Measuring the effectiveness of answers in Yahoo! Answers

Chua, Alton Yeow Kuan; Banerjee, Snehasish

2015

Chua, A. Y. K., & Banerjee, S. (2015). Measuring the effectiveness of answers in Yahoo! Answers. Online information review, 39(1), 104-118.

https://hdl.handle.net/10356/107409

https://doi.org/10.1108/OIR-10-2014-0232

Measuring the Effectiveness of Answers in Yahoo! Answers

Alton Y. K. Chua and Snehasish Banerjee

Abstract

Purpose – This study investigates the ways in which effectiveness of answers in Yahoo! Answers, one of the largest community question answering sites (CQAs), is related to question types and answerer reputation. Effective answers are defined as those that are detailed, readable, superior in quality, and contributed promptly. Five question types that were studied include factoid, list, definition, complex interactive, and opinion. Answerer reputation refers to the past track record of answerers in the community.

Design/Methodology/Approach – The dataset comprises 1,459 answers posted in Yahoo! Answers in response to 464 questions that were distributed across the five question types. The analysis was done using factorial analysis of variance.

Findings – The results indicate that factoid, definition and opinion questions were comparable in attracting high quality as well as readable answers. Although reputed answerers generally fared better in offering detailed and high quality answers, novices were found to submit more readable responses. Moreover, novices were more prompt in answering factoid, list and definition questions.

Originality/value – By analyzing variations in answer effectiveness with a twin-focus on question types and answerer reputation, this study explores a strand of CQA research that has hitherto received limited attention. The findings offer insights to users and designers of CQAs.

Keywords Community question answering, Answer effectiveness, Question types, Answerer reputation

Paper type Research paper

Introduction

Recent years have witnessed the rise of collaborative information seeking applications known as community question answering sites (CQAs). Essentially, CQAs allow online users to ask and answer various types of questions in natural language, browse the corpus of already answered questions, rate the quality of answers, as well as vote for the best answers (Deng *et al.*, 2011; Qu *et al.*, 2012). They serve as avenues for users to tap into the wisdom of crowds (Surowiecki, 2004).

To most users, the value of CQAs lies almost solely in the effectiveness of answers returned. Effective answers are typically those which are sufficiently detailed, readable, superior in quality, and returned promptly. Detailed answers that come with expanded explanations generally connote a greater sense of credibility than those that are terse (Jeon *et al*., 2006). Likewise, readable answers could better meet information needs than those that are difficult to read (Ghose and Ipeirotis, 2011; Toba *et al*., 2014). Answers that are superior in quality would always be of greater relevance than those that are inept (Blooma *et al*., 2012). Furthermore, prompt answers are almost always perceived as being more effective than those that are late (Mamykina *et al*., 2011). There is little wonder why much CQA research has been trained on answer effectiveness (e.g., Blooma *et al*., 2012; Jeon *et al*., 2006; Kim & Oh, 2009).

As an extension to current research on answer effectiveness, this study hypothesizes that obtaining an effective answer is a joint function of both askers and answerers. For one, answer effectiveness in CQAs could be predicated by the types of questions posted by askers. In particular, questions asked in CQAs could be classified into five types, namely, factoid, list, definition, complex interactive, and opinion (Dang *et al*., 2007; Lin and Katz, 2006; Voorhees, 2004, 2005). Factoid, list and definition type questions that tend to elicit objective responses could be answered more effectively vis-à-vis complex interactive and opinion questions that might require a fuzzy discourse.

The other part of the equation on answer effectiveness lies with answerers. Most CQAs have a reputation point system to recognize answerers whose submissions are endorsed by the community. It is therefore in the interest of answerers to offer effective answers. Nonetheless, answerers with high reputation scores (henceforth, reputed answerers) and those with low reputation scores (henceforth, novice answerers) might not always attract the same level of attention. This is because CQA users' perception towards answer effectiveness could potentially be clouded by the reputation of answerers (Agichtein *et al*., 2009; Jeon *et al*., 2006). If so, reputed answerers would continue to attract endorsements while novice answerers would remain mostly ignored even if contributions from both are equally compelling. However, current research has yet to shed light on whether reputed answerers consistently outperform novices in offering effective answers.

Hence, stemmed from an asker-answerer perspective, this study seeks to investigate the ways in which answer effectiveness in Yahoo! Answers is related to question types and answerer reputation. Specifically, Yahoo! Answers was chosen for investigation for being one of the largest CQAs (Adamic *et al*., 2008; Jin *et al*., 2013). Moreover, it uses an extensive reputation point system to summarize the past track record of its users, thereby making it appropriate for this study.

By analyzing variations in answer effectiveness with a twin-focus on question types and answerer reputation, this study explores a strand of CQA research that has received limited attention thus far. To the best of our knowledge, prior studies have rarely explicated such subtleties in answer effectiveness. This study therefore has potential implications for different stakeholders in the CQA community. In particular, askers could be guided on the types of questions to post in CQAs to maximize chances of obtaining effective answers. Answerers could glean insights from this study on the ways to write answers in order to establish their reputation in the community. For CQA designers, this study could highlight areas of improvements in the design of CQA websites so that each answer, regardless by whom it was submitted, will be weighed according to its own merit.

The remainder of this article is structured as follows. The following section reviews the literature on the three key themes, namely, answer effectiveness, question types, and answerer reputation. The Methods section explains the procedure for data collection, measurement and analysis. The results are presented next. This is followed by a discussion of the results. Finally, the article concludes by highlighting its implications and limitations.

Literature Review

Answer Effectiveness

For the purpose of this study, effective answers refer to those that are detailed, readable, superior in quality, and contributed promptly (Blooma *et al*., 2012; Jeon *et al*., 2006; Mamykina *et al*., 2011; Toba *et al*., 2014). The extent to which answers posted in CQAs are detailed with expanded explanations serves a crucial proxy for answer effectiveness (Jeon *et al*., 2006; Toba *et al*., 2014). Level of details has long been known as a key indicator of effective writing (Larkey, 1998). However, it could vary across question types (Jeon *et al*., 2005). Questions eliciting objective responses could be answered more effectively without being overly detailed compared with questions that entail fuzzy discourse. The level of details in answers might also vary across the reputation of answerers. For example, expert answerers with high reputation in the community could submit terse yet effective answers.

