word-of-Mouth Elasticity," *Journal of Marketing*, 79(2), 19-39.

Figure 1– Time Frame of the RCT

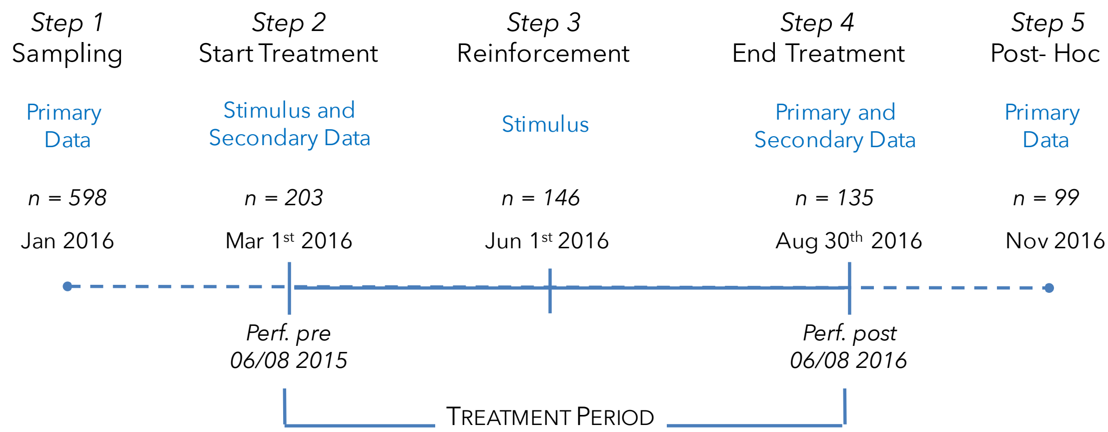


Figure 2 – Average Treatment Effect of Online Review Text Analytics

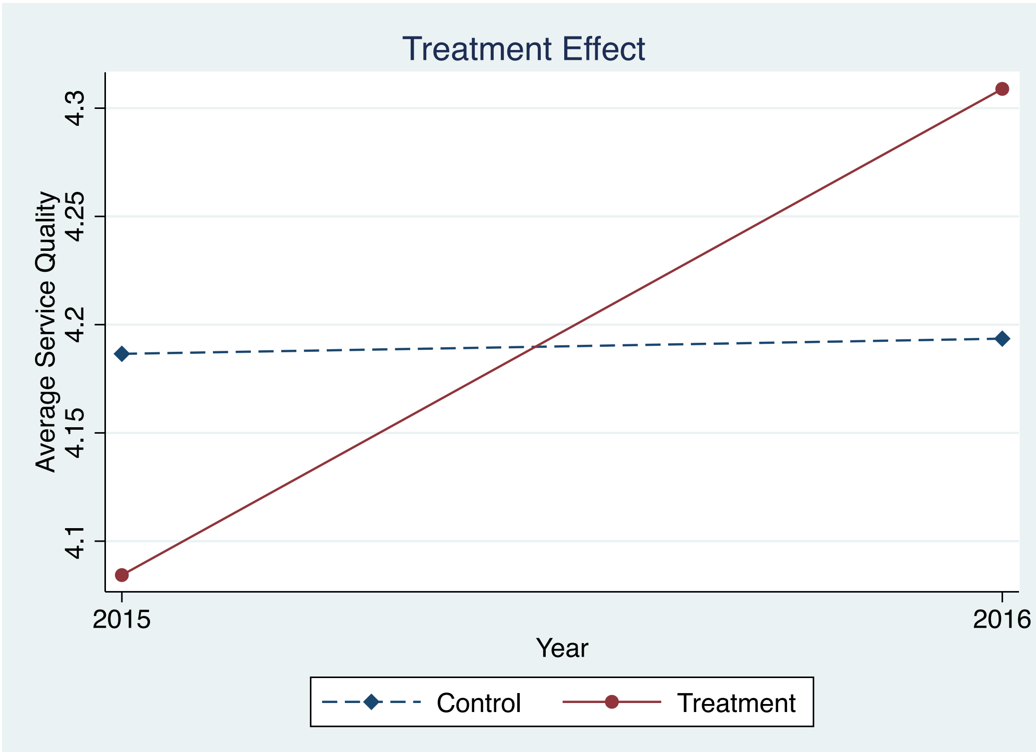


Figure 3 – Average Treatment Effect of Online Review Text Analytics: The Moderating Role of Managerial Accountability

