Quality of Health Answers in Social Q&A

Sanghee Oh **Yong Jeong Yi** **Adam Worrall**
College of Communication and Information, Florida State University
PO Box 3062100, Tallahassee, FL 32306-2100
shoh@cci.fsu.edu, yjy4617@my.fsu.edu, apw06@my.fsu.edu

ABSTRACT
The purpose of the current study is to investigate perceptions regarding the quality of online health answers in social Q&A. The current study differs from previous studies by focusing on the topic of health, comparing the evaluations of users against experts. Three groups of participants -- librarians, nurses, and users of Yahoo! Answers -- were invited to assess the quality of health answers posted in Yahoo! Answers. Forty participants from each group reviewed a total of 400 health answers, rating them with a 5-points Likert scale according to 10 evaluation criteria: accuracy, *completeness, relevance, objectivity, source credibility, readability, politeness, confidence, knowledge*, and *efforts*. Findings indicated that there was no significant difference of the quality ratings between librarians and nurses. There was, however, significant difference between those two expert groups (librarians and nurses) and users. Librarians and nurses rated the quality of answers lower on most of the evaluation criteria than users. This research will help librarians and nurses better understand how laypeople, such as their patrons and patients, evaluate online health information in social contexts, leading to the offering of better health information services to these audiences.

Keywords
Social Q&A, online health information, online health answers, quality evaluation, evaluation criteria.

INTRODUCTION
Thanks to Web 2.0, people can interact with anonymous others who have different levels of expertise and experiences through various channels of social media, such as blogs, wikis, and social network services. They ask questions and answer one another, producing a large volume of health information. Fox and Jones (2009) reported 41% of patients have consulted ratings, reviews, or comments related to health issues given by others in online news groups, websites, or blogs. This percentage will increase as more people use mobile or wireless devices for interpersonal communication. The most critical concern with this trend is the quality of information shared in social contexts. Are online comments or answers accurate enough for people to consult with their health problems? How reliable are they? Despite the popularity of sharing health information in social contexts, little is known about the quality of health information in social contexts and its influence on people's health care decisions.

Therefore, the current study investigates the quality of health answers in social questioning and answering (social Q&A). Social Q&A is a Web-based service allowing people to ask and answer one another in many different topic areas, including health. It is free and easy to access and use. People can benefit from the varying levels of knowledge, expertise, and experiences of others. Social Q&A has grown with incredible speed over the past few years (Gazan, 2011; Hitwise, 2008; Shah, 2011); health is a popular topic. Health answers in social Q&A are user-generated information and content, dynamically created in response to requests by people who want information, advice, opinions, and experiences from others who are interested in or have health issues. To examine the quality of health answers in social Q&A, the current study proposes two research questions.

- How good is the quality of health answers in social Q&A?
- To what extent do librarians, nurses, and users differ in their assessments of the quality of health answers?

The purpose of the current study is twofold. First, we attempt to identify a set of criteria specific to assessing the quality of *online health answers* and test it with a set of health answers obtained from Yahoo! Answers. Second, we investigate different perceptions of the quality of health answers by three distinct groups: librarians, nurses, and users of social Q&A.

BACKGROUND
There have been many empirical studies on identifying and developing a set of criteria assessing the quality of health information on the Internet. Eysenbach, Powell, and Kuss (2002) selected 79 journal articles on the evaluation of health websites or webpages from a comprehensive list of academic databases and found accuracy, completeness, readability, design, and references were the most frequently used criteria in evaluations. Ambre et al. (1997) developed a set of evaluation criteria in areas of credibility, content, disclosure, links, design, interactivity, and caveats of health

ASIST 2012, October 28-31, 2012, Baltimore, MD, USA.

websites. The e-Health Ethics Initiative (Rippen & Risk, 2000) and Health on the Net Foundation (Health on the Net Foundation, 1997) also proposed a set of criteria—"e-Health Code of Ethics" and "HONcode"—for promoting the appropriate use and evaluation of online health information. Overall, these guidelines are practical and provide ways for people to assess health information available on health care sites or services on the Internet. The contexts of information they observe are limited to websites, webpages, or documents, and do not reflect health information exchanged in social contexts. On the other hand, Stvilia, Mon, and Yi (2009) used a set of criteria in assessing and comparing the quality of health information obtained from two different types of resources, health websites and email transcripts, with health-related inquiries asked by users at the Internet Public Library (IPL), developing a model for online consumer health information quality.

