Online Q&A in STEM Education: Curating the Wisdom of the Crowd

Chirag Shah and Rich Gazan

The proposed research project will integrate crowdsourced information and the wisdom of librarians and other information professionals to enhance the experience of STEM (Science, Technology, Engineering, and Math) learners, thus addressing IMLS's strategic goal of providing engaging and empowering learning experiences. The outcomes will provide mechanisms through which current informational services (e.g. virtual referencing) could obtain broader applicability and sustainability with new applications in education contexts. Through a series of studies involving STEM students and educators (teachers, librarians), this research will: (1) address the gap in our knowledge about how online Q&A fits the larger context of information-seeking behaviors, specifically for STEM students; (2) construct methods and metrics for assessing and using user-generated content; and (3) develop new approaches and tools for incorporating subject and search expertise to help students and educators in determining reliability of online sources. This work will require a highly skilled and diverse team of researchers and practitioners. Thus, the project will include PI Shah (information-seeking behavior, online Q&A, statistical analysis, system building); PI Gazan (social Q&A, STEM education, library services); a distinguished set of scholars serving as advisors; consultants and assessors for evaluating educational outcomes; and the largest homework Q&A service, Brainly, as an industry partner.

1 Statement of Need

The constantly changing landscape of online information, especially the flood of questionable information generated by non-experts, creates a dire need for much better tools than are currently available to guide and instruct students about evaluating and using online information. STEM education especially depends on students having not only access to trustworthy sources of hard facts but also the skills and tools to evaluate the reliability of information as they study the rigorous, evidenced-based fields most critical to addressing the nation's practical problems—from rebuilding infrastructure to tackling climate change. The proposed project will integrate crowdsourced information (user-generated in online environments) and the wisdom of librarians and other information professionals to give STEM learners the systems, skills, and tools they need to evaluate online Q&A information before incorporating these sources into their coursework. This section summarizes the need for the proposed research and development work, supported by a brief overview of relevant literature, and gives an outline of the specific work to be done following a set of research questions, as well as possible outcomes for theory and practice.

1.1 Motivation and background

The Association of College and Research Libraries (ACRL) *Information Literacy Standards* Task Force (ACRL, 2012) recommended that the next iteration of the standards must acknowledge the role of students as content creators and curators, and acknowledge the value of information in diverse formats. Online Q&A services allow people to post a question and receive responses from multiple participants in an online Q&A community. Online Q&A has become an increasingly popular method for students to seek, create, and curate information. Libraries have attempted to adapt by moving their reference services online, often referred to as virtual reference (VR). However, studies show that regardless of reliability, VR services have not been as successful or sustainable as Web-based Q&A services (Goldberg, 2010; Henderson & Bosch, 2010), and it is instructive to look beyond traditional LIS literature to understand underlying trends.

Turnitin.com is primarily a plagiarism detection service that analyzes student papers uploaded by subscribing institutions, but their analysis also includes identifying the sources students cite in their papers. A 2011-2012 analysis of 28 million course papers submitted to Turnitin (2013) revealed that social networking and other user-generated content sites were cited in 23% of the papers written by students in higher education institutions (see Figure 1). The same paper also lists online Q&A sites such as Yahoo! Answers and WikiAnswers as second only to Wikipedia among sources used by students (p.6).

Higher Ed Sources	%
Homework & Academic	33%
Social Networking & Content Sharing	23%
Paper Mills & Cheat Sites	19%
Encyclopedias	14%
News & Portals	10%
Shopping	1%
Total	100%

Figure 1: Results table from Turnitin.com white paper: The Sources in Student Writing (Turnitin, 2013, p.9).

This trend is especially concerning for STEM learners, where the need to scaffold understanding with factual, trustworthy

information is paramount. **What separates STEM education from other disciplines** is the need to blend various scientific data and methods, while evaluating the source and applicability of each, to address real-world problems. Rigorous STEM training focuses on structured inquiry-based and problem-oriented learning, using factual information and induction/deduction techniques to arrive at proofs or conclusions, as well as producing verifiable or testable hypotheses and findings. While almost any student could benefit from having appropriate support with online content assessment, the specific characteristics of STEM education, including the need to apply systematic and analytical thinking to problem solving, present a unique set of challenges and opportunities (Fairweather, 2009).

While STEM learners clearly need help assessing the quality of their online information sources, these students may also be discovering valuable content at online Q&A sites appropriate to their learning objectives, content not acknowledged or accepted by instructors or VR librarians due to its nontraditional nature. Meaningful STEM education is not simply discovering and applying facts: it requires understanding the process of inquiry and the conversations surrounding those facts, which is the essence of online Q&A (Gazan, 2011). The ACRL Information Literacy Standards for Science and Engineering/Technology (ACRL, 2015) require that students incorporate and comparatively assess a broad range of formal and informal information resources. STEM competencies fundamentally incorporate information literacy and can be integrated into undergraduate education via library instruction (Scaramozzino, 2010).