Besides, a readable answer could enhance clarity, comprehension, retention and thereby, be deemed effective by a wider audience (Ghose and Ipeirotis, 2011). Effective answers posted in CQAs are generally easy to read (Toba *et al*., 2014). However, the extent to which answer readability varies in relation to the ways questions are phrased or across the reputation of answerers is largely unknown.

Answer quality has piqued substantial scholarly attention in recent years (e.g., Blooma *et al*., 2012; Jeon *et al*., 2006; Kim and Oh, 2009). This is perhaps because the quality of answers that are posted in CQAs without any gate-keeping process could range from excellent to abysmal (Agichtein *et al*., 2008; Suryanto *et al*., 2009). While some answerers could submit high quality answers out of altruism or intention to establish reputation, others might post sub-standard responses out of boredom or to have fun (Chen and Sin, 2013). Even though the overall quality of answers posted in popular CQAs is acceptable, the quality of specific answers differs drastically (Su *et al*., 2007).

Promptness is another essential determinant for answer effectiveness (Mamykina *et al*., 2011; Shah, 2011). Most askers who fail to receive prompt answers from CQAs tend to turn to alternative sources to meet their information needs. Moreover, they are unlikely to return to the CQAs to check if their questions had been answered (Kitzie and Shah, 2011). To an asker, the value of a high-quality answer will greatly be undermined if it was posted after a long lag-time.

Question Types

Studies on question types in CQAs are generally rare. Among the few studies, scholars have identified several types of questions posted in CQAs. For example, Harper *et al*. (2008) identified three question types, namely, factual, opinion and advice. Factual questions sought objective data. Opinion questions were meant to elicit others' thoughts on a given topic. Advice questions solicited recommendations to address the personal situation of askers. In yet another study, Nam *et al*. (2009) identified additional question types such as procedural and task-oriented. The former sought procedure to handle a given situation while the latter solicited the details about a given task. More recently, Westbrook (2014) grouped questions into four types, namely, advice, binary, explanation, and fact. Advice questions were meant to seek solutions to personal problems. Binary questions asked for a choice between two options. Explanation questions sought detailed responses while fact questions solicited factual answers. Conceivably, these question types are not always mutually exclusive. For instance, the lines between opinion and advice questions as well as between procedural and task-oriented questions are often blurred.

Therefore, for the purpose of this study, questions commonly asked in CQAs were classified into five mutually exclusive types, namely, factoid, list, definition, complex interactive, and opinion. These have been identified in prior studies as possible ways of phrasing questions (Dang *et al*., 2007; Lin and Katz, 2006; Voorhees, 2004, 2005). Moreover, this taxonomy of question types is more comprehensive than those suggested by Harper *et al*. (2008), Nam *et al*. (2009), and Westbrook (2014). Specifically, factoid questions refer to those that are meant to seek factual answers, for example, "*Where was FIFA world cup 2014 held?*" List questions require collections of multiple related answers, for example, "*What are the planets in the solar system?*" Definition questions are meant to elicit the meaning of a term or a concept, for example, "*What is fractional distillation?*" Complex interactive questions are the ones set in a specific context with a series of sub-questions built on the previous questions, for example, "*What is diabetes mellitus? What is the treatment for diabetes mellitus?*" Opinion questions are meant to seek advice and instructions about a specific phenomenon, for example, "*How do you change font in windows 7?*"

Answerer Reputation

The fact that CQAs allow users to post content without any editorial control is both a boon and a bane. On the one hand, by harnessing the wisdom of everyone on the CQA community, it serves as a viable platform for users to meet their information needs (Surowiecki, 2004). On the other hand, the voluntary and participative nature of CQAs allow for the dissemination of sub-standard answers by any answerers in the community (Suryanto *et al*., 2009). Most CQAs overcome this problem by using a reputation point system to recognize users' participation in the community.

Reputation of information source has long been found to influence readers' judgment (Chaiken, 1980). A piece of information contributed by a reputed source is generally viewed more favorably vis-à-vis one that is shared by an amateur source (Wathen and Burkell, 2002). This perception bias has also been found to persist in the context of CQAs (Agichtein *et al*., 2009; Jeon *et al*., 2006). Hence, answers from a reputed answerer tend to attract more endorsements through the reputation point system than those from a novice. In such a scenario where the rich get richer and the poor get poorer, reputed answerers and novice answerers rarely compete on a level-playing field even if both submit equally effective answers. Hitherto, little scholarly attention has been trained on this issue. Hence, it is a timely attempt to analyze answer effectiveness in CQAs as a function of both question types and answerer reputation.

Methods

Dataset

The dataset for this study was created by three research associates (henceforth, coders), who held graduate degrees in Information Science with more than two years of professional experience. Moreover, they were familiar with the use of CQAs. The data collection process involved three steps, namely, identifying questions, posting questions in Yahoo! Answers, and harvesting answers. It lasted from July, 2011 to April, 2012.

The first step was to identify some 600 questions of the five question types, namely, factoid, list, definition, complex interactive and opinion. For this purpose, questions were retrieved from two popular CQAs, namely, WikiAnswers and Answerbag (300 from each), which consistently attract substantial user-base (Shachaf and Rosenbaum, 2009). Specifically, questions were drawn from five categories that included entertainment, sports, computers, science and health. These categories were chosen given that they are commonly available across most CQAs, and attract active participation. The questions garnered were jointly coded into the five question types. The process of garnering and coding questions was iterated to yield a corpus of 600 questions uniformly distributed across the five question types and the five categories.

The second step was to post the identified questions in Yahoo! Answers with randomized timings as much as possible. Specifically, Yahoo! Answers was chosen because it is not only one of the largest CQAs but also represents one of the most active collaborative information seeking and knowledge sharing communities (Adamic *et al*., 2008; Jin *et al*., 2013). Its large user-base provides an ideal setting to investigate answer effectiveness. Furthermore, Yahoo! Answers tracks the past record of users using an extensive reputation point system. In particular, it gives points to users who respond to questions, vote for answers, and whose answers are selected as best answers. The availability of this metadata makes the site appropriate for this study.

The third step was to harvest answers attracted by the posted questions over a window of four days. Specifically, a window of four days was considered because Yahoo! Answers by default uses it as the upper threshold before unanswered questions expire (Yang *et al*., 2011). Moreover, most answering activities in CQAs take place soon after questions are posted (Mamykina *et al*., 2011; Shah, 2011). Therefore, questions that fail to attract answers in the first four days are unlikely to receive any responses later. Among the 600 questions posted in Yahoo! Answers, 464 questions attracted a total of 1,459 answers, all of which were harvested and analyzed. For each answer, the time elapsed in minutes between posting questions and receiving answers was recorded. The reputation scores of the corresponding answerers were also retrieved.

