# CONSISTENCY OF ONLINE CONSUMERS' PERCEPTIONS OF POSTED COMMENTS: AN ANALYSIS OF TRIPADVISOR REVIEWS

## Foo Sheng Khoo, Phoey Lee Teh & Pei Boon Ooi

*Department of Computing and Information Systems*
*Sunway University, Malaysia*

*kfoosheng@hotmail.com; phoeyleet@sunway.edu.my;*
*peiboono@sunway.edu.my*

## ABSTRACT

Ratings and comments play a dominant role in online reviews. The question, thus, arises as to whether or not there is any consistency in consumer perception of the reviews, and how future choices might be influenced. We analysed 2000 comments of 20 different hotels posted on TripAdvisor to determine if the comments posted by previous guests of a hotel influence the decisions of potential guests. Two hundred human raters were asked to consider 20 reviews and to rate a hotel based on the reviews. The Cohen Kappa coefficient was used to evaluate the degree of agreement on the hotel quality as determined by the human raters and the star rating given by the original reviewer. The results showed a high consistency between the human raters' evaluation and the reviewers' star rating. This research reveals the importance of website feedback such as TripAdvisor in influencing consumer choice.

**Keywords:** Cohen Kappa, comments, degree of agreement, rating, online reviews.

## INTRODUCTION

Social media is a rich platform to seek out and share products, experiences and information (Chan, Lacka, Yee, Lim, 2015). According to social psychology perspectives, consumers tend to review online comments before purchase decisions (Wong, 2015) or during the employment-seeking process

(Sander, Teh, & Sloka, 2015). The users provide feedback through comments (Siersdorfer, Chelaru, Pedro, Altingovde & Nejdl, 2014). The online review has become a notable strategy to share products worldwide and has been termed as the online word-of-mouth (WOM) (Zhang, Guo, & Goes, 2013). In addition, the feedback tracked online also contributed to opinion mining (Raut & Londhe, 2015) in which a product or service improvement is required (Leskovec, 2011). With the advancement of information technology, in the otel industry, various tools could be adopted and used to capture the feedback, fulfil greater customer expectations and thus increase customer relationship (Hassan, Hussain, & Saibu Rahman, 2013).

There are various ways to provide feedback (Kauffman, Laf, Lin, & Chang, 2009). For example, rating or ranking a post is one of the common ways in the social media. TripAdvisor (Wyndham, 2015), YouTube (Alec, 2013) and Metacritic (Dietz, 2015) use rating as their review mechanism. YouTube, for instance, uses symbols such as *Thumb Up* or the *Like* button to indicate the degree of agreement, support or helpfulness of the shared information or links. Consumers provide feedback in the social media along with comments and ratings for products and services they consumed. The reviews, thus become an issue that either provides a helpful or non-helpful manner of rating (Krestel & Dokoohaki, 2011). Bermingham (2014) suggested that comments are useful information for formulating a business strategy (Chan et al., 2015) and facilitating a marketing research pertaining to customer behaviour and supporting the management disciplines. Evidently, user reviews and ratings are capable of influencing the consumers' decision-making; thus, making them worthy to be investigated.

Chen, Guo, Tseng and Yang (2011) stated that a prestigious user's comment received a higher number of followers. They also found that the quality of a comment judged editorially is almost uncorrelated with the rating it received. This research question arises on the consistency in the value expressed in the reviews with the human rater opinion on the perspective of quality of review toward the hotel. Krestel and Dokoohaki (2011) postulated that some of the comments and ratings are inaccurate, irrelevant or inconsistent. Zhang, Guo and Goes (2013) revealed the same information, i.e. some of the reviews are inconsistent and some are irrelevant. If one could rank the reviews based on the relevancy of a user's favour, it will assist in understanding more about the products with reviews. Therefore, to identify if a reader's response has any relevance to the level of a user's favour, we need to determine whether or not comments posted by previous guests of the hotels has any relevancy and thus influence the decisions of potential guests.

## RELATED WORK

TripAdvisor is a popular online travel website in the world (Vásquez, 2011). It has also assisted the traveller in the decision-making of their hotel stay (Matos-rodríguez, 2014). A useful comment increased the website's reputation and it served as a precious asset to the website and the user's gratification assessment often times appeared in the form of ratings (Moghaddam & Martin, 2011). Not only that, Chen et al., (2011) stated that it also generated a higher support in the web community. This is further corroborated when experienced customers agreed on the statement or judgment of the overall quality of products and thus influenced other buyers' purchase decisions (Klan & Ries, 2014).

Web communities allowed comments in both objects and written format. Written comments consumed time in interpretation as compared to objects. According to Samsudin, Puteh, Hamdan and Ahmad Nazri (2013), noisy texts is a common phenomenon in online reviews and it affects data mining exercise. Also, comments may be irrelevant or casual (Zhang et al., 2013). In text analysis techniques, the content analysis tool is used to determine the positive and negative emotions from blog texts (Gill, French, Gergle, & Oberlander (2008). According to Gill et al. (2008), a positive writer tends to use positive words frequently. Gill showed positive and negative emotional concept words in their table to determine the significant differences between means.

The mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) was used to test emotional intelligence (EI) (Abe, 2011). Students' performance rated by their supervisors were associated with positive emotive words, whereas students' perceived benefits and practical courses were less associated with positive emotive words. Several types of research have been performed on the emotive word in various means. For instance, in a cognitive study, Filippi, Ocklenburg, Bowling, Heege, Gunturkun, Newen and de Boer (2016) stated that humans typically combine linguistic and non-linguistic information to comprehend emotions. In the same study, language evolution was discussed, however, it was not specific to emotive word processing. On the other hand, in the developmental science, Li and Yu (2015) suggested children's emotive word comprehension (EWC) developed with age. A psycholinguistic study by Sereno, Scott, Yao, Thaden, and O'Donnell (2015) showed a significant interaction between participant mood and word emotionality. However, there is limited study on comments related to emotive words used in online comments.