The sheer number of students' online Q&A citations makes both quality assessment of Q&A content and a more nuanced understanding of the role of Q&A content in STEM education worth investigating. At the same time, incorporating crowdsourced information into education must be done with caution. Library practitioners and scholars are in a unique position: both are accustomed to evaluating socially constructed information, but also understand how specific communities define quality information in diverse ways. The purpose of this project, similarly, sits at a unique intersection of LIS, education, and social sciences. This research investigates the social phenomenon of people constructing and seeking information through online Q&A; uses theories and expertise from LIS to evaluate and curate this information; and applies the resulting knowledge to STEM education by augmenting students' learning experiences with curated wisdom of the crowd. Finally, the project brings all of these theoretical and practical outcomes back to the libraries to help transform their informational, especially VR, services in light of students' increasing use of crowdsourced information.

Rather than designing incremental improvements to VR services, the PIs propose to leverage and investigate students' demonstrated use of social Q&A (SQA) and community-based Q&A (CQA) services in learning environments, with the goal of identifying areas of potential crossover that would result in better library and educational services. Specifically, the proposed project will: (1) investigate the use of online Q&A services and develop a theoretical framework that contextualizes them within the broader domain of information behavior; (2) develop methods for evaluating the relevance, quality, and impact of content created and shared in Q&A environments; and (3) test the outcomes of the first two stages in STEM education environments via a formative design-evaluation cycle.

Application Scenario: Imagine a high school student, Sarah, who has a question "Given the reaction Fe2O3 + 3 CO ---> 2 Fe + 3 CO2, what is the molar ratio for CO to CO2?" If she searches about this on Google, she gets thousands of results that match any of the keywords in her question, which will be almost impractical to go through. It is possible that some of these results are coming from a site like Yahoo! Answers (YA) where someone has already asked such a question. When Sarah visits such a link, she encounters many answers to the question that vary greatly in size, quality, and relevance. Current systems such as YA rely on the "Wisdom of the Crowd" to address quality and authenticity issues, but as shown by many (e.g. Kim & Oh, 2008; Shah & Kitzie, 2012), such reliance may be misplaced. Sarah will have to depend on her own judgment to determine which answer to use, if any, so she still does not have a reliable answer to her question. A trained professional such as a teacher or a librarian could help, but they may not be immediately available – physically or virtually. But what if the expertise of trained professionals is captured in some way and available to Sarah when she needs and even at times when she does not recognize such need? An automatic assessment tool, one of the outcomes of this project, integrated in Sarah's browser could do this and assist her in both evaluating and using such crowd-generated information (see Appendix C for a mockup of this tool). Thus, the proposed project's central tenet is that we cannot stop the students from searching information online, but we can help them assess its quality and credibility, encouraging them to extend that experience by consulting a teacher or a VR librarian.

1.2 Project goals

This research project is aimed at investigating people's online Q&A behavior and the quality assessment method applied by experts to augment learning experiences of students, specifically in STEM fields. The project will address the following research questions (RQs).

RQ1: How can online Q&A activities be explained and investigated as information-seeking behavior? We know enough from the literature (covered later) about Q&A activities as well as information seeking, but lack knowledge about explaining Q&A as a part of a larger information-seeking behavior.

RQ2: What are appropriate quality metrics for online Q&A content, specifically related to STEM fields?

RQ3: How would the addition of online Q&A content and quality assessment help meet the rigorous requirements of STEM education? Specifically, what effects will an automatic assessment helper for Q&A content have on STEM students' information gathering, evaluation, and synthesis activities?

The proposed work to address these RQs will produce the following outcomes:

1. A new theoretical framework to understand and explain online Q&A behavior;
2. An extension of this framework to existing models of information seeking and behavior;
3. A methodology to assess content quality in online environments;
4. A method and a tool – a Q&A content assessment system – to incorporate curated Q&A content into STEM education; and
5. Empirical data about the effects of incorporating Q&A content into STEM education.

1.3 Q&A services and STEM education: What we know, what we have, and what is missing

Online Q&A services provide a way to ask a question in natural language, as opposed to relying on keywords when using a search engine. The question 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 (Shah et al., 2009). Online Q&A services are comprised of websites that vary by the genre of content exchanged, the types of people who respond to the content (i.e. other users, subject experts, reference librarians), and the way in which content is disseminated (i.e. one-to-one, one-to-many, many-to-many). Examples of Q&A sites/services include Yahoo! Answers, Answerbag, WikiAnswers, Quora, StackOverflow, and Internet Public Library (IPL).

Online Q&A provides a unique context for information exchange. These exchanges occur in cyberspace, which facilitates new approaches to information interaction and has changed both our perceptions and expectations of how information exchange should occur (Marchionini, 2010). At the same time, the exchange itself remains between people, albeit mediated by systems. This exchange provides opportunities to interact with others, while at the same time offering challenges in how information is expressed, understood, and used, both within and outside of the online Q&A site in which it was originally disseminated. Each community tends to have different content, practices, and standards for what constitutes an appropriate answer, so it is useful to distinguish Q&A services at several levels.

