International Review of Research in Open and Distributed Learning

"Remember to Hand Out Medals": Peer Rating and Expertise in a Question-and-Answer Study Group

Marisa Ponti

Cite this article

Article abstract

This article reports on an exploratory study of giving medals as part of a peer rating system in a question-and-answer (Q&A) study group on Python, a programming language. There are no professional teachers tutoring learners. The study aimed to understand whether and how medals, awarded to responses in a peer-based learning environment, can work as a mechanism to assess the value of those responses when traditional markers of expertise are not always clearly defined and identifiable. Employing a mixed-method approach, the analysis examined (a) the content of the answers that were awarded medals and their perceived immediate value and (b) the nature of the networked relationships resulting from participants' interactions. The findings suggest that the peer rating system makes visible what the participants find immediately valuable and allocates a form of recognition that extends the "legitimation code", which refers to the credentials that make someone competent and worthy of recognition.

"Remember to Hand Out Medals": Peer Rating and Expertise in a Question-and-Answer Study Group

Marisa Ponti
University of Gothenburg, Sweden

Abstract

This article reports on an exploratory study of giving medals as part of a peer rating system in a question-and-answer (Q&A) study group on Python, a programming language. There are no professional teachers tutoring learners. The study aimed to understand whether and how medals, awarded to responses in a peer-based learning environment, can work as a mechanism to assess the value of those responses when traditional markers of expertise are not always clearly defined and identifiable. Employing a mixed-method approach, the analysis examined (a) the content of the answers that were awarded medals and their perceived immediate value and (b) the nature of the networked relationships resulting from participants' interactions. The findings suggest that the peer rating system makes visible what the participants find immediately valuable and allocates a form of recognition that extends the "legitimation code", which refers to the credentials that make someone competent and worthy of recognition.

Keywords: Gamification; expertise; non-formal education; peer rating; question-and-answer site

Introduction

In formal education, one of the long-established roles of teachers is to act as "gatekeepers", that is as central knowledge filters who direct students to content they have vetted based on established quality criteria and who evaluate students' work through institutionalized assessment procedures. Built on the assumption of the scarcity of information and expertise (Weller, 2011), the model of traditional gatekeeper oversight is made untenable in a networked environment where participants share information from a plurality of sources they have direct access to. Furthermore, emergent technologies are also claimed to disrupt the notion that learning should be controlled by traditional gatekeepers, as information and "knowledgeable others" are available on online networks (Kop, Fournier, & Fai Mak, 2011). This shift in information provision suggests circumstances under which uncredentialed sources can provide valuable information. The use of tools for crowdsourced evaluation can be seen as harnessing the ability of those providing useful information, even though they lack special training, credentials, or established offline reputation. It can be argued that the use of these tools can problematize the concept of expertise by calling into question several of the indicators upon which people have commonly relied to signal expertise, as well as the notions about the evaluation of which information is most valuable (Pure et al., 2013).

This article reports on an exploratory study of giving medals as part of a crowdsourced evaluation in a question-and-answer (Q&A) study group on Python, a programming language. Medals awarded to best responses denote a relationship between two or more study group participants. The goal of the study was to understand whether and how medals, awarded to responses in a peer-based learning environment, can work as a mechanism to assess the value of those responses when traditional markers of expertise are not always clearly defined and identifiable. For example, many participants do not provide information about their job titles and credentials, therefore their peers cannot focus on traditional markers of source expertise as the basis for assessing the value of responses. The overall goal is to develop a broader understanding of whether the use of medals legitimizes the answers provided by responders, regardless of their status and characteristics, and participates in changing the "legitimation code" (Maton, 2000), which refers to the credentials that make someone competent and worthy of recognition.

Related Work

To place the study presented here in context, an overview of social network question asking services and studies of their use are presented. The phenomenon of gamification in education and what research has revealed about it is then discussed.

Social network question asking.

Recently several scholars in the field of Computer Supported Cooperative Work (CSCW) have investigated social media question asking in a variety of platforms where participants ask and answer questions reciprocally. Participants can be identified either by using pseudonyms or by

their real names. They ask questions to their own social circle (Panovich, Miller, & Karger, 2012), strangers (Nichols & Kang, 2012) or topical experts (Mamykina et al., 2011). Motivations for participation in these sites can be either intrinsic (such as altruism or the desire to learn) or extrinsic (such as gaining reputation-enhancing benefits or rewards) (Mamykina et al., 2011). A stream of research has studied participants asking questions in general-purpose online social networks for information seeking. For example, Morris, Teevan, and Panovich (2010) conducted a survey study of how people used their status updates on Facebook and Twitter and found a majority of questions asking for recommendations and opinions. Panovich, Miller, and Karger (2012) examined the role of tie strength in responses to status message questions on Facebook, while Paul, Hong and Chi (2012) analyzed questions asked on Twitter and found that 42% of questions were rhetorical. Social media question asking in specific communities, such as the blind (Bigham et al., 2010) and mobile phone users (Lee, Kang, Yi, Yi, & Kantola, 2012), has also been a topic of inquiry.

This review is necessarily partial but reveals an apparent shortage of studies of social network question asking in educational settings. In education one of the most popular Q&A sites is Piazza (Barr & Gunawardena, 2013), which supports students in courses. Aritajati and Narayanan (2013) have also turned attention to a question and answer board called Green Dolphin, designed to facilitate students' collaborative learning of programming.

Crowdsourced evaluation and gamification in education.

Several Q&A sites (e.g., StackoverFlow and Yahoo! Answers) use systems for crowdsourced evaluation. Typically, they let large numbers of individuals, who usually do not know each other, assess the value of the provided answers. An evaluation often depends on one's interpretation of what is "valuable" in the provided answers. To motivate engagement in asking and providing good answers several Q&A sites use gamification mechanisms to retain and reward participants. The term *gamification* began to be used in the media in 2010 (Radoff, 2011) and has been defined as the use of game elements and game mechanics applied to non-game settings in order to improve user experience and engagement of participants (Deterding, Sicart, Nacke, O'Hara, & Dixon, 2011). These elements can include levels to define statuses of participants, scoreboards or points to track progress towards goals, badges to reward achievements, and leaderboards to compare progress with peers. The aim of gamification is to extract and adapt the game elements that make games enjoyable and fun and use them for learning (Simões, Redondo, & Vilas, 2012). Therefore, Simões, Redondo, and Vilas argue, students learn not by playing specific games, but they learn as if they were playing a game. In education gamification has also been considered as potentially useful for improving student engagement and motivation in classroom and online settings because it can allow teachers to situate learners in authentic environments where they can practice their skills, gain immediate feedback, earn recognition for doing well and feel good for overcoming a challenge (Kapp, 2012, p. 22). Although there is a vast literature on using games for learning, there is an apparent dearth of studies on the use of gamification in Q&A sites (e.g., 12), and in education (Ioana Muntean, 2011; Paisley, 2013). The study this article reports on adds to

the knowledge of social networking question asking in a study group by presenting findings from a mixed-method study that illustrate how participants used gamification elements to assess each other's responses.

