



### Average Scores Integration in Official Star Rating Scheme

Journal:	<i>Journal of Hospitality and Tourism Technology</i>
Manuscript ID	JHTT-07-2017-0050.R3
Manuscript Type:	Refereed Article
Keywords:	eWOM, Hotels, Average Scores, Star Rating Systems, Online Reviews

SCHOLARONE™  
Manuscripts

When revising your paper, please prepare this report explaining how you have responded to each reviewer's comments and suggestions specifically.

#### REVIEWER 1

Suggestions/comments from the Reviewer	Response from the Author(s)
<p>1. Make sure to follow JHTT author guidelines closely: <a href="http://www.emeraldgrouppublishing.com/products/journals/author_guidelines.htm?id=jhtt">http://www.emeraldgrouppublishing.com/products/journals/author_guidelines.htm?id=jhtt</a> For example, when there are three or more authors, you need to use Adam et al., XXXX (or Adam et al., XXXX) format for the first time and after.</p>	<p>The paper has been duly reviewed to ensure that the JHTT author guidelines have been met. Particular attention has been made to the in-text citation format for when three or more authors.</p>
<p>2. Revisit the Introduction to highlight the research purpose, research problem, and significance of the study. Please explain these in application to both hotels and hotel guests. Revisit the Discussion and Conclusions sections one more time to better answer the "So What" question. There should be four sub-sections under this section: (1) Conclusions, (2) Theoretical Implications, (3) Practical Implications and (4) Limitations and Future Research.</p>	<p>Explicit mention to the research purpose, research problem and significance (i.e. implications of the study) for both hotels and guests have been included in the introduction. We have revisited the Discussion and Conclusion sections to better answer the "So What" question and have divided the section in the requested 4 subsections.</p>
<p>3. Cross check all references within text with your reference list. You may like to add more recent and relevant references published in recent months/years in JHTT or other journals.</p>	<p>A thorough check that all the in-text citations are in the reference list has been made, and additional references from work published in recent months/years in relevant journals (including JHTT) have been added.</p>
<p>4. Keep your article between 6000 and 7000 words including references, tables and figures.</p>	<p>The article is 6501 words including abstract, figures and references.</p>
<p>5. Proofread your article one more time and also you may ask a technical writer/copy editor to proofread it for you. After the manuscript is accepted, we will not ask you to proofread it again. In short, after I send you an official acceptance e-mail, you will not be able to make any further changes in your manuscript.</p>	<p>We have proofread the article one more time and asked a technical colleague to proofread it for us.</p>
<p>6. Submit a clean version of your paper. You don't need to show/highlight all the changes made in the paper. I will read its final version anyway.</p>	<p>OK, thank you. We are submitting a clean version of the paper as suggested.</p>
<p>7. Include a brief report showing how you have responded to the above requests. You don't need to show/highlight all the changes made in the paper. I will read its final version anyway.</p>	<p>We are including this report in response to this request.</p>
<p>8. Finally, take the above suggestions seriously when revising the article. Failure to do so may result in an unfavorable decision as we do not really have time to work on an article multiple times.</p>	<p>We have taken all these comments very serious and made all the amendments to the best of our abilities. We hope that this version of the paper fulfils all the requirements for its publication in the JHTT.</p>

## Average Scores Integration in Official Star Rating Scheme

### Abstract

Purpose of this paper: Evidence suggests that Electronic Word of Mouth (eWOM) plays a highly influential role in decision-making when booking hotel rooms. The number of online sources where consumers can obtain information on hotel ratings provided has grown exponentially. Hence, a number of companies have developed Average Scores to summarise this information and to make it more easily available to consumers. Furthermore, Official Star Rating schemes are starting to provide these commercially developed Average Scores to complement the information their schemes offer. This raises questions regarding the robustness of these systems, which this paper addresses.

Design/methodology/approach: Average Scores from different systems, as well as the scores provided by one rating site were collected for 200 hotels and compared.

Findings: Findings suggested important differences in the ratings and assigned descriptive word across websites.

Research limitations/implications (if applicable): The results imply that the application of Average Scores by Official organisations is not legitimate and identifies a research gap in the area of consumer and star rating standardization.

What is original/value of paper: The paper is of value to the industry and academia related to the examination of rating scales adopted by major online review tourism providers. Evidence of malpractice has been identified and the adoption of this type of scales by Official Star Rating schemes is questioned.

**Keywords:** eWOM, Hotels, Average Scores, Star Rating Systems, Online Reviews

### 1. INTRODUCTION

Social media in general, and Electronic Word of Mouth (eWOM) in particular, have completely restructured business consumer relationships. eWOM refers to non-formal online communication about products, services or their sellers (Litvin et al., 2008). A growing number of websites allow consumers to post their opinions about products, contributing to eWOM. This information can then subsequently be accessed by other buyers and considered at the time of making purchases.

Due to the growing number of eWOM sites, and overwhelming amount of information, systems synthesizing this information for consumers have emerged. These systems produce *average scores* from the different eWOM sites.

This information is highly valuable, and this is the reason why a number of official bodies, namely hotel star accreditation systems are starting to integrate these average scores into their own accreditation systems (Blomberg-Nyard and Anderson, 2016). However, the problem is that the process by which these average scores are produced lack transparency, which is one of the fundamental characteristics of star accreditation systems.