Q&A services could be broadly classified into online and face-to-face, with traditional reference service in libraries being an example of the latter. Within online Q&A services, two types are prominent – vertical and horizontal. Vertical online Q&A services are focused on a specific topic. Examples of vertical Q&A, also referred to as online forum, include StackOverflow (http://stackoverflow.com/) for programmers and PRIUSchat (http://priuschat.com/) for Toyota Prius owners. Horizontal Q&A services typically cover a broad range of topics instead of being organized around just one topic. Examples of horizontal Q&A services include Yahoo! Answers (http://answers.yahoo.com/) and WikiAnswers. (http://wiki.answers.com/). Community-based, collaborative, and social Q&A are defined as peer-based services, distinct from expert-based Q&A (Choi & Shah, In Press).

Virtual Reference (VR) services offered by IPL and many libraries today are examples of expert-based Q&A. While VR services typically offer better quality than peer-based Q&A services (Shah & Kitzie, 2012), their model for service provision could be unsustainable (Hersberger, 2010, p. 12). This circumstance has led to several private, expert-based Q&A services, such as Google Answers, to shut down (Shah et al., 2008), and publicly funded VR services to struggle (Henderson & Bosch, 2010). Peer-based services, on the other hand, provide a user-driven, sustainable model of operation, albeit often compromising quality and authenticity (Oh et al., 2009; Shah, 2011).

The relative pros and cons of expert and peer-based Q&A services open up new opportunities to create hybrid solutions. The keys to creating such hybrid solutions for STEM education are to: (1) gain knowledge about assessing user-generated content in online Q&A; and (2) use that knowledge to augment/support STEM students' online information seeking. The current project addresses #1 with the help of experts (librarians) and #2 by developing a new tool, a Q&A content assessment system, for testing and deployment in libraries and STEM classrooms.

Several studies in the literature have tried to assess content relevance and quality in online Q&A. Panovich et al. (2012), for instance, performed a study of seeking and sharing information within social networking sites (e.g. Facebook) and pointed out that stronger ties tend to provide stronger contributions to social networking users' overall knowledge, meaning that close network friends tend to provide better quality of information than weaker ties do. Both Kim and Oh (2009) and Shah and Kitzie (2012) performed studies of online Q&A users' conceptions of relevance criteria, finding that while identified criteria overlap with criteria found in traditional relevance literature, the role of socio-emotional/affective criteria achieves a degree of primacy not expressed in other contexts. Similarly, the findings from a recent study by Worrall and Oh (2013) indicate that social and emotional support comprises a critical factor in assessing the quality of information among Yahoo! Answers users seeking health-related information.

Far outnumbering user-based studies of relevance, systems-based studies of online Q&A services investigate how textual and non-textual features of content exchanged contribute to askers' perceptions of relevance, quality, and satisfaction. These studies employ both quantitative and qualitative approaches to: examine text retrieval-based algorithms to extract answer features (Shah & Pomerantz, 2010; Arai & Handayani, 2013; Cai & Chakravarthy, 2013); predict asker satisfaction using proxies, such as Mechanical Turk (https://www.mturk.com/) (Liu et al., 2011); identify and filter deceptive answers promoting products or services by employing a user preference graph (Li et al., 2013); compare content quality, scope of corpora, response time, and usability of interfaces (Chua & Banerjee, 2013a); analyze answer quality and speed among different question types (Chua & Banerjee, 2013b); and employ qualitative approaches such as content analysis to generate best answer selection criteria (Kim & Oh, 2009). Qualitative approaches have also been applied to evaluate asker satisfaction with VR services along efficiency and effectiveness measures, such as being provided with the correct answer (Pomerantz et al., 2006; Shachaf & Horowitz, 2008).

While the proposed project will stand on the foundation of these works and their findings, it is important to note that the type of content exchanged on online Q&A sites is often dominated by advice and opinion-oriented questions as opposed to fact-based ones (Kim et al., 2008), and user-identified motivations within SQA are often found to be 'being driven by altruism' (Nam et al., 2009; Oh, 2012). As noted before, **STEM-related content is likely to be factual, empirically driven, and verifiable**. Therefore, while online Q&A, crowdsourced information, and social media have received a lot of attention lately, when it comes to quality assessment for the content in these areas, STEM education presents a unique set of challenges yet to be addressed. The current project will focus on STEM and provide generalizable outcomes for other disciplines.

1.4 Proposed approach

This project will address the research questions listed in Section 1.2 with three phased studies, each addressing a specific RQ. A summary is given in Table 1, with more details in Section 3.

Table 1: Overview of how the objectives of theory and practice will be met through research studies.