The consideration of emotive words is fundamental in the affective language theoretical formation (Hinojosa, Rincón-Pérez, Romero-Ferreiro, Martínez-García, Villalba-García, Montoro & Pozo, 2016). However, emotive words are not the only means to express emotion. Sentiment analysis is moving beyond basic text mining and extracting opinions from the text alone; emoticons and other means of expressing sentiment are now part of the analysis. Researchers revealed that emoticons are not just for fun, in fact, they serve as an additional and valuable supplement in communication methods (Huang, Yen, & Zhang, 2008). In addition, the latest study by Teh, Rayson, Pak, Piao, and Yeng (2016) discovered that emoticons have significant abilities to reverse the polarity of comments. It further corroborates that the processing and understanding of emotions expressed by the comments is vitally important in sentiment analysis and studies.

According to Zhang, Zhang, and Yang (2016), travellers' rating behaviour is affected by a number of expert reviews on a particular hotel, level of reviewers' expertise, and the website recognized expertise. However, in the comment analysis, Fong, Lei, and Law (2016) pointed out that a review is not necessarily consistent with its sentiment. In their study, only one hotel's reviews were analysed and studied. To address the gap, we investigated and compared the consistency between the rating provided by readers of the comments and the original reviewers' rating of 20 hotels. In addition, this study assessed, compared, and investigated the consistency of the human raters and online comments, without revealing the hotel star rating by the original reviewers to the human raters.

## METHODOLOGY

In this study, we selected TripAdvisor as our main platform to aggregate the comments. The reason for this selection was because of its worldwide popularity. TripAdvisor is an American travel website founded in February 2000 (TripAdvisor, 2015). It has since expanded to 45 countries, translated into 28 languages with 280 million users and it has accumulated almost 170 million comments. Users create a personal profile, with reviews and ratings (Matos-rodríguez, 2014) and share their experiences with others. The TripAdvisor bubble rating is shown as a 5-point Likert scale: 1= "Terrible", 2= "Poor", 3= "Average", 4= "Very Good" and 5= "Excellent" (TripAdvisor, 2015). Hypotheses were formulated to examine the relationship between items, online reviews, and the human raters as well as its inter-relativity and the degree of agreement of the review samples and it rating by 200 human raters.

## DATA AND EXPERIMENTS

### Hypothesis Development

Two hypotheses were developed to examine the consistency between the readers' positive and negative ratings and the original reviewers' rating without revealing the hotel star rating by the original reviewers.

**Hypothesis 1:** There is a significant relationship between the TripAdvisor positive rating and the human raters' rating.

**Hypothesis 2:** There is a significant relationship between the TripAdvisor negative rating and the human raters' rating.

### Questionnaire Design

We recruited human raters to evaluate the sampled reviews, based on the selected top ten (10) frequently used positive and negative emotive words, regardless of which star rating they belonged to in order to examine the reviews' quality level to of the hotel. This was to identify if the reviews have the potential to attract new customers or return customers based on the consistency of the original rating reviews and the quality of readers' opinions.

Table 1

*Questions Categorized for Set A Questionnaire.*

| Hotel star-rating | Section 1 | | Section 2 | |
|---|---|---|---|---|
| | Question quantity | Positive emotive words | Question quantity | Negative emotive words |
| 2 | 3 | Loved, Like, Enjoyed | 3 | Disappointed, Disappointment, Disappointing |
| 3 | 3 | Relaxing, Love, Rest | 3 | Fault, Annoying, Upset |
| 4 | 2 | Happy, Enjoy | 2 | Sadly, Regret |
| 5 | 2 | Relax, Liked | 2 | Disappoint, Worrying |

Each questionnaire comprised a set of 20 sample comments; the total number of questions was thus 40 (there were two set of questionnaires). Of the 20 comments in each questionnaire, ten comments were positive in nature and ten were negative; each comment contained one of the emotive words given

in Tables 1 and 2. Correspondingly, for each category of the star-rated hotels, there was a total of ten comments given in the two questionnaires; five positive comments and five negative comments.

Table 2

*Questions Categorized for Set B Questionnaire.*

| Hotel star-rating | Section 1 | | Section 2 | |
|---|---|---|---|---|
| | Question quantity | Positive emotive words | Question quantity | Negative emotive words |
| 2 | 2 | Happy, Enjoy | 2 | Sadly, Regret |
| 3 | 2 | Relax, Liked | 2 | Disappoint, Worrying |
| 4 | 3 | Relaxing, Love, Rest | 3 | Fault, Annoying, Upset |
| 5 | 3 | Loved, Like, Enjoyed | 3 | Disappointed, Disappointment, Disappointing |

Two sets of questionnaires were designed for this study in order to have a balanced sampling of the selected reviews of the hotel from all rankings and polarity of words used. Each questionnaire contained a set of 20 sample comments: the total number of questions was thus 40 (there were two sets of questionnaires). For the 20 comments in each questionnaire, ten comments were positive in nature and ten were negative: each comment contained one of the emotive words given in Tables 1 or 2. Correspondingly, for each of the different categories of star-rated hotels there was thus an overall total of ten comments given in the two questionnaires: five positive comments and five negative comments.

Eventually, each set of questionnaire consisted of two (2) sections. Section 1 was sampled with reviews that contained positive emotive words from Table 4, each with a sample from the listed top 10 emotive words. Similarly, section 2 with the sample of reviews contained negative emotive words accordingly from the list in Table 5. Each section had 10 reviews, with the top ten emotive words appearing in the positive and negative reviews.

**Sampling Techniques**

TripAdvisor categorized the hotels into four different star ratings; 2-Star (lowest rating hotel), 3-Star, 4-Star, and 5-Star (highest rating hotel). Five "Top Traveller's choice hotels" (as defined by TripAdvisor) were selected from each star category. For each hotel selected, 100 most recent comments were extracted (as of 30 June 2016).

