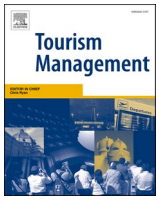
Tourism Management 85 (2021) 104280

Contents lists available at ScienceDirect

Tourism Management

journal homepage: http://www.elsevier.com/locate/tourman



Effects of Booking.com's new scoring system

Juan Pedro Mellinas [a,*], Eva Martin-Fuentes [b]

[a] Facultad de Ciencias de La Empresa, Universidad Politécnica de Cartagena, Calle Real, 3, 30201, Cartagena, Spain
[b] Business Administration Department. University of Lleida, Campus de Cappont, C/ Jaume II, 73, 25001, Lleida, Spain

ARTICLE INFO

Keywords:
Booking.com
Scores
Reviews
Scale
Ratings

ABSTRACT

Booking.com provides a massive database compiling millions of reviews about thousands of accommodations worldwide that hotel managers and academics have extensively consulted during the past decade. In 2019–2020, however, the famous website changed several aspects of its methods of calculating hotel scores, the most important one being a change from its peculiar 2.5–10 scale to a more conventional 1–10 scale. Such novelties may cause changes in hotel scores that do not reflect changes, if any, in customer satisfaction. This article offers an initial investigation into the nature and consequences of those changes that professionals and academics should consider to avoid errors in future studies that involve using Booking. com's database.

1. Introduction

With a database of approximately 180 million verified reviews from real guests (Booking, 2020), Booking.com positions itself as the website that collects the most hotel reviews (Murphy, 2017), even more than TripAdvisor. In turn, since 2010, the website's massive amount of data has attracted the attention of various authors, who continue using the information in this database, with increasing frequency (Mariani et al., 2020; Phillips et al., 2020). Indeed, a quick bibliographic search for "Booking.com reviews" in a top journal such as *Tourism Management* returns only one article per year between 2013 and 2014 but six or seven between 2015 and 2019.

In 2015, one article publicized that Booking. com's rating scale was not a 0–10 or 1–10 scale, as assumed in 12 academic articles already published, but a controversial 2.5–10 scale (Mellinas et al., 2015). In addition, the article explained that the final score for each hotel derives from the arithmetic mean of the six items evaluated, with four scoring options represented by four smiley faces respectively equivalent to scores of 2.5, 5, 7.5, and 10.

Subsequently, other authors have investigated the effects of Booking. com's scale on frequency distributions (Mariani & Borghi, 2018) and distortions that the scoring system may cause in overall hotel scores (Martin-Fuentes et al., 2020; Mellinas et al., 2016; Parra et al., 2018). In their studies, these authors compared Booking. com's scores for hotels with scores on other websites, including Priceline, Agoda, Travel Republic, and HRS. Their results suggested that hotels with low ratings scored far higher on Booking.com than in systems with a 0–10 or 1–10 scale, medium-scoring hotels scored slightly higher, and high-scoring ones had similar scores in all systems. Somehow, they were trying to predict what would happen if Booking.com had a more conventional rating scale. In any case, such differences, in addition to the different scales used, could be explained by other factors, including a different method of calculating the final score, each website's different user profile, and different ways of presenting questionnaires.

In September 2019, Booking.com (2019) announced the introduction of changes to its rating system: "Now guests select an 'overall' score themselves. Smiley faces are replaced by a new sliding scale of 1–10 for this score." On another Booking.com page (Booking.com, 2020), the website details how the questionnaire now looks, which is basically identical to the previous one, but adding a first question "How was your stay?" with a sliding scale of 1–10. The rest of the questionnaire is identical and the smileys with a scale of 2.5–10 are still used.

In 2020, however, despite once calculating the global score for each hotel based on reviews posted in the previous 24 months, Booking.com announced the following:

> "To prevent you from getting negatively affected by a loss of guest reviews, Booking.com will gradually extend the lifetime of your existing reviews from two to three years. This means reviews that were about to expire as of July 31, 2020 will remain for an extra year." (Booking.com, 2020)

The announcement, made in 2019 within Booking. com's so-called "Partner Hub", has received more than 300 comments from owners of accommodations, most of whom expressed their disagreement with the

* Corresponding author.
E-mail addresses: juan.mellinas@upct.es (J.P. Mellinas), eva.martin@udl.cat (E. Martin-Fuentes).

https://doi.org/10.1016/j.tourman.2020.104280
Received 22 August 2020; Received in revised form 28 November 2020; Accepted 5 December 2020
Available online 17 December 2020
0261-5177/© 2020 Elsevier Ltd. All rights reserved.