In social Q&A research, several studies have examined the quality evaluation of answers, but the method of answer evaluation has varied. Liu, Bian and Agichtein (2008) developed algorithms to predict user satisfaction as an indicator for evaluating the quality of answers, inviting paid raters in Amazon's Mechanical Turk (MTurk)[1] to test the effectiveness of the algorithms. Recently, Shah and Pomerantz (2010) adapted a set of criteria developed by Zhu et al. (2009). As with Liu et al. (2008), they recruited Amazon MTurk raters for the evaluation and analyzed various features from questions, answers, and users to evaluate the quality of answers. Harper, Raban, Rafaeli, and Konstan (2008) recruited undergraduates as raters and compared the quality of answers in social Q&A to digital reference services and expert Q&A. They found social Q&A provides a higher quality of answers than other services, due to answerers' concern as articulated in the answer. Kim and Oh (2009) also found questioners preferred emotional overtones in evaluating answers. Kim (2010) interviewed Yahoo! Answers users, asking about criteria for assessing the credibility of answers. She found users apply message-related criteria (e.g., accuracy, clarity, logic, grammar) with greater frequency than source-related criteria (e.g., references, expertise).

For the quality evaluation of health answers in the current study, a set of 10 criteria was selected from the literature. Table 1 shows the 10 criteria and associated statements given to participants when rating the quality of each.

[1] Amazon's Mechanical Turk (MTurk) (mturk.com) is an online tool, which connects researchers to raters who are willing to provide their input for a small compensation.

Criteria	Statements
Accuracy	The answer provides correct information.
Completeness	The answer includes everything. There is nothing to add.
Relevance	The answer is relevant to the question.
Objectivity	The answer provides objective information.
Readability	The answer is easily readable.
Source Credibility	The source of information is authoritative.
Politeness	The answerer is polite.
Confidence	The answerer is confident in the answer.
Empathy	The answerer expresses his or her empathy to the questioner.
Efforts	The answerer puts effort into providing this answer.

Table 1. Health answer evaluation criteria

METHOD

A total of 400 health-related questions and associated answers were selected at random from the 25 Health categories of Yahoo! Answers during April 2011. The topics cover all kinds of diseases and associated conditions, dental and optical care, alternative medicine, diet, and fitness. The same set of 400 questions and associated answers were reviewed by each of three groups of participants: librarians, nurses, and Yahoo! Answers users (hereafter users). Librarians were recruited online through several email lists, including lists run by the Medical Library Association and Ask-a-Librarian, and a public contact list of librarians in Florida and Georgia public or health science libraries. Nurses were recruited from several Advanced Nurse Practitioner Councils in Florida and graduate students from the Florida State University College of Nursing. Users who had posted health-related questions at least once during April 2011 were invited to participate in the study.

Recruiting continued until 40 participants were in each group. Each participant reviewed 10 questions and associated answers. Within the groups, the 400 questions and associated answers were assigned to each participant at random to control for the individual differences of the participants. A total of 400 health answers were evaluated by each group of participants, using the 10 criteria in Table 1, scoring on a scale ranging from 1 (lowest) to 5 (highest), with an option of Not Applicable. During pretesting, we found members of the two expert groups, librarians and nurses, took longer to review answers and were more thorough than users. Thus, librarians and nurses received $30 Amazon.com gift cards for spending 60-90 minutes on evaluation, while users received a $10 Amazon.com gift card for spending less than 30 minutes on average.

RESULTS

Quality Rating Distribution by Gender and Age

In total, 119 participants (40 librarians, 40 nurses, and 39 users[2]) rated the quality of 400 health answers in Yahoo! Answers. Among them, 98 (82.4%) were female and 21 (17.6%). Most of the participants were female; this may be influenced by librarians and nurses being female dominated professions, although females were also a majority of user participants (see Table 2). Among all participants, there were no statistical differences in rating criteria between males and females.