Overview of the Study Group MIT 619

MIT 619 *A Gentle Introduction to Python* is a study group for individuals interested in learning the basic notions of Python. No prerequisites are supposed to be necessary aside from high school algebra. In this peer-based learning environment organized as a Q&A site there are no professional teachers or topical experts tutoring and assessing learners. The responsibility to run study groups is distributed across all the participants. Participants are expected to engage and help their peers. They ask questions, report problems and difficulties, share coding assignments, and link to external materials (Figure 1). In July 2014 the group had 2,113 registered members and 515 questions. The study group is named after the massive open online course (MOOC) with the same title (http://mechanicalmooc.org/) but is available as a standalone study group as well.

The study group is divided into two major sections. The section on the left side contains open questions and is updated in real time. The section on the right side contains closed questions (e.g., askers can close questions when they are satisfied with provided answers, or to be able to ask another question because they can have only one question open at a time). Closed questions can be as short as one question and one answer or as long as 30 posts. Each closed question has its own section that displays a top-level question followed by one or more answers, the best answers and the awarded medals (if any), and the names of askers and responders with their overall SmartScore. To credit responders who provide good answers the study group has a code of conduct encouraging participants to reward helpers with medals. When askers, or other participants, find an answer valuable they can click the "Best Response" button to reward it with a medal. The code of conduct allows each participant to give only one best response for each question but each question can have best responses by several different participants. When a response is marked as a best response the responder receives a medal. The number of medals received is one metric contributing to the SmartScore, a scoreboard that reflects the effort and ability participants apply to helping and being supportive of their peers. At the time of this writing, the overall SmartScore comprises participants' scores in three categories: teamwork, problem solving, and engagement. The scores for each of these three categories permit recognition based not only on medals awarded to answers, but also on the quality of interaction (being prompt and polite, for example), the willingness to contribute to solving problems, and the time spent on site. Members are ranked on their SmartScore and these ranks are visible not only to all registered members, but also to people who are not registered.

As a newcomer, a participant starts as Hatchling in all the three categories. For each category, new titles (e.g., Rookie, Recruit, and Life Saver) are issued when scores reach 20, 50, 75, and 90, with a participant holding a special and unique title if s/he scores 100 for a category.

Figure 1. An example of question and answer.

Theoretical Perspective and Method

Inspired by a sociotechnical sensitivity, this study investigates the sociomaterial agency of medals to examine how they work as a mechanism to assess the value of best responses when traditional markers of expertise are not always clearly defined and identifiable. Following Leonardi (2012), sociomaterial agency is meant as the ways in which the entanglement of social phenomena (e.g., values, norms, discourses) and material phenomena (e.g., technologies) acts. Putting materiality at the center of the analysis allows attention to be given to the devices that come to shape rating and influence the activity to which the rating refers. Such is the importance of materiality that, as Pollock (2012) suggested, there can be no rating without the devices of rating. Pollock argued that it is only through working with these devices that ranking organizations can produce and communicate ratings. This view emphasizes the material and distributed character of rating, which means that this action is not performed by individuals alone, but through the enrollment of a variety of material artifacts. Similarly, in this article, rating is not a capacity to assess that inherent within humans, but a capacity performed through the relationship of humans and material artifacts.

In this study data were gathered from multiple sources, including questions posted in the study group and online documentation. Data sources and data analysis procedures are discussed in detail below.

Sampling and Data Collection

Virtual ethnography was conducted through non-covert, non-participant observation for 16 months, from the start of the Q&A site in October 2012 until mid-January 2014. During the period of observation data were collected exclusively from the closed questions section. Although closed questions can be closed without being answered, generally they are answered by at least one participant and often by several participants, therefore, it is possible to see if helpers giving

best responses were credited with medals. In contrast, open questions can remain unanswered, waiting for someone to give answers. At the time of completing this study there were 302 closed questions out of a total of 475 questions. In order to make the analysis manageable, participants who asked closed questions were sampled. Sampling can be problematic when conducting online research (Fricker, 2008) and, similarly, some issues had to be addressed in this study. Members join the group based on an interest in learning Python and little information is required when registering to the study group. Registration typically involves asking for the individual's name, school and a little description, but few participants provide this information about themselves in their profile and even this information may be questionable because there is no guarantee that participants provide accurate demographic or personal information. This issue makes it difficult to generate a sample frame but it is somewhat less of a concern because this study is non-probabilistic. Therefore, a homogeneous sample was chosen (Patton, 2002). After examination of the entire corpus of closed questions, the sample only included the participants who, at the time of conducting the observation, were at the level of Hatchling at least in two categories, whose questions received at least one response rated as best response and whose responder was awarded medals. Excluded from the sample were all the participants with different titles as well as closed questions unrelated to Python and closed questions followed by answers with no medals. Finally, 132 closed questions were included in the homogenous sample. Other than sharing the Hatchling title, the participants included in the sample came from a vast range of backgrounds and ages, although with a substantial gender difference in that there was a prevalence of men.

In addition to observation of closed questions, data were also collected through publicly available online documentation, especially that provided by OpenStudy.com, the organization running the platform used by the study group. This documentation included relevant questions and answers in OpenStudy Feedback, videos, and blog entries, and provided a basis for understanding the formal rules and structure of the study group.

Analysis

Two methods of inquiry were used to analyze the data: qualitative content analysis and social network analysis (hereafter referred to as QCA and SNA). QCA was used to examine the content of the answers considered best responses and their perceived immediate value (Wenger, Trayner, & de Laat, 2011), which is defined as the value provided by good answers that can be directly and immediately applied in the learning process. The analysis was conducted on indicators of such value, which were manifest in the content of the answers. Rourke et al. (2003) defined manifest as the content that is at the surface of communication and is therefore easily observable. The value provided by best responses was not seen as an attribute intrinsic to these answers but as a relational property that depended on peer evaluation and recognition through awarding medals. This means that medals were treated as markers of the perceived "value" of responses, according to medal givers.

