Utilizing Content Moderators to Investigate Critical Factors for Assessing the Quality of Answers on Brainly, Social Learning Q&A Platform for Students: A Pilot Study

Erik Choi, Michal Borkowski, Julien Zakoian, Katie Sagan, Kent Scholla, Crystal Ponti, Michal Labedz, Maciek Bielski
Brainly
79 Madison Ave, New York, NY 10016
erik.choi@brainly.com, michal.borkowski@brainly.com, julien.zakoian@brainly.com,
katie.sagan@brainly.com, kent.scholla@brainly.com, crystal.ponti@brainly.com,
michal.labedz@brainly.com, maciek.bielski@brainly.com

ABSTRACT
In this paper, we present data findings from the pilot study focusing on utilizing content moderators from Brainly, a social learning Q&A platform, to assess the quality of answers. Because it can be argued that Brainly users who actively moderate contents may have better contextual understandings of how users interact with each other through question-answering activities, and which answers are more likely relevant and appropriate to a question in a context of Brainly. The findings indicate that helpfulness, informativeness, and relevance are the most critical factors that have impacts on the quality of answers. Further content analysis also identified two new criteria: 1) descriptiveness - evaluating how well answers provide descriptive summaries through detailed and additional information, and 2) explicitness - clearly constructing answers to reduce vagueness of what information answerers intend to provide to satisfy an asker’s need.

Keywords
Information Seeking Behavior, Social Q&A, Information Quality, Human Assessments, Educational Information.

INTRODUCTION
Social question-answering (SQA) services are used to facilitate social information seeking and sharing behaviors in online communities where users are able to ask a question, while others voluntarily answer in order to satisfy an asker’s information need. SQA is designed for users to “ask a question in natural language, as opposed to relying on keywords when using a search engine, which is then read and responded to by other people, who deliver personalized answers tailored to the asker’s information need rather than a summary list of documents displayed on a search engine results page (SERP)” (Shah et al., 2014, p.669). In other words, SQA is “exemplifying the Web 2.0 model of user-generated and user-rated content” (Gazan, 2011, p.2302).

Given the fast growth and popular uses of SQA services in recent years, various types of SQA services have been emerged, ranging from general topic-based Q&A services (e.g., Yahoo! Answers, etc.) to specific subject-based Q&A services (e.g., Stack Overflow, etc.). A rich body of research has also been conducted to understand various aspects of SQA. According to Shah, Oh, and Oh (2009), previous research on SQA mainly focused on either user- and/or content-based studies; the former is more likely to understand motivations to ask and answer a question as well as user satisfaction, and the latter is to investigate user-generated contents (e.g., question type, question formulation, answer quality, etc.).

Since many SQA services are based on the wisdom of crowds (Surowiecki, 2004) that facilitates users’ participations to seek and share appropriate answers to a question, previous research studies have focused on understanding what constitutes a high quality of answer. To do that, previous research utilized human assessments such as Amazon Mechanical Turkers (Shah and Pomerantz, 2010) and a panel of blinded judges (Harper et al., 2008) or content analysis (e.g., Fichman, 2011; Kim & Oh, 2009), in order to investigate a variety of textual- and non-textual features of answers for gaining a better understanding of the quality of answers.

However, previous studies focusing on the answer quality may be limited to understanding how SQA users actually
perceive the quality of answers in such context where they ask and answer a question because previous studies mainly used outside members of SQA (e.g., Amazon Mechanical Turkers, etc.) who might have a lack of contextual understanding of SQA when measuring the answer quality. In this sense, it can be argued that SQA users who actively ask and answer a question may share similar discourses and interests, and have better contextual understandings of: 1) how users interact with each other through question-answering activities, and 2) which answers are more likely relevant and appropriate to a question. Therefore, the current study is focused on investigating how active SQA users who moderate contents make evaluative judgments to assess the quality of answers in a context of SQA.

BACKGROUND
Social Question-Answering (SQA)
Social question-answering (SQA) services are “online information sources where people identify their information need, formulate the need in natural language, and interact with one another to receive answers that satisfy their information need in virtual environments” (Choi & Shah, in press). Shah, Oh, and Oh (2009) identified three primary components of SQA services: a way to ask a question, a way to answer a question, and the participation around such question-answering activities. As they, and others, recognized – it is the participation around Q&A activities that is the most defining characteristic of such SQA services. SQA services represent a Web 2.0 application that capitalizes on collaboration between communities of users for information exchange (Shachaf & Rosenbaum, 2009).

Brainly: Social Learning Network for Students
The current study used Brainly to recruit content moderators to assess the quality of answers. Brainly is a community-based social learning Q&A platform in which students (i.e., middle school, high school) and educators are able to ask and answer a question to solve homework problems in 16 subjects (e.g., mathematics, chemistry, biology, physics, etc.). Brainly generates more than 40 million unique visitors per month in over 35 countries as of May 2015. Similar to other SQA services, registered users are allowed to ask a question to specify their homework problems and their information needs, while other users voluntarily answer a question - providing their help to solve an asker’s homework problems.

