

## Learning with comments: An analysis of comments and community on Stack Overflow

Subhasree Sengupta  
Syracuse University  
[susengup@syr.edu](mailto:susengup@syr.edu)

Caroline Haythornthwaite  
Syracuse University  
[chaythor@syr.edu](mailto:chaythor@syr.edu)

### Abstract

*Stack Overflow (SO) has become a primary source for learning, how to code, with community features supporting asking and answering questions, upvoting to signify approval of content, and comments to extend questions and answers. While past research has considered the value of posts, often based on upvoting, little has examined the role of comments. Beyond value in explaining code, comments may offer new ways of looking at problems, clarifications of questions or answers, and socially supportive community interactions. To understand the role of comments, a content analysis was conducted to evaluate the key purposes of comments. A coding schema of nine comment categories was developed from open coding on a set of 40 posts and used to classify comments in a larger dataset of 2323 comments from 50 threads over a 6-month period. Results provide insight into the way the comments support learning, knowledge development, and the SO community, and the use and usefulness of the comment feature.*

### 1. Introduction

Stack Overflow (SO) is a highly successful question and answer (Q&A) site covering a wide range of topics on computer programming. In May 2017, an answer on SO to the question of how many users there are on SO indicated 829,905 users, of whom 564,682 were registered (i.e., have signed up with SO giving identifiable user information), 265,189 unregistered, and 32 moderators. Current estimates indicate that users on SO have asked 18 million questions, provided 27 million answers, and 74 million comments (June 4, 2019 on <https://data.stackexchange.com/>). This makes SO smaller than general Q&A forums such as Quora (300 million monthly visitors), or Reddit (542 million monthly visitors; 234 unique visitors), but larger and more diverse than other programming forums.

Engagement on SO is an example of ‘learning in the wild’ [20], i.e., an informal and non-formal self-organizing learning environment where crowds of participants ask, answer, comment, correct, argue, and make the effort to present information in informed and accessible ways, while also monitoring content, value, and appropriate behavior. While based on crowd participation, these environments are also communities: epistemic communities, based on a common orientation to a particular knowledge domain [44]; discourse communities, understanding and employing particular language and genre of communication [45]; and communities of practice, with common goals and orientations [34].

These knowledge-focused peer production communities engage in continuously emerging interpretation, clarification, and explanation of knowledge, while maintaining a focus on accuracy, referencing, and the practice of the domain of knowledge. For example, in Reddit AskHistorians, contributions are written to meet a general reader’s understanding and provide references as support for arguments, and for further reading; learners gain understanding of new areas, and also how the study of history is conducted (historiography) [12].

In addressing learning in SO, we follow the ‘learning in the wild’ emphasis on analysis of conversation and interaction and how this supports learning and community. This approach takes, as its starting point the theoretical perspective of social learning [29, 30], and its more recent investigation in online forums as social learning analytics [31]:

“[T]he focus of social learning analytics is on processes in which learners are not solitary, and are not necessarily doing work to be marked, but are engaged in social activity, either interacting directly with others (for example, messaging, friending or following), or using platforms in which their activity traces will be experienced by others (for example, publishing, searching, tagging or rating).

Social Learning Analytics ... draws on the substantial body of work demonstrating that new skills and ideas are not solely individual achievements, but are developed, carried forward, and passed on through interaction and collaboration.” [31, p. 5]

Interaction builds social networks, and as such this research is also predicated on concepts from social network analysis, and its application to learning networks [32, 33]. This perspective looks to interactions among network members analyzing the way, these interactions build social structures that support and sustain the community. This includes structures of norms and rules, social roles and reputation systems, and network outcomes.