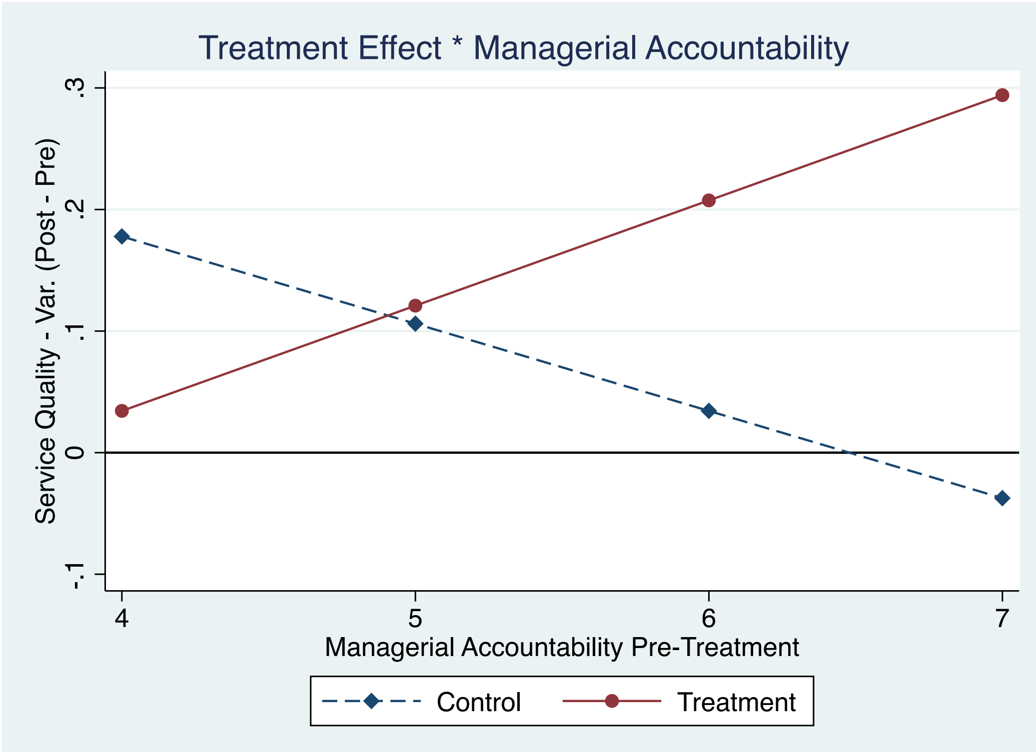


Figure 4 – Average Treatment Effect of Online Review Text Analytics: The Moderating Role of Prior Firm Performance