Such a step-wise data collection process was preferred over scraping data directly from Yahoo! Answers. This facilitated controlling for the time of question posting. As indicated earlier in the second step of data collection, questions were posted in Yahoo! Answers with randomized timings as much as possible. As a result, the potential confounding effects of different time zones, time of the day, and day of the week on attracting answers from users around the globe were minimized.

Measures and Analysis

This study analyzes answer effectiveness across question types and answerer reputation. Answer effectiveness, was operationalized in terms of four components that include level of details, readability, quality and promptness. First, level of details was measured using answer length in words (Jeon *et al*., 2006). After all, lengthy answers could be perceived as being sincere, trustworthy and hence, effective.

Second, readability of answers was measured using readability indicators such as Gunning-Fog Index (FOG), Automated Readability Index (ARI) and Flesch-Kincaid Grade Index (FKG). In particular, FOG relies on average sentence length and proportion of words with more than two syllables, ARI is based on average word length and sentence length, while FKG depends on average sentence length and average number of syllables per word. Lower values in these indicators represent a more readable review. Readability of each answer was calculated as the arithmetic mean of the three readability indicators. This was necessary to take into account the strengths of all the three (Ghose and Ipeirotis, 2011).

Third, quality of answers was measured in terms of three criteria, namely, content quality, cognitive quality, and socio-emotional quality (Chua and Banerjee, 2013; Kim and Oh, 2009). Content quality refers to the content-richness of answers. It is enhanced by factors such as reasonableness and soundness of answers. Reasonableness refers to the credibility of an answer while soundness measures its completeness (Blooma *et al*., 2011). Cognitive quality is a measure of answers' ability to pique the cognitive cues of users' knowledge. It is enhanced by factors such as understandability and novelty. Understandability refers to the comprehensibility of an answer while novelty measures its ability to trigger creative thinking (Kelly *et al*., 2007; Kim and Oh, 2009). Socio-emotional quality is a measure of interpersonal relationships and emotions as reflected through answers. Gratitude, appreciation and empathy are some forms of emotions commonly expressed in CQAs to thank others for sharing their knowledge or providing emotional support (Kim and Oh, 2009). The three coders were employed to rate answers indicating the extent to which they agreed that entries were rich in content quality, cognitive quality and socio-emotional quality (1 = strongly disagree, 5 = strongly agree). While rating answer quality, answerer reputation of the given answer was concealed from the coders to minimize biases.

Fourth, promptness of answers was measured by calculating the time elapsed in minutes between posting a question and receiving an answer based on system timestamp. Generally, the popularity of Yahoo! Answers stems from its prompt turnaround time for answers (Shah *et al*., 2008). An answer received promptly could be more effective than one obtained after a long delay.

Question types were determined by agreement among coders, as indicated earlier in the first step of the data collection procedure. Answerer reputation was measured on the basis of points earned by the answerers in Yahoo! Answers (Blooma *et al*., 2012). Following standard practices to dichotomize a continuous variable (e.g., Allen, 1998; Cortese and Lustria, 2012; Gurrea, et al., 2013), a median-split was used to classify answerers as either reputed or novice.

Finally, two-way factorial Analysis of Variance (ANOVA) was used to analyze answer effectiveness as a function of question types and answerer reputation. This statistical procedure is appropriate to analyze ways in which two factors are related to an outcome (Darwin, 2008; Rezaei, and Zakariaie, 2011). This study intended to investigate the ways in which two factors, namely, question types and answerer reputation, were related to answer effectiveness, which is conceived as level of details, readability, quality and promptness of responses. Therefore, a 5 (question type: factoid, list, definition, complex interactive, and opinion) x 2 (answerer reputation: reputed and novice) two-way factorial ANOVA was used to disinter the extent to which the following varied across question types and answerer reputation: (1) level of details in answers, (2) answer readability, (3) answer quality, and (4) answer promptness. When a statistically significant relationship was detected for question types, Tukey's Honestly Significant Difference post hoc test (henceforth, post-hoc test) was used to identify the specific question types that differed from one another (Darwin, 2008).

Results

Inter-coder Reliability

As indicated in the Methods, this study relied on coding for two purposes. The first was to ascertain question types. All the 600 questions that were identified for posting in Yahoo! Answers were coded into the five question types, namely, factoid, list, definition, complex interactive and opinion, by the three coders. The mean pair-wise inter-coder reliability in terms of Cohen's Kappa was 0.94, indicating non-chance level of agreement.

Moreover, coding was used to rate answer quality. The three coders were asked to score answers on a scale of 1 to 5 in terms of the three quality criteria, namely, content quality, cognitive quality, and socio-emotional quality. For this purpose, the coders familiarized themselves with the three criteria and independently rated a set of randomly selected 200 answers. The quality of each answer was calculated as the arithmetic mean of the scores for the three criteria. The mean pair-wise inter-coder reliability among the coders in terms of Cohen's Kappa was 0.82, indicating non-chance level of agreement. Thereafter, the remaining 1,259 answers in the dataset were distributed among the coders in a non-overlapping fashion so that each could code comparable number of entries.

Descriptive Statistics

As indicated earlier, a median-split of reputation points was used to classify answerers as either reputed or novice. The overall nature of reputation points in the dataset is summarized as follows: Mean = 25337.50, SD = 72867.76, Median = 2075, Mode = 336, Min = 0, Max = 874575, Skewness = 5.86, Kurtosis = 43.16. In general, there were relatively more answerers with low reputation points compared with those with high reputation points. As a result, it is not surprising that even though the highest reputation point was 874575, the median was relatively low (2075). Based on the median-split, answerers with reputation scores above 2075 points were deemed as reputed answerers while the rest were reckoned as novices.

The dataset comprised a total of 1,459 answers (factoid = 332, list = 273, definition = 254, complex interactive = 290, opinion = 310). Of these, 729 were contributed by reputed answerers while the remaining 730 were posted by novice answerers.