This raises questions about the robustness of these scores. Therefore, the purpose of this study is that one of examining the robustness of these scores. To do this, three key questions will be addressed: i) whether the different average scores provided by license software are resulting in equivalent or different scores. In the case of a difference, hoteliers with strong lobby capacity could support the adoption of that system which provides them with a more favorable score. And an example of this can already be found in the work by UNWTO (2014), which explains that the system developed for Norway has not been applied due to opposition

1  
2  
3 41 from major hotel chains. ii) whether the "average score" provided by these software systems is  
4 42 different to those provided by the metasearch websites. In the event that the differences are  
5 43 minimal, the adoption of licensed pieces of software, which are likely to be quite costly, would  
6 44 not be justifiable. Finally, iii) if significant differences among systems were observed, it would  
7 45 be of interest to determine the reasons.

8  
9 46 The results will have strong implications for hotels and hotel guests. Findings suggesting lack of  
10 47 robustness would justify the position of those hotels challenging the integration of these  
11 48 average scores from online reviews into official star classification accreditations. For hotel  
12 49 guests it would imply that consumer information and possibly new policies would be required  
13 50 to ensure consumer protection.

## 16 51 **2. LITERATURE REVIEW**

17  
18 52 eWOM is more influential than traditional WOM due to its speed, convenience, ability to reach  
19 53 many and the lack of human pressure which influences face to face communication (Sun et al.,  
20 54 2006). Additional reasons for consumer attention to eWOM relate to the expectation that  
21 55 receiving information may decrease the time and effort taken in making a decision and/or  
22 56 contribute to the outcome of making a more satisfying decision (Schiffman and Kanuk, 2000).

23  
24  
25 57 This breadth of eWOM scope and ease access can deeply affect a company's performance.  
26 58 Companies who adequately manage eWOM can have a competitive advantage (Loureiro and  
27 59 Kastenholz, 2011). This is the reason why companies are increasingly seeking to understand  
28 60 the factors that influence the use of eWOM and its impact.

29  
30  
31 61 The influence of eWOM on tourism has been studied extensively. Femback and Thomson  
32 62 (1995), Wang et al., (2002), and Yoo et al., (2009) suggest that travel online reviews are  
33 63 perceived as similar to recommendations by friends and relatives, and as more trusted  
34 64 information than the official one. Gretzel et al., (2006), and Wang et al., (2002) suggest that  
35 65 the reason for this impact is that social media decreases uncertainty, and it provides a sense of  
36 66 belonging into virtual travel communities. Other studies argue that the influence of eWOM on  
37 67 purchase decisions, may also be related to how the ratings position them on the ranking of  
38 68 searching results (Reino and Massaro, 2016). Thus, those establishments with the highest  
39 69 scores appear first in the searching results.

40  
41  
42 70 In the hotel sector, online reviews and ratings have a significant impact on potential  
43 71 consumers and their purchase decisions, both in terms of number of bookings (Onur Taş,  
44 72 2012), possibility to increase room price (Ye et al., 2009; and Anderson, 2012) and increment  
45 73 of occupancy rate (Viglia et al., 2016).

### 48 74 **2.1 Average Score**

49  
50 75 An increasingly growing number of Internet sites provide access to hotel scores. The average  
51 76 score facilitates the access to this information from across sites, synthesizing it into a single  
52 77 web. Furthermore, it enables drawing comparisons between a larger number of hotels and the  
53 78 analysis of their evolution over time.

1  
2  
3 79 Currently, there are mainly two business models behind the implementation of average score  
4 80 systems for commercial purposes. The first one refers to commercialisation tools that help  
5 81 hoteliers understanding their online reputation. This is the case of ReviewPro  
6 82 ([www.reviewpro.com](http://www.reviewpro.com)), TrustYou ([www.trustyou.com](http://www.trustyou.com)) and Olery ([www.olery.com](http://www.olery.com)). These  
7 83 software vendors perform a compilation of reviews from dozens of sources of information on  
8 84 hotels, offering detailed reports on valuations of hotels by department, by keyword, client  
9 85 profile, etc. It is specialized software, which has an economic cost for businesses, but can  
10 86 provide valuable information. The services offered by these companies generally relate to 1)  
11 87 pulling together reviews and ratings in one dashboard from different review platforms; 2)  
12 88 integrating and weighting scores through an algorithm based on a number of variables and  
13 89 providing a holistic score, typically on a scale from 1-100; 3) comparing hotel performance  
14 90 within a group, or a competitive set; 4) live monitoring whereby push notifications can be set  
15 91 for ratings below a predefined threshold; and 5) the provision for hoteliers to respond to  
16 92 reviews in one platform. However, a license is required to use this software (Henses, 2015).

17  
18  
19  
20 93 The level of information provided prior payment differs across systems. For example, TrustYou  
21 94 only facilitates certain information for free (i.e. the global average score) but most features are  
22 95 restricted and only available after payment. However, Olery does not provide access to any  
23 96 information without having taken on the software license previously. ReviewPro did not use to  
24 97 allow free access to their scores. However, from 2015 onwards, the company started providing  
25 98 free access through some official organisations. Further details can be found in the following  
26 99 section.