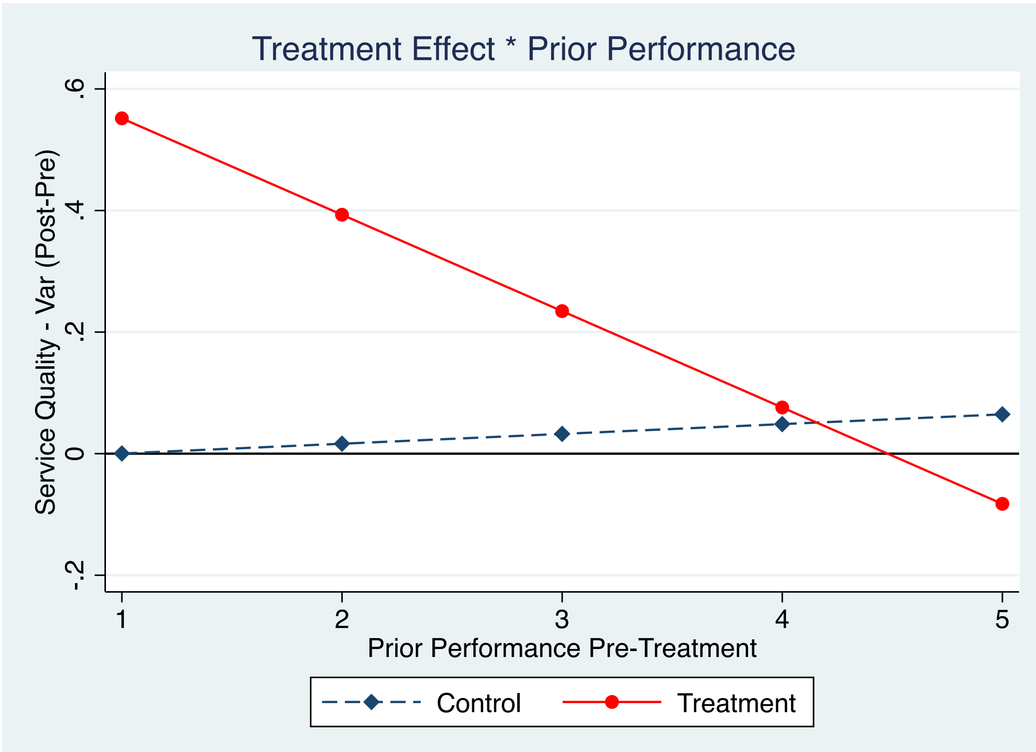


Table 1 - Stratified Random Assignment of Participant Hotels across Treatment and Control Conditions

Location	Business locations		Tourist locations*	
Hotel Class	Less than 30 rooms	More than 30 rooms	Less than 30 rooms	More than 30 rooms
Control condition				
Class 1-2	10	2	7	-
Class 3	15	21	17	7
Class 4-5	1	8	1	13
Total (102)	26	31	25	20
Treatment				
Class 1-2	10	3	4	-
Class 3	15	19	19	7
Class 4-5	1	9	2	12
Total (101)	26	31	25	19

* *Tourist locations include seaside, mountain and lakes.*

Table 2 - Difference-in-Difference Analysis of Treatment Effect

Online Ratings	Parameter estimate	Standard error*	P>|t|
Treatment (vs. Control)	-.107	.116	0.36
Year (2016 vs. 2015)	.020	.065	0.75
Treatment × Year	.235	.101	0.02
Constant	4.19	.076	0.00

*N = 262; *Robust Clustered Standard Errors*

Table 3 - Mixed Model Analysis of Treatment Effect

Online Ratings	Parameter estimate	Standard error*	P>|t|
Treatment (vs. Control)	-.102	.114	0.37
Year (2016 vs. 2015)	.007	.060	0.91
Treatment × Year	.218	.095	0.03
Constant	4.186	.078	0.00

Random-effects Parameters	Parameter Estimate	Standard error*	[95% Conf. Interval]	
Hotel: Id				
Var (_cons)	.294	.052	.207	.416
Var (Residual)	.084	.014	.059	.117

*N = 262; Group Variable: Hotel; Number of groups = 134, *Robust Clustered Standard Errors*

Appendix A

Measures Used in Research

Credibility of reviews (new single-item measure), January 2016 and December 2016; used for sample selection

1. In general, online customers' reviews are credible and correct.

Decision Making Confidence (Sniezek, 1991), January 2016 ($\alpha = 0.93$) and August 30th 2016 ($\alpha = 0.90$); used for sample selection

1. I feel very confident about the decisions I make based on online customers' reviews
2. By using online customers reviews I feel I have made the right decisions
3. I strongly believe in the decisions I make based on online customers' reviews

Perceived Usefulness (Chenhall and Morris 1986), January 2016 ($\alpha = 0.90$) and August 30th 2016 ($\alpha = 0.90$); used for sample selection

1. The information from online consumers' feedback is very useful to me in my job/decision making
2. The information from online consumers' feedback is very helpful to me in my job/decision making

Managerial Accountability (Royle et al. 2005), August 30th 2016 ($\alpha = 0.85$); used for heterogeneous effect of managerial accountability

1. I am accountable for some really important programs and projects at work
2. The work I do, and am accountable for, is central to the overall effectiveness of my organization
3. I am held very accountable for my actions at work

Prior Firm Performance (new measure), March 2016 ($\alpha = 0.87$); used for heterogeneous effect of past performance based on ratings (on a scale from 1-5) of the occupancy rate of the hotel in June 2015, July 2015, and August 2015.