	Theory	*Practice*
Study-1 for RQ1 (Year-1)	Understanding of online Q&A behavior with a possibility of extending existing information seeking/behavior models	Preliminary data for supporting prototype development of a Q&A augmentation system
Study-2 for RQ2 (Year-2)	Construction of a new framework and a methodology for assessing content quality in community-generated content	Derivation of useful features contributing to content quality for automatic evaluation
Study-3 for RQ3 (Year-3)	Incorporation of theories from LIS and learning sciences to create a new framework that extends students' learning experiences using Q&A support in learning environments	Development, distribution and field-testing of new tools and systems that use Q&A augmentation for STEM education

2 Impact

2.1 Research impact

This research and development project will add substantially to our understanding of information-seeking practices, by people generally and by STEM students in particular, in online environments through asking and answering questions. This new knowledge will be grounded by empirical data about assessing content quality from diverse sources and embedded in an important information-seeking domain: STEM education. Knowledge gained through this project could help us extend and/or change existing models of human information behavior (e.g., Belkin, 1980; Kuhlthau, 1991; Wilson, 1999), and shed light on broader areas of how people generate, evaluate, and use content via social media. Finally, through its application to education, the project will impact interdisciplinary research that bridges LIS, computer science, and learning sciences.

2.2 Practical impact

Through an integrated research and education plan, this research program will contribute directly and significantly to support for online content assessment in higher education. Because the PIs will be testing and evaluating the proposed tool specifically in classroom and library instruction environments, that tool, released as open-source software as part of the project's dissemination plan, will be ready to be used and further developed by others in higher education. To engage other developers in building further modules for the system, the PIs will host the system on a service such as GitHub with an appropriate Creative Commons or MIT license. The PIs will report on the project at various scholarly meetings attended by others in library and information science education, such as the Annual Conference of ALISE. It is also anticipated that the outcomes could be adapted across a broad range of STEM education environments, including high schools and colleges going beyond Rutgers, Hawaii, and their libraries and classrooms. With the help of the project's industry partner – Brainly.com (an online Q&A service for students and educators) – the PIs will take the outcomes of this project to more practice-oriented contexts, including online education (MOOCs, continuing and professional training).

2.3 Strategic impact

The project will provide a valuable contribution to our national goal of encouraging and better supporting STEM education, and making curated information more accessible and applicable to larger audiences. This also addresses IMLS's strategic goal #1 of providing engaging and empowering learning experiences. The proposed project will help connect people with professionals, search experts with domain experts, and information seekers with information facilitators in an effective and meaningful way: promoting and supporting crowdsourced contributions, assessed by experts. Due to the unique combination of computational and social elements of this project that incorporate human-centered collaborative/social aspects, the research is expected to attract women and students from underrepresented communities, addressing the national initiative for social participation (Shneiderman, 2009). See Section 4 for specific details.

3 Project Design

This section provides details for the three main studies that will address the three RQs for the proposed project. These studies include investigating how end-users (STEM students) retrieve and evaluate crowdsourced information (Study-1), investigating how experts assess content quality (Study-2), and then, using the knowledge gathered from these two studies, building and testing a Q&A content assessment system (Study-3).

3.1 Study-1 for RQ1: Log data collection and user interviews

To understand how and why STEM students engage in online Q&A activities, a live user study will be conducted. In this study, three types of log data will be collected to elicit information about online Q&A users' questioning behaviors: (1) an access log, (2) a transaction log, and (3) an incident log. An access log will include user's IP address and the timestamp when they visit online Q&A sites, as well as the content of users' specific requests, including their questions posted within online Q&A sites.

A transaction log is an automatic data collection method that captures information such as user start/end time, protocol commands, text content, topic, keyword, etc. (Rice & Borgman, 1983). The goal of a transaction log is to "gain a clearer understanding of the interactions among searcher, content, and system or the interactions between two of these structural elements based on whatever research questions drive the study" (Jansen, 2006, p.409). The research questions in this project focus on STEM-related information-seeking behaviors performed in the context of online Q&A. Although log data collection may provide insights into how and when people decide to employ one of the online Q&A services for seeking information, a transaction log will not record "the

reasons of the search, the searcher motivations, or other qualitative aspects of use" (Jansen, 2006, p.424). Therefore, an incident log will also be collected for investigating online Q&A users' behaviors.

An incident log, also referred to as a diary method, is designed to capture "little experiences of everyday life that fill most of our working time and occupy the vast majority of our conscious attention" (Wheeler & Reis, 1991, p. 340). The benefits of an incident log are that this data collection method allows research participants to report their natural events and experiences naturally and spontaneously (Reis, 1994). This data collection method has gained in popularity in the LIS field to understand human-centered perspectives of information behaviors in various contexts (e.g., see Byström & Järvelin, 1995; Hansen & Järvelin, 2005; Kuhlthau, 1991, 1993). Moreover, an incident log is designed with "the short term between event occurrence and recording, hence, less subject to memory lapses and retrospective messaging, as may be the case with interviews" (Hyldegård, 2006, p. 154), which mitigates concerns of the potential incomplete recall in a transaction log.