Assessments of Answer Quality
As mentioned previously, evaluating the quality of answers and identifying critical factors in measuring the quality has been researched in previous studies related to SQA services. For instance, Shah and Pomerantz (2010) used Amazon Mechanical Turkers to rate the quality of each answer for a given question and developed a prediction model in order to assess the quality of answers on SQA services, and the recent study by Liu et al. (2015) also proposed a new training approach to investigate the relationships of non-textual features and unlabeled data to predict the best quality of answers. However, few studies have investigated an approach of utilizing active SQA users’ (especially, ones who are assigned to moderate content) perceived judgments of assessing answers in order to examine the quality of answers in the SQA research literature.

METHOD
Framework of Assessing Answer Quality
In order to investigate how SQA users assess the quality of answers, the current pilot study adapted the expert dimensions and metrics for the quality of answer identified in the previous study by Zhu et al. (2009) - informativeness, politeness, readability, relevance, truthfulness, originality, objectivity, helpfulness, and expertise. The study used the guidelines from different social Q&A services such as Answerbag, WikiAnswers, and Yahoo! Answers, to identify the nine criteria to be considered as exerts’ advice and suggestion.

Data Collection
A survey was developed to measure the nine criteria of assessing the quality of answers; all items were taken on a 5-point Likert scale (1-Ver important factor for the quality of answer to 5-Very important factor). Additionally, in order to identify a potential emergence of any dimensions used as a subjective guidance to assess the quality of answers, the survey also included open-discussion sections for six sets of question-answer where the survey participants were required to provide their perceived judgments on each answer to the given question.

Brainly moderators in a US market were selected in this pilot study. They are a group of active users who are chosen based on having a comprehensive answering history and demonstrate good communication skills with other Brainly users. In terms of assuring the answer quality, they are able to delete wrong, poor, and spam answers with additional explanations to answerers, approve appropriate answers, warn and ban users who behave against policies, as well as participate in the forum to share their experiences on content moderations.

A survey was administered electronically, and the survey link was distributed through the internal newsletter email and forum. The initial distribution ran from May 22nd to 29th, 2015, and the follow-up reminder was sent out in the week after the first recruitment message.

RESULTS
A total of 25 Brainly moderators participated in this pilot study, and the sample population was fairly equal between male students (n=11, 44%) and female students (n=14, 56%). In terms of the level of current education level, most
Brainly moderators are between 7th and 12th grades. Of the participants, 17 participants (68%) reported they have been a Brainly moderator for between 1 and 6 months, and the majority of them (n=19, 76%) indicated they have moderated more than 300 answers since they became a Brainly moderator. In addition, it was found that mathematics is the core subject the survey participants most frequently moderate contents on Brainly (n=17, 68%).

In regard to the criteria that participants apply to assess the quality of answers on Brainly, it was found that helpfulness, which is about sharing useful and helpful answers to a question, is the most critical factor (Mean=4.92, S.D.=0.28), followed by relevance focusing on how answers are connected and appropriate to a question (Mean=4.83, S.D.=0.41). Informativeness focusing on providing correct information related to the question, and being informative and accurate (Mean=4.82, S.D.=0.37) is the second critical factor. Politeness focusing on being free of personal information, no personal attacks or mean comments, and addressing the asker in a respectful manner (Mean=4.71, S.D.=0.48) is the third critical factor. Originality focusing on how answers are original and not plagiarized from other sources (Mean=4.67, S.D.=0.75) is the fourth critical factor. Readability focusing on using correct spelling and grammar, and having limited uses of slang, abbreviations, or instant messaging type style (Mean=4.23, S.D.=0.67) is the fifth critical factor. Truthfulness getting answers from trusted members (Mean=4.13, S.D.=0.81) is the sixth critical factor. Objectivity focusing on a neutral point of view (Mean=3.88, S.D.=0.85) is the seventh critical factor. Expertise showing a high level of professional or subject-level knowledge (Mean=3.17, S.D.=0.90) is the eighth critical factor.