Service Offering Improvement (new single-item measures): December 2016; used for assessing the mechanism underlying service quality improvement:

1. In the last six months, the level of services in my hotel offering has significantly improved
2. In the last six months, the price/quality ratio of my offering has significantly increased.

Appendix B

Ex-Ante Power Analysis

To examine the adequacy of the sample size of 203 in our RCT, we conducted a preliminary power analysis for repeated measure designs (Muller et al. 1992).

This analysis requires assumptions about the anticipated effect size, the distribution (i.e., mean and standard deviation) of the criterion variable, the correlation between the repeated observations of the criterion variable, and the potential dropout rate.

First, we assume a small-to-medium effect of our treatment (i.e., Cohen's d = 3.5). Second, the TripAdvisor average numeric ratings for Italian hotels is 4 (standard deviation = 0.55). Third, we assume a correlation (i.e., r = .60, the correlation in our final sample = .62) between "pre" and "post" treatment measures. Fourth, we assume an attrition rate of 30%, which is reasonable for a randomized field trial involving small companies.

Based on these assumptions, we require a sample size of 166 hotels. Thus, our initial sample size of 203 appears to be adequate.

Appendix C

Sample Selection: Comparison of Key Characteristics across Participant (n=215) and Non-Participant (n=383) Hotels

Hotel characteristics	Participant	Non-Participant	Difference
Segment			$\chi^2_{(2)}$ =.42; p=.81
Budget	19%	21%	
Mid-price/Suite	57%	57%	
Boutique	24%	22%	
Location			$\chi^2_{(1)}$ =.38; p=.54
Urban	50%	47%	
Tourist	50%	53%	
Size			$\chi^2_{(1)}$ =.05; p=.82
<30 rooms	52%	51%	
>30 rooms	48%	49%	

Respondent characteristics	Participant	Non-Participant	Difference
Education			$\chi^2_{(3)}$ =3.87; p=.28
Primary	35%	29%	
Secondary (Hotel)	43%	44%	
Secondary (General)	19%	20%	
Tertiary	3%	6%	
Age	50.0	48.1	t=1.91; p=.06
Work experience (yrs)	3.2	3.5	t=1.80; p=.07

Perceptions on reviews (1-7)	Participant	Non-Participant	Difference
Credibility of online reviews	3.9	3.7	t=1.64; p=.10
Reviews are informative (score)	4.0	3.8	t=1.60; p=.11
Reviews are informative (text)	3.9	4.1	t=1.38; p=.17
Decision confidence	4.1	4.0	t=0.39; p=.70
Usefulness of online reviews	4.7	4.4	t=1.99; p=.05
Lack of time for online reviews	3.3	3.6	t=1.49; p=.14
Lack of staff for online reviews	3.5	3.6	t=0.90; p=.37

Appendix D

Panel: Control Group

Hotel

Report 1st March 2016

During the period February 1st, 2015- February 1st, 2016 your hotel received:

72 Reviews on TripAdvisor

The Ratings of Your Customers

In summary, how customers judge your offering

General 4

Location	4.5	Services	4.5
Rooms	4.5	Cleanliness	4.5
Sleeping quality	4.5	Price/Quality	4

Analysis of Ratings

Trends and Variations of Ratings

Topics	Travel Appeal Index	Var.	Estimation	Reviews	Last Review
General	84.80/100	– 0.00	4.24/5	73	12 days ago
Cleanliness	80.60/100	– 0.00	4.03/5	8	24 days ago
Price/Quality	74.60/100	– 0.00	3.73/5	12	24 days ago
Sleeping Quality	95.80/100	– 0.00	4.79/5	14	25 days ago
Rooms	81.80/100	– 0.00	4.09/5	10	29 days ago
Location	85.40/100	– 0.00	4.27/5	12	35 days ago

*

Recensito 19 agosto 2016 tramite dispositivo mobile

Pernotto

Abbiamo dormito una notte in questo hotel, stanza bella e pulita personale della reception molto cordiale. Buona colazione ma il personale addetto poco attento. Ambiente silenzioso e confortevole. Ci siamo trovati bene consigliabile a chiunque voglia dormire in un hotel a quattro stelle. Più