27  
28  
29  
30 100 The second type of systems relate to metasearch sites. These provide a comparison of hotels  
31 101 with a summary of their inventory available through different online booking platforms.  
32 102 Examples include Kayak ([www.kayak.com](http://www.kayak.com)), Trivago ([www.trivago.com](http://www.trivago.com)), Skyscanner  
33 103 ([www.skyscanner.com](http://www.skyscanner.com)) or HotelsCombined ([www.hotelscombined.com](http://www.hotelscombined.com)). Along with price and  
34 104 other information, metasearch websites show an average valuation of the property based on  
35 105 the reviews of guests on other sites. The information provided by metasearch websites is  
36 106 freely available and its data has been referred to in a number of research papers in the field.  
37 107 This is the case of Schamel, 2012; Pouplana, 2014; and Pesonen and Palo-oja, 2010.

## 38 108 **2.2 Official Use of the Average Score**

39  
40  
41  
42 109 UNWTO (2014) and Blomberg-Nyard and Anderson (2016) suggested the need to integrate  
43 110 online review ratings into hotel classification. This would be done to complement the  
44 111 quantitative measures offered by hotel classification schemes with the qualitative information  
45 112 provided by online reviews. In their investigation of online reviews and star ratings, these  
46 113 authors found no correlation, and that they serve complementary purposes. Furthermore,  
47 114 these authors refer to two industry studies (one by the National Tourism Development  
48 115 Authority of Ireland and another one by Tourism Ireland), which concluded that online ratings  
49 116 are considered more important than star ratings by both consumers and hoteliers.

50  
51  
52  
53 117 A number of official organisations (including tourism boards, accrediting bodies and trading  
54 118 organisations) have started providing access to Average Scores. They provide it in collaboration  
55 119 with commercialisation tools such as those outlined in the previous section. An example is the  
56 120 Australian Star Rating scheme. They facilitate this information through their own website

1  
2  
3 121 ([www.starratings.com.au](http://www.starratings.com.au)) and use ReviewPro average rating scores about the hotels listed to  
4 122 complement the star rating information. The Switch Hotel Association is also providing similar  
5 123 type of data through their website (<https://hotels.swisshoteldata.ch/>) and have developed this  
6 124 through a collaboration with TrustYou. Though according to Thiessen (2013) and Edgumbe  
7 125 (2014), Abu Dhabi was the first destination to incorporate online ratings and feedback into its  
8 126 classification system. This was done in collaboration with Olery, a tourism data management  
9 127 provider. An additional destination providing this type of information is Bahrein (Hensens,  
10 128 2015). And Norway, with its model developed by QualityMark Norway, and Germany are  
11 129 working towards the development of similar data (UNWTO, 2014), though limited details are  
12 130 available.

13  
14  
15  
16 131 The approach of resorting this type of service to an existing software company seems  
17 132 understandable. After all, it involves a high volume of data and it would be an overwhelming  
18 133 task for Official Organisations if wanting to do it by themselves. The problem is that these  
19 134 specialised software companies fail to provide an explanation of the methodology used. When  
20 135 requested to the software companies, this is referred to as a "secret algorithm" or a  
21 136 "proprietary algorithm". However, when these scores are adopted by Official Bodies they  
22 137 acquire an "official" character, and it would be reasonable to expect that the algorithms  
23 138 become publicly available, providing transparency in their rating process.

24  
25  
26 139 Academic literature has previously suggested the importance of standardizing this type of data  
27 140 prior to its aggregation (i.e. Chaves et al., 2012; Tano et al., 2014; and Viglia et al., 2016).  
28 141 Nevertheless, even these studies fail to reach a consensus on the establishment of an agreed  
29 142 methodology.

### 31 143 **3 METHODOLOGY**

32  
33  
34 144 Secondary data (quantitative and qualitative), has been used to address the three research  
35 145 questions of this study, which are:

- 36 146
- 37 147 1) Are there variations in the average scores provided by the different licensed  
38 148 software companies?
  - 39 149 2) Are there differences between the average score provided by software systems and  
40 150 metasearch websites?
  - 41 151 3) In the event that significant differences were observed, what are the reasons?  
42 152

43 153 Quantitative Data was collected and analysed in order to answer questions 1 and 2. The most  
44 154 reasonable approach would be to take a random sample of hotels and to check the scores  
45 155 obtained by them in each of the different systems (including both licensed software companies  
46 156 and metasearch websites). However, as noted in the previous section, of the 3 licensed  
47 157 software companies, only TrustYou provides free access to average scores for all their listed  
48 158 establishments. ReviewPro scores for Australia are freely available through the Australian  
49 159 Rating Scheme (<https://www.starratings.com.au/>) and access to ReviewPro data for Spain was  
50 160 available through Travel Advisors ([www.traveladvisorsguild.com](http://www.traveladvisorsguild.com)). Therefore, the sample was  
51 161 developed taking into account these limitations. Data was extracted from the two licensed  
52 162 software companies for which data was available (i.e. TrustYou and ReviewPro) and four hotel  
53 163 metasearch (HotelsCombined, Trivago, Skyscanner and Kayak) for Australian and Spanish  
54 164 hotels only. In addition, data from Booking.com (one major online booking systems that

165 provides verified opinions of hotels) was also collected for the same establishments, to reveal  
166 any possible differences between aggregating systems and single scoring sites.