In open, online Q&A and learning forums, norms, rules and procedures emerge associated with knowledge and community practice. Roles include *moderators* who manage rules, and *braiders* who weave together threads from others' answers [28]; recognitions include *flairs* granted for merit-based contributions (Reddit). Roles and recognitions emerge from practice: e.g., participants who are known for finding previous answers to current questions have recently been recognized with a "FAQ finder" flair in Reddit AskHistorians [12]. Roles and reputations are built through contributions, and thus they make network connections of different types that accord cohesiveness and support to the community, like flairs in Reddit, and reputation points in SO. Roles, recognitions, and reputation systems provide structure in the network that forms the character and practice of a particular community of practice [16, 34].

An important network outcome in learning sites is trust in the knowledge gained through the site, and that in turn can be based on trust in the knowledge exchange practices. In SO, as in other such communities, any knowledge hierarchy that exists is created and recognized from within, and trust is built through a recognition system. The *reputation* built through SO activity leads to *privileges* in the community. In SO, reputation is recognized by accumulating points based on others' acknowledgement of the value of a question or answer (points do not accrue for comments). Points are awarded according to the number of upvotes on a question (5 points) or answer (10 points; downvotes subtract points), an answer 'accepted' as the best solution (15 points), and answering a question with an associated bounty (various amount of points; reputation points placed on a question as a way to elicit answers; (<https://stackoverflow.com/help/whatsreputation>)).

The privileges lead to increased access to the workings of SO. This allows participants to strengthen their commitment to the community through more and different kinds of contributions (in social network terms, this increases their relational multiplexity, which is an indicator of tie strength). The privilege, most relevant to the current study is, being allowed to post comments; a privilege provided only to those with 50 reputation points, and thus is available only to those demonstrating some competence and knowledge about coding and

community practice. Privileges also include permission to post to SO chat, add to the community wiki, vote posts up or down, flag posts for moderator evaluation, be named to a site status, and access to moderator and analytic tools (<https://stackoverflow.com/help/privileges>). The result is a self-organizing system where recognition of reputation opens doors to further activities that both support the community and increase the reputation of the individual.

## 2. Research Questions

As the importance of these open, online, knowledge-exchange initiatives increase, a number of questions arise about how conversational practices sustain participation, valued knowledge exchange, and community commitment. Here we examine these practices in relation to SO, and specifically regarding SO comments.

Three crucial conversational features define practices in SO: the *question* around which the discussion is centered; the *answer* or answers; and the *comments* on both questions and answers. Comments in SO serve as "temporary post-it notes on questions and answers". By design, comments can only be posted below questions and answers, and thus, comments are always associated with the discussion around the question or answer(s) in a post. As noted, posting comments is a privilege, open only to SO users with some reputation gained through SO participation (<https://stackoverflow.com/help/privileges/comment>). Comments clarify and enrich the content conveyed through questions and answers. Examining comments is particularly relevant when considering SO as a learning site because comments go beyond the question or answer to show the process of learning and knowledge construction.

Prior scholarship on SO has analyzed questions and answers extensively, and how these enrich discussion quality. However, the contribution of comments in SO discussions has not been thoroughly analyzed. This study provides initial insight into the typology of comments in SO, and the value these brings to the SO community. The overall research question for this study is:

- How do comments support learning and community in SO?

To evaluate the nature of commenting on SO, a content analysis was carried out to identify the types of comments in SO, and then applied to a larger dataset of comments to address the questions:

- What are the main categories or types of comments observed in SO threads?

- How are comments used in relation to SO questions, and SO answers?

This provides the opportunity to examine how comments contribute to discussion threads on SO, and how commenting practices contribute to learning and community on SO. Results also help to investigate whether comments are a necessary design feature for SO, and thus a potentially useful feature for other, similar sites. More widely, this study adds to our understanding of online learning and community practices in open, online forums.

### 3. Background Research

A number of studies on groups, virtual communities, and online learning sites, provide insight into the workings of these open knowledge-sharing initiatives. These studies show that online groups function much like offline groups, learning and co-constructing the norms, roles and processes that support the continued maintenance of the group. The dual function of achieving learning goals, and of maintaining a functioning community underpin the analysis conducted here. SO, as a successful site, can be expected to show characteristics that support production – of quality questions and answers – as well as mechanisms for the well-being of the community. The following addresses relevant research relating to achieving learning and community goals.