Across question types, level of details was the highest for answers to complex interactive questions (54.75 ± 71.30) while factoid questions were found to lie at the other end of the spectrum (46.17 ± 65.59). Answer readability was the best for opinion questions (6.53 ± 6.96) whereas list questions lagged behind in the rear (12.54 ± 19.34). Answer quality was the highest for factoid questions (3.92 ± 0.67) while complex interactive questions were found to lie at the other end of the spectrum (3.56 ± 1.03). Answer promptness was the best for definition questions (307.81 ± 806.74) whereas list questions appeared to attract answers with the longest turnaround time (443.51 ± 1089.11).

Across answerer reputation, level of details was greater for answers contributed by reputed answerers (58.07 ± 75.21) compared with novices (40.03 ± 55.13). In terms of answer readability, novice answerers (7.87 ± 10.77) fared better than reputed answerers (9.39 ± 15.01). With respect to answer quality, reputed answerers (3.85 ± 0.80) appeared to outperform novices (3.73 ± 0.93). In terms of answer promptness, both reputed (369.26 ± 911.24) and novice answerers (369.06 ± 810.80) appeared to exhibit comparable performance. The descriptive statistics of the dataset are presented in Table 1.

Table 1: Descriptive statistics of the dataset.

Answer effectiveness	Question types					Answerer reputation	
	Factoid	**List**	**Definition**	**Complex**	**Opinion**	**Reputed**	**Novice**
Level of details	46.17 ± 65.59	48.22 ± 70.58	49.10 ± 59.10	54.75 ± 71.30	47.46 ± 65.05	58.07 ± 75.21	40.03 ± 55.13
Readability	6.71 ± 6.23	12.54 ± 19.34	9.02 ± 7.40	9.05 ± 18.52	6.53 ± 6.96	9.39 ± 15.01	7.87 ± 10.77
Quality	3.92 ± 0.67	3.89 ± 0.81	3.85 ± 0.85	3.56 ± 1.03	3.74 ± 0.91	3.85 ± 0.80	3.73 ± 0.93
Promptness	372.97 ± 904.00	443.51 ± 1089.11	307.81 ± 806.74	341.83 ± 667.20	375.45 ± 793.58	369.26 ± 911.24	369.06 ± 810.80

Inferential Statistics

With respect to level of details, answers did not significantly differ across question types. However, level of details in answers differed significantly across answerer reputation, F(1, 1449) = 26.50, p < 0.001. Reputed answerers consistently appeared to offer lengthier answers than novices across all question types. The interaction between question types and answerer reputation was non-significant. Figure 1 depicts the variation of level of details in answers across question types and answerer reputation.

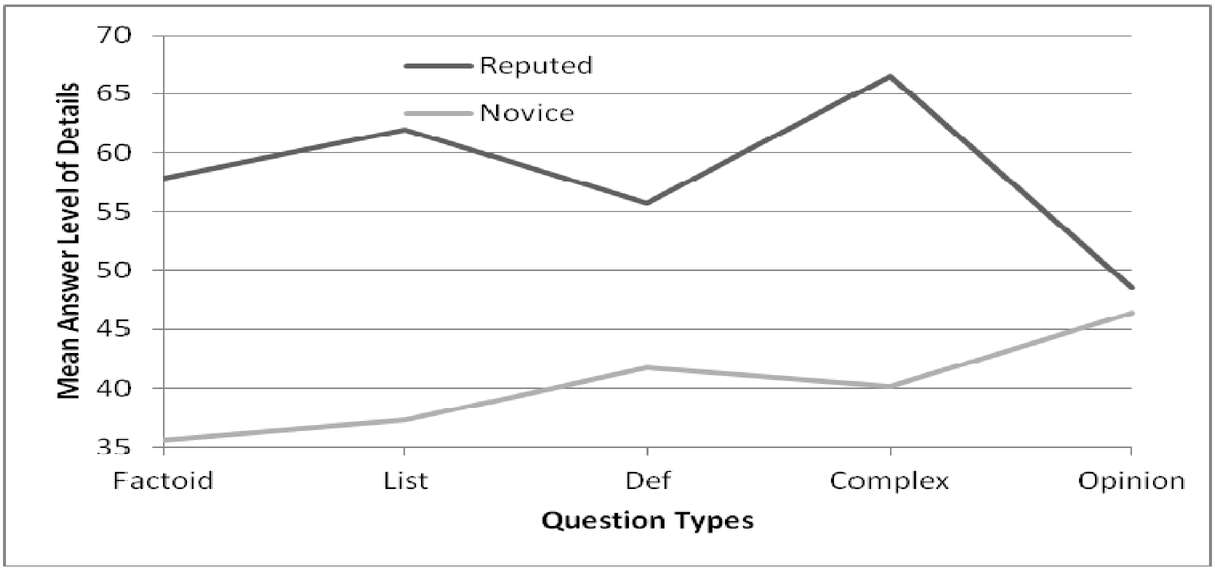

Figure 1: Variation of level of details in answers across question types and answerer reputation.

With respect to readability, answers differed significantly across question types, F(4, 1449) = 10.29, p < 0.001. Answer readability was the best for opinion questions. Furthermore, the post-hoc test showed no statistically significant difference in readability between answers to opinion questions and responses to factoid, definition and complex interactive questions. On the other hand, answers to list questions were generally found wanting in terms of readability. Answer readability also differed significantly across answerer reputation, F(1, 1449) = 6.27, p < 0.01. Answers contributed by novices were generally more readable than those posted by reputed answerers. However, the interaction between question types and answerer reputation was non-significant. Figure 2 depicts the variation of answer readability across question types and answerer reputation.