new system, mostly due to the drop in scores that they had observed as a result. In addition, on the Partner Hub, one such owner posted "We urge BDC [Booking Dot Com] to go back to the previous system where the final review is a calculation based on the category reviews" and received 200 comments in response, in most of which various owners complain about the drop in scores with the new system (Booking.com Partner Hub, 2019). Some examples of the new situation are provided in the comments:

> "(Vicki Webber) We too are experiencing unfair difficulties with the new rating system.
>
> EXAMPLE: ***. 10 for all the individual scores and then a 9 overall?
>
> EXAMPLE: ***. 9.8 for all the individual scores and then an 8 overall?
>
> This new system is flawed and simply does not work."
>
> "(pibomarco) But when you will receive 7,5 for all individual scores and a 9 overall in this case you will probably not complain.) Sometimes you will receive a higher overall score and sometimes higher individual scores."

Table 1 lists different situations that may cause Booking. com's scores for each hotel to increase or decrease due to the mentioned changes. However, because we do not know which of the listed elements will have more or less weight, we do not know whether the new system can make high and low scores rise or fall.

Scholars have paid great attention to whether hotel scores on Booking.com will indeed drop, as well as to what extent, due to the change in the system as predicted (Martin-Fuentes et al., 2020; Mellinas et al., 2016) and as the comments of some hoteliers in online forums suggest.

Because scores along with reviews using the former system will remain active until September 2022, the most orthodox way to analyze changes would be to take the score for a sample of hotels from September 2019 (i.e., all scores from the old system) and repeat the operation in September 2022 such that all reviews use the new system. Collecting large volumes of online reviews through a big data and analytics approach could offer accurate results on the variations produced during those three years. Doing so, however, would require a waiting period, and variations could then stem from both the new system and changes in the conditions of each hotel after 3 years. Beyond that, the exceptional situation affecting the period — that is, the COVID-19 pandemic — could also affect the scores.

Against that background, the purpose of this article is to report a preliminary study that involved estimating real variations that may occur in Booking. com's scores for hotels around the world, without having to wait 2 years to obtain definitive results. We anticipate that such information is of great interest not only to hoteliers, who seem to have begun detecting the effects of changes in the system and have shown their concern, but also to the scientific community that uses the database with increasing regularity.

Table 1
Possible situations with Booking.com's new scoring system.

	Very dissatisfied	Very satisfied
Might affect negatively	Very dissatisfied customers can assign a rating of 1 or 2 instead of 2.5. If the guest had a horrible experience, then evaluating some objectively positive aspect (e.g., location or staff) no longer raises the average.	Customers who are very satisfied in all aspects can assign a maximum score on a scale of 4 smileys (i.e., equal to 10), but when offered a scale of 1–10, they can select 8 or 9 if they feel that the service was great but not perfect.
Might affect positively	The guest may consider that though the hotel is of low quality, the overall value for the price is high, thereby encouraging the guest to assign a high score in the overall rating.	It is not necessary to assign the maximum score to all parameters in order to obtain a score of 9 or 10.

2. Methodology

Booking.com provides the reviews of each hotel on two webpages, which have nearly identical content but different formats, as can be observed at the following URLs for the same hotel:

- Reservation website: https://www.booking.com/hotel/us/pennsylvania-new-york.html
- Reviews website: https://www.booking.com/reviews/us/hotel/pennsylvania-new-york.html

Since Booking.com implemented its new system in September 2019, we have tracked the scores received by a random sample of hotels and observed that the system's implementation has been gradual — that is, scores calculated with the former system (i.e., scores with decimals) have been combined with others seemingly calculated with the new system (i.e., scores in whole numbers) on the reservation website. However, on the reviews website, all scores shown were calculated with the old system until April, when scores with the new system began appearing.

In March 2020, Booking.com updated its webpage describing how reviews are collected (Booking.com, 2020). Since then, all reviews registered on the reservation website have had scores calculated with the new system, albeit with the same reviews — same hotel, user, and date — being shown on the reviews website, accompanied by scores calculated with the former system. Such divergence indicates, with total precision, the difference between scores obtained with the old and new systems for hundreds of reviews and hotels.