**Sampling Population/Selection Criteria for Respondents**

The questionnaires were distributed to 200 students aged between 22 to 30 years old. Table 3 shows 93 or 88.5% respondents reported that they sometimes, often, or always, referred to reviews and booked hotels online. Approximately 88.5% responded that they often or always booked a hotel online. 95 respondents (47.5%) stated that they frequently commented on the hotel-booking website, 54 (27%) respondents indicated they sometimes commented on the website, followed by 39 (19.5%) respondents who admitted that they always commented on the website, and 12 respondents (6%) said they rarely commented on the website.

Table 3

*The Frequency of Referring, Booking, and Writing Comments*

|  | Never | Rarely | Sometimes | Often | Always |
|---|---|---|---|---|---|
| How often do you refer and book a hotel online when you travel? | 0% | 0.5% | 11% | 42% | 46.5% |
| How often do you write a review at a hotel booking website? | 0% | 6% | 27% | 47.5% | 19.5% |

**Sampling-size Calculation/Justification**

Two hundred (200) questionnaires were distributed. Responses from the returned questionnaires were collected and the input was keyed into SPSS for statistical analysis. The Pearson Correlation was run to determine the relationship between the two variables (Evans, 1996) and the Inter-reliability was examined to evaluate the agreement between the two classifications. The Cohen's Kappa value ranged from 0 to 1.0 (Cohen, 1960; Fleiss et al., 2003) with a higher score indicating a higher strength of agreement between the wo classifications. We recorded the participants' rating to of the reviews based on their feelings/perceptions. The average time spent to answer the questionnaire was 10 minutes, at a private university in Malaysia. The respondents answered one of the questionnaires randomly and were not allowed to participate twice.

**Data Analysis Strategy**

A total frequency of 279 emotiove words was categorized by means of Wmatrix (Rayson, 2008) and the word "happy" and "happily" appeared at least once in every hotel review for all rankings. When all the reviews were analysed again, the top 10 frequently used words were recorded. Table 4 and Table

5 show all the top 10 frequently used positive and negative emotive words. Loved (0.059%), like (0.056%), disappointed (0.026%) and disappointment (0.011%) were the four top positive and negative emotive words.

Table 4

*Top 10 Frequently Used Positive Emotive Words*

| No. | Top 10 positive emotion words | Frequency | Relative frequency [1] (Between comments of the 20 hotels) |
|-----|-------------------------------|-----------|------------------------------------------------------------|
| 1   | Loved    | 47 | 0.059% |
| 2   | Like     | 45 | 0.056% |
| 3   | Enjoyed  | 34 | 0.042% |
| 4   | Happy    | 31 | 0.038% |
| 5   | Enjoy    | 29 | 0.036% |
| 6   | Relaxing | 28 | 0.035% |
| 7   | Love     | 27 | 0.034% |
| 8   | Rest     | 22 | 0.027% |
| 9   | Relax    | 21 | 0.026% |
| 10  | Liked    | 19 | 0.024% |

Table 5

*Top 10 Frequently Used Negative Emotive Words*

| No. | Top 10 negative emotive words | Frequency | Relative frequency[1] (Between comments of the 20 hotels) |
|-----|-------------------------------|-----------|-----------------------------------------------------------|
| 1   | Disappointed   | 21 | 0.026% |
| 2   | Disappointment | 9  | 0.011% |
| 3   | Disappointing  | 8  | 0.010% |
| 4   | Fault          | 8  | 0.010% |
| 5   | Annoying       | 6  | 0.007% |
| 6   | Upset          | 5  | 0.006% |
| 7   | Sadly          | 3  | 0.004% |
| 8   | Regret         | 3  | 0.004% |
| 9   | Disappoint     | 3  | 0.004% |
| 10  | Worrying       | 3  | 0.004% |

[1]*Relative frequency does not sum to 100% as it only reports frequency of the 10 most used positive/ negative emotive words.*

**RESULTS**

**Respondent Rating of Impression after Reading the Reviews**

Figures 1, 2, 3, and 4 show the respondents' rating – from poor to excellent – of the opinions gained after reading the whole sample of reviews. These reviews were selected based on the most used emotive words. Figures 1 and 2 are the results plotted from questionnaire set A, while Figures 3 and 4 are the results plotted from Set B questionnaire.



| | PW1a | PW2a | PW3a | PW4a | PW5a | PW6a | PW7a | PW8a | PW9a | PW10a |
|---|---|---|---|---|---|---|---|---|---|---|
| ▪ Poor | | | | | | | | | | |
| ▪ Fair | | | | | | | | | | |
| ▪ Good | 8 | 8 | 4 | 6 | 6 | 6 | 6 | 3 | 1 | 2 |
| ▪ Very Good | 33 | 55 | 39 | 32 | 40 | 40 | 45 | 28 | 50 | 35 |
| ▪ Excellent | 59 | 37 | 57 | 62 | 54 | 54 | 49 | 69 | 49 | 63 |

*Abbreviations: PW1a = Comment containing 1st Positive Emotive Words for Set A questionnaire; PW2a = Comment contains Second Positive Emotive Words for Set B questionnaire, etc.