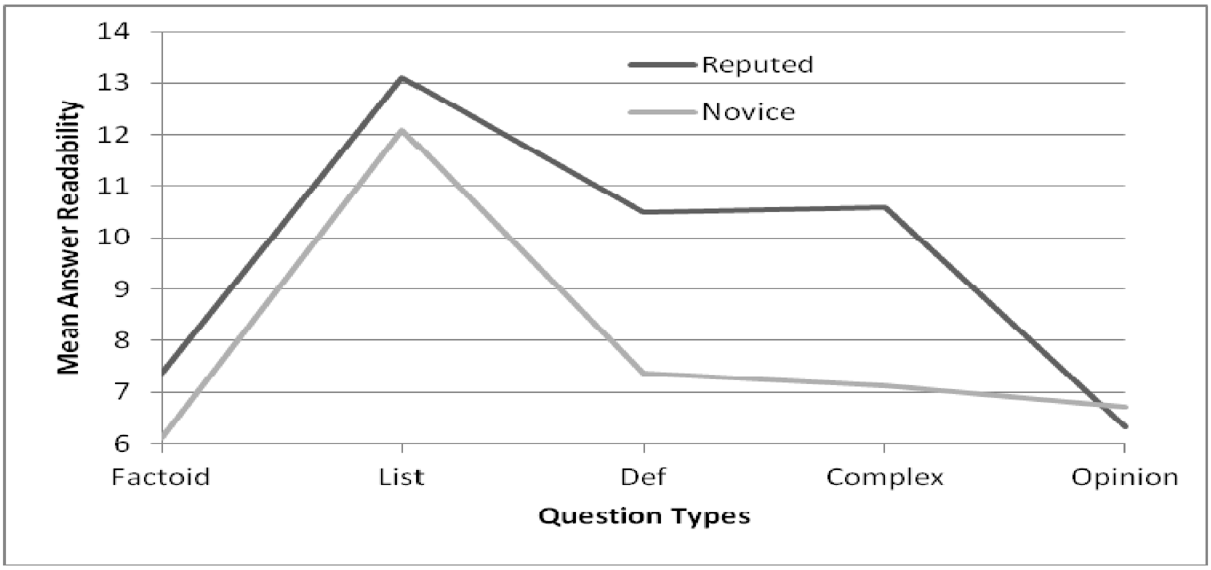

Figure 2: Variation of answer readability across question types and answerer reputation.

With respect to quality, answers differed significantly across question types, F(4, 1449) = 9.00, p < 0.001. In particular, answer quality was the best for factoid questions. Furthermore, the post-hoc test showed no statistically significant difference in quality between answers to factoid questions and responses to list, definition and opinion questions. On the other hand, answers to complex interactive questions were generally found wanting in terms of quality. Answer quality also differed significantly across answerer reputation, F(1, 1449) = 8.58, p < 0.01. Answers contributed by reputed answerers were generally better in quality than those contributed by novices. However, the interaction between question types and answerer reputation was non-significant. Figure 3 depicts the variation of answer quality across question types and answerer reputation.

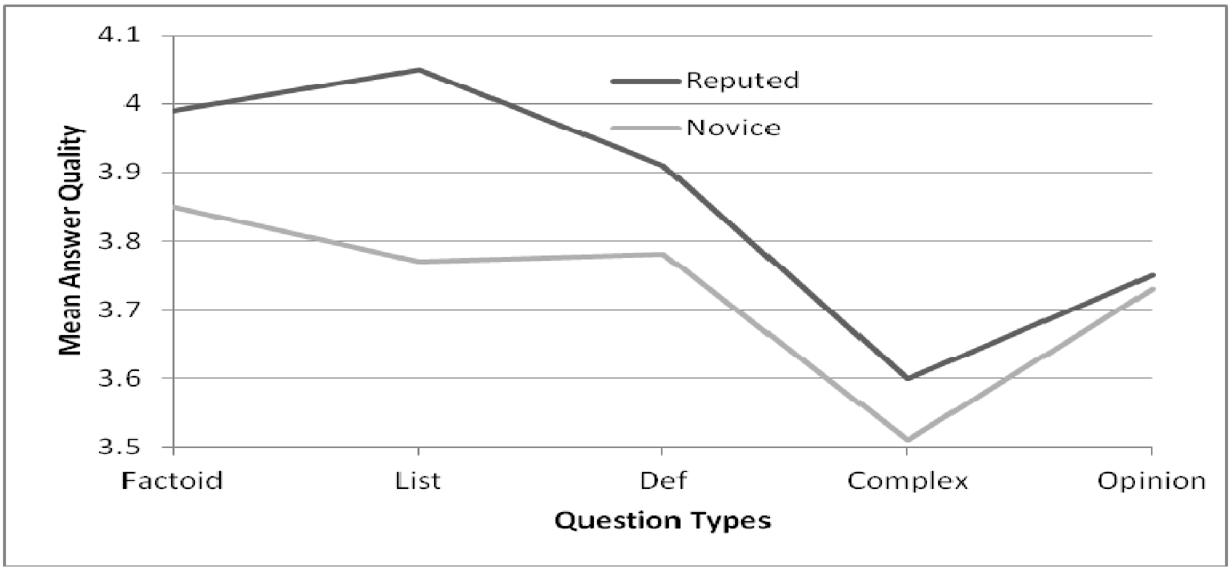

Figure 3: Variation of answer quality across question types and answerer reputation.

With respect to promptness, answers differed significantly across neither question types nor answerer reputation. Nonetheless, there was a statistically significant interaction between question types and answerer reputation, F(4, 1449) = 2.56, $p < 0.05$. Reputed answerers appeared to outperform novices in providing prompt answers for complex interactive and opinion questions. On the other hand, novices fared better in offering prompt answers in response to factoid, list and definition questions. Figure 4 depicts the variation of answer promptness across question types and answerer reputation.

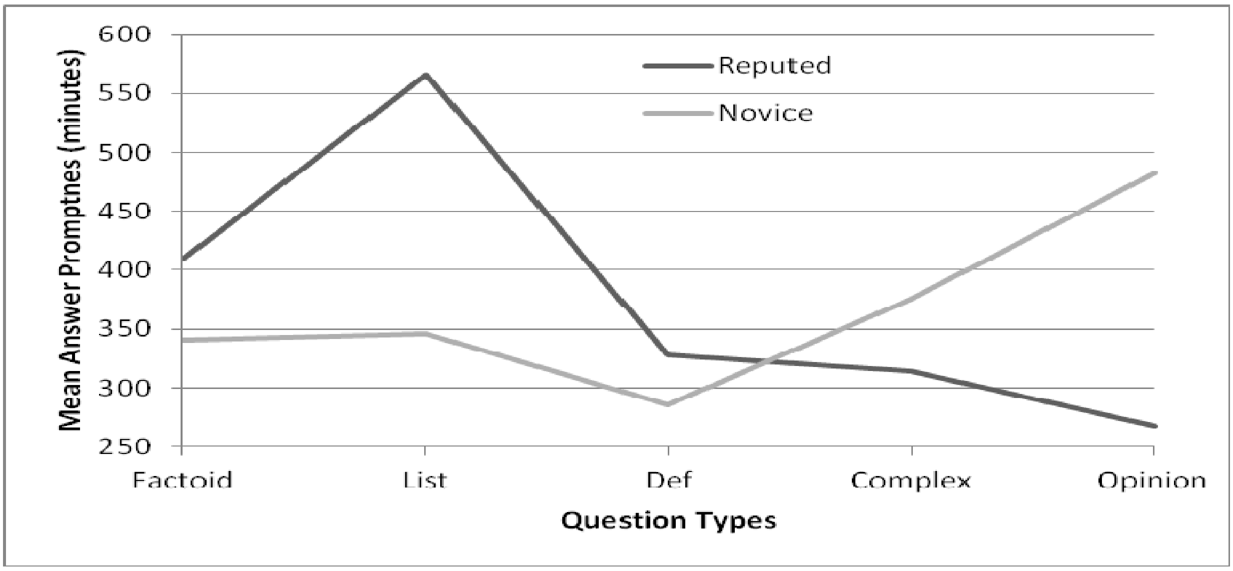

Figure 4: Variation of answer promptness across question types and answerer reputation.