A mix of pre-existing and inductive codes was used to analyze instances of immediate value provided by best responses. Pre-existing codes from two schemes were brought in; the first

scheme was developed by Mason (1991, p. 168) and includes seven broad categories to analyze interactions in an online educational forum. The categories include use of personal experience related to a course theme, reference to appropriate materials outside a course and tutors acting as facilitator, among others. The second scheme was developed by Anderson et al. (2001, p. 8) and includes six categories to analyze facilitation of discourse. The categories include drawing in participants and prompting discussion, setting a climate for learning, and encouraging or reinforcing student contributions, among others. These two schemes were chosen because previous observations of the whole corpus of closed questions gave a sense of what was relevant there. Based on those observations, the two schemes provided a set of broad categories that did not constrain openness to new concepts suggested by the data. In coding the data a single post was chosen as a unit of analysis, being less time consuming and facilitating unit reliability (Rourke et al., 2003). Data were entered in HyperResearch software for open coding of best responses as exhibiting one or more indicators of each of the categories used in the analysis. The same content was coded on two different occasions to determine intra-coder reliability (Johnson & Christensen, 2012).

SNA was performed as second inquiry method because it offers the potential to describe the nature of networked relationships resulting from the flow of information and influence found among participants' interactions. In preparation for SNA, simple descriptive statistics (counts, percentages, averages) were conducted. For example, participants' asking and responding and giving medal behaviors provided data about participation. Questions were addressed such as: who are the active answerers among the Hatchlings? Who are the top responders and who do they reply to? Who are the most knowledgeable participants? What are the common social roles that occur among the participants? In this study the purpose was to better understand the nature of the relationships between Hatchlings and participants bearing other titles, and whether participants' networked positions reflected a distribution of expertise and knowledge. To accomplish this centrality was used, as a measure of prominence based on the number of mutual and non-reciprocal ties that participants have with each other. Participants' overall centrality was calculated by combining measures of in-degree centrality, which are counts of inbound ties with other participants, and out-degree centrality, which are counts of outbound ties. These measures, when considered separately, are indicators of network prestige (in-degree centrality) and influence (out-degree centrality). In the examined Q&A site prestige can be used to measure the number of people who responded to a person or the number of people who gave medals to a person. In contrast, influence can be used to measure the number of people a person responded to, or the number of people a person gave medals to. NodeXL software was used to measure degree centrality, in-degree centrality and out-degree centrality to quantify participants' interactions, including giving best responses to questions from Hatchlings and non-Hatchlings and giving medals to those best responses.

Results

First, the results from the analysis of measurable behaviors, such as numbers of questions and best responses and awarded medals were gathered and used to analyze the structure of the network. The findings from the qualitative QCA of the sampled closed questions are then reported.

Participant Asking, Responding and Giving Medals

Asking Questions. 132 questions were asked by Hatchlings and 48 were asked by non-Hatchlings (this broad category includes participants with all the other titles). Hatchling askers were 77 and non-Hatchling askers were 24. Of the 77 Hatchling askers 39% asked only one question (median = 1 question/asker, average = 1.7 question/asker) and provided 52 questions. Figure 2 shows the distribution of questions per Hatchling askers.
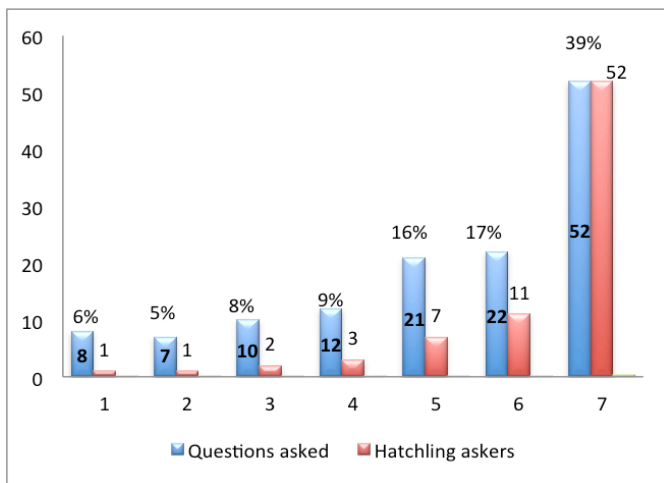

Figure 2. Percentages of questions asked by hatchlings.

48 questions asked by 24 non-Hatchling askers and followed by best responses were also analyzed to explore how questions were distributed across these askers. 27% asked only one question (median = 1 question/asker, average = 2 questions/asker), but one participant also counted for 23% as shown in Figure 3.
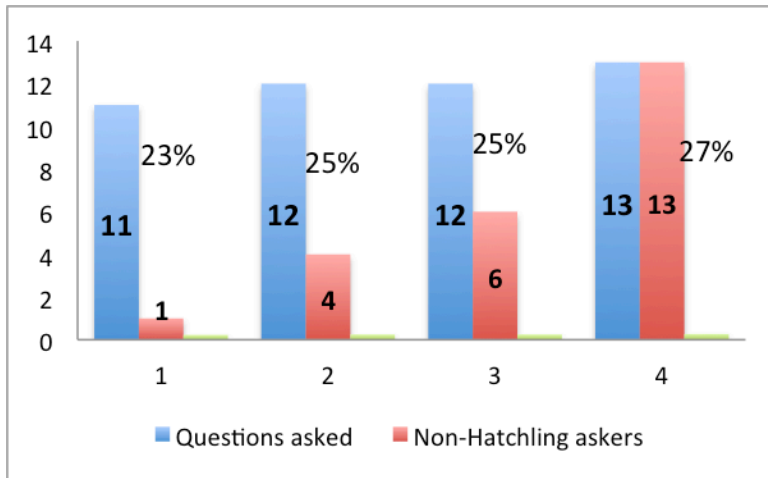
Figure 3. Percentages of questions asked by non-hatchlings.

Best Responding. 76 participants gave 164 best responses (median = 1 best response/answerer, average = 2.1 best responses/answerer) to the 132 questions asked by Hatchlings. Figure 4 shows the distribution of best responses across responders per title.
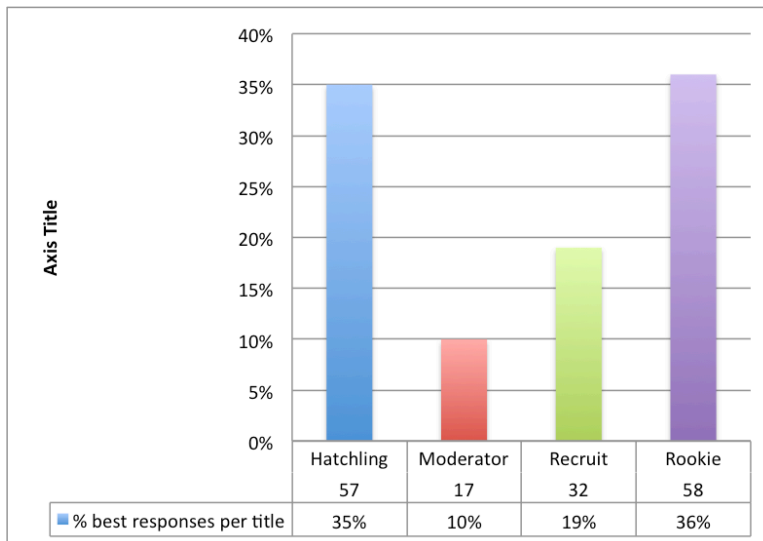

Figure 4. Best responders to hatchling questions per title.