167 A random stratified sampling approach was adopted. The search took place in an aleatory  
168 manner but fulfilling the following two conditions: they all have information emerging from at  
169 least 7 web sources and they all have at least 200 opinions in each of the webs. This criterion  
170 was established to ensure that the comparison of data across sites would be feasible and  
171 consistent. Searches through the database were undertaken in the major cities of each country  
172 (19 in total) in January 2016. One hundred Spanish hotels and the same number of Australian  
173 ones listed in ReviewPro were selected.

174 Data needed to be standardised. All the systems under study have one of the following types  
175 of scales: a scale with 100 as the maximum score and whole numbers or a scale with 10 as the  
176 maximum score and decimal values. Both scales are comparable when multiplying by 10 the  
177 values of the second scale. Therefore, this approach was taken.

178 It was decided that in the event that answering question 3 was applicable, this would be done  
179 by undertaking an analysis of the descriptions provided by the different systems in their own  
180 websites and the characteristics of their reviewing scores systems. However, given that these  
181 systems provide limited information about the way their means are calculated, it may not be  
182 possible to undertake a systematic evaluation of the same systems included in the previous  
183 analysis. Therefore, a heuristic approach has been adopted, in which different systems,  
184 including those in the previous part of the analysis, but also others, have been included. This  
185 relates to the two licensed software companies for which data was available (i.e. TrustYou and  
186 ReviewPro); and the four hotel metasearch sites (i.e. HotelsCombined, Trivago, Skyscanner,  
187 Kayak); as well as Booking.com. Additionally, data from Olery, which is the aggregator  
188 providing the ratings for Abu Dabi, as well as Agoda and TravelRepublic (both OTAs), which are  
189 included in the two UNWTO documents advocating for the integration of these systems in  
190 official star rating schemes (UNWTO, 2014), have also be included in the analysis. Finally, as  
191 previously explained, Olery was not included in the previous parts of the analysis due to the  
192 lack of public accessibility to their average scores.

193

#### 194 **4 RESULTS**

195 Data from 200 hotels from the 7 different systems was collected, producing a total of 1,400  
196 average rating scores. The typical number of reviews from which each of these average scores  
197 emerged was 397,124. However, the specific numbers varied across systems. For example, it  
198 was noted that Trivago's average scores were based upon over three times the number of  
199 reviews used to produce ReviewPro's average scores (666,984 reviews in Trivago vs. 212,972 in  
200 ReviewPro).

201 The results show different average scores across systems. Only one of the establishments in  
202 the sample of 200 hotels received the same scores across systems. Furthermore, while in 20  
203 cases the differences were just one point, in the remaining cases (180), these are two or more

204 points. In addition, there are 67 cases (33.5%) in which the difference is equal to or greater  
205 than 5 points and two of the cases show the very substantial difference of 9 points.

#### 206 4.1 Differences Among Licensed Software Providers

207 Table 1 shows the overall comparison of average scores received by the hotels in the sample.  
208 This data shows that the difference between ReviewPro and TrustYou is limited (lower than 1).  
209 The t-test analysis corroborates that the difference is small by suggesting that is not significant  
210 ( $p>0.05$ ). However, it should be noted that the scores differ for the large majority of the  
211 establishments (73%). The difference is determined by 62% of the establishments for which  
212 the scores are higher in ReviewPro, and 11%, for which the scores are higher in TrustYou. Only  
213 54 of the hotels (27%) display exactly the same scores.

214  
215 TABLE 1 GOES HERE

216  
217  
218 TABLE 2 GOES HERE

219 Additionally, these differences significantly affect the rankings of hotels, as shown in table 2.  
220 When reordering the ranking of 100 Australian hotels with the TrustYou scores, only 15 occupy  
221 the same position, while 45 worsen it and 40 improve it. Of those improving their position,  
222 there are 14 that improved it in 5 positions or more, and 15 which worsened it in 5 positions or  
223 more. Similarly, when the ranking of Spanish hotels is reordered with the scores given by  
224 TrustYou, only 13 establishments stay in the same position, 45 improve it and 42 worsen it,  
225 with 25 hotels seeing their position changed by 5 or more positions.

226 Therefore, in terms of answering the first question of this study, i.e. "are there differences in  
227 the average scores provided by the different licensed software companies?", the answer is *yes*,  
228 *there are differences in the average scores provided by licensed software companies*. And these  
229 differences do not only relate to the average scores obtained by establishments throughout  
230 the different systems, but also to their position in the ranking of results. This has important  
231 implications for the hotel industry, as it may imply that powerful corporations will lobby for  
232 their countries' star classification systems to adopt those systems which benefit their position  
233 in the rankings.

234 As explained in the literature review section, not only are scores important in determining  
235 consumer's choice (Öğüt and Onur Taş, 2012; Ye et al., 2009; Anderson, 2012; Viglia et al,  
236 2016) but also the position in the rankings in which hotels appear when users undertake a  
237 search (Reino and Massaro, 2016). Therefore, based on the differences presented across  
238 system, there is a strong reason to believe that the performance of hotels will vary depending  
239 on which licensed software is adopted. This suggest that despite requiring payment,  
240 integrating licensed software providers, instead of other type of free aggregators (e.g.  
241 metasearch sites), offers no guaranteed of the reliability of the average scores that they



242 provide. And given the lack of transparency of both these systems, it is not possible to find out  
243 which one of the two providers is reliable, if any.