*Figure 1*. Respondent rating after reading the sample review (Positive words: Set A questionnaire).
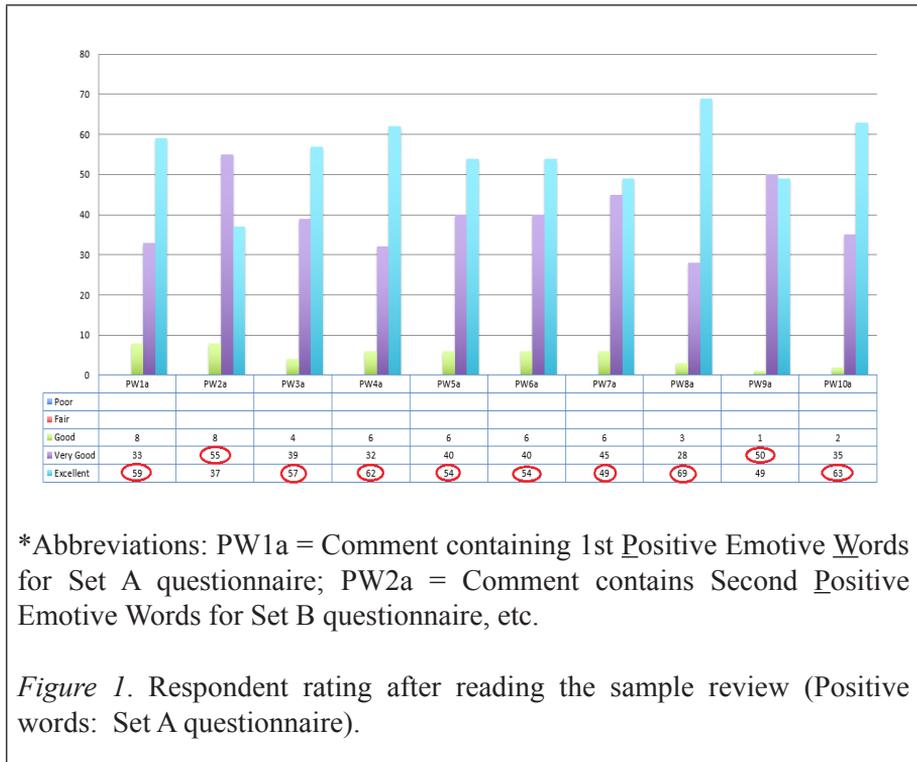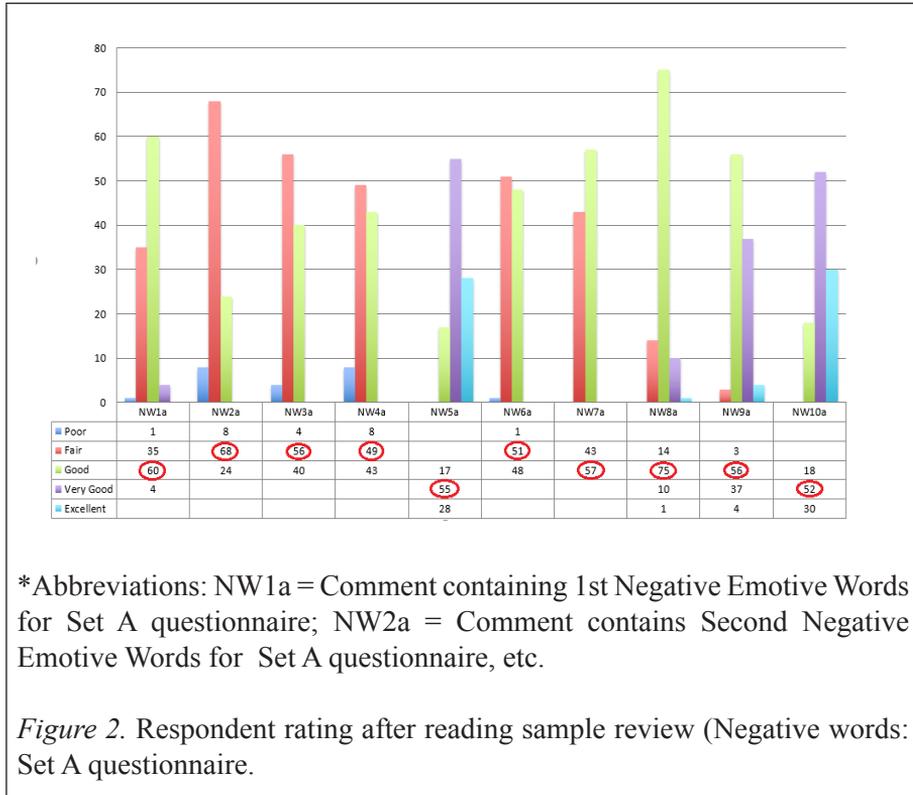
Figure 1 shows the rating from respondents after they read the review samples from questionnaire Set A. The circle indicates the highest rank of the positive emotive words. Apparently, eight out of ten of the positive words were rated "Excellent", and only two were rated "Very Good". This result suggested a coherent positive word used in the comments of the reader's perception. However, the usage of negative emotive words was not negatively coherent to the rating of impression. For example, in Figure 2, there were only 4 rated as "Fair" by the respondents in set A – NW2a (68%), NW3a (56%), NW4a (49%) and NW5a (51%). The other 4 rated as "Good" had a higher number of

respondents, namely NW1a (60%), NW7a (57%), NW8a (75%), and NW9a (56%). This sums up to 248 higher respondents who rated better impression. Even if it is a negative use of emotive words, the value of the reviews appeared incoherent.



| | NW1a | NW2a | NW3a | NW4a | NW5a | NW6a | NW7a | NW8a | NW9a | NW10a |
|---|---|---|---|---|---|---|---|---|---|---|
| ▪ Poor | 1 | 8 | 4 | 8 | | 1 | | | | |
| ▪ Fair | 35 | 68 | 56 | 49 | | 51 | 43 | 14 | 3 | |
| ▪ Good | 60 | 24 | 40 | 43 | 17 | 48 | 57 | 75 | 56 | 18 |
| ▪ Very Good | 4 | | | | 55 | | | 10 | 37 | 52 |
| ▪ Excellent | | | | | 28 | | | 1 | 4 | 30 |

*Abbreviations: NW1a = Comment containing 1st Negative Emotive Words for Set A questionnaire; NW2a = Comment contains Second Negative Emotive Words for Set A questionnaire, etc.