To collect these three types of log data, this study will use a plug-in for Web browsers called Coagmento (http://coagmento.org), already developed and tested by PI Shah and his team (see González-Ibáñez & Shah, 2011; Shah, 2010, for a detailed description of Coagmento). Coagmento will serve as a client-level log data collection tool. In addition, this tool will have a sidebar that will present necessary elements for the participants to keep a diary of their question-answering incident each time they post or encounter a question or answer within online Q&A sites (see Appendix A). At the end of the month, each participant will be interviewed to elicit information about how and when they use online resources for educational needs and how they understand/assess their content quality.

Based on past experiences of conducting such studies (e.g., Choi & Shah, In Press) and the needed statistical power for analysis, a total of 40 students from STEM classes will be recruited using a stratified sampling method. The strata for such sampling will be derived using an approach similar to that described by Choi et al. (2013), where an initial survey of a number of participants provides groups based on behavioral patterns. Each participant will be paid $100 for using the provided browser plug-in for recording access, transaction, and incident logs for about four weeks and for being interviewed on the phone at the end of the study. To motivate the participants, eight (20%) of the most active participants will receive an additional $50 per person. Such extrinsic motivation has been found to be very effective for user studies (e.g., Shah & Marchionini, 2010).

A combination of techniques – both quantitative and qualitative – will be used on the data collected. These analyses will help the researchers understand: (1) the motivations for information seekers to post questions online; (2) the expectations for an asker from an online Q&A service; (3) how an asker perceives and assesses gathered information (answers) in response to his/her questions; and (4) how such user-generated information gets used in the larger context of information seeking. In other words, the findings from this study will add to our knowledge about the user side of the online Q&A coin. To understand the other side – the experts' assessment of content – the following study will be conducted.

3.2 Study-2 for RQ2: Building services for content evaluation using expert assessors (librarians)

Strategies for automated recognition of question content have been developed to address the problem of how to route questions to virtual reference librarians. Pomerantz and Lankes (2003) identify characteristics of questions that make them more and less appropriate for automated recognition, and propose a taxonomy space as a model for question "triage" in a virtual reference environment. They conclude that a hybrid approach involving both automated and human evaluation could provide a better solution. The PIs plan to adapt these and other findings toward the development of a statistical model to evaluate Q&A answer content. To build such a model, the PIs will obtain quality assessments on social media content, specifically answers from an online Q&A service, and perform a triage of features that contribute to content quality.

Two kinds of assessments will be obtained: objective and subjective. The former will be gathered from crowdsourcing and the latter will be elicited from expert assessors. For obtaining objective assessments for answers using crowdsourcing, the researchers will recruit subjects from Amazon's Mechanical Turk (http://www.mturk.com). Half of the data for this assessment will be collected from Yahoo! Answers, the largest online Q&A service, and the other half from Brainly, the industry partner for this project. Using corresponding tools (APIs for Yahoo! Answers and direct access for Brainly), a total of 10,000 answers will be collected along with their questions from the STEM-related categories including Physics, Chemistry, Computers & Internet, Consumer Electronics, and Science & Mathematics. Note that multiple answers may be

connected to the same question. These 10,000 Q&A pairs will be posted using MTurk for assessment of each answer along 13 objective criteria as described in Shah & Pomerantz (2010). Each MTurk worker will be given 100 pairs, to be completed in about two hours. Three different MTurk workers will assess each pair, thus requiring a total of 300 workers. See Appendix B for an example of how such work assignment looks like.

To obtain subjective assessment about content quality from expert assessors (librarians), a smaller dataset will be used. Specifically, random sampling will be used to extract one tenth of the dataset described above, thus yielding a total of 1,000 Q&A pairs. These pairs will be further divided in 10 subsets. Each of these subsets will be given to an expert assessor – preferably a subject librarian. Each of the 10 assessors will be asked to review the 100 Q&A pairs assigned to them to: (1) provide a subjective assessment about the quality of an answer given that question; and (2) be interviewed to discuss how they approached this task of quality assessment. Each of the assessors will be paid $100 for their participation.

To build a statistical model by triaging objective assessments of the crowd and subjective assessments of the experts, a technique similar to earlier works by PI Shah (e.g., Shah & Pomerantz, 2010; Choi et al., 2013) will be used. Specifically, expert assessments will be used as the gold standard to build a predictive model using regression on objective assessments of the 13 identified features. Using regression analysis (see Appendix D), the most influential features contributing to content quality will be extracted, and crosschecked once again with expert assessments – **a part that has been missing from prior works in this area**. Once such a model is built with reasonable prediction accuracy, it will be shared with the same expert assessors to obtain their feedback. Further testing and revision of this model will then be possible via bootstrapping and *n*-fold cross-validation techniques for training-testing. PI Shah has relevant experience in this area and will be responsible for this part.

3.3 Study-3 for RQ3: Evaluation in STEM learning environments (simulated in lab)

Once the statistical model for content quality assessment is prepared, it will be integrated in the Coagmento tool described earlier. A mockup of what this may look like is shown in Figure 2 (see Appendix C for a detailed mockup). As shown, the tool will provide a visual representation of various dimensions that contribute to content quality using on-the-fly analysis of information, specifically answers to a question on a peer-based Q&A service, displayed in the browser by applying the model built through the previous study. This tool will then be evaluated specifically for its usage among students in lab-based (simulated) instructional setting.