244

#### 245 **4.2 Differences Among License Software Systems and Metasearch Websites**

246 Looking back at table 1, this shows that there are not important differences in the average  
247 scores produced across systems, except Trivago's. Trivago's results are lower than all the  
248 others. As an illustration of this, the difference between Trivago (which provides the lowest  
249 score) and ReviewPro (which provides the highest) is of 2.01 points.

250 TABLE 3 GOES HERE

251 Independent T-test was carried out on all paired systems (see table 3), and the results  
252 suggested significant differences ( $p < 0.05$ ) across a considerable number of paired systems.

253 In terms of differences between the licensed systems and the metasearch sites, these exist.  
254 Review-pro presents significant differences when compared with Trivago ( $p < 0.001$ ),  
255 Skyscanner ( $p < 0.05$ ), Kayak ( $p < 0.05$ ). Trustyou is the one presenting differences with the least  
256 number of other systems, and this is only with trivago ( $p < 0.05$ ). And this gives reasons to  
257 believe that of the two licensed systems, this may be the most reliable one.

258 With regards to the reliability of the metasearch systems, Trivago shows significant differences  
259 with the two licensed systems (ReviewPro and TrustYou,  $p < 0.001$  and  $p < 0.05$  respectively),  
260 and also with one of the metasearch systems, i.e. HotelsCombined ( $p < 0.01$ ). However,  
261 HoterlsCombined only presents differences with Trivago ( $p < 0.01$ ). Similarly, Skyscanner and  
262 Kayak, only present differences with one other systems. Interestingly, they both present  
263 differences with ReviewPro (both  $p < 0.05$ ).

264 Based on these analysis, there is a strong reason to believe that the average scores provided  
265 by HotelsCombined, Skyscanner and Kayak (all freely available), may offer similar reliability, at  
266 least, to those provided by TrustYou, and more trustworthiness than those provided by  
267 ReviewPro. This has important implications in terms of their possible integration by official  
268 rating schemes, which relates to the fact that both TrustYou and ReviewPro are licensed  
269 systems, and their use requires payment; while the metasearch sites are freely available.

270 Consequently, in answer to the second question of this study: *yes, there are significant*  
271 *differences between the average score provided by these software systems and those by*  
272 *metasearch websites*. However, further analysis suggested that differences may not  
273 necessarily be related to the lower reliability of metasearch websites. This has important  
274 implications for hoteliers and also for star rating classification systems, which may be  
275 considering the adoption of average scores. Taking this into consideration, and the fact that  
276 metasearch websites are free and licensed systems require a subscription, the adoption of the  
277 latter may be difficult to justify.

278

#### 279 **4.3 Possible Influencing Factors**

280 An analysis of the descriptions offered through the different systems' websites has suggested  
281 possible factors generating these differences. Information looked for included sources of data;  
282 any possible difference related to data which is excluded (e.g. old reviews); and differences in  
283 the way the data may be transformed to new scales.

284 As explained in the methodology section, this analysis included the same systems which were  
285 examined in the previous two subsections (i.e. TrustYou, ReviewPro, HotelsCombined, Trivago,  
286 Skyscanner, Kayak and Booking.com), and this was completed with others such as Olery, Agoda  
287 and TravelRepublic. However, no description of the way this data is sourced and/or produced  
288 has been found for HotelsCombined, Trivago, Skyscanner or Kayak. Some details about  
289 ReviewPro were found through the StarRating Australia website, as well as some others about  
290 TrustYou, and these form part of the analysis below. Additionally, data about Olery, Agoda and  
291 TravelRepublic has also been included to support the analysis and discussion.

292 Through the analysis, the first difference that has been noted, relates to the time periods for  
293 which online reviews and ratings are kept. During the examination of the sites, it was noted  
294 that Booking.com deletes all the reviews which are older than 24 months, while other sites  
295 don't (e.g. TripAdvisor which is one of the sources of Olery). Olery explains on their site (Olery,  
296 n.d.) that "only the reviews from the last three months are taken into account, and newer  
297 reviews are weighted a bit more heavily than older reviews". However, no explanation of how  
298 these weights are generated is provided.

299 The second difference identified relates to the frequency of updates. As described by  
300 StarRating Australia on their site (Star Rating, n.d.) "...the Travellers' Rating on  
301 starratings.com.au [which is provided by Reviewpro.com] will be updated monthly..." however,  
302 all the other systems fail to provide any information regarding the frequency with which their  
303 ratings are updated.. This suggest differences in the systems and websites' policies about  
304 keeping old reviews. And this is likely to have an important effect on the results.

305 The third difference identified relates to the use of different scales. As explained by UNWTO  
306 (2014), the variety of websites that collect scores use different scales, and these must be  
307 transformed to a single scale. For example, Booking.com uses a 2.5-10 scale (Mellinas et al.,  
308 2016), Agoda uses a 2-10 and TravelRepublic uses a 0-10. And interestingly, UNWTO (2014)  
309 assumes that Booking.com and Agoda use a 1-10 scale; error that could lead to inaccuracies in  
310 final average score or divergences between different score aggregators. The process by which  
311 each average score aggregator undertakes this transformation could well determine that  
312 difference in the results. ReviewPro conducted and published an analysis of the scales used by  
313 the different sites where the reviews are originated. However, this is not the case of all the  
314 other systems, which have not provided any type of explanation to the process by which  
315 ratings are transformed. And even ReviewPro fails to make full disclosure.