Gender	Librarians	Nurses	Users	Total
Female	34 (85.0%)	37 (90.2 %)	27 (69.2%)	98 (82.4%)
Male	6 (15.0%)	3 (7.3%)	12 (30.8%)	21 (17.6%)
Total	40 (100.0%)	40 (100.0%)	39 (100.0%)	119 (100.0%)

Table 2. Gender distribution of librarians, nurses, and users

The average age of the participants was 38.01 (SD = 14.38). The youngest participant was 18 years old (because only adults were included in the study); the oldest participant was 70 years old. The average age of users was younger than librarians and nurses (See Table 3).

Gender	Librarians	Nurses	Users
Youngest	24	23	18
Oldest	70	64	48
Average	46.45	41.72	25.54

Table 3. Age distribution of librarians, nurses, and users

A series of bivariate linear regression analyses were conducted to evaluate the differences in rating criteria by age. For librarians and nurses, there were no statistical differences in rating criteria by age. For users, there were statistical differences in *accuracy* (B = -.34, SE = .014, p = .020), *relevance* (B = -.37, SE = .011, p = .002), and *readability* (B = -.038, SE = .010, p = .001), but no other criteria. Younger users rated higher than older participants in *accuracy*, *relevance* and *readability*. Younger users could be more open and generous in evaluating answers because they are accustomed to using and obtaining information from social Q&A sites or other similar venues on the Internet.

Table 4 shows descriptive statistics of the quality ratings by all participants, sorted by the means, in order from highest to lowest. (We discuss differences between the groups of participants in the next section.) The grand mean of all 10 criteria was 3.16. The top four criteria of *confidence*, *politeness*, *readability*, and *relevance* were higher than the grand mean. *Confidence* and *politeness* are socio-emotional criteria attached to personal feelings or beliefs, while *readability* and *relevance* are content-related criteria. Findings indicate participants do not have tendencies to rate higher in either content-centered or socio-emotional criteria. However, the most-used content-centered criteria in previous studies, such as *accuracy*, *completeness*, *objectivity*, and *source credibility*, were ranked lower.

Rank	Quality Criteria	*N*	Mean	SD
1	Confidence	119	3.86	.72
2	Politeness	118	3.63	.69
3	Readability	119	3.63	.80
4	Relevance	119	3.58	.74
5	Empathy	118	3.15	.74
6	Objectivity	118	2.98	.89
7	Accuracy	119	2.97	.82
8	Efforts	118	2.92	.75
9	Source Credibility	119	2.46	1.02
10	Completeness	118	2.40	.82

Table 4. The overall ranking of quality criteria by mean ratings

Quality Rating Distribution across Groups: Librarians, Nurses and Users

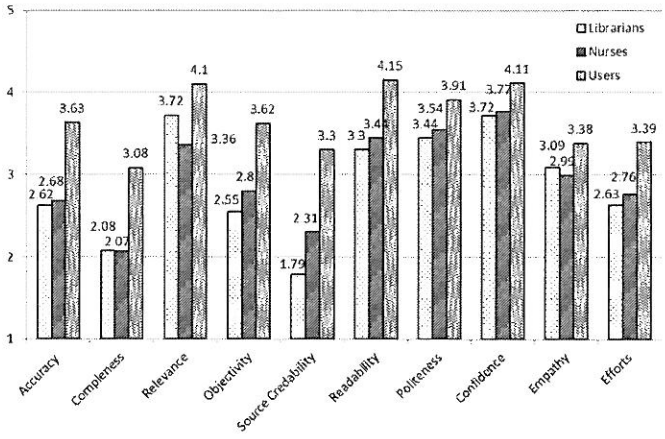

Figure 1. Comparison of the means of quality ratings for librarians, nurses, and users

Overall, the differences in ratings between users and the other two groups, librarians and nurses, were clear. Figure 1 shows the mean distribution of the criteria ratings across the groups. For all of the criteria, users rated the quality of

[2] We aimed 40 for users, but had to drop one because we weren't able to reach one during the period of data collection.

answers higher than librarians and nurses. A one-way ANOVA was performed to see the statistical differences in ratings across the three groups; Table 5 shows the results.