*Figure 2*. Respondent rating after reading sample review (Negative words: Set A questionnaire.

Reviews were extracted to examine how the negative words were used; NW5a and NW10a above were extracted as examples.

Overall, the words used conveyed a "very good" impression and "excellent" rating to the readers.

We repeated the investigative steps on Set B questionnaires. Figure 3 displays a similar consistency. The top ten positive emotive words used in the reviews were highly impressive to the readers. 6 questions were rated as "Very Good"; PW1b (48%), PW2b (53%), PW3b (54%), PW4b (48%), PW7b (50%) and PW9b (53%). The others rated with "Excellent" were PW5b (51%), PW6b (60%), PW8b (51%) and PW10b (60%). Overall, the positive emotive words appeared to give a positive impression to the readers consistently.
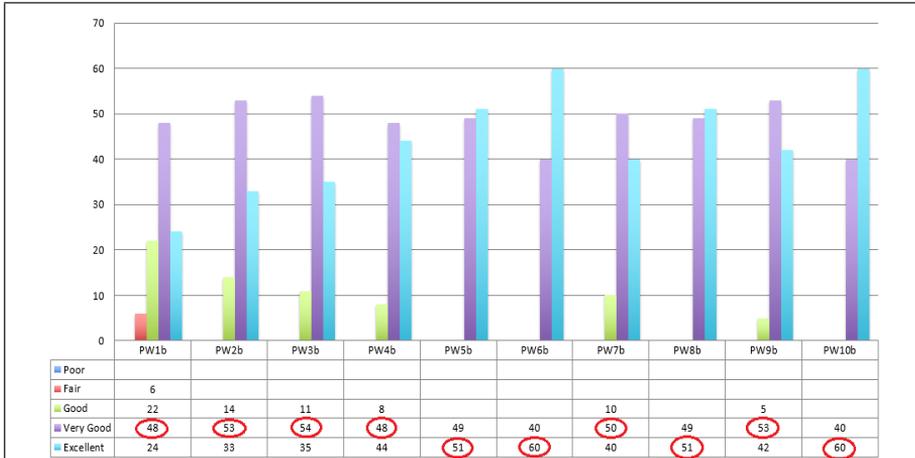
*Abbreviations: PW1b = Comment containing 1st Positive Emotive Words for Set B questionnaire; PW2b = Comment contains Second Positive Emotive Words for Set B questionnaire, etc.

*Figure 3.* Respondent rating after reading sample review (Positive words: Set B questionnaire)
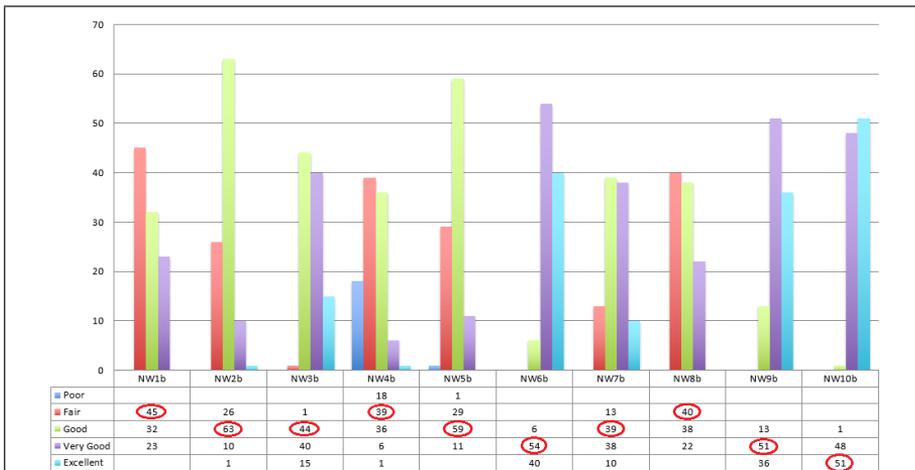


*Abbreviations: NW1b = Comment containing 1st Negative Emotive Words for Set B questionnaire ; NW2b = Comment contains Second Negative Emotive Words for Set B questionnaire, etc.

*Figure 4.* Respondent rating after reading sample review (Negative words: Set B questionnaire).

Set B with the negative words used has not being coherent to the respondent rating, thus has been rated 3 times as "Fair", 4 times as "Good", 2 times as "Very Good" and 1 time as "Excellent". We further examined the overall sentiment value of the reviews as in the following example:

**TripAdvisor Reviews Consistency Comparison**
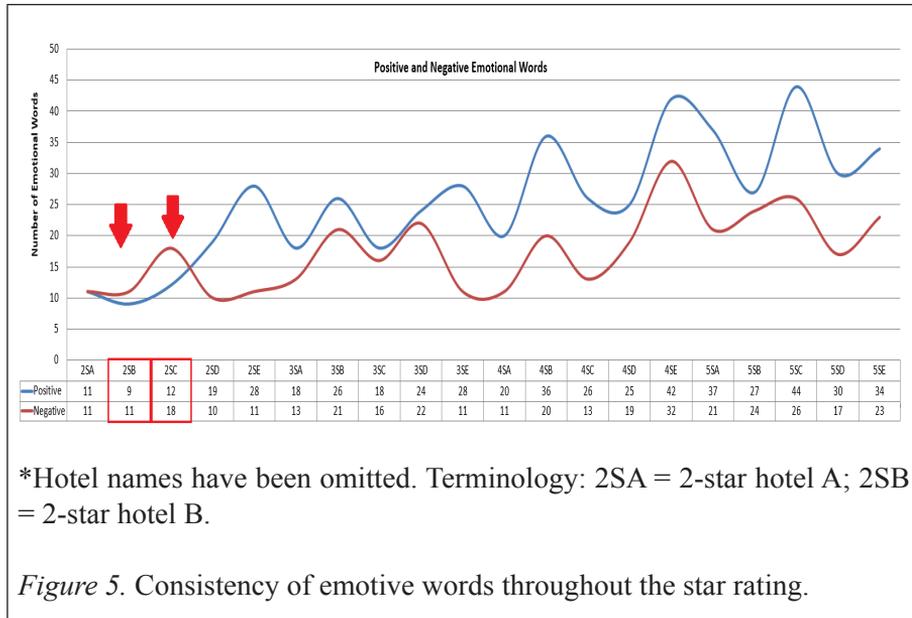


*Hotel names have been omitted. Terminology: 2SA = 2-star hotel A; 2SB = 2-star hotel B.