Grazie, Franco I

Risposta da SIMONA T, Proprietario presso Hotel Airone

Risposta inviata 22 agosto 2016

Gent.mo sig.Franco, la sua recensione è stata quantomai preziosa e la ringrazio per aver trovato il tempo per lasciare la sua ottima testimonianza. Sarei felice di poterla accogliere in un momento di minor affollamento, come appunto è il mese di Agosto, quando l'Hotel è al... Più

Recensito 17 agosto 2016

bellissimo hotel

Un bellissimo hotel con tutti i confort possibili. Camere moderne e ben curate con cassaforte, frigo bar, televisore molto grande, aria condizionata e un bagno abbastanza grande. Ottima anche la spa che si trova la quarto piano con piscina, sauna, minipalestra ecc ecc. Prezzo nella... Più

Grazie, EdoardoPuri16

Risposta da SIMONA T, Proprietario presso Hotel Airone

Risposta inviata 18 agosto 2016

Gent.mo sig.Edoardo, grazie per aver recensito "con lode" tutti i nostri servizi, compreso il prezzo che è in linea con gli altri Hotel della zona che però sono decisamente più vetusti. La location del nostro Hotel è centrale, in una zona residenziale, accanto al Tribunale... Più

Recensito 16 agosto 2016 tramite dispositivo mobile

LA PERFEZZIONE

Beh è davvero un hotel con i fiocchi. Educazione e servizio di primo ordine. Poi la struttura è super curata fresca in ogni angolo e sopratutto profumata qui si respira aria di nuovo neanche fosse appena stato costruito. Davvero davvero un ottimo posto credo che... Più

Grazie, Daniele A

Risposta da SIMONA T, Proprietario presso Hotel Airone

Risposta inviata 18 agosto 2016

Gent.mo sig.Daniele, l'Airone sarà anche un hotel "con i fiocchi", ma lei è un cliente a 5 stelle!!! Credo che sia impossibile scrivere una recensione più bella della sua, grazie, siamo commossi ed orgogliosi, tant'è che l'ho stampata e fatta vedere a tutto il personale... Più

Recensito 15 agosto 2016 tramite dispositivo mobile

Piccola delusione!!

Posso capire il periodo 14 e 15 Agosto stress pienone e tutto quello che volete....ma la gentilezza in alcuni casi non deve mancare MAI. Fai colazione e senti i dipendenti che discutono tra di loro e devi alzarti te dal tavolo per ordinare un cappuccino......parliamo... Più

*Sample of review texts

Appendix D (Contd.)

Panel: Treatment Group

Hotel

1st March 2016

During the period
February 1st, 2015- February 1st, 2016
your hotel received:

19 Reviews on TripAdvisor

The Ratings of Your Customers

In summary, how customers judge your offering

General 3.5

Location	4.5	Services	4.5
Rooms	4.5	Cleanlinness	4.5
Sleeping quality	4.5	Price/Quality	4

Analysis of Ratings

Trends and Variations of Ratings

Topics	Travel Appeal Index	Var.	Estimation	Reviews	Last Review
General	68.40/100	— 0.00	3.42/5	18	31 days ago
Cleanliness	95.20/100	— 0.00	4.76/5	3	85 days ago
Price/Quality	87.20/100	— 0.00	4.36/5	8	85 days ago
Sleeping Quality	97.00/100	— 0.00	4.85/5	3	85 days ago
Rooms	68.45/100	▾ -0.13	3.74/5	3	140 days ago
Location	88.65/100	▾ -0.13	4.75/5	4	140 days ago

*

Recensito 3 gennaio 2016

Almeno era pulita...

Positivo: una sola, era molto pulito negativo(s) : - La colazione era di qualità scadente - Due giorni di fila (tra cui il 1° gennaio!!!!), il personale di pulizia era al lavoro al... 8:00 AM, che bussano alle porte e cercando di entrare nelle... Più

Esprimi un giudizio sulla traduzione. Dai un punteggio

Recensito 17 dicembre 2015

Ottimo rapporto qualità prezzo

Posizione strategica e comoda nel centro di Legnano con parcheggio gratuito annesso. Per soli 50 sesterzi ho pernottato in questo Hotel molto silenzioso, pulito ed abbastanza moderno. Personale gentile. Alla fine non manca nulla e costa molto nemo di altri in zona. Unico piccolo difetto... Più

1 Grazie, Bruno G

Recensito 30 novembre 2015 tramite dispositivo mobile

Da ritornare

Il personale maschile molto gentile. Buona pulizia. Camera confortevole... a parte il cuscino... troppo morbido. Colazione discreta.. considerato che c'è una signora non molto carina.. nel comportamento. Suggerirei di riscaldare meglio la sala. Per il resto buono.