316 The fourth difference identified relates to the sources from which comments and scores are  
317 extracted. For example, ReviewPro suggests that they take into consideration social sites like  
318 Facebook while that is not the case of all the others. On the other hand, Olery for example  
319 takes into consideration the unverified ratings from TripAdvisor, while others such as  
320 ReviewPro and TrustYou do not. Given that TripAdvisor holds over 300 million reviews, their  
321 addition/inclusion to the data is expected to have an important impact on the results. The case

1  
2  
3 322 of Olery is interesting, as they suggest that "...ratings from Booking.com are weighted more  
4 323 heavily in the GEI calculation than those from TripAdvisor..." (Olery, n.d.), due to the fact that  
5 324 ratings from Booking.com are verified. This suggests further problems for data standardisation  
6 325 across systems though.

7  
8 326 These differences have important implications for consumers, as it suggests that the  
9 327 information that they receive is not robust and may be calling for the need to develop policies  
10 328 to ensure consumer protection.

## 11 329 **5 CONCLUSIONS**

12 330

13 331 The tourism industry is entering a process of introducing online hotel ratings official as part of  
14 332 the standard hotel classification, either in an integrated or complementary form. The  
15 333 aforementioned report UNWTO (2014), and the first experiences in Australia, Abu Dhabi,  
16 334 Switzerland, Germany, Bahrain and Norway show that this practice has already started to be  
17 335 deployed at a practical level. The intention of this type of initiatives is that one of providing an  
18 336 official status to this type of score. This type of status will automatically grant it a higher level  
19 337 of recognition than those rankings currently provided directly by private entities such as  
20 338 TripAdvisor, Expedia or Booking.com. It is also explained in the article that slight variations on  
21 339 online ratings of hotels significantly influence their occupancy levels. Therefore, this article has  
22 340 aimed to find out whether the data provided by the different average score aggregators is  
23 341 robust enough to make this type of adoption legitimate.

24 342

25 343 The results have shown that there are important differences not only in the rating scores  
26 344 obtained by the hotels but also on the rankings in which the hotels appear when undertaking a  
27 345 search. And this may have a more important effect than the rating itself, as it acts as a filter  
28 346 when searching for hotels.

29 347 The analysis has demonstrated that there are variations on the average scores provided by the  
30 348 different licensed software providers. Furthermore, they have also manifested differences  
31 349 between the group of licensed software providers and that one made up by metasearch  
32 350 websites. Further analysis of the systems has concluded that variations may be determined by  
33 351 the differences in the time periods during which the data is collected; variations in the  
34 352 frequency of data capture; inconsistency in the type of scale adopted and the process by which  
35 353 these are transformed; and differences in the sites which each aggregator includes. Therefore,  
36 354 it seems that the data provided by average score aggregators (either licensed or free  
37 355 metasearch sites) is not robust enough for their adoption by official organisms, as they stand  
38 356 at present.

39 357

### 40 358 **5.1 Theoretical Implications**

41 359

42 360 This study builds from extant knowledge on eWOM impact and use by hotel organisations, as  
43 361 well as by official star rating schemes for hotels. The findings expanded on this knowledge by  
44 362 analysing and explaining their practices.

45 363

## 364 **5.2 Practical Implications**

365 The findings about the differences in the rankings has important implications for the hotel  
366 industry, as it may imply that powerful corporations will lobby for their countries' star  
367 classification systems to adopt those systems which benefit their position in the rankings.

368  
369 The differences in the ratings have important implications for consumers, as it suggests that  
370 the information that they receive is not robust and may be calling for the need to develop  
371 policies to ensure consumer protection.

372  
373 Additionally, findings in the lack of robustness of these average scores also have important  
374 implications for star rating classification systems, which may be considering the adoption of  
375 average scores. This study opposes the calls made by UNWTO (2014) to support the  
376 integration of these systems in official rating scales until further transparency is provided by  
377 these systems. Furthermore, in the event of ignoring this recommendation and deciding to  
378 integrate average scores as they currently operate, it is herewith argued that the practice of  
379 integrating a paid system such as ReviewPro or TrustYou does not present sufficient  
380 justification. The comparison across systems provided through this study has given no  
381 indication to believe that the average scores produced by these two licensed systems offers  
382 additional reliability than the average scores provided by metasearch websites such as  
383 HotelsCombined, Kayak and Skyscanner, which are all freely available.

384

## 385 **5.3 Limitations**

386 The limitations of this study mainly relate to data access. For the last part of the analysis,  
387 information about certain systems was not available. Therefore, further sites needed to be  
388 integrated in the analysis to supplement this.

## 389 **5.4 Future Research**

390

391 Despite the issues of lack of transparency and robustness presented by these systems, there  
392 are already a few cases of their integration in star rating systems by official organisations. If so  
393 far this issue has not generated any controversy it could be due to the low implementation of  
394 the system and the relative lack of attention paid to it by consumers. At the time when this  
395 practise becomes generalised, surely hoteliers will be interested in finding out how the score is  
396 assigned to their establishment. The only way to avoid problems is to establish a transparent  
397 system for calculating overall scores, with an algorithm of public and open character. It would  
398 also be advisable that the criteria used were previously agreed with hoteliers, thus facilitating  
399 its subsequent acceptance by the sector.