*Figure 5.* Consistency of emotive words throughout the star rating.

Figure 5 shows the consistency in the number of positive emotive words used throughout the higher star ranking hotels. The positive emotive words increased consistently with the star ranking. It appeared that reviewers expressed further with more positive emotive words. However, there were only two pointers in 2SB (Hotel B of 2-star rating) and 2SC (Hotel C of 2-star rating) with negative words higher than positive words, which provided us with a comparison between these comments. The results analysis revealed that there were more negative emotive words used than positive emotive words. The semantic tag-set of positive emotive actions in 2SB was 9 and the negative emotive action was 11. The 2SC positive emotion actions are 12 and negative emotive actions were 18. In addition, the moves of the positive and negative lines were consistent.

The top ten positive words from Set1A and Set2A were correlated using Pearson's correlation. The result indicated that there is no significant relationship between Set1A and Set2A (r=.408, p>.05). The same procedure

was repeated with the top ten negative words from Set1B and Set2B (r=.397, p>.05). The results showed that there are no significant relationships between both sets of positive or negative words. Therefore, the words are not associated and thus are used independently in the survey.

Table 6

*Correlation Table*

| Rating | Human raters |
|---|---|
| TripAdvisor positive rating | .690* |
| TripAdvisor negative rating | 818** |

* *p<.05, ** p<.01*

**Hypothesis 1:** There is a significant relationship between the TripAdvisor positive rating and the human raters' rating.

The result indicated that there is a positive significant relationship between the TripAdvisor positive rating and the human raters (r= .690, p<.05). Furthermore, the result suggested that a similar positive rating in TripAdvisor was reflected by the human raters. This further emphasizes the importance of online comment as it reflects the human raters rating and views of the accommodation or services used.

**Hypothesis 2:** There is a significant relationship between the TripAdvisor negative rating and the human raters' rating.

The result indicated that there is a significant positive relationship between the TripAdvisor negative rating and the human raters (r= .818, p<.05). The result indicated that a similar negative rating in TripAdvisor was associated with human raters' rating.

Table 7

*CrossTabulation of TripAdvisor Rating Versus Human Raters*

| | | Rating | | | | |
|---|---|---|---|---|---|---|
| | | Fair | Good | Very Good | Excellent | Total |
| Fair | Count | 1 | 6 | 0 | 0 | 7 |
| | % within human rater | 14.3 | 85.7 | 0 | 0 | 100 |
| | % within TripAdvisor rating | 100 | 46.2 | 0 | 0 | 17.5 |

(continued)

| | | Rating | | | | |
|---|---|---|---|---|---|---|
| | | Fair | Good | Very Good | Excellent | Total |
| Good | Count | 0 | 6 | 2 | 0 | 8 |
| | % within human rater | 0 | 75 | 25 | 0 | 100 |
| | % within TripAdvisor rating | 7 | 46.2 | 15.4 | 0 | 20 |
| Very Good | Count | 0 | 1 | 9 | 2 | 12 |
| | % within human rater | 0 | 8.3 | 75.0 | 16.7 | 100 |
| | % within TripAdvisor rating | 0 | 0 | 15.4 | 684.6 | 32.5 |
| Excellent | Count | 0 | 0 | 2 | 11 | 13 |
| | % within human rater | 0 | 0 | 15.4 | 84.6 | 100 |
| | % within TripAdvisor rating | 0 | 0 | 15.4 | 84.6 | 32.5 |

Cohen's Kappa was run to determine if there is an agreement between the two raters' judgments to understand better the degree to which the two raters gave their ratings in their judgment of *Fair*, *Good*, *Very Good* and *Excellent*. Table 7 reported the human raters' and TripAdvisor ratings; one respondent gave a *Fair* rating as agreed by both raters. Six (6) respondents gave a *Good* rating as agreed by both raters, and nine (9) respondents gave a *Very Good* rating as agreed by both raters. Eleven (11) respondents gave an *Excellent* rating as agreed by both raters. Meanwhile, there were 13 respondents (i.e. 6 + 1 + 2 + 2 + 2 = 13) with whom the two raters did not agree on the rating.

Cohen's Kappa ($\kappa$) was reported to be .553 ($p <.05$). This was the proportion of agreement *over and above* chance agreement; Kappa ($\kappa$) .553 represented a moderate strength of an agreement. In conclusion, there was a moderate agreement between the two raters' judgments; $\kappa$ = .553 ($p<.05$).

## DISCUSSION

The findings above are important for the decision-making process and further support the importance of the new electronic word-of-mouth way of marketing (Vermeulen & Seegers, 2008). Pre-connection and perception were established even before the visit (Brown, 2012). It is consistent with the findings of Ling, Beenen, Ludford, (2005), that is, individuals are inclined to contribute to the online community when they see a similarity in opinion and are more likely to share their personal views. The online reviews could influence consumers' attitudes and thus affect their decision-making in consuming the services and/or products. Classification techniques were used in the hotel reviews to suggest helpful yet contrasting reviews to end-users (O'Mahony & Smyth, 2010). According to human behaviour, individuals are confined to social

norms, thus, they would refer to online comments before making decisions. The users do not just look at hotel options; 77% reviewed the TripAdvisor reviews before actually choosing a hotel (PhoCusWright, 2014). The online hotel reviews increased the average chances of an individual in choosing a certain hotel service and promoting a greater awareness on the lesser-known hotels (Vermeulen & Seegers, 2008). On the other hand, negative comments lead to a higher conformity effects and gain higher acceptance by reviewers (Lee, Park, & Han, 2008). Thus, it is essential for business managers to employ a dedicated monitoring mechanism and take note of the negative comments before they go viral and create negative publicity for their services (Dean, 2004).