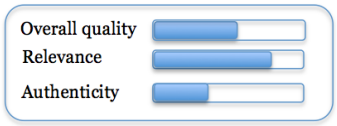

Figure 2: Mockup of automatic content assessment tool.

There are several precedents in the literature when it comes to presenting information about content quality to the users. Hargittai et al. (2010) found that young adults associate a high ranking in a search results list with content credibility. Kim et al. (2011) found that undergraduates specifically use others' reactions to evaluate Q&A content, and the proposed automatic content assessment tool would present this crucial data in a compact, usable form. Schwartz and Morris (2011) presented users with augmented Web pages and search results lists, where usually invisible credibility features such as site traffic and awards were presented visually alongside the content, and found that the visualizations increased the accuracy of users' credibility assessments twofold. Thus, the proposed method/tool for providing content assessment information is based on prior research methodology and results.

The lab test using this new tool will involve 40 STEM learners (students), paid $20/participant. Once again, to provide extrinsic motivation, eight (20%) of the best performing participants will be given an extra $20 per person. The participants will be given a homework-like task that requires gathering and evaluating information, and will be asked to use online Q&A services to complete the task. Using the tool, the researchers will provide the participants with online Q&A content augmented with automatic credibility assessment results, and investigate the extent to which their assessment of the content credibility changes. Critically, an open-ended response element will be included that allows subjects to express elements of online Q&A credibility they identify relevant to STEM education. These elements will be compiled in a master list that complements the 13 identified features from Shah & Pomerantz (2010).

Once lab-tested, the new tool will be experimentally deployed into existing information literacy instruction sessions run by the Leeward Community College Library in Hawaii. These instruction sessions are integrated with students' coursework and designed to strengthen their critical thinking and information quality assessment skills, so these sessions are an ideal environment in which to test the tool.

3.4 Project outcomes and evaluation

The results of various qualitative and quantitative analyses on lab study data will include: (1) an updated content assessment system that will be made available to STEM students, educators, and the public; (2) a list of guidelines for supporting and enhancing integration of social and crowdsourced information seeking activities in STEM education; and (3) a set of instruments and measures for studying, evaluating, and integrating Q&A content into STEM education, linked with existing Student Learning Outcomes (SLOs) and ACRL information literacy standards.

To measure the tool's effectiveness in STEM classroom instruction, interviews will be held with **STEM instructors** in which the automatic content assessment tool will be demonstrated and the master list of user-provided credibility criteria will be reviewed. Instructor feedback will be solicited regarding how to best integrate the tool into their classroom STEM instruction, and how to map the findings to both course-specific and institutional SLOs in STEM courses. To measure the tool's effectiveness in library instruction, similar interviews will be held with **STEM subject librarians**, specifically relating to teaching credibility evaluation, information literacy, and critical thinking, as well as mapping the findings to the ACRL Information Literacy Competency Standards for Higher Education (ACRL, 2000). Feedback on improvements to the tool will also be solicited. Assessment of student understanding is built into each information literacy instruction session conducted by the Leeward CC library, which provides a built-in method of evaluation for this phase of the study. This multifaceted approach will allow STEM students, instructors, and librarians to mutually engage in the question of how and when online Q&A content is appropriate in STEM education, as suggested by Veletsianos et al. (2013).

Feedback from these interviews will be used to improve the tool in a formative design-evaluation cycle, as well as to develop guidelines for its use in educational settings. The newly revised system will be deployed through Sakai, an online course management system, with the help of the Center for Teaching Advancement & Assessment Research (CTAAR) at Rutgers University. Students in multiple STEM classes will use the automatic credibility tool while working on information seeking and synthesis projects over a period of 6-8 weeks. User behavior will be analyzed through log data analysis. During the course of this study, the researchers will obtain various forms of information from the students, including demographic information, a pre-project questionnaire (including a pre-test), and a post-project questionnaire (including a post-test). A few of the subjects, chosen with the help of their teachers, will also be interviewed in focus groups at the end of their projects. It is expected that about **300 undergraduates from diverse institutions**, including Rutgers University, University of Hawaii, and Leeward Community College, will participate in this study. We follow the findings of Scaramozzino (2010), that the most effective STEM literacy instruction is tightly integrated with specific departmental and course student learning outcomes developed in consultation with instructors, but we provide in Appendix E a sample curriculum for one of these courses and how/where Q&A-STEM assessment activities could happen. The data generated from these instruments as well as logging tools will be analyzed using quantitative and qualitative approaches. At the end of the second and the third year of this project, an **independent external assessor** will be brought to Rutgers University, who will help us with evaluating project outcomes, especially related to STEM education. The recruitment of this assessor(s) as well as the plans for assessments will be aided by CTAAR at Rutgers.