On March 27, 2020, we attempted to create a large database of such reviews by looking for ones posted during March. However, because most countries had recently announced the closure of borders and hotels due to the COVID-19 pandemic, the task proved exceptionally difficult. Ultimately, we compiled a database of hotels in cities where the lockdown arrived somewhat later: London, Las Vegas, New York City, Miami, Rio de Janeiro, and Moscow. In sum, the database comprised 19 hotels with 400 total reviews registered in March and with scores from both the old and new systems. When hotels began to open in some countries in June 2020, we tried to expand that sample of 400 reviews. However, it was impossible because we noticed that Booking.com now showed the same scores on "Reservation website" and "Reviews website".

To check the normality of the data, a Kolmogorov-Smirnov test was performed, which confirmed that the Booking.com ratings were not following a normal distribution, as can be seen in the results section. Then, we performed a nonparametric kernel density estimator test following the same procedure as Mariani and Borghi (2018), which is used with different populations (e.g. old and new Booking.com score), and when they do not follow normality, we can construct a density function for each sample according to this criterion and classify the new individual simply by assigning it to the population where there is the highest density value, which will mean that over that, there is more influence of that population.

3. Results

The average score for the sample of reviews obtained with the old system was 7.606 and with the new system 7.062 (Table 2). The difference of 0.544 was slightly higher than the difference of 0.470 determined in a comparative study of hotel ratings on Booking.com and Priceline.com (Mellinas et al., 2016) and substantially higher than differences calculated when hotel scores on Booking.com were compared

Table 2
Descriptive statistics of the Booking.com scoring system.

	N	Min	Max	Mean	SD	Skewness Stat.	Error	Kurtosis Stat.	Error
New scoring system	400	1	10	7.062	2.64	-.878	.122	-.040	.243
Old scoring system	400	2.5	10	7.606	2.03	-.678	.122	-.293	.243

with scores on other databases: 0.130 on Travel Republic, 0.151 on Atrápalo, and 0.338 on HRS (Parra et al., 2018).

Although the frequency distribution calculated for the old system (Fig. 1) coincided with the one obtained by (Mariani & Borghi, 2018) with a large dataset of hotels in London, we identified important changes when using the new system's frequency distribution (Fig. 2). In particular, the number of totally dissatisfied customers (i.e., score of 1.7%) equaled that of partly dissatisfied ones (i.e., scores of 2, 3, or 4, 7.9%). Thus, when customers were dissatisfied, in most cases they were extremely dissatisfied.

Along similar lines, the percentage of reviews with scores less than 5 rose from 9.6% with the old system to 14.9% with the new one. By contrast, however, the percentage of customers who gave scores from 9 to 10 was highly similar between the old and new systems (33.3% and 33.8%, respectively), with the particularity that, in the old system, the maximum score of 10 was assigned by 18% of reviewers and 22% in the new system. Reviews that currently assign a score of 10 mostly corresponded not only with scores of 10 in the previous system but also with lower scores of 8.8, 9.2, and 9.6, which confirms a situation anticipated in Table 1 and as previously suggested (Martin-Fuentes, Mateu, & Fernandez, 2018) when proving that high-rated hotels on Booking.com would benefit from a scoring system that assigns ratings directly without using the arithmetic mean of several items.

The new system has also significantly affected scores from totally or partly dissatisfied customers. Among them, those ones who assigned scores of less than 5 in the old system gave nearly 2 points less on average with the new system, which also corroborates what is suggested in Table 1.

A Kolmogorov–Smirnov test was performed to determine whether normality existed in the difference of means between the scores in the old system and scores in the new one. The results confirmed that the difference of means did not follow a normal distribution ($p < 0.001$). In response, a nonparametric test to compare mean scores for related samples, the Wilcoxon signed-rank test, was performed; its results indicated that 106 scores in the old system were lower in the new system, that 223 scores in the old system were higher, and that 71 scores were identical. Such results also revealed significant differences between scores in the old versus the new system ($p < 0.001$), with lower scores in the latter.

Last, based on previous research (Mariani & Borghi, 2018) and following the same methodology, we performed a nonparametric kernel density estimator test.

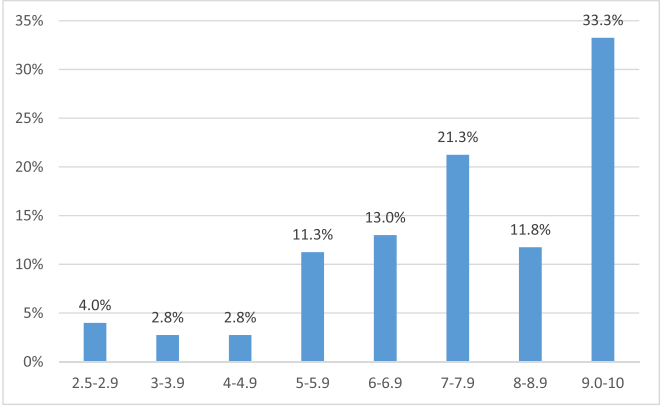


Fig. 1. Distribution of Booking.com hotel scores with the old system.