This study fills in the gaps in the studies conducted by Fong et al. (2016) and Dincer and Alrawadieh (2017). The former focused on one hotel and the latter only focused on 424 negative reviews of the hotels. Our study encompassed both positive and negative reviews from 20 hotels. Although studies on TripAdvisor are common, none have assessed as extensively as we have. Indeed, at present, general research in online review and consumer behaviour is still in its infancy. Therefore, it is sensible to monitor the online reviews (both in text and also the moticon format) in order to improve hotel management and customer satisfaction in order to realise the full potential of effective complaint management.

## CONCLUSION

Our study concluded that the online reviews and the human raters were consistent and the result indicated a moderate level of agreement of judgments. The human raters' evaluation and the reviewers' star rating were consistent. It corroborates the importance of comments and opinions by previous guests, regardless of the hotel's star rating. The analysis of the human raters' agreement with the original reviewers' comments indicates that the comments are reliable and a dependable source of reference prior to making a decision. The hotel star rating does not influence the guests' opinion; even the lowest star-rated hotel is still perceived as good by the guests. Future studies should consider examining whether there is any difference in the area of satisfaction (rated with excellent) among hotels of different ratings. It could offer some advice to the hotel industry about consumer preference irrespective of the hotel star rating. This research has also validated the importance of online reviews of business websites, such as TripAdvisor, in influencing consumer decision-making.

## REFERENCES

Abe, J. A. A. (2011). Positive emotions, emotional intelligence, and successful experiential learning. *Personality and Individual Differences*, *51*(7), 817-822. http://doi.org/10.1016/j.paid.2011.07.004

Alec (2013). *The 25 worst YouTube comments ever*. Retrieved 9 December 2015 from http://runt-of-the-web.com/worst-youtube-comments

Bermingham, A. (2014). *Would you dare? One business' risky approach to negative online reviews - Utter Digital*. Retrieved 12 December 2015 from http://utterdigital.com/digital-marketing/would-you-dare-one-business- risky-approach-to-negative-online-reviews/

Brown, B. (2012). Beyond recommendations: Local review websites and their impact. *ACM Trans. Comput.-Hum. Interac, 19*, 4, Article 27, DOI = 10.1145/2395131.2395134 http://doi.acm.org/10.1145/2395131.2395134

Chan, H. K., Lacka, E., Yee, R. W. Y., & Lim, M. K. (2015). The role of social media data in operations and production management. *International Journal of Production Research*, 1–11. http://doi.org/10.1080/00207543.2015.1053998

Chen, B., Guo, J., Tseng, B., & Yang, J. (2011). User reputation in a comment rating environment. *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, 159-167.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37-46.

Dean, D. (2004). Consumer reaction to negative publicity: Effects of corporate reputation, response, and responsibility for a crisis event. *Journal of Business Communication*, *41*(2), 192-211.

Dietz, J. (2015). *2015 Game publisher rankings - Metacritic*. Retrieved 7 December 2015 from http://www.metacritic.com/feature/game-publisher-rankings-for-2014-releases [Accessed September 10, 2015].

Dincer, M. Z., & Alrawadieh, Z. (2017). Negative word of mouse in the hotel industry: A content analysis of online reviews on luxury hotels in Jordan. *Journal of Hospitality Marketing & Management*, 10.1080/19368623.2017.1320258

Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Pacific Grove, Calif: Brooks/Cole Pub Co.

Filippi, P., Ocklenburg, S., Bowling, D. L., Heege, L., Güntürkün, O., Newen, A., & de Boer, B. (2016). More than words (and faces): Evidence for a Stroop effect of prosody in emotion word processing. *Cognition and Emotion*, *9931*, 1-13. https://doi.org/10.1080/02699931.2016.1177489

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions* (3rd ed.). London: Wiley.

Fong, L. H. N., Lei, S. S. I., & Law, R. (2016). Asymmetry of hotel ratings on TripAdvisor: Evidence from single- versus dual-valence reviews. *Journal of Hospitality Marketing & Management*, *00*(00), 1–16. https://doi.org/10.1080/19368623.2016.1178619

Gill, A. J., French, R. M., Gergle, D., & Oberlander, J. (2008). The language of emotion in short blog texts. *Proceedings of The 2008 ACM Conference on Computer Supported Cooperative Work*. USA: San Diego, doi>10.1145/1460563.1460612

Hassan, M. G., Hussain, F., & Saibu Rahman, M. B. (2013). Exploring usefulness of CRM and IT in Malaysian hotel industry: A qualitative approach. *Journal of Information and Communication Technology*, *12*, 21-37.

Hinojosa, J. A, Rincón-Pérez, I., Romero-Ferreiro, M. V., Martínez-García, N., Villalba-García, C., Montoro, P. R., & Pozo, M. A. (2016). The Madrid Affective Database for Spanish (MADS): Ratings of dominance, familiarity, subjective age of acquisition and sensory experience. *PLoS ONE* 11(5), e0155866. doi:10.1371/journal.pone.0155866

Huang, A. H., Yen, D. C., & Zhang, X. (2008). Exploring the potential effects of emoticons, *Information & Management*, *45*(7), 466-473.