Criteria	F	df	p-value
Accuracy	27.10	118	.000*
Completeness	29.03	117	.000*
Relevance	18.42	118	.000*
Objectivity	20.28	117	.000*
Source Credibility	34.79	118	.000*
Readability	15.81	118	.000*
Politeness	2.90	117	.059
Confidence	3.54	118	.032*
Empathy	3.00	117	.054
Efforts	13.08	117	.000*

* ($p < .05$)

Table 5. Mean comparison of criteria ratings across librarians, nurses, and users

As seen in Table 4, there were significance differences for ratings of *accuracy*, *completeness*, *relevance*, *objectivity*, *source credibility*, *readability*, *confidence*, and *efforts* across the three groups. There were no statistical differences significant at the $\alpha = .05$ level for ratings of *politeness* and *empathy*. Participants across the three groups may share common understanding of how an answerer is polite or shows empathy in answers, but this may not be true because these two criteria would be statistically significant at the $\alpha = .10$ level. Further investigation of these two criteria is necessary in future studies.

Follow-up Tukey HSD post hoc tests were conducted to evaluate pairwise differences among the means. The tests confirmed users rated *accuracy*, *completeness*, *relevance*, *objectivity*, *source credibility*, *readability*, *confidence*, and *efforts* significantly higher than librarians and nurses ($p < .05$). There were no significant differences in most of the ratings between librarians and nurses, except *source credibility*; nurses rated this criterion higher than librarians. Librarians seem more thorough in evaluating the quality of sources used in providing health answers, because they have been trained to search for and select authoritative resources through their education and careers.

DISCUSSION

The nature of health answers differs from other kinds of health information available from health websites. Health answers are dynamically created by people who are responding to others with immediate needs for health information. Social Q&A sites are social places where people interact with one another by consulting others about their health problems, sharing factual information, personal experiences, advice, and emotional support. During the process of interacting, people may have their feelings or personal beliefs given to health answers. The current study proposed a set of 10 criteria, composed of content-centered criteria—*accuracy*, *completeness*, *objectivity*, *readability*, *relevance*, and *source credibility*—and socio-emotional criteria—*confidence*, *efforts*, *empathy*, and *politeness*.

Overall, *confidence* and *politeness* were the two highest rated criteria. Kim, Oh, and Oh (2007, 2009) and Kim and Oh (2009) identified socio-emotional criteria to be very important for and highly valued by users in evaluating answers they receive to questions they ask in social Q&A sites. Our study implies socio-emotional criteria may be of similarly high value for librarians, nurses, and other experts. Use of socio-emotional criteria in assessing the quality of health answers may be important to all because of common recognition of the value of the social nature of health answers. When comparing the criteria ratings among librarians, nurses, and users, there were statistical differences in ratings of all criteria, except two socio-emotional criteria, *politeness* and *empathy*. This may reflect participants across the groups having common interpretation and valuing of these two socio-emotional criteria, although their evaluation of them differed from the other two socio-emotional criteria, *confidence* and *efforts*. An in-depth study of the nature and use of socio-emotional criteria in social Q&A, and how they may be similar or different for different groups, would be needed to explore this further. Besides social Q&A, socio-emotional criteria could be applicable in evaluating the quality of information (e.g. postings, comments, or reviews) from other types of social media. For example, Gooden and Winefield (2007) identified social support as a major element of discussions in an online group for cancer survivors, implying socio-emotional criteria may be highly valued across social media settings.

Two main factors related to participants' demographics, age and gender, were tested to observe the statistical differences in rating health answers using the criteria. There was no statistical difference between male and female participants. By analyzing comments created by participants in online communities, Preece (1999) found differences in drafting comments by gender; female participants made more empathic comments, while comments by male participants were more factual. It seems gender difference is not reflected when evaluating the quality of information, but may be reflected when writing comments.