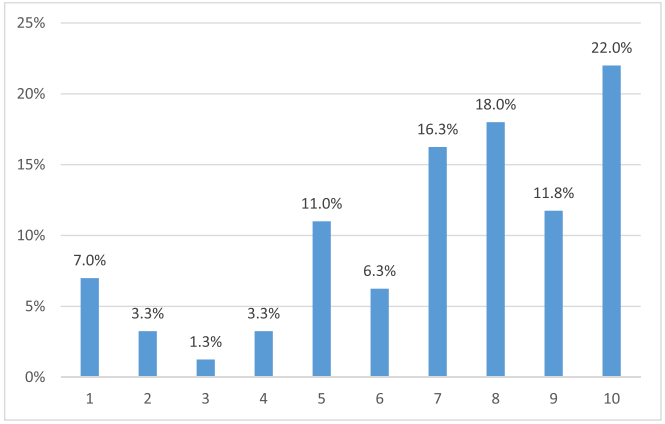


Fig. 2. Distribution of Booking.com hotel scores with the new system.

As shown in Fig. 3, although both distributions were left-skewed, they statistically differed.

4. Conclusions

Differences between Booking. com's old and new scoring systems, at least for the hotel scores analyzed, were closer to the half-point predicted by (Mellinas et al., 2016) when comparing with scores of U.S. hotels on Priceline.com than in the study of Parra et al. (2018) and Martin-Fuentes et al. (2020), in which differences of two-to three-tenths were observed.

Our findings seem to confirm that the differences will be significant, close to one for hotels with low scores, and possibly minimal in hotels with high scores. However, due to the small size of our sample, the results obtained do not allow us to offer an accurate estimate of the variations that will occur, which can be overcome only by collecting large volumes of online reviews through a big data and analytics approach (M. Mariani et al., 2018). Furthermore, these variations could also be affected by the sociodemographic characteristics of the users or the use of different devices (mobile vs. desktop) (M. M. Mariani, Borghi, & Gretzel, 2019).

Factors listed in Table 1 that could favor the lowering of scores for very dissatisfied customers seem to outweigh the sole factor that could make them rise. However, it remains somewhat unclear which ones weigh more for very satisfied customers: factors favoring the rise in scores or ones causing them to drop.

Therefore, hotels with currently low and medium scores should expect substantial drops in their scores. However, those scores will not correspond to drops in customer satisfaction or changes in consumers' perceptions due to new tourism regimes amid the COVID-19 pandemic. For that reason, attempting to take corrective measures to solve a seeming problem that does not in fact exist would be futile, if not damaging.

Although a significant drop in a hotel's score can have dramatic consequences in a normal situation, hoteliers should conduct a rational analysis taking into account all available data, including that in this paper. However, we fear that many hoteliers (with low-scoring properties) may be concerned when they find that their scores drop by almost one point, while those of other hotels (with high scores) are hardly