Regarding the best responses to the questions asked by non-Hatchlings, out of the 64 best responses provided by 33 answerers to the 48 questions asked 12 Hatchlings gave 13 best responses, representing 21% of the total, as shown in Figure 5.
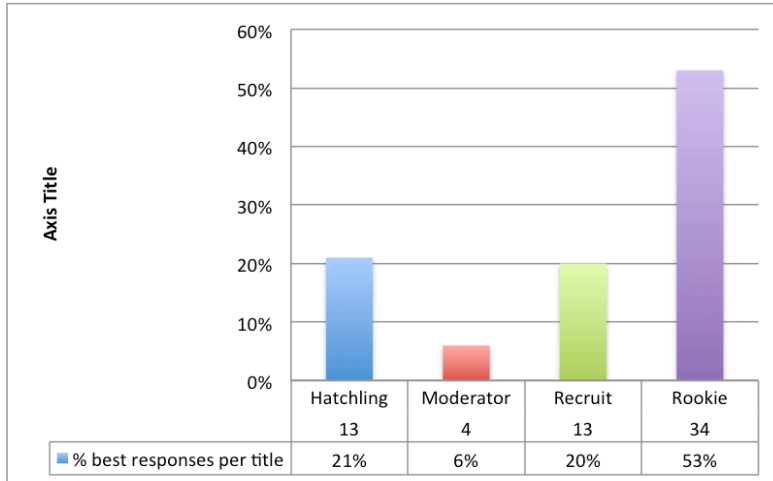
Figure 5. Best responders to non-hatchling questions per title.

Giving Medals. 76 best responders to the questions asked by Hatchlings received 208 medals (median = 1 medal/best responder, average = 2.7 medals/best responder). 33% of medals were given to 43 Hatchling responders, who represent 57% of all responders, as shown in Figure 6. Medal givers played three different roles; they could have: 1) asked a question, 2) participated in another person's question, or 3) just observed another person's question.
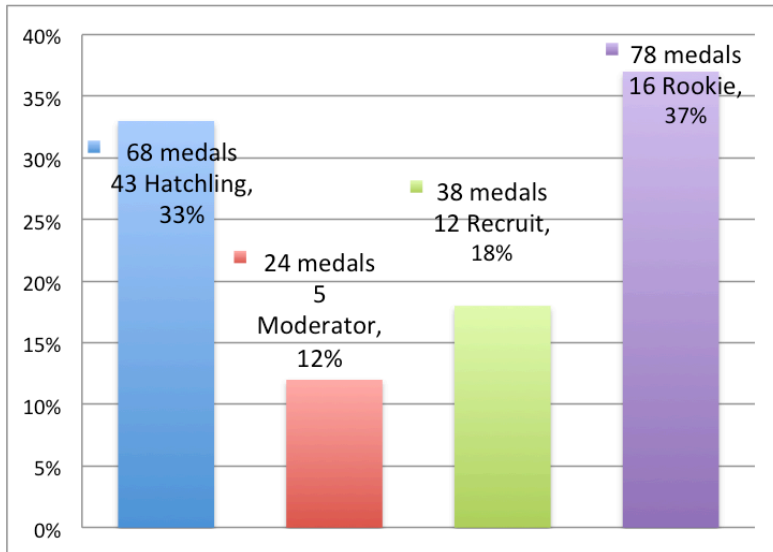

Figure 6. Medals received by best answerers to hatchling questions.

Figure 7 shows that 12 Hatchlings out of the 33 best responders to the questions asked by non-Hatchlings received 17% of the 82 awarded medals.
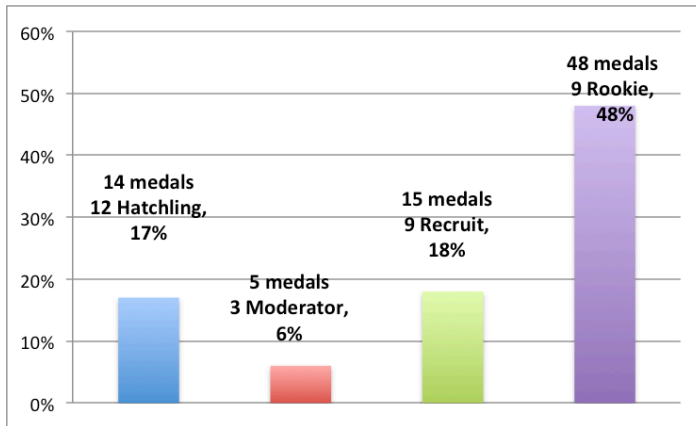

Figure 7. Medals received by best answerers to non-hatchling questions.

Value of Best Responses

Table 1 shows the codes, percent of coded best answers, and representative answers for the 22 categories resulting from the content analysis.

Table 1

Codes and Representative Answers

Codes	Percent/ Frequency	Examples
Giving help for fixing coding errors or solving exercises	26% (141)	"I can't give you the deep theory, but if you try this, you'll see what's going on: x ="012345"…"
Giving examples of code	9% (47)	"[x**3 for x in range (1,11)]"
Giving help about the functioning of the MOOC	8% (43)	"At this early stage, peer review is maybe not so important unless you are a complete programming beginner. Later, with more complicated assignments, peer review is a good way to get new ideas…"
Drawing on one's own experience	8% (41)	"I tried to make the queue a list at first but then I couldn't decide how to control the FIFO bit so I made it a string instead…"
Setting climate for learning	7% (39)	"Welcome to OpenStudy!!! If you ever need help with a question, just post it in the correct section and I'm sure you'll get help very soon…"
Acknowledging and reinforcing peer contribution	6% (32)	"Thanks! I wasn't aware you could use the "for x in list" structure inside a list bracket, but that's pretty nice…"
Drawing in peers and prompting interaction	6% (30)	"In the first example, can you determine the number, when multiplied by itself (or squared) gives you 16?"