**3.1.1. Learning Goals.** Research on groups (e.g., [22, 23, 25]) has provided much insight into processes that support work outcomes and has been used to inform research on many forms of online initiatives (e.g., [10, 14] on online learning). For sites creating a product, such as an online encyclopedia (e.g., Wikipedia), open access maps (e.g., OpenStreetMap), or online book reviews (e.g., LibraryThing, Goodreads), the goal is easily identified in association with the product. However, it is less clear what the ‘product’ is in a Q&A learning and knowledge-exchange environment, and hence what the specific goal is for the site; yet, the orientation of the group to an outcome still remains, i.e., the outcome of providing valuable answers to participant questions. In studies of motivations to contribute to open, online initiatives, orientation to the overall purpose of the site, to create open access maps, to learn and disseminate knowledge about history [8,12], appears as a strong motivator for participation. This is particularly so for participants who are not strongly tied into the community and thus for whom the connection to purpose is the major and perhaps only motivator for engagement [8,15,16].

Knowledge production in such a site depends on making arguments in the right form for the forum and for the subject matter under discussion. One well-known model for learning discussions is the five phase Interaction Analysis Model [36]: sharing or comparing information; discovery and exploration of dissonance; negotiation of meaning; testing and modification of proposed synthesis; and agreement statements or applications of newly constructed meaning. Another take on knowledge building addresses three kinds of talk in exploratory dialogue [37]: *Disputational*, “characterised by disagreement and individualised decision making”; *Cumulative*, “in which speakers build positively but uncritically on what the others have said”; and *Exploratory*, “in which partners engage critically but constructively with each other's ideas” ([37] p. 146).

Studies form a social learning analytics perspective have built on [37] to explore online interaction as a step toward automated analysis. For example, one study developed codes for interaction in MOOCs with categories of challenge, evaluation, extension and reasoning [35]; Another [18, 20] coded more generally for interaction and argumentation in Q&A posts in four ‘Ask’ subreddits, identifying eight categories of posts: Explanation (with disagreement, with agreement, and with neutral presentation); Socializing (with negative or positive intent); Providing References; Information Seeking; and Community Rules and Norms.

While beyond the scope of this paper to explore coding schemas and argumentation in depth, the work here builds on this past work in considering the kinds of practices that support knowledge development in online learning conversations, and SO comments.

**3.1.2. Community Goals.** As online communities, conversations in support of community practice are as important as those around the topic of the site, and vital for building a successful *community of practice* (CoP; Wenger, 1998). Functioning CoPs establish ways of bringing new participants on board, adhering to a community practice, and maintaining a focus on the purpose of the community. Maintaining a CoP includes engaging lurkers, novices and experts, and supporting transitions across these roles as new participants learn to be members of the community. Lurking can be a stage of community entry, observing how the community operates by engaging in *legitimate peripheral participation* [27]. New participants observe how conversations happen online, how rules of the community are defined, and the way these are policed and transgressors sanctioned [26], while also learning how to become visible in the community [6,7].

Recognition and reward systems, such as the SO reputation system described above, provide a way for communities to distinguish among the many functions

within the community, and for individuals to recognize what is considered important in this community. As participants become more engaged, they move through stages of joining, maintaining presence, and eventually disengaging from the online community [14]. Disengaging can mean leaving a role, e.g., as a moderator may step down from that role (particularly in an intense environment such as Reddit; [12]), or an individual may leave the community altogether due to a change in interest, career or life stage [17, 38].

SO as a community can be expected to demonstrate interactions serving to support community with attention to the kind of learning and knowledge exchange important for the site (relating to the epistemic community, [44]), modeling and using appropriate language and genre [45], and creating and modeling community practices [34].