The methodological approach of the health answer evaluation in the current study is unique. Experts in search (librarians) and subject (nurses) were invited to evaluate the quality of answers, and their perceptions on the answers' quality were compared to perceptions of users of social Q&A. At the beginning of the study, we assumed there may be a gap between the two expert groups and the users in assessing the quality of health answers. The results confirmed the assumption; users rated answers higher than librarians and nurses in almost all criteria. Educational

efforts to reduce the gap between experts and users in evaluating health information, as previously encouraged by Bibel (2008), Gillaspy (2005), Orban (2005), Stvilia et al. (2009), and Thompson & Thompson (2007) should be made, helping users better select and use health answers when making their health decisions and improve their health literacy.

Another important finding from the study is that there were no statistical differences between librarians and nurses in almost all criteria, although there may be gaps in medical knowledge and expertise between the two groups. The current study found librarians and nurses have consensus when evaluating the quality of health answers. They may serve their users and patients in different ways, provide different services, and face different time and workload pressures (Gillaspy, 2005; Harris, Henwood, Marshall, & Burdett, 2010), but they can help their users and patients assess the quality of health information in the same manner.

The current paper and study have a few limitations. We reported the results of quality evaluation based on the two demographic characteristics of participants – sex and age – and compared them by the three groups. Other background information of participants, such as level of education, working experiences as nurses or librarians, could be important factors to analyze and need to be further studied. Also, Yahoo! Answers questions and associated answers were reviewed in the current study because it is the most popular and representative social Q&A sites. The quality of answers from other social Q&A services may be different from one another and would need to be evaluated considering the characteristics of each.

CONCLUSION

The use of social media for health information has increased with substantial speed over the past few years. As information professionals, we are concerned with guiding and helping users to obtain reliable health information. To serve users better, we must first understand the contexts of social media by exploring what kinds of information users want, discuss, and obtain from social media. For example, Answer Board Librarians[3] have been active in participating in social Q&A and providing answers for users. The current study took a further step from exploration to evaluation of the quality of health answers. The findings confirmed there is a gap between librarians and users in their perceptions of the quality of health answers. The next step of action should develop strategies and programs for educating users in using social media to find health information, to reduce the gap.

Nurses are also an important group of experts who provide health information to their patients. The current study found nurses' perceptions when assessing the quality of health information did not differ much from librarians' perceptions. Both groups of experts may be in the same stages of serving their users and patients in an era of social media. A possible collaboration between the two groups could establish synergy in user education for locating reliable health information.

The current study analyzed the quality evaluation distribution by gender and age and across the participant groups. An in-depth analysis across the participant groups will be followed by comparing the criteria ratings in accordance with 1) participants' background information, such as level of education or working experiences as a nurse or a librarian; and 2) the characteristics of the answer content, such as length of answers and citing of sources.

[3] http://answerboards.wetpaint.com

REFERENCES

Ambre, J., Guard, J. R., Perveiler, F. M., Renner, J., & Rippen, H. (1997). *Criteria for assessing the quality of health information on the Internet.* Retrieved from http://www.mlanet.org/tech_is/meb/criteria.pdf

Bibel, B. (2008). Public libraries and health literacy. In M. Kars, L. Baker, & F. Wilson (Eds.), *The Medical Library Association guide to health literacy* (pp.197-208). New York, NY: Neal-Schuman.

Eysenbach, G., Powell, J., Kuss, O., & Sa, E. (2002). Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. *Journal of the American Medical Association, 287*, 691-2700. doi:10.1001/jama.287.20.2691

Fox, S. & Jones, S. (2009). *The social life of health information* (Report). Washington, DC: Pew Internet and American Life Project. Retrieved from http://www.pewinternet.org/~/media/Files/Reports/2009/PIP_Health_2009.pdf

Gazan, R. (2011). Social Q&A. *Journal of the American Society for Information Science and Technology, 62*, 2301–2312. doi:10.1002/asi.21562

Gillaspy, M. L. (2005). Factors affecting the provision of consumer health information in public libraries: The last five years. *Library Trends, 53*, 480–495.