Codes	Percent/ Frequency	Examples
Giving help for fixing technical concerns	5% (26)	"We've noticed Codecademy can be a little flaky sometimes. Try a different browser or come back later and try…"
Linking or referring to relevant material outside the MOOC	4% (24)	"According to this is possible http://stackoverflow.com/questions/4583367/how-to-run-multiple-python-version-on-windows and there is similar article if you are using Linux…"
Referring to participants by name	4% (21)	"Great to see you all online. Want to stress what @e.mccormick said above. The resources are persistently available…"
Linking or referring to material in the MOOC	4% (19)	"The sequence for the course is available here http://mechanicalmooc.wordpress.com/sequence/…"
Communication serving purely social function	3% (14)	"You're welcome"
Using the study group effectively	2% (12)	"For medals, click on the Best Response button. To become someone's fan, hover over their screenname and there's a 'Become a Fan' button…"
Presenting content related to Python	2% (11)	"…Python is modelled after the Fortran line; there is a clear distinction between expressions and statements…"
Linking to examples of code in online compilers	2% (10)	"I'm not very excited about my solution to that problem, but here it is: http://ideone.com/yioELL"
Giving feedback on solutions of exercises	1% (6)	"Looking back at your code I can see that you assigned P_1 = 'Player_1 right before your if statements, thus P_1 would never be rock/paper/scissors. Getting rid of those 2 lines of code and assigning the variable directly to your raw inputs would have you working."
Assessment of the efficacy of the process	1% (5)	"…I also think the MOOC definitely should have included more lectures for these exercises – it's a big leap from week 6 to week 7 without the lecture I think"
Complimenting and expressing appreciation	1% (5)	"Yeah Buddy! I'm so excited guise."
Explaining a concept relating to programming	1% (4)	"Floats are imprecise. Due to rounding errors in binary conversion, floating point math has some issues. They do not exist in integer math. Therefore integers are more precise."
Giving general information	1% (4)	"This is Python study group"
Referring to others' posts	1% (4)	"Check other messages. There was a thread sharing their nims code. At least 3 people did submit working examples…"
Self-disclosure	0% (2)	"… I'm just a lowly electrician who is studying to be an electrical engineer and also like computers."

Characterizing the Network

Not all the participants in the study group asked questions, nor did they all respond to questions. A bow tie structure analysis was used to examine the general structure of the study group network. According to this structure, initially proposed by Broder et al. (2000), the Web structure

resembles a bow tie, including four components: Core, In, Out, and "Tendrils" and "Tubes". Following the example provided by Zhang, Ackerman, and Adamic (2007), who used the bow tie structure to analyze a Java Forum structured as a Q&A site, the central core contains participants that more frequently help each other. In the case of the examined study group the Core subset is conceptualized as a more strongly connected component in which one can reach many participants from each participant in the subset by following the question containing the questioner-answerer thread. The In component contains participants that usually only ask questions. The Out component contains participants that usually answer questions only from participants in the Core. Other participants, the Tendrils and the Tubes, either respond only to questions posed by the In participants, or to questions responded only by the Out participants.

Table 2

Comparison between the Study Group and the Java Forum

	Core	In	Out	Tendrils	Tubes
Web	27.7%	21.2%	21.2%	21.5%	0.4%
Java Forum	12.3%	54.9%	13.0%	17.5%	0.4%
Study Group	11.0%	56.6%	6.2%	N.A.	N.A.

The results in Table 2 indicate that the study group looks similar to the Java Forum. Similarly to this forum, the study group also has a relatively larger In component and a relatively smaller Core than the Web. This indicates that in this study group only about 11% of participants asked and answered questions more actively. For example, only about 9% of Hatchling askers (7 out of 76) asked at least 4 or more questions, while 52 askers (68%) asked just one question. Furthermore, only 5.3% (4 out of 76) of best responders to Hatchling questions gave 9 or more responses to several different askers, while 54 (71%) best responders provided just one best response. The same four participants also received most medals by participants. Most Hatchling askers only asked questions (63 out of 76 (83%)) whereas only 10 out of 33 (30.3%) non-Hatchling askers only asked questions. As the Core of askers, both Hatchlings and non-Hatchlings, is quite small, the Out component is also very small, only 5 out of 76 participants (6.5%) giving best responses more frequently to questions from the Core Hatchling askers, and only 2 out of 33 participants (6.0%) giving best responses more frequently to questions from the Core Non- Hatchling askers. There are no Tendrils or Tubes. In line with what Zhang, Ackerman, and Adamic (2007) found in their study, these results also indicate that the study group is not a place where people help each other reciprocally, rather it is a help seeking network where askers are more likely to come to look for help from volunteer helpers. These results are confirmed by social network analysis. By connecting participants who asked questions to those who gave them best answers, and participants who gave medals to those who provided best answers, two graphs were created,

which were called QA network and MG network respectively. Using the NodeXL software, the structure of these two networks was analyzed. The analysis shows scale-free networks in which only few nodes are more linked, while most nodes are only very sparsely connected. For example, nearly 85% of the nodes in the QA network (responders to Hatchling askers) have both very low out-degree and in-degree values (when out-degree = 1, in-degree = 0 and vice versa). The highest out-degree value of a node in this network is 9 and the highest in-degree value of a node is 8.

While a high value of centrality indicates that a node can reach the others on short paths, or that a node lies on many short paths so that removing that node would lengthen the paths between many pairs of nodes, in these networks high values of centrality of a few more connected nodes do not have this meaning. Apparently, connectivity is not a proxy for centrality. No nodes seem to link together different parts of the networks. The non-reciprocity shows that the tie between any two given nodes seems to be valid for the "here and now", which means that instead of everybody helping each other equally, there are a few more active participants who answer more questions and reward questions with medals while the majority of participants answer one or two questions and give one or two medals. This result is visible in the participants' ego networks. Each ego network consists of a specific participant, the ties to other participants that person interacts with directly, and the interactions among those participants. Figures 8 and 9 show the graphs of ego networks of some randomly sampled participants from the QA and the MG networks.
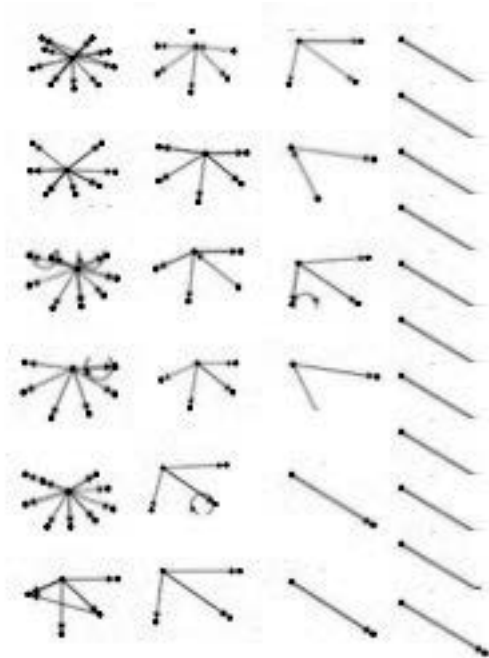

Figure 8. Sampled ego networks of question askers.