4 Diversity Plan

PI Shah's home institution, Rutgers University, was ranked first in the country in campus ethnic diversity by U.S. News & World Report, and Co-PI Gazan's University of Hawaii was ranked sixth in the same study. Within Shah's primary affiliation at Rutgers SC&I, 40% of Ph.D. students identify as being of nonwhite ethnic background, and more than 75% are women. Both PIs will actively recruit students from underrepresented groups to participate in the research, and will engage with target institutions with diverse populations throughout the project. As an integral part of these activities PI Shah will be involved in the Rutgers Future Scholars program, which includes a minority, disadvantaged, and first-generation college population. Co-PI Gazan will work with students, instructors, and librarians at Leeward Community College, which has the highest enrollment of Native Hawaiian students on the island of Oahu, to develop, implement, and test the project. The PIs will regularly evaluate the outreach and recruiting activities, talks, and presentations. Following Goode (2008), they will distribute a personal invitation letter to attendees of these events and will include a

survey to gather information about perceptions and attitudes toward graduate studies in the computer and information science fields. The PIs will also use surveys partially derived from the questionnaire developed by the WGBH Educational Foundation (New Image, 2009), and include questions regarding the participants' image of computer and information science activities, their likelihood to choose this field as a career path, and how or whether participating in the activity changed their attitude. Interviews with participants will further expose the utility of the activity and ways in which it can be improved.

5 Project Resources: Budget, Personnel, and Management Plan

5.1 Budget

The proposed project will span three years, from 9/1/2016 until 8/31/2019. The total estimate of costs related to personnel, participant support, and project management is $704,210, of which $204,837 **(nearly 30%) will be cost-shared** by the participating institutions. The type of qualitative, in-depth, and systems-design work proposed in the grant project's three phases will require highly trained, knowledgeable personnel to collect and analyze the data.

It is worth mentioning that the project will rely as much as possible on graduate student research assistants and student hourly employees because of the cost-effectiveness and learning opportunities these positions provide for hands-on analysis and acquisition of important analytical skills.

5.2 Personnel

Chirag Shah, Ph.D., is an Assistant Professor, Rutgers University, School of Communication & Information. Shah has a strong background in building and evaluating information systems. He has been working in the field of SQA for several years, resulting in several substantial contributions and recognitions, including a small grant from the IMLS to look at how VR and SQA services may be combined to create synergic solutions for question-answering. That work resulted in design ideas for a hybrid system, which will be instrumental for the current project. He will devote 21% of his time during the academic year or a summer month or a combination.

Rich Gazan, Ph.D., is an Associate Professor, University of Hawaii at Manoa, with a dual appointment in the Department of Information & Computer Sciences and Library & Information Science Program. He serves as Chair of the LIS Program, and has studied interdisciplinary scientific collaboration in two NASA Astrobiology Institute grant projects. He helped develop Answerbag, the first SQA site in the US, and has published extensively in the area of SQA. He will devote 20% of his time during September through May and one month during the summer to this project.

Prof. Nick Belkin, Prof. Marie Radford, Prof. Dan O'Connor, and Prof. Michael Lesk from the Department of LIS at Rutgers University will serve as the internal advisors on the project. Dr. Belkin is a distinguished scholar in the field of LIS, whose work in information seeking/behavior has inspired and trained many generations. He will help the PIs make connections between information seeking and STEM education. Dr. Radford is a prominent researcher, educator, and practitioner in the field of reference librarianship (physical and virtual), and will be able to advise on issues related to creating collaborative services for the libraries and providing sustainable models for virtual referencing. Dr. O'Connor specializes in data analytics as well as evaluation of library and information systems and services. He will help the PIs with analysis of user-generated content and development of evaluative measures. From Bell Labs to the NSF and academia, Dr. Lesk is known for his work with natural language processing, digital curation, and preservation. He will specifically advise the PIs on accessing and storing digital content as well as interfacing it with library users and students.

In addition, **Dr. Lynn Connaway from OCLC, Dr. Stephen Freeland from University of Maryland Baltimore County, Head Librarian Wayde Oshiro from Leeward Community College**, and **Joel Downs from CBS Interactive** will serve as the external advisors on the project. Dr. Connaway, a senior research scientist at OCLC, has collaborated with PI Shah on an IMLS-funded project involving virtual reference (VR) and social Q&A services. She will help the PIs in continuing to push the boundaries of VR and other library services with the help of social and crowdsourced information, making libraries more relevant and applicable to educational practices. Dr. Freeland directs an interdisciplinary studies degree program supporting a broad range of STEM learners, and Mr. Oshiro leads a library that designs and conducts more than 100 library instruction and information literacy sessions each semester: both will provide access to target STEM learners. Joel Downs founded Answerbag, the first social Q&A site in the US that also received SXSW award for Best Educational Resource in 2004, and will provide a valuable industry perspective. The project will also have an industry

partner, **Brainly.com**, an online Q&A service for students and educators. PI Shah is already collaborating with Brainly to investigate issues of content and user assessments in online Q&A. As a part of this project, Brainly will provide the PIs access to their data and tools, as well as providing advice on implementing the findings and outcomes in a widely accessible service through libraries and other sites.