Discussion

Three key findings can be gleaned from the results. First, factoid, definition and opinion questions are comparable in attracting high quality as well as readable answers. In one of the related studies, Harper *et al*. (2009) suggested that answer quality to questions seeking fuzzy discourse could be worse compared with that to questions seeking objective responses. In contrast, this study found that even opinion questions attracted answers that were as good as those submitted in response to factoid and definition questions. In light of this finding, CQAs appear as a robust collaborative information-seeking platform.

However, answers to list questions were found wanting in terms of readability. Since list questions such as "*Please recommend some good science fiction films.*" seek sets of multiple related items, answers to such queries could contain long lists of comma-separated entries. This in turn might result in lengthy sentences in answers, thereby compromising readability. However, on delving deeper into some of the answers to list questions, it was found that even though they had poor readability scores, they were not too difficult to be read by humans. Nonetheless, it seems that askers should ask list questions that do not entail overly long sets of entries to maximize chances of obtaining readable responses. Alternatively, they could consider referring to other information sources such as official websites to seek answers to list questions. Moreover, answers to complex interactive questions were found wanting in terms of quality. For example, complex interactive questions such as "*What are the rules in golf? How to play?*" comprise a series of related sub-questions, and are cognitively challenging to answer. Hence, askers should break up a complex interactive question into distinct stand-alone queries to attract effective answers. This finding bears a striking similarity to the methods of designing questionnaires in social science research. To elicit unambiguous responses, double-barreled questions should always be avoided in questionnaires (Berg & Lune, 2004; Palanisamy, 2014). Likewise, to elicit high quality answers, it appears that complex interactive questions should be avoided in CQAs.

Second, reputed answerers fared better than novices in offering detailed and high quality answers. The former consistently posted lengthier answers with greater levels of discourse. Even though answers could well be verbose or consist of text pasted from other sources, lengthy answers are generally perceived as being more effective vis-à-vis those that are sketchy. Moreover, answers submitted by novices were generally richer in content quality, cognitive quality as well as socio-emotional quality compared with responses posted by novices. While prior studies merely suggested that reputed answerers attract more attention than novices in CQA communities (Agichtein *et al.*, 2009; Jeon *et al.*, 2006), this finding now sheds more light by showing that the former indeed offered more effective answers than the latter. Answerer reputation is thus a useful heuristic to gauge answer effectiveness in CQAs.

However, reputed answerers were found to submit less readable answers than their novice counterparts. A possible explanation for this stems from reputed answerers' inclination of self-presentation. Users are often inclined to raise their online self-esteem through self-presentation (Krämer and Winter, 2008; Mehdizadeh, 2010). Specifically in CQAs, users could self-present by displaying their linguistic competence, which is known to enhance perceptions of credibility (Ghose and Ipeirotis, 2011). It is possible that such a device is mostly used by established CQA users, who seek to entrench their position in the community. As a result, reputed answerers could use sophisticated language in answers with lengthy words and sentences to impress the online community. This in turn could take a toll on readability. In contrast, novices are perhaps not overly driven by intention to self-present. Hence, answers posted by them were found to be relatively more readable. Nonetheless, it would require more scholarly investigation to unravel the exact reasons for which reputed answerers submit less readable responses vis-à-vis novices.

Third, novices appear to maintain the momentum in CQAs by providing prompt answers to factoid, list and definition questions. Prior research indicates that if CQA users fail to receive answers promptly, they would turn to alternative sources of information instead of returning back to the CQAs (Kitzie and Shah, 2011; Shah, 2011). Thus, a responsive community is crucial for the viability of CQAs (Mamykina *et al.*, 2011). Novices appear to play a key role by offering prompt answers to factoid, list and definition questions, likely because these questions are generally easy to answer. Hence, they do not adequately pique the interests of reputed answerers, who perhaps prefer the challenge of answering complex interactive and opinion questions.

Besides being prompt, novices' answers to factoid, list and definition questions were found to be more readable than those from reputed answerers. A readable answer is crucial to enhance clarity, comprehension and retention (Ghose and Ipeirotis, 2011), and will certainly be well-received. This finding thus offers insights into the possible way newbies mature in CQA communities. For a start, novices, could perhaps focus on factoid, list or definition questions that are relatively easy to answer. The promptness and the readability of their responses would earn them endorsements, which could help them enhance their reputation and grow in the community.

Conclusion

This study investigated the effectiveness of answers in Yahoo! Answers by incorporating question types and answerer reputation. Effective answers were defined as those that are detailed, readable, superior in quality, and contributed promptly. Five types of question were considered, namely, factoid, list, definition, complex interactive and opinion. Answerers were classified into reputed or novice based on median-split of their reputation points in Yahoo! Answers. The results indicate that factoid, definition and opinion questions were comparable in attracting high quality as well as readable answers. Although reputed answerers generally fared better in offering detailed and high quality answers, novices were found to submit more readable responses. Moreover, novices were more prompt in answering factoid, list and definition questions.

By analyzing answer effectiveness with a twin-focus on question types and answerer reputation, this study explored a territory of CQA research that has hitherto been relatively uncharted. The findings have implications for askers, answerers and designers of CQAs. For askers, the findings suggest that most question types tend to attract answers with comparable details, readability, quality and promptness. However, answers to list questions could lack readability while responses to complex interactive questions could suffer in terms of quality. Therefore, it appears that users should ask factoid, definition and opinion questions to maximize chances of attracting effective answers. On the other hand, they should not expect highly readable answers in response to list questions, especially those that entail a long set of entries. Furthermore, they should consider breaking up complex interactive questions, which comprise a series of sub-questions, into independent queries to maximize chances of obtaining effective answers in CQAs.