As with the sampled ego networks in the Programming category showed by Adamic et al. (2008), the most active participants among questions askers are "answer people" because most of their ties, who are the participants they are helping, are not connected with each other. Similarly, Figure 9 also shows that the most active users are "medal givers" because most of their ties are not connected. In some cases Figure 9 also shows recursive nodes, that is links from nodes to themselves, which means that these participants were both medal givers and medal receivers.
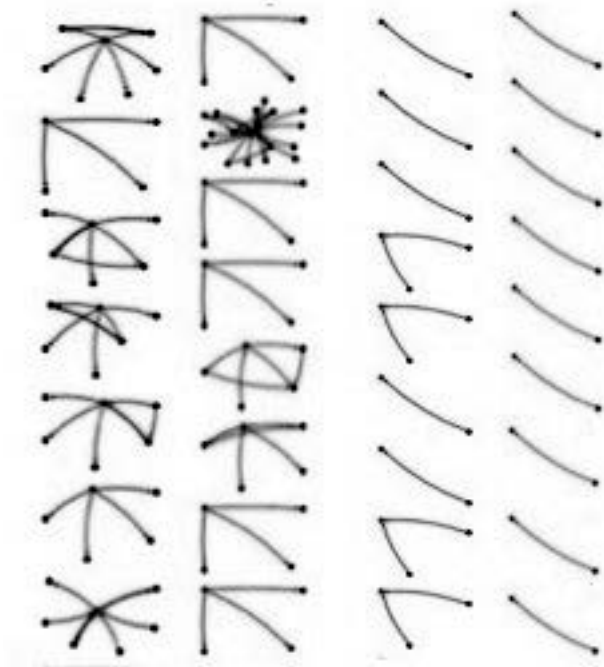

Figure 9. Sampled ego networks of medal givers.

These two graphs suggest casual or opportunistic participation in the group in which members take advantage of all the resources available through the study group and do not interact more regularly with other participants once they have received help from active "answer people" and awarded them medals. These results are consistent with the findings in Table 2, confirming that the study group is not a place where people help each other reciprocally but a help seeking network where askers are more likely to come to look for help from volunteer helpers.

Discussion

The theoretical lens used in this study focuses on the sociomaterial agency of the medals used to reward best responses. Arguably, putting materiality in the foreground allows the bringing into

focus of the values that underlie the peer rating system and an understanding of the influence this system can exert.

Four main claims can be drawn from the results of this study. First, the peer rating system makes visible what the participants find immediately valuable. Unsurprisingly, the results show that the most useful answers were those helping to fix coding errors and solve exercises. Two other large categories of valuable answers included responses where participants applied their own experience to learning Python and gave examples of code, either by linking to external documents or by copying and pasting code into the study group. These results correspond partially to previous research on a web-based reciprocal peer review system by Cho and Schunn (2007), who showed that learners benefit from receiving feedback from others with similar experiences, especially when the review process is scaffolded, anonymous, and reciprocal. In this study group, however, the review process is not always anonymous (although the majority of the study group members provide fictitious names and scarce personal information) and does not imply reciprocity.

By making good responses recognizable, the rating system makes more visible what the participants can gain from each other and what they can achieve by helping each other. Giving medals to best responses may act as a scaffold (Clark, 1997), that is, an external structure that participants can use to learn what makes a good response. This external structure is a constituent of rating the perceived quality of responses and is involved in a distributed cognition. Therefore, it can be argued that the use of medals arranges rating to help participants in their cognitive actions. The medals awarded to good responses act as "tokens of appreciation" and partake of a mechanism aimed at supporting motivation, engagement, and commitment to participation in the study group. It can be questioned, however, whether this form of reward also supports instrumental learning (Greeno, Collins, & Resnick, 1996), given that, in the absence of a teacher, a good response is often one that provides a quick solution, with no apparent obligation to develop deep thinking.

Second, unlike formal education where an accredited intermediary, such as the teacher, controls the definition of what counts as good work, and "assessment confirms that the tutor is in the position of holding specialist and superior knowledge" (Jones, 1999), it can be suggested that in this study group the rating system allocates a form of recognition that extends the "legitimation code" [3], which refers to the credentials that make someone competent and worthy of recognition. The results show that the participants asking questions – or just observing or contributing feedback – decide on their own and award medals accordingly. A medal can be considered a marker of the quality of a response because the technical nature of the subject helps ensure that the content is "verifiable": An answer either solves a coding problem or it does not, and this can be determined quickly. It can be suggested that this more immediate "verifiability", supported by the use of medals, influences the perceived value of responses provided by the participants and contributes to building confidence in what counts as a good response when no credentialed teacher or expert acts as the arbiter of their validity. It can be argued that the more

immediate "verifiability" of responses problematizes the notion of credentialed expertise. In this Q&A site where participants want to learn a technical subject, even responders without legitimate attributes (e.g., being a teacher or an accredited expert) can provide valuable responses. In such circumstances, these people can be seen as experiential experts (Pure et al., 2013) when they competently report about their own experience of fixing coding errors and solving exercises. Their first-hand experience is the basis of their expertise, even though they lack formal credentials.

Third, the results suggest that the peer rating system can provide clues, in the form of medals and titles and score of responders, which can help judge the value of external sources linked or referred to by responders. In this Q&A site every participant can access a plethora of external resources on Python and it is claimed that everyone can provide valuable knowledge in the form of suggestions, experience and coding. The large amount of information to be filtered and evaluated is likely to make models of traditional gatekeeper oversight untenable. However, participants may find it difficult to contextualize and judge the quality of these resources if online source information, for example, author(s), is unavailable or difficult to identify. Furthermore, there are no established criteria for selecting and suggesting resources. When there is this "context deficit" (Pure et al., 2013), the peer rating system can be suggested as cluing participants in to the value and quality of resources for learning a specific topic, especially when responders have first-hand experience of the resources they link or refer to. The peer rating system can also be a strategy to help judge the reliability of a responder over time, and/or the reliability of provided information across multiple sources.