The PIs will meet (physically or virtually) with the internal advisors once a month during the academic year, and twice a month during the summer. Each of the internal advisors has agreed to commit 5% of their time to the project. The PIs will, similarly, communicate frequently with the external advisors via emails and phone calls. The PIs will arrange for an annual meeting with the advisory board at the ASIST conference and have informal meetings with them at other professional gatherings. The advisory board will be able to stay current with the project's progress using the project Website.

In addition, to help the PIs with studies 2 and 3 and in general have ready access to various library services, **Interim VP of Information Services and University Librarian Jeanne Boyle** has graciously agreed to be available for this project. Her staff, and Rutgers Libraries in general, will be available to the PIs as they proceed with collecting and evaluating user-generated content, and deploy various services in libraries, laboratories, and classrooms. Specifically, **Laura Palumbo, Chemistry & Physics Librarian and Science Data Specialist**, will be available to assist the PIs with issues related to STEM learners. **Dr. Monica Devanas, the Director of Faculty Development and Assessment Programs from the Center for Teaching Advancement & Assessment Research (CTAAR)** at Rutgers University, will be available to help the PIs with assessment-related issues. Dr. Devanas will especially assist the PIs in defining assessment rubrics for various studies, and hiring independent assessors to help with evaluation of project outcomes.

5.3 Management plan

PI Chirag Shah will direct the overall research and project management with the help of co-PI Rich Gazan. The responsibility for the basic research components is divided between the PIs. Two PhD students – one at Rutgers University and one at the University of Hawaii – will be hired as fellows for the duration of the project. These students will be responsible for helping Shah and Gazan design and conduct user studies, collecting and analyzing data, and writing and disseminating project results. An hourly student will be hired during the summer months to be responsible for programming various components needed for the user studies, as well as data cleaning and processing. The PhD students will work directly with the hourly student, guiding him/her according to the project requirements.

As overall project director, Shah will be responsible for ensuring timely reporting to IMLS and adequate wider dissemination to the scholarly community. Both Shah and Gazan, with the help of their PhD students, will also be responsible for building resources such as a project Website and open-source distribution of the tools and services developed as a part of this project.

6 Communication Plan

Results from this project will be disseminated to diverse research and professional communities, to government and industry, and to the general public. More specifically, the results will be communicated through presentations at relevant conferences and workshops such as the ASIST annual meeting, and ACM SIGIR and ACM CSCW conferences. The PIs have also co-organized and served on several panels at ASIST and other conferences that attract both academic and professional audiences relating to VR and SQA topics, and will continue doing so. Results will also be written for journals such as the JASIST, IP&M, and Reference Services Quarterly. To broaden the impact, the advisory board will help identify K-12 school and public library journals/conferences to disseminate the findings from this project. The PIs and PhD students will also present interactive demonstrations to librarians and learners for dissemination, outreach, and formative evaluation. These demonstrations will take place at various libraries, starting with the collaborating libraries for this project (Rutgers, University of Maryland Baltimore County, and Leeward Community College in Hawaii), and then expanding to other venues with the help of the advisors (such as Connaway from OCLC) and partners.

Both Rutgers and Hawaii maintain dedicated offices for public relations to disseminate information about important projects, publications, and recognitions to local media outlets, professional and research communities, and the public. The project will also have dedicated Facebook, YouTube, and Twitter channels available through the project Website (see examples of such outlets on http://infoseeking.org run by PI Shah), which will help in communicating with STEM educators outside of the LIS field.

Project Timeline	Year One (2016-2017)												Year Two (2017-2018)												Year Three (2018-2019)											
	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S
GENERAL ADMINISTRATION																																				
General Administrative Oversight and Assessment																																				
Advisory board meeting																																				
External assessor's visit																																				
Investigating online Q&A behaviors (RQ1)																																				
Consolidate preliminary findings and tools for capturing online Q&A behaviors																																				
Prepare tools and database for conducting Study-1																																				
Study-1: Log data collection																																				
Interviews with Study-1 participants																																				
Analyses of log and interview data from Study-1																																				
Content assessment (RQ2)																																				
Collect a total of 10,000 STEM-related Q&A pairs from Brainly and Yahoo! Answers																																				
Prepare Human Intelligent Tasks (HITs) using Q&A pairs and post on Mturk (Study-2)																																				
Librarians providing expert assessment for a total of 1,000 Q&A pairs (Study-2)																																				
Interviews with the librarians (Study-2)																																				
Q&A model applied to STEM education (RQ3)																																				
Development of automatic content assessment tool																																				
Lab experiments to test the effectiveness of content assessment technique (Study-3)																																				
Project outcomes and evaluation																																				
Revision of the tool and services to prepare for deployment																																				
Classroom experiments with the revised tool																																				
Student Learning Outcomes and ACRL-based interviews with students, teachers, librarians																																				
Reports and Dissemination																																				
Website, blogs, social media presence																																				
Complete all analyses, prepare interim reports and final report																																				
Attend professional conferences																																				
Dissemination through publications																																				