Besides, askers should not be too overly influenced by answerer reputation in assessing answer effectiveness. It is conceivable that CQA users could rely on the heuristic of answerer reputation to ascertain the effectiveness of answers, especially for those in response to difficult questions. However, the results indicate that even though novice answerers could lag behind reputed answerers in terms of level of details and quality of answers, the former consistently contribute more readable responses. Furthermore, novices seem to be more prompt than reputed answerers in responding to factoid, list and definition questions.

For answerers, the findings suggest that novices need to contribute more detailed answers of better quality to enhance their reputation in the community. Furthermore, they need to be more prompt in responding to complex interactive and opinion questions. For reputed answerers who had already established their standing in the community, the findings suggest that they do not have to succumb to the pressure of self-presentation. In so being, they would be able to post more readable answers.

For CQA designers, the findings offer implications to fine-tune the reputation point system. As indicated earlier, reputed and novice answerers might not always compete on a level-playing field because askers' perception towards the effectiveness of answers could be influenced by answerer reputation. Most CQAs facilitate earning points based on levels of activity and answer quality. For example, Yahoo! Answers rewards users based on actions such as answering questions, voting for answers or having responses selected as best answers. However, it does not consider other factors such as level of details, readability and promptness of answers. The inclusion of such facets might allow for a more comprehensive evaluation of answerer reputation. Given that novices offer more readable answers than reputed answerers, and are more responsive to several types of questions, such a revised reputation point system might result in fairer CQA platforms for novices and experts alike.

Furthermore, even though most CQAs display the best answer to a given question more conspicuously and allow users to sort answers based on promptness, users seldom have the liberty to sort answers based on level of details or readability of answers. It is conceivable that different users might look for answers with varying levels of details or readability. Hence, the design of CQAs could be improved by allowing users the flexibility to sort answers based on such factors. This in turn might allow users make better use of CQAs to effectively meet their information needs.

This study is constrained by two key limitations that future research should address. First, the dataset is limited by the English language version of Yahoo! Answers. It is possible that English native speakers were more likely to become reputed compared with non-English native speakers (Ozmutlu *et al*., 2003). Future research could consider investigating the role of answerer reputation in Yahoo! Answers by including several other languages to allow for better triangulation. Second, the questions in the dataset were drawn from five categories that included entertainment, sports, computers, science and health. Future research could analyze the interplay of question types and answerer reputation in shaping answer effectiveness by drawing data from more categories to obtain greater insights.

How to cite this work

Chua, A.Y.K. & Banerjee, S. (2015). Measuring the Effectiveness of Answers in Yahoo! Answers. Online Information Review, 39(1) 104-118.

References

Adamic, L.A., Zhang, J., Bakshy, E. and Ackerman, M.S. (2008), “Knowledge sharing and Yahoo Answers: Everyone knows something”, *Proceedings of the International Conference on World Wide Web*, ACM, New York, pp. 665-674.

Agichtein, E., Castillo, C., Donato, D., Gionis, A. and Mishne, G. (2008), “Finding high-quality content in social media”, in *Proceedings of the International Conference on Web Search and Web Data Mining*, ACM, New York, pp. 183-194.

Agichtein, E., Liu, Y. and Bian, J. (2009), “Modeling information seeker satisfaction in community question answering”, *ACM Transactions on Knowledge Discovery from Data*, Vol. 3 No. 2, pp. 10:1-10:27.

Allen, B. (1998), “Designing information systems for user abilities and tasks: An experimental study”, *Online Information Review*, Vol. 22 No. 3, pp. 139-153.

Berg, B.L. and Lune, H. (2004), *Qualitative research methods for the social sciences*, Vol. 5, Pearson, Boston.

Blooma, M.J., Goh, D.H.L. and Chua, A.Y.K. (2012), “Predictors of high-quality answers”, *Online Information Review*, Vol. 36 No. 3, pp. 383-400.

Chaiken, S. (1980), “Heuristic versus systematic information processing and the use of source versus message cues in persuasion”, *Journal of Personality and Social Psychology*, Vol. 39 No. 5, pp. 752-766.

Chen, X. and Sin, S.C.J. (2013), “‘Misinformation? What of it?’Motivations and individual differences in misinformation sharing on social media”, in *Proceedings of the American Society for Information Science and Technology*, available at http://www.asis.org/asist2013/proceedings/submissions/posters/23poster.pdf (accessed 14 July 2014).

Chua, A.Y.K. and Banerjee, S. (2013), “English versus Chinese: A cross-lingual study of community question answering sites”, in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, IAENG, Hong Kong, pp. 368-373.

Cortese, J. and Lustria, M. L. A. (2012), “Can tailoring increase elaboration of health messages delivered via an adaptive educational site on adolescent sexual health and decision making?”, *Journal of the American Society for Information Science and Technology*, Vol. 63 No. 8, pp. 1567-1580.

Dang, H.T., Lin, J. and Kelly, D. (2007), “Overview of the TREC 2006 question answering track”, in *Proceedings of the Text REtrieval Conference*, pp. 99-116.

Darwin, C. (2008), “Continuous variables: Analysis of variance”, in Peat, J. K. and Barton, B. (Eds.), *Medical Statistics: A Guide to Data Analysis and Critical Appraisal*, Blackwell Publishing, Massachusetts, pp. 108-155.

Deng, S., Liu, Y. and Qi, Y. (2011), “An empirical study on determinants of web based question-answer services adoption”, *Online Information Review*, Vol. 35 No. 5, pp. 789-798.

Ghose, A. and Ipeirotis, P.G. (2011), “Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics”, *IEEE Transactions of Knowledge and Data Engineering*, Vol. 23 No. 10, pp. 1498-1512.

Gurrea, R., Orús, C., and Flavián, C. (2013), “The role of symbols signalling the product status on online users' information processing,” *Online Information Review*, Vol. 37 No. 1, pp. 8-27.

Harper, F.M., Raban, D., Rafaeli, S. and Konstan, J.A. (2008), “Predictors of answer quality in online Q&A sites”, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, New York, pp. 865-874.

Harper, F.M., Moy, D. and Konstan, J.A. (2009), “Facts or friends? Distinguishing informational and conversational questions in social Q&A sites”, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, New York, pp. 759-768.

Jeon, J., Croft, W.B. and Lee, J.H. (2005), “Finding similar questions in large question and answer archives”, in *Proceedings of the International Conference on Information and Knowledge Management*, ACM, New York, pp. 84-90.