The results suggest circumstances under which participants that are not considered experts in a traditional sense – as they lack special training and credentials – can be in the position to provide valuable information. In this Q & A site, responders are often pseudo-anonymous, and thus reliance on the information they impart can be less significant to determine the value of their responses than their credentials. The peer rating system is thus suggested as a strategy to shift the focus of participants from the status of responders to the information provided by responses. In virtue of this shift of focus, good responses given by the participants who consider themselves beginner programmers have been awarded medals as those provided by the participants with more experience in programming. This evidence can also be related to the peer-based learning approach adopted by the designers of the study group which promotes the participation and engagement of all members of the group over prior accreditation. Promoting participation over prior accreditation can be seen as a way to acknowledge "expertise somewhere within the distributed learning environment and making sure students know how to access and deploy it" (Jenkins et al., 2006, p. 68). The study group is arranged to motivate participants to engage in asking and responding to questions. Askers and responders can be seen as "*distributed* cognitive engines" (Clark, 1997, p. 68). Peer support is about tapping into external resources or experience from the participants whose knowledge may be useful in solving a particular problem. According to this understanding, as Lévy (2000) would argue, everyone knows something, nobody knows everything and what any one person knows is available to the whole group. Valuable information and knowledge can be provided by people with more experience with coding, but also by

beginners to programming. In this distributed cognitive environment medals can be seen as markers of quality that suggest which responses can be trusted. However, it can be argued that medals play this role more effectively depending on how participants give medals. For example, the use of medals does not avoid the potential for abuse or competition, such as creating fake accounts to inflate medal counts or colluding with friends to gain unearned medals. To avoid this abuse, the challenge would be to develop an algorithm that evaluates the medals awarded to participants to compute an overall credibility for each individual.

Finally, the results of SNA indicate that authority over this network is negligible: higher status based on higher SmartScore conveys no ability to shape the participation of others, nor can question posers steer their requests towards higher status participants. Participants self-select on both sides, providing an example of the virtues of open systems in the allocation of limited resources: time and expertise. The findings also suggest that self-selected best contributors do not need to be credentialed authorities on a topic but they can either act as hubs, pointing to authoritative external resources on a topic (Kleinberg, 1999), or they can act as experiential authorities (Pure et al., 2013). As argued by Pure et al., digital technologies are problematizing the notion of expertise because they change how information is disseminated and consumed. Whilst expertise has traditionally been considered the domain of a small group of individuals, now a host of web resources calls into question several indicators which people rely upon to signal expertise, such as credentials and job title. The open study group operates in a culture of linking and borrowing from other sources. Participants can consult directly with a number of external resources, representing different types of expertise ranging from tutorials produced by universities to software repositories put together by communities of programmers. The structure of the web itself – through the use of search engines and hyperlinks – encourages the cross-validation of information across multiple sources in a way that de-emphasizes the notion of expertise as being invested in a single entity and opens up the possibility of value assessment being derived from peer rating.

Whilst a peer rating system holds the promise of being a mechanism to assess the value of responses and to support experiential expertise when traditional markers of expertise are not always clearly defined and identifiable, the risks of relying on such a system should also be considered. For instance, although the peer rating system can provide clues, in the form of medals, and titles and score of responders, which can help judge the value of responses, exclusive reliance on these indicators can be problematic. These clues can lead to responses being judged based on surface elements that are not necessarily indicative of information quality. Another risk may be that of equating the SmartScore with expertise or knowledge. Research found that people often rely on the endorsement of others when evaluating a piece of information, without engaging in independent evaluation (Metzger et al., 2010).

In the absence of credentialed experts, the promise held by the peer rating system in a learning environment relies on learners being able to interpret clues, especially medals, provided by unknown peers and which can be ambiguous. It has been found that information consumers who

have been shown to rely on the average star rating, without considering critical complementary information about the numbers of rating provided, can misinterpret simple aggregated commercial ratings (Flanagin et al., 2011).

Conclusion

In conclusion, from this study it is apparent that a peer rating system holds the potential to be a mechanism to assess the value of responses and support experiential expertise when traditional markers of expertise are not always clearly defined and identifiable. The examined situation seems to be in line with a transition from authority being vested only in credentialed intermediaries to a more bottom-up concept of expertise based on the experiences and contributions of many people. In contemporary social media environments, credentialed expertise can be complemented by other forms of authority, including experiential authority, which can be supported by unique features of social software. A group of people learning a programming language can be one of those situations in which credentialed expertise may be less important to peers than experiential authority. Useful answers originated from a diversity of individuals reporting on their own practicing of Python, even though none may be an expert in the traditional sense. Arguably, social software tools influence the notion of expertise by extending the range of individuals that can supply relevant information. For example, rating systems can make the contributions of uncredentialed individuals as relevant as those of credentialed and trained people (Pure et al., 2013). However, it is challenging to predict how these new forms of evaluation will develop further and what their implications will be. Consonant with the spirit of the 10 ideas for the 21st Century Education (Hampson et al., n.d.), evidence from this study supports the idea that learners informally advise and help each other and that schools should recognize the potential of harnessing these mentoring skills, to help students to work alongside teachers and take on roles that enable them to do things traditionally assigned to teachers, such as assessment.

Given the innovative nature of the study group, both from a pedagogical and technological perspective, it is argued that some important insights emerge from the results of this study. However, the limitations of this research are also recognized. It is an exploratory study and has, therefore, offered insights into the role played by a peer rating system in one particular Q&A site on a technical subject. Further examination is needed, for example, to map out the richness and complexity of giving medals, by studying this behavior from the standpoint of participants. Furthermore, it would be worth examining whether and under which circumstances the rating system harnesses the collective intelligence of a wide range of geographically dispersed and diverse learners, and dispenses with the need for external validation of knowledge by accredited intermediaries.

Acknowledgments

References

Anderson, T., Rourke, L., Garrison, R., & Archer, W. (2001). Assessing teaching presence in a computer conferencing conferencing context. *JALN,* 5(2), 1-17.

Adamic, L., Zhang, J., Bakshy, E, & Ackerman, M. S. (2008). Knowledge sharing and Yahoo Answers: Everyone knows something. *Proceedings of of the 17th International Conference on World Wide Web (WWW'08)* (pp. 665-674). New York: ACM Press.

Aritajati, C., & Narayanan, N. H. (2013).Facilitating students' collaboration and learning in a question and answer system. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work companion* (pp. 101-106). New York: ACM Press.

Barr, J., & Gunawardena, A. (2012). Classroom salon: A tool for social collaboration. *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education (SIGCSE '12)* (pp. 197-202)*.* New York: ACM Press.

Bigham, J.P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R.C., Miller, R., Tatarowicz, A., White, B., White, S., & Yeh, T. (2010). VizWiz: Nearly real-time answers to visual questions. *Proceedings of the 23rd annual ACM symposium on User Interface Software and Technology* (pp. 333-342). New York: ACM Press.