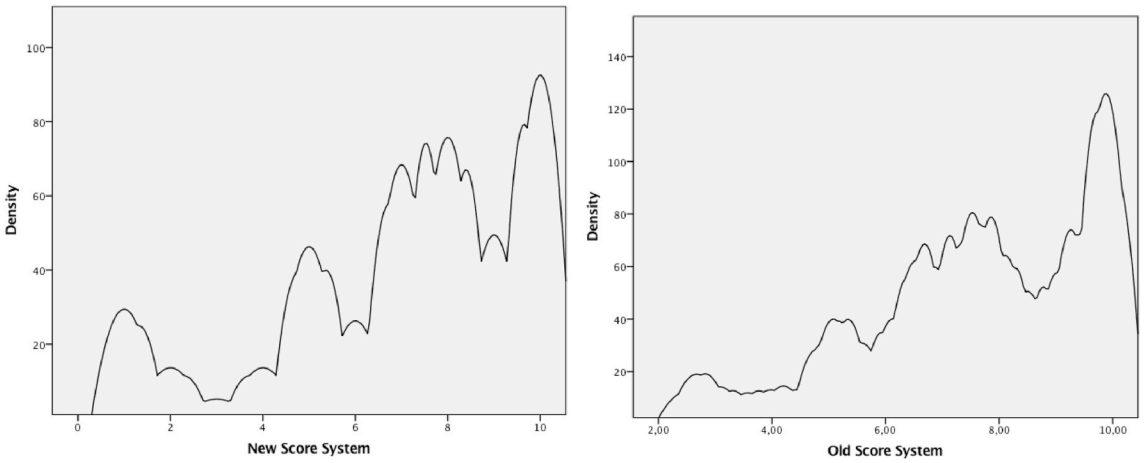


Fig. 3. Kernel density estimator test for the new and old scoring systems on Booking.com.

affected by this change. It can be difficult for them to understand that Booking.com is implementing a new system that is more rational and more in accordance with market standards and that they do not intend to harm a specific group of hoteliers.

For academics, this article can serve as an initial reference for future studies on changes that Booking. com's new scoring system may cause. The preliminary data offered here, unlike the data obtained in the future, are distorted neither by the traumatic experience of COVID-19 nor by real changes that may occur in each hotel in the near future.

From 2015 to 2020, the scientific community has worked intensively with databases provided free of charge by Booking.com. In most cases, scholars have taken into account the 2.5–10 scale used by the website and thus avoided errors committed prior to the publication of the article that made that unexpected scale known (Mellinas et al., 2015). In the coming years, academics who continue examining these reviews should bear in mind that the system now uses a 1–10 scale and should calculate scores based on reviews from the last 3 years to prevent substantial statistical errors. Studies based on comparing hotel ratings on Booking.com (e.g. Nicolau et al., 2020) at an initial date and a later that do not take into account the new scale could give totally wrong results.

In that light, it should be taken into account that, because Booking.com now uses reviews from the past 3 years to calculate final scores, these final scores derive from the average of individual scores obtained with the old system (2.5–10) and the new one (1–10). Therefore, all studies based on hotel ratings on Booking.com collected between September 2019 (i.e., first scores with the new system) and September 2022 (i.e., 3 years since the total application of the new system) will have an important methodological limitation that must be clearly indicated.

Once the new system is completely implemented on Booking.com, research using ratings can continue to be carried out with any methodology previously used with the old system such as OLS regression analysis (M. M. Mariani, Borghi, & Kazakov, 2019), Tobit regression models (M. M. Mariani et al., 2020), or Support Vector Machine (Martin-Fuentes, Mateu, & Fernandez, 2018), depending on the goal of each research.

For example, researchers who collect hotel scores on Booking.com in October 2021 should consider that those scores originate from the average user scores between October 2018 and October 2021. That circumstance implies the following situations:

- From October 2018 to September 2019, scores represent the old system;
- From September 2019 to March 2020, scores represent a mix of the old and new systems; and
- From March 2020 to October 2021, scores (will) exclusively represent the new system.

Ignoring both the new scale and using reviews made with different scales to calculate the mean could cause systematic errors in results obtained during the next 3 years. If authors, reviewers, and editors remain unaware of that possibility, then articles with erroneous results could be published in academic journals. This problem could be extended to other review platforms, which can also make changes and not advertise them. Hoteliers and academics must fully know the characteristics of the information they obtain from the Internet, to avoid errors in their analysis.

Impact statement

Booking.com altered in 2019–2020 how its calculates hotel scores, from using a 2.5–10 scale to a 1–10 scale, from using scores based on the average of six items to user-generated global scores, and from deleting reviews after 2 years to deleting them after 3 years.

Hotels around the world, especially those with medium and low scores, are going to see their scores significantly reduced by these changes. Therefore, hoteliers have to analyze whether drops in their scores stem from the deterioration of their services or from changes to Booking. com's methods. Attempting to take corrective measures to solve a seeming problem that does not in fact exist would be futile, if not damaging.

Academics should also consider these changes when working with scores obtained from Booking.com database through the end of 2022.

Credit author statement

Juan Pedro Mellinas contributed to conception, Methodology, data collection, building the literature review and writing the manuscript. Eva Martin-Fuentes contributed to the literature review, data análisis and writing the manuscript.

Declaration of competing interest

None.

Acknowledgments

This work was partially funded by the Spanish Ministry of the Economy and Competitiveness: research project TURCOLAB ECO 2017-88984-R. This research article has received a grant for its linguistic revision from the Language Institute of the University of Lleida (Spain) (2020 call).