Table 1 describes the details of the criteria applied to assess the quality of answers among the survey participants.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Measurement</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helpfulness</td>
<td>Sharing useful and helpful answers to a question</td>
<td>4.92</td>
<td>0.28</td>
</tr>
<tr>
<td>Relevance</td>
<td>Focusing on how closely answers are connected and appropriate to a question</td>
<td>4.83</td>
<td>0.41</td>
</tr>
<tr>
<td>Informativeness</td>
<td>Providing correct information related to the question, and being informative and accurate</td>
<td>4.82</td>
<td>0.37</td>
</tr>
<tr>
<td>Politeness</td>
<td>Being free of personal information, no personal attacks or mean comments, and addressing the asker in a respectful manner</td>
<td>4.71</td>
<td>0.48</td>
</tr>
<tr>
<td>Originality</td>
<td>Focusing on how answers are original and not plagiarized from other sources</td>
<td>4.67</td>
<td>0.75</td>
</tr>
<tr>
<td>Readability</td>
<td>Using correct spelling and grammar, and having limited uses of slang, abbreviations, or instant messaging type style</td>
<td>4.23</td>
<td>0.67</td>
</tr>
<tr>
<td>Truthfulness</td>
<td>Getting answers from trusted members</td>
<td>4.13</td>
<td>0.81</td>
</tr>
<tr>
<td>Objectivity</td>
<td>Focusing on a neutral point of view</td>
<td>3.88</td>
<td>0.85</td>
</tr>
<tr>
<td>Expertise</td>
<td>Showing a high level of professional or subject-level knowledge</td>
<td>3.17</td>
<td>0.90</td>
</tr>
</tbody>
</table>

Table 1. Expert dimensions and metrics used for assessing the quality of answers in the current pilot study.

Brainly’s answering guidelines are provided to help answerers improve their answers by ensuring they are helpful, informative, relevant, and accurate. In addition, Brainly’s answering guidelines are provided to help answerers improve their answers by ensuring they are helpful, informative, relevant, and accurate.

While informativeness is to measure how correct information is conveyed in an answer, descriptiveness may evaluate how well answers provide descriptive summaries or analyses through additional information to help askers solve their homework problems.

In addition, it was found that explicitness was another criteria emerged from the survey participants’ comments when they assessed the quality of answers in the study (n=20, 13.33%). This finding indicates that answers should be clearly constructed to reduce vagueness of what information answerers intend to provide to satisfy an asker’s need.

**DISCUSSION**

Findings from this pilot study identified that helpfulness, informativeness, and relevance are the most important factors that have critical impacts on improving the quality of answers in a context of Brainly. This indicates that since Brainly is designed to seek and share educational information that helps users solve their homework problems, information is mainly to be shared in order to satisfy an asker’s cognitive needs. As the previous study (Choi & Shah, in press) indicates that learning through self-education is one of the significant motivations to use SQA that constitutes a social practice of learning-teaching formations, the findings also highlighted that sharing helpful and informative answers that are relevant to an
asker’s question would be a critical factor for helping conceptualize and improve an asker’s process of self-education on Brainly. Future studies, to gain a better understanding of what constitutes a high level of informativeness, relevance, and helpfulness in a context of educational information, may focus on conducting content analyses that investigate textual- and non-textual features of questions and answers, as well as in-depth interviews to investigate how users perceive information acquired through question-answering interactions for their learning processes. The findings also yield practical implications to improve SQA services for sharing the high quality of answers. For instance, SQA may focus on developing new features (e.g., FAQ page, guideline, etc.) that effectively inform users on how to write high quality answers.

Another interesting finding identified from a content analysis in this study is that answers should be more descriptive to share additional information for better learning processes among askers on Brainly. One survey participant provided his insights into sharing additional information for the high quality of answers when he assessed one of the mathematics answers in the study:

“Multiple choice questions are easy to just give the answer, but the answerer also provided a reason why that answer was correct and not the others.”

This may signify how the quality of answers is comprised based on: 1) question formats (e.g., open-ended vs. multiple-choice question), and 2) subject categories (e.g., mathematics vs. chemistry). Therefore, future studies may include those non-textual features into other features to develop a holistic and refined framework for quality evaluation in a context of SQA (Gazan, 2011).

It was found that expertise is the least important factor for assessing the quality of answers among the survey participants. This result may be prevalent since users already distinguish a context of Brainly in which peer-to-peer interactions most likely occurred to seek and share information, not interactions with teachers or professionals. However, as Gazan (2011) argues SQA services “can be considered worthy complements to professional resources in terms of answer quality” (p.2307), Brainly users may utilize information received to their question as additional information sources from their peers, which may be a timely complement to other information and knowledge from professional sources (e.g., teachers), for their learning processes.

CONCLUSION

The goal of the current pilot study is to investigate a way of utilizing content moderators to assess the quality of answers on SQA. Brainly, a social learning Q&A for students was used to recruit a total 25 content moderators in a US market who actively moderate answers to improve question-answering interactions to create a knowledge base for users’ homework problems. The current pilot study was carried out as part of a research project that will attempt to scale up the study to a larger sample of Brainly content moderators from different markets (e.g., Russia, France, Indonesia, etc.), which has the potential for exploring different research opportunities such as validating and generalizing the findings from the pilot study, investigating any differences in applying criteria to measure the quality of answers in different cultures and markets, as well as gaining more insights into any new criteria that content moderators may apply when they assess the quality of answers on Brainly.

REFERENCES