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., & Wiener, J. (2000). Graph structure in the Web. *Computer Networks,* 33(1-6), 309-320.

Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers and Education*, 48(3), 409–426.

Clark, A. (1997). *Being there: Putting brain, body and world together again*. Cambridge, MA: MIT Press.

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., & Dixon, D. (2011). Gamification: Using game-design elements in non-gaming contexts. *Proceedings of CHI' 11 Extended Abstracts on Human Factors in Computing Systems* (pp. 2425-2428). New York: ACM Press.

Flanagin, A.J., Metzger, M.J., Pure, R., & Markov, A. (2011). User-generated ratings and the evaluation of credibility and product quality in ecommerce transactions. *Proceeding of 44th Hawaii International Conference on System Sciences (HICSS)*. DOI: 10.1109/HICSS.2011.474.

Fricker, R. (2008). Sampling methods for web and e-mail surveys. In N. Fielding, R. M. Lee, & G. Blank (Eds.), *The SAGE handbook of online research methods* (pp. 195–217). London: Sage.

Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). In D. C. Berliner, & R. C. Calfee (Eds.), *Cognition and learning. Handbook of educational psychology* (pp. 15 – 46). New York: MacMillan.

Hampson, M., Patton, A., & Shanks, L. (n.d.). *10 Ideas for 21st century education*. Innovation Unit. Retrieved from http://www.innovationunit.org/sites/default/files/10%20Ideas%20for%2021st%20Century%20Education.pdf

Ioana Muntean, C. (2011). Raising engagement in e-learning through gamification. *Proceedings of the 6th International Conference on Virtual Learning ICVL* (pp. 323 – 329). Retrieved from http://icvl.eu/2011/disc/icvl/documente/pdf/met/ICVL_ModelsAndMethodologies_paper42.pdf

Lee, U., Kang, H., Yi, E., Yi, M.Y., & Kantola, J. (2012). Understanding mobile Q&A usage: An exploratory study. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3215 – 3224). New York: ACM Press.

Lévy, P. (2000). *Collective intelligence: Mankind emerging world in cyberspace.* New York: Perseus.

Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education.* San Francisco, CA: Wiley.

Kleinberg (1999). Hubs, authorities, and communities. *ACM Computing Surveys (CSUR) 31*(4), article 5. Doi:10.1145/345966.345982.

Kop, R., Fournier, H., & Fai Mak, J. S. (2011). A pedagogy of abundance or a pedagogy to support human beings? Participant support on massive open online courses. *International Review of Research in Open and Distance Learning, 12*(7), 74-93. Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/1041/2025.

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

Mason, R. (1991). Analysing computer conferring interactions. *Computers in Adult Education and Training* 2(3), 161–173.

Maton, K. (2000). Languages of legitimation: The structuring significance for intellectual fields of strategic knowledge claims. *British Journal of Sociology of Education, 21*(2), 147–167. DOI:10.1080/713655351.

Metzger, M.J., Flanagin, A.J., & Medders, R. (2010. Social and heuristic approaches to credibility evaluation online. *Journal of Communication, 60*(3), 413–439.

Morris, M.R., Teevan, J., & Panovich, K. (2010). What do people ask their social networks, and why? A survey of status message Q&A behavior. *Proceedings of ACM CHI 2010 Conference on Human Factors in Computing Systems* (pp. 1739–1748). New York: ACM Press.

Nichols, J., & Kang, J-H. (2012). Asking questions of targeted strangers on social networks. *Proceedings of ACM CSCW'12 Conference on Computer-Supported Cooperative Work* (pp. 999–1002). New York: ACM Press.

Panovich, K., Miller, R., & Karger, D. (2012). Tie strength in question & answer on social network sites. *Proceedings of ACM CSCW'12 Conference on Computer-Supported Cooperative Work* (pp. 1057–1066). New York: ACM Press.

Paisley, V. (2013). Gamification of tertiary courses: An exploratory study of learning and engagement. In H. Carter, M. Gosper, & J. Hedberg (Eds.), *Proceedings of the 30th ASCILITE Conference* (pp. 671–675). Macquarie University, AU.

Patton, M.Q. (2002). *Qualitative evaluation and research methods.* London: Sage.

Paul, S.A., Hong, L., & Chi, E.H. (2012). Is Twitter a good place for asking questions? A characterization study. *Proceedings of ICWSM 2011* (pp. 1–4). AAAI Press.

Pure, R. A., Markov, A. R., Mangus, J. M., Metzger, M. J., Flanagin, A.J., & Hartsell, E. H. (2013). Understanding and evaluating source expertise in an evolving media environment. In T. Takševa (Ed.), *Social software and the evolution of user expertise: Future trends in knowledge creation and dissemination* (pp. 37–51). Hershey, PA: IGI Global.

Radoff, J. (2011). *Gamification.* http://radoff.com/blog/2011/02/16/gamification. Retrieved March 21, 2013. (Web log message).

Rourke, L., Anderson, T., Garrison, D.R., & Archer, W. (2003). Methodological issues in the content analysis of computer conference transcripts. In D. R. Garrison, & T. Anderson (Eds.), *E-learning in the 21st century* (pp. 129–152). London: Routledge Falmer.

Simões, J., Redondo, R. D., & Vilas, A. F. (2013). A social gamification framework for a K-6 learning platform. *Computers in Human Behavior, 29*(2), 345–353.

Wenger, E., Trayner, B., & de Laat, M. (2011). *Promoting and assessing value creation in communities and networks: A conceptual framework.* Rapport 18. Ruud de Moor Centrum, Open University, The Netherlands.

Jenkins, H., Clinton, K., Purushotma, R., Robison, A.J., & Weigel, M. (2006). *Confronting the challenges of participatory culture: Media education for the 21st century*. Chicago, IL: The MacArthur Foundation.

Jones, C. R. (1999). Taking without consent: Stolen knowledge and the place of abstractions and assessment in situated learning. In C. Hoadley & J. Roschelle (Eds.), *Proceedings of the Computer Support for Collaborative Learning (CSCL) 1999 Conference*, article 35. Mahwah, NJ: Lawrence Erlbaum Associates.

Johnson, B., & Christensen, L. (2012). *Educational research: Quantitative, qualitative and mixed approaches*. Thousand Oaks, CA: Sage.

Weller, M. (2011). A pedagogy of abundance. *Spanish Journal of Pedagogy, 249*, 223–236.

Zhang, J., Ackerman, M. S., & Adamic, L. (2007). Expertise networks in online communities: Structure and algorithms. *Proceedings of the 16th international conference on World Wide Web* (pp. 221-230). New York: ACM Press.