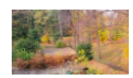
Grazie, Mel A

Questa recensione è stata tradotta dall'inglese con traduttore automatico Di cosa si tratta?
Mostra traduzione automatica? Sì No

Recensito 28 novembre 2015

Piccolo Hotel Conveniente in Legano, Italia

Ho trascorso una settimana in questo piccolo hotel e abbiamo apprezzato l'ampio parcheggio, il prezzo ragionevole le camere pulite e la bella vista dalla mia camera. Se siete a Legnano o nelle vicinanze è una scelta eccellente.



Esprimi un giudizio sulla traduzione. Dai un punteggio | Migliora la traduzione

Recensito 13 ottobre 2015

AVOID

Google Traduttore

We were on business and stayed in this hotel for the location. Save your money and the stress of having... Più

1 Grazie, lucia903

*Sample of reviews texts

Hotel

1st March 2016

The level of sentiment of each topic

Topics	Number of opinions	Sentiment	Level of sentiment
Hotel: general	41	85.02/100	
Room	23	45.95/100	
Staff: general	23	72.90/100	
Location	14	74.44/100	
Appearance	11	100.00/100	
Room: bathroom	9	100.00/100	
Restaurant	9	64.31/100	
Services	6	100.00/100	
Interior design	3	31.36/100	
Room: bed	3	100.00/100	
Room: entertainment	3	0.00/100	
Reception	2	100.00/100	
Meals	2	100.00/100	
Room: forniture	2	100.00/100	
Room: technology	2	33.82/100	
Room: structure	1	100.00/100	
Room: bed elements	1	100.00/100	
Room: structure of bathroom	1	100.00/100	
Room: bathroom forniture	1	0.00/100	

The most influential topics of the reviews

The most influential topics that emerged from the semantic analysis

The general level of the sentiment is 81,09%

Appendix E

Threats from Sample Attrition

We experienced some attrition during the six-month period of the RCT: 68 hotels (37 in the control condition and 31 in the treatment condition) did not reply to our intermediate and final surveys (33.7%) resulting in a usable sample of 135 hotels. Moreover, as the purpose of our intermediate surveys was to ensure that hotel managers were participating in our RCT, it is likely these hotel managers were not affected by our stimuli, as they did not comply with our treatment protocol (our robustness check #7 confirms this assumption).

To investigate sample attrition, we first estimated a Probit model (Fitzgerald, Gottschalk, and Moffitt 1998) to assess whether the attrition was systematic and could be explained by any observable characteristics. The results indicated that the extent of missing cases was homogeneous across control and treatment conditions (b = -.18; p = .35) and was independent of hotel class ($\chi^2_{(2)} = 1.61$; *ns*) and location ($\chi^2_{(3)} = 5.86$; *ns*). Only one observable characteristic was significantly associated with the likelihood of attrition, the hotel's size (b = .57; p < .02); large hotels (more than 30 rooms) were 19% more likely to drop out from the RCT.

Both missing and retained hotels in the RCT had similar pre-treatment (b = -.02; p = .93) and post-treatment (b = -.23; p = .31) service quality. Finally, sample attrition did not depend on the attitude of the hotel's managers toward online consumer reviews (measure in Appendix A) measured by the extent to which the managers perceives online reviews as being credible and correct (b = -.01; p = .81).

Table E-1: Probit Model to Assess Sample Attrition

Attrition	Parameter estimate	Standard errors*	P>|z|
Treatment	-.140	.196	0.48
Typology (Budget)			
Mid-price/Suite	-.001	.289	0.99
Boutique	-.253	.364	0.49
Location (Urban)			
Seaside	-.395	.338	0.24
Mountainside	-.462	.234	0.05
Lakeside	-.683	.376	0.07
Size	*.563*	*.234*	*0.02*
Pre-treatment performance	-.006	.187	0.98
Post-treatment performance	-.262	.222	0.24
Credibility of online reviews	-.006	.064	0.92
Constant	.236	.838	0.79

N = 191; *Robust Standard Errors