References

Booking. (2020). *Reviews: How does it work? Booking.Com*. https://www.booking.com/reviews.en-gb.html.

Bookingcom. (2020). *How is my guest review Score generated? Booking.Com for partners*. https://partner.booking.com/en-us/help/guest-reviews/how-my-guest-review-score-generated.

Booking.com Partner Hub. (2019). *Hosts asking for justice about the new review system*. Booking.Com for Partners. https://partner.booking.com/en-gb/community/partner-feedback/hosts-asking-justice-about-new-review-system.

Mariani, M., Baggio, R., Fuchs, M., & Höepken, W. (2018). Business intelligence and big data in hospitality and tourism: A systematic literature review. *International Journal of Contemporary Hospitality Management, 30*(12), 3514–3554. https://doi.org/10.1108/IJCHM-07-2017-0461

Mariani, M. M., & Borghi, M. (2018). Effects of the Booking.com rating system: Bringing hotel class into the picture. *Tourism Management, 66*, 47–52. https://doi.org/10.1016/j.tourman.2017.11.006

Mariani, M. M., Borghi, M., & Gretzel, U. (2019). Online reviews: Differences by submission device. *Tourism Management, 70*, 295–298. https://doi.org/10.1016/j.tourman.2018.08.022

Mariani, M. M., Borghi, M., & Okumus, F. (2020). Unravelling the effects of cultural differences in the online appraisal of hospitality and tourism services. *International Journal of Hospitality Management, 90*, 102606. https://doi.org/10.1016/j.ijhm.2020.102606

Martin-Fuentes, E., Mateu, C., & Fernandez, C. (2018). Does verifying users influence rankings? Analyzing Booking.Com and tripadvisor. *Tourism Analysis, 23*(1), 1–15. https://doi.org/10.3727/108354218X15143857349459

Martin-Fuentes, E., Mellinas, J. P., & Parra-Lopez, E. (2020). Online travel review rating scales and effects on hotel scoring and competitiveness. *Tourism Review*.

Mellinas, J. P., Martínez María-Dolores, S.-M., & Bernal García, J. J. (2015). Booking.com: The unexpected scoring system. *Tourism Management, 49*, 72–74. https://doi.org/10.1016/j.tourman.2014.08.019

Mellinas, J. P., Martínez María-Dolores, S.-M., & Bernal García, J. J. (2016). Effects of the Booking.com scoring system. *Tourism Management, 57*, 80–83. https://doi.org/10.1016/j.tourman.2016.05.015

Murphy, C. (2017). *Report: 78% of all online hotel reviews come from the top four sites*. Revinate https://learn.revinate.com/blog/report-78-of-all-online-hotel-reviews-come-from-the-top-four-sites.

Nicolau, J. L., Mellinas, J. P., & Martín-Fuentes, E. (2020). The halo effect: A longitudinal approach. *Annals of Tourism Research, 83*, 102938. https://doi.org/10.1016/j.annals.2020.102938

Parra, E., Mellinas, J. P., Martínez María-Dolores, S.-M., Bernal Garcia, J. J., & Gutiérrez-Taño, D. (2018). Effects of reviews scales on hotel online reputation. *Turitec*, 98–116, 2018.

Phillips, P., Antonio, N., de Almeida, A., & Nunes, L. (2020). The influence of geographic and psychic distance on online hotel ratings. *Journal of Travel Research, 59*(4), 722–741. https://doi.org/10.1177/0047287519858400

Juan Pedro Mellinas. He holds a PhD in Business Administration; a MSc Tourism Planning and Management and a BA in Business Administration. Currently he is lecturer for the Department of Business Management at the Universidad Politécnica de Cartagena. He has experience working in international corporations and as entrepreneur for 15 years. His research focuses on online reviews in websites like Booking and TripAdvisor among others. He has published in Tourism Management, Annals of Tourism research and Tourism Review, among other journals.

Eva Martin-Fuentes. She holds an international PhD in Engineering and Information Technologies; a MSc in Tourism Planning and Management; a BA in Advertising and Public Relations; and a BA in Tourism. She is lecturer for the Department of Business Management at the University of Lleida (Spain) where she has been recently recognized with the Teaching Excellence Award for the areas of tourism management and social media. She has been working as a tourism manager at the Tourism Board of Lleida and on events organization at the University of Lleida. She has published in Journal of Hospitality and Tourism Management, International Journal of Hospitality Management and Tourism Review, among other journals.