Jeon, J., Croft, W.B., Lee, J.H. and Park, S. (2006), “A framework to predict the quality of answers with non-textual features”, in *Proceedings of the International SIGIR Conference on Research and Development in Information Retrieval*, ACM, New York, pp. 228-235.

Jin, X.L., Zhou, Z., Lee, M.K. and Cheung, C.M. (2013), "Why users keep answering questions in online question answering communities: A theoretical and empirical investigation", *International Journal of Information Management*, Vol. 33 No. 1, pp. 93-104.

Kelly, D., Wacholder, N., Rittman, R., Sun, Y., Kantor, P., Small, S. and Strzalkowski, T. (2007), "Using interview data to identify evaluation criteria for interactive, analytical question-answering systems", *Journal of the American Society for Information Science and Technology*, Vol. 58 No. 7, pp. 1032-1043.

Kim, S. and Oh, S. (2009), "Users' relevance criteria for evaluating answers in a social questions and answers site", *Journal of the American Society for Information Science and Technology*, Vol. 60 No. 4, pp. 716-727.

Kitzie, V. and Shah, C. (2011), "Faster, better, or both? Looking at both sides of online question answering coin", in *Proceedings of the American Society for Information Science and Technology*, available at https://asis.org/asist2011/posters/180_FINAL_SUBMISSION.pdf (accessed 15 July 2014).

Korfiatis, N., García-Bariocanal, E. and Sánchez-Alonso, S. (2012), "Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content", *Electronic Commerce Research and Applications*, Vol. 11 No. 3, pp. 205-217.

Krämer, N.C. and Winter, S. (2008), "Impression management 2.0: The relationship of self-esteem, extraversion, self-efficacy, and self-presentation within social networking sites", *Journal of Media Psychology: Theories, Methods, and Applications*, Vol. 20 No. 3, pp. 106-116.

Larkey, L.S. (1998), "Automatic essay grading using text categorization techniques", in *Proceedings of the International SIGIR Conference on Research and Development in Information Retrieval*, ACM, New York, pp. 90-95.

Lin, J. and Katz, B. (2006), "Building a reusable test collection for question answering", *Journal of the American Society for Information Science and Technology*, Vol. 57 No. 7, pp. 851-861.

Mamykina, L., Manoim, B., Mittal, M., Hripcsak, G. and Hartmann, B. (2011), "Design lessons from the fastest Q&A site in the west", in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, New York, pp. 2857-2866.

Mehdizadeh, S. (2010), "Self-presentation 2.0: Narcissism and self-esteem on Facebook", *Cyberpsychology, Behavior, and Social Networking*, Vol. 13 No. 4, pp. 357-364.

Nam, K.K., Ackerman, M.S. and Adamic, L.A. (2009), "Questions in, knowledge in? A study of Naver's question answering community", in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, New York, pp. 779-788.

Ozmutlu, S., Ozmutlu, H.C. and Spink, A. (2003), "Are people asking questions of general web search engines?", *Online Information Review*, Vol. 27 No. 6, pp. 396-406.

Palanisamy, R. (2014), "The impact of privacy concerns on trust, attitude and intention of using a search engine: An empirical analysis", *International Journal of Electronic Business*, Vol. 11 No. 3, pp. 274-296.

Qu, B., Cong, G., Li, C., Sun, A. and Chen, H. (2012), "An evaluation of classification models for question topic categorization", *Journal of the American Society for Information Science and Technology*, Vol. 63 No. 5, pp. 889-903.

Rezaei, A. and Zakariaie, M. (2011), "Exploring the impact of handcraft activities on the creativity of female students at the elementary schools", *International Education Studies*, Vol. 4 No. 1, pp.127-133.

Shachaf, P. and Rosenbaum, H. (2009), "Online social reference: A research agenda through a STIN framework", in *Proceedings of the iConference*, 2009, Chapel Hill, NC, available at http://www.ideals.illinois.edu/bitstream/handle/2142/15209/shachaf-rosenbaum_iconf091.pdf (accessed 14 July 2014).

Shah, C. (2011), "Effectiveness and user satisfaction in Yahoo! Answers", *First Monday*, Vol. 16 No. 2, available at http://firstmonday.org/ojs/index.php/fm/article/view/3092/2769 (accessed 7 July 2014).

Shah, C. Oh, J.S. and Oh, S. (2008), "Exploring characteristics and effects of user participation in online social Q&A sites", *First Monday*, Vol. 13 No. 9, available at http://firstmonday.org/ojs/index.php/fm/article/view/2182/2028 (accessed 7 July 2014).

Su, Q., Pavlov, D., Chow, J.H. and Baker, W.C. (2007), "Internet-scale collection of human-reviewed data," in *Proceedings of the International Conference on World Wide Web* , ACM, New York, pp. 231-240.

Surowiecki, J. (2004), *The wisdom of crowds*, Anchor Books, New York.

Suryanto, M.A., Lim, E.P., Sun, A. and Chiang, R.H. (2009), "Quality-aware collaborative question answering: Methods and evaluation", in *Proceedings of the International Conference on Web Search and Data Mining*, ACM, New York, pp. 142-151.

Toba, H., Ming, Z.Y., Adriani, M. and Chua, T.S. (2014), "Discovering high quality answers in community question answering archives using a hierarchy of classifiers", *Information Sciences*, Vol. 261 No. 10, pp. 101-115.

Voorhees, E.M. (2004), "Overview of the TREC 2003 question answering track", in *Proceedings of the Text REtrieval Conference*, pp. 54-68.

Voorhees, E.M. (2005), "Overview of the TREC 2004 question answering track", in *Proceedings of the Text REtrieval Conference*, pp. 52-62.

Wathen, C.N. and Burkell, J. (2002), "Believe it or not: Factors influencing credibility on the web", *Journal of the American Society for Information Science and Technology*, Vol. 53 No. 2, pp. 134-144.

Westbrook, L. (2014), "Intimate partner violence online: Expectations and agency in question and answer websites", *Journal of the Association for Information Science and Technology*, DOI: 10.1002/asi.23195.

Yang, L., Bao, S., Lin, Q., Wu, X., Han, D., Su, Z. and Yu, Y. (2011), "Analyzing and predicting not-answered questions in community-based question answering services", in *Association for the Advancement of Artificial Intelligence*, pp. 1273-1278.