Kauffman, R. J., Laf, H., Lin, H-C., & Chang, Y-S. (2009). Do textual comments and existing orders affect consumer participation in online group-buying? *Proceedings of the 42ⁿᵈ Hawaii International Conference on System Sciences*. Hawaii, 1-10, 10.1109/HICSS.2009.165

Klan, F. & Ries, B. K. (2014). Recommending judgment targets for rating provision. *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, 327–334. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6927642.

Krestel, R., & Dokoohaki, N. (2011). Diversifying product review rankings: Getting the full picture. In *Proceedings - 2011 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2011*, *1*, 138–145. http://doi.org/10.1109/WI-IAT.2011.33

Lee, J., Park, D-H., & Han, I. (2008). The effect of negative online consumer reviews on product attitude: An information processing view. *Electronic Commerce Research and Applications, 7*(3), 341-352.

Leskovec, J. (2011). Social media analytics: Tracking, modelling and predicting the flow of information through networks. *World Wide Web Internet and Web Information Systems*, 277–278. Retrieved 9 May 2015, from http://dl.acm.org/citation.cfm?id=1963309.

Li, Y., & Yu, D. (2015). Development of emotive word comprehension in Chinese children from 2 to 13 years old: Relationships with valence and empathy. *PLoS ONE*, *10*(12), 1–19. https://doi.org/10.1371/journal.pone.0143712

Ling, K., Beenen, G., Ludford, P., Wang, X., Chang, K., Li, X., Cosley, D., Frankowski, D., Terveen, L., Rashid, A. M., Resnick, P., & Kraut, R. (2005). Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication*, 10, 00. doi:10.1111/j.1083-6101.2005.tb00273.x

Matos-rodríguez, Y. E. (2014). *Online traveller ratings: Impact on determining the top 25 hotels*. Retrieved 10 August 2016, from http://scholarsarchive.jwu.edu/mba_student/38/

Moghaddam, S., & Martin, E (2011). ILDA: Interdependent LDA model for learning latent aspects and their ratings from online product reviews. *Proceedings of the Annual ACM SIGIR International conference on Research and Development in Information Retrieval (SIGIR- 2011).*

O'Mahony, M. P., & Smyth, B. (2010). A classification-based review recommended. *Knowledge-based Systems,* 23, 323-329.

PhoCusWright. (2014). Hotel guests read 6-12 reviews before booking, says new TripAdvisor survey. *Hotelmarketing'com*. Retrieved 9 December 2015, from http://hotelmarketing.com/index.php/content/article/hotel_guests_read_6_12_reviews_before_booking_says_tripadvisor_survey#

Raut, V. B., & Londhe, D. D. (2015). In: Guerrero J. E. (ed.), *Proceedings of International Conference on Computational Intelligence and Communication Networks (CICN),* India: Bhopal, 556-559.

Rayson, P. (2008). From keywords to key semantic domains. *International Journal of Corpus Linguistics*, *13*(4), 519-549.

Samsudin, N., Puteh, M., Hamdan, A. R., & Ahmad Nazri, M. Z. (2013). Normalization of noisy texts in Malaysian online reviews. *Journal of Information and Communication Technology*, 12, 147-159.

Sander, T., Teh, P. L., & Sloka, B. (2015). Use of social network site's profile for the employment-seeking process. *Advances in Intelligent Systems and Computing*, *353*, 1023–1032. http://doi.org/10.1007/978-3-319-16486-1

Sereno, S. C., Scott, G. G., Yao, B., Thaden, E. J., & O'Donnell, P. J. (2015). Emotive word processing: Does mood make a difference? *Frontiers in Psychology*, *6*, 1-13. https://doi.org/10.3389/fpsyg.2015.01191

Siersdorfer, S., Chelaru, S., Penro, J. S., Altingovde, I. S., & Nejdl, W. (2014). Analyzing and mining comments and comment ratings on the social web. *ACM Transactions on the Web*, 8(3), 1-39.

TripAdvisor. (2015). All about your TripAdvisor bubble rating | TripAdvisor Insights. *TripAdvisor*. Retrieved 3 January 2016 from https://www.tripadvisor.com/TripAdvisorInsights/n2640/all-about-your-tripadvisor-bubble-rating

Teh, P. L., Rayson, P., Pak, I., Piao, S., & Yeng, S. M. (2016). Reversing the polarity with emoticons. In E. Métais, F. Meziane, M. Saraee, V. Sugumaran, & S. Vadera (Eds.), *Natural Language Processing and Information Systems: 21st International Conference on Applications of Natural Language to Information Systems, NLDB 2016, Salford, UK, June 22-24, 2016, Proceedings* (pp. 453–458). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-41754-7_48

Vásquez, C. (2011). Complaints online: The case of TripAdvisor. *Journal of Pragmatics*, *43*, 1707–1717.

Vermeulen, I, E., & Seegers, D. (2008). Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism Management,* 1-5, doi:10.1016/j.tourman.2008.04.008

Wong, Y. J. (2015). The psychology of encouragement: Theory, research, and applications. *The Counselling Psychologist*, *43*(2), 178–216. http://doi.org/10.1177/0011000014545091

Wyndham. (2015). Wyndham Midtown 45 TripAdvisor Reviews & Ratings | Hotel Reviews in New York, NY 10017-3301. Retrieved 19 December 2015 from http://www.wyndham.com/hotels/new-york/new-york/wyndham-midtown-45/hotel-reviews-ratings

Zhang, Z., Guo, C., & Goes, P. (2013). Product comparison networks for competitive analysis of online word-of-mouth. *ACM Transactions on Management Information Systems*, *3*(4). Retrieved 9 January 2016 from http://dl.acm.org/citation.cfm?id=2407744.

Zhang, Z. Q., Zhang, Z., & Yang, Y. (2016). The power of expert identity: How website-recognized expert reviews influence travellers' online rating behaviour. *Tourism Management*, 55, 15-24.