400

401 Therefore, future research should be undertaken in the development of a robust methodology  
402 to ensure the robustness of the aggregation techniques used by these sites. In this process, it  
403 may still be advisable to count with the collaboration of software suppliers (Olery, ReviewPro,  
404 TrustYou), since the amount of information that needs collecting and the underlying processes  
405 may require using their technology. It would also be desirable to try avoiding the disparity of  
406 criteria by countries currently exists with regard to systems of hotel classification (stars,

407 diamonds, etc ...), so that the initiatives led by UNWTO, EU or other transnational organisms,  
408 are more suitable.

409

410 Finally, consumer perceptions on the practice of aggregating this data to official star rating  
411 classifications would also be very valuable.

412

#### 413 REFERENCES

414 Anderson, C. (2012), "The Impact of Social Media on Lodging Performance", *Cornell Hospitality Report*, Vol. 12 No.  
415 15, pp. 6-11.

416 Blomberg-Nyard, A. and Anderson, C.K. (2016), "United Nations World Tourism Organization Study on Online Guest  
417 Reviews and Hotel Classification Systems: An Integrated Approach", *Service Science*, Vol. 8 No. 2, pp. 139-  
418 151.

419 Chaves, M.S, Gomes, R., and Pedron, C. (2012), "Analysing Reviews in The Web 2.0: Small and Medium Hotels in  
420 Portugal", *Tourism Management*, Vol. 33 No. 5, pp. 1286-1287.

421 Edgcumbe, D. (2014), "Abu Dhabi Hotel Ratings to Go Social in 2014", available at:  
422 <http://www.hoteliermiddleeast.com/19456-abu-dhabi-hotel-ratings-to-go-social-in-2014/> (accessed 18 May  
423 2018).

424 Femback, J. and Thomson, B. (1995), "Virtual Communities: Abort, Retry, Failure? Presented as "Computer-  
425 Mediated Communication and the American Collectivity" at the Annual Meeting of the International  
426 Communication Association, Albuquerque, New Mexico, May 1995..

427 Gretzel, U., Fesenmaier, D. and O'Leary, J. (2006), "The Transformation of Consumer Behaviour", in Buhalis, D. and  
428 Costa, C. (Eds.), *Tourism Business Frontiers: Consumers, Products and Industry*, Elsevier Butterworth-  
429 Heinemann, Oxford, pp. 7-16.

430 Hensens, W. (2015), "The Future of Hotel Rating", *Journal of Tourism Futures*, Vol. 1 No. 1, pp. 69-73.

431 Litvin, S.W., Goldsmith, R.E. and Pan, B. (2008), "Electronic Word-of-Mouth in Hospitality and Tourism  
432 Management", *Tourism Management*, Vol. 29 No. 2008, pp. 458-468.

433 Loureiro, S.M. and Kastenholz, E. (2011), "Corporate Reputation, Satisfaction, Delight, and Loyalty Towards Rural  
434 Lodging Units in Portugal", *International Journal of Hospitality Management*, Vol. 30 No. 2011, pp. 575-583.

435 Mellinas, J. P., Martinez, M.D., & Garcia, J. J. B. (2016), "Evolución de las Valoraciones de los Hoteles Españoles de  
436 Costa (2011-2014) en Booking.com". *PASOS. Revista de Turismo y Patrimonio Cultural*, Vol. 14 No. 1, pp.  
437 141-151.

438 Ögüt, H., and Onur Taş, B. K. (2012), "The Influence of Internet Customer Reviews on The Online Sales and Prices in  
439 Hotel Industry", *The Service Industries Journal*, Vol. 32 No. 2, pp. 197-214.

440 Olery (n.d.), "Guest Experience Index", available at: <http://www.olery.com/guest-experience-index/> (accessed 18  
441 May 2018).

442 Pouplana, J.R. (2014), "Do you Care?: Metasearch Your Presence in Marketing, Essential for Good Metasearch  
443 Tecnohotel strategy", *Professional Magazine Hotels and Restaurants*, Vol. 463, pp. 18-19.

444 Pesonen J., Palo-oja O.M. (2010), "Comparing Internet commerce adoption between the Finnish and the European  
445 independent accommodation companies", in Gretzel, U., Law, R., Fuchs, M. (Eds), *Information and  
446 Communication Technologies in Tourism 2010*. Springer, Vienna.

447 Reino, S., and Massaro, M.R. (2016), "Assessing the Value of Hotel Online Reviews to Consumers", in Katsoni, V. and  
448 Stratigea, A., *Tourism and Culture in the Age of Innovation*, Springer International Publishing, London.