Harris, R., Henwood, F., Marshall, A., & Burdett, S. (2010). "I'm not sure if that's what their job is": Consumer health information and emerging "healthwork" roles in the public library. *Reference and User Services Quarterly, 49*, 239–252.

Harper, F. M., Raban, D., Rafaeli, S., & Konstan, J. A. (2008). Predictors of answer quality in online Q&A sites. *Proceedings of the 26th annual SIGCHI conference on human factors in computing systems (CHI '08, Florence, Italy)* (pp. 865–874). New York, NY: ACM. doi:10.1145/1357054.1357191

Health on the Net Foundation. (1997). *HON code on conduct (HONCode) for medical and health Websites* (version 1.6). Retrieved from http://www.hon.ch/HONcode/Conduct.html

Hitwise. (2008). *U.S. visits to question and answer websites increased 118 percent year-over-year* Retrieved from http://www.hitwise.com/news/us200803.html

Kim, S. (2010). Questioners' credibility judgments of answers in a social question and answer site. *Information Research, 15*(1). Retrieved from http://informationr.net/ir/15-2/paper432.html

Kim, S., Oh, J. S., & Oh, S. (2007). Best-answer selection criteria in a social Q&A site from the user oriented perspective. In A. Grove (Ed.), *Proceedings of the 70th Annual Meeting of the American Society for Information Science and Technology: Joining Research and Practice: Social Computing and Information Science.* Silver Spring, MD: ASIST. doi:10.1002/meet.1450440256

Kim, S., & Oh, S. (2009). Users' relevance criteria for evaluating answers in a social Q&A site. *Journal of the American Society for Information Science and Technology, 60*, 716-727. doi:10.1002/asi.21026

Kim, S., Oh, S., & Oh, J. S. (2009). Evaluating health answers in a social Q&A site. In A. Grove (Ed.), *Proceedings of the American Society for Information Science and Technology* (Vol. 45). Silver Spring, MD: ASIST. doi:10.1002/meet.2008.14504503134

Liu, Y., Bian, J., & Agichtein, E. (2008). Predicting information seeker satisfaction in community question answering. *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '08)* (pp. 483–490). New York, NY: ACM. doi:10.1145/1390334.1390417

Orban, E. (2005). Health literacy: Challenges and implications for consumer health librarians. *Indiana Libraries, 24*(3), 2-6.

Preece, J. (1999). Empathic communities: Balancing emotional and factual communication. *Interacting with Computers, 12*(1), 63-77. doi:10.1016/S0953-5438(98)00056-3

Rippen, H., & Risk, A. (2000). e-Health Code of Ethics. *Journal of Medical Internet Research, 2*(2), 1-5. doi:10.2196/jmir.2.2.e9

Shah, C. (2011). Measuring effectiveness and user satis satisfaction in Yahoo! Answers. *First Monday, 16*(2). Retrieved from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index. php/fm/article/view/3092/2769

Shah, C., & Pomerantz, J. (2010). Evaluating and predicting answer quality in community QA. *Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval (SIGIR '10)* (pp. 411–418). New York, NY: ACM. doi:10.1145/1835449.1835518

Stvilia, B., Mon, L., & Yi, Y. J. (2009). A model for online consumer health information quality. *Journal of the American Society for Information Science and Technology, 60*, 1781-1791. doi:10.1002/asi.21115

Thompson, S., & Thompson, A. (2007). Skeptical medical reference: Helping patrons find critical resources for consumer health issues. *Library Philosophy and Practice.* Retrieved from http://www.webpages.uidaho.edu/~mbolin/thompson2.htm

Zhu, Z., Bernhard, D., & Gurevych, I. (2009). *A multi-dimensional model for assessing the quality of answers in social Q&A sites* (Technical Report No. TUD-CS-2009-0158). Darmstadt, Germany: Technische Universitat Darmstadt. Retrieved from http://tuprints.ulb.tu-darmstadt.de/1940/1/TR_dimension_model.pdf