- 1  
2  
3 449 Schamel, G. (2012), "Weekend vs. Midweek Stays Weekend. Modelling Hotel Room Rates in a Small Market",  
4 450 *International Journal of Hospitality Management*, Vol 31 No. 4, pp. 1113-1118.
- 5 451 Schiffman, L.G., and Kanuk, L.L. (2000), *Consumer Behavior* (7th ed), Prentice Hall, UpperSaddle River.
- 6  
7 452 Star Rating (n.d.), "Introducing the Traveller's Rating", available at: [https://www.starratings.com.au/docs/default-](https://www.starratings.com.au/docs/default-source/default-document-library/faqs---travellers-39-rating.pdf?status=Temp&sfvrsn=0.07011642376892269)  
8 453 [source/default-document-library/faqs---travellers-39-](https://www.starratings.com.au/docs/default-source/default-document-library/faqs---travellers-39-rating.pdf?status=Temp&sfvrsn=0.07011642376892269)  
9 454 [rating.pdf?status=Temp&sfvrsn=0.07011642376892269](https://www.starratings.com.au/docs/default-source/default-document-library/faqs---travellers-39-rating.pdf?status=Temp&sfvrsn=0.07011642376892269) (accessed 18 May 2018).
- 10  
11 455 Sun, T., Youn, S., Wu, G., and Kuntaraporn, M. (2006), "Online Word-of-Mouth: An Exploration of Its Antecedents  
12 456 and Consequences", *Journal of Computer-Mediated Communication*, Vol. 11 No. 4, pp. 1104-1127.
- 13  
14 457 Tano Gutierrez, D., Parra López, E., Gonzalez Wetherill, Y. (2014), "The Influence of Identity Authentication in The  
15 458 Online Evaluations of The Hotels, paper presented at AECIT XVIII Congress, 26-28 November 2014,  
16 459 Benidorm.
- 17  
18 460 Thiessen, T. (2013), "Abu Dhabi folds reviews into hotel ratings", *Hotel News Now*, available at:  
19 461 <http://www.hotelnewsnow.com/articles/21034/Abu-Dhabi-folds-reviews-into-hotel-ratings> (accessed 18  
20 462 May 2018).
- 21 463 UNWTO (2014), "Online Guest Reviews and Hotel Classification Systems: And Integrated Approach", UNWTO,  
22 464 Madrid.
- 23  
24 465 Vermeulen, I.E. and Seegers, D. (2009), "Tried and Tested: The Impact of Online Hotel Reviews on Consumer  
25 466 Consideration", *Tourism Management*, Vol. 30, pp. 123-127.
- 26  
27 467 Viglia, G., Minazzi, R. and Buhalis, D. (2016), "The Influence of E-Word-of-Mouth on Hotel Occupancy Rate",  
28 468 *International Journal of Contemporary Hospitality Management*, Vol. 28 No. 9, pp. 2035-2051.
- 29  
30 469 Wang, Y., Yu, Q., and Fesenmaier, D.R. (2002), "Defining the Virtual Tourist Community: Implications for Tourism  
31 470 Marketing", *Tourism Management*, Vol. 23 No. 4, pp. 407-417.
- 32  
33 471 Ye, Q., Law, R., Gu, B., and Chen, W. (2011), "The Influence of User-Generated Content on Traveler Behavior: An  
34 472 Empirical Investigation on The Effects of E-Word-of-Mouth to Hotel Online Bookings", *Computers in Human*  
35 473 *Behavior*, Vol. 27 No. 2, pp. 634-639.
- 36 474 Yoo, K. H., Lee, Y. J., Gretzel, U. and Fesenmaier, D. (2009), "Trust in Travel-Related Consumer Generated Media" in  
37 475 Höpken, W. Gretzel, U. and Law, R. (Eds.), *Information and Communication Technologies in Tourism 2009*,  
38 476 Springer-Verlag, Vienna.

Table 1. Average Scores Received by the Hotels

TYPE OF SYSTEM	LICENSED SYSTEMS		METASEARCH SITES				INDIVIDUAL SCORE PROVIDER	AVERAGE
	REV.PRO	TRUSTYOU	H.COMB	TRIVAGO	SKY	KAYAK	BOOKING	
SCORE	82,79	82,03	82,24	80,78	81,74	81,66	82,02	81,89
NUMBER OF REVIEWS	1064,86	2265,44	1883,60	3334,92	2110,88	2245,82	993,83	1985,62

Table 2. Position in the TrustYou Ranking Compared to ReviewPro

	EQUAL	IMPROVE	WORSEN	IMPROVED 5 OR MORE POSITIONS	WORSEN 5 OR MORE POSITIONS
AUSTRALIA	15	45	40	14	15
SPAIN	13	45	42	13	12



Table 3. Independent T-Test

	REV.PRO	TRUSTYOU	H.COMB	TRIVAGO	SKY	KAYAK
REV.PRO		0.138155262	0.3203985	<b>0.0001</b>	<b>0.04031</b>	<b>0.0275</b>
TRUSTYOU	0.138155262		0.7127957	<b>0.01464</b>	0.56911	0.46395
H.COMB	0.320398513	0.712795673		<b>0.00924</b>	0.37337	0.29911
TRIVAGO	<b>0.000101917</b>	<b>0.014637139</b>	<b>0.0092429</b>		0.05953	0.08723
SKY	<b>0.040311756</b>	0.569113884	0.3733748	0.05953		0.86782
KAYAK	<b>0.027502688</b>	0.463949955	0.2991134	0.08723	0.86782	
BOOKING	0.116537488	0.975653735	0.6835171	<b>0.01229</b>	0.57495	0.46543