### 3.2. Understanding the value of posts in SO

Prior work on engagement in SO have addressed two aspects: (1) the types of users of SO, concentrating of frequency and length of engagement with the site, and (2) what makes a SO thread popular or valued. In relation to this study, the types of users are relevant for understanding the roles of teacher and learner, and the nature of audience in terms of expertise, but less relevant for the current study with its focus on comments. Thus, the literature discussed here relates to value of posts (for more on users in SO, see [11, 21]).

Methods of evaluating value in posts have included: evaluating models to predict the long-lasting value of a post, using a combination of features related to the activity on the post, temporal dynamics observed, and the reputation of community members associated with the post [2]; examining answers for presentation quality, affect, and temporal features such as average response time, number of follow-up comments and time elapsed before the first answer was posted [9]; developing an automated approach using LDA to find the popular topics in posts [5]; and examining unanswered questions [4].

A core part of SO discussions is the computer code, and code is used both in questions and answers. Code provides an excellent, community-relevant base for launching and addressing questions. Nasehi et al [24] suggest that code written with good descriptions provides the most value in SO discussions. Their study identified the main types of questions and answers as follows: Questions fit four main types: (1) Debug/corrective: dealing with problems in the code under development; (2) Need to know: questions regarding possibility or availability of (doing) something; (3) How to do something: questions regarding how to implement something; (4) Different solution: questioner has

working code yet is seeking a different approach. Valued answers (upvoted) fit eight types: (1) Providing concise code; (2) Expanding the code in the questions; (3) Discussing the code or software limitations; (4) Providing code with detailed steps; (5) Highlighting essential steps in previous answers; (6) Discussing alternate possible solutions; (7) Linking to extra resources; and (8) Expanding discussions in the comments.

This stream of work provides insight into the kinds of knowledge exchange that supports the SO community. In continuing to complete the picture of SO interaction, it is then worthwhile to see how comments add value in relation to questions and answers.

## 4. Method

In keeping with the orientation to conversation and interaction, underpinned by social learning theory, and a social network perspective, comments were examined for the kinds of information they convey, and how this supports knowledge exchange practices and community. Using a content analysis approach [39, 40], categories were derived through an open coding process. This approach was considered most applicable given the lack of previous analysis of comments but was informed by previous work on open online exchanges. The process led to discovery and then classification of comments according to the information conveyed, its format, and its presentation (including affect). As described below, nine categories of exchange were identified and then applied to a larger data set.

### 4.1. Category Derivation

To derive categories of comments, two rounds of coding were conducted. A set of 40 posts with a total of 990 comments were examined, selected randomly from all SO posts from October to November 2018, and retrieved using the python wrapper for the Stack Overflow API. The term *post* here refers to the full Q&A thread from first creation to last contribution, including the question, answer, and all comments associated with the thread.

The data was coded qualitatively using the guidelines in [43] by one of the authors. First, 20 posts were selected randomly from the set of those that passed a threshold criterion of having at least one comment, and the content categories were developed based on this first iteration of content category generation. In the second iteration, the content categories were validated across another set of 20 posts selected by the same threshold criterion. As a final check, the content categories were validated to ensure that when considered in aggregate

across all the posts, no new categories emerged, and reapplied across all 40 posts. The resulting nine categories were then used for coding the larger set of posts as described below.

## 4.2. Data Collection and Sampling

The google cloud-based database was used for data collection (<https://cloud.google.com/bigquery/>). This database has an archive of SO posts, complete to September 2016 (at the point of data collection). For each year, the data is stored in four datasets, each covering three months (January to March, April to June, July to September, October to December). The metadata available for each dataset includes question id, question title, creation date, count of answers, count of comments, edit history, owner of the post, number of views, and score (the difference between the number of upvotes and downvotes on the post).

Since the central aim of this study is to understand the value of SO comments, posts selected for the study had to have at least five comments, and at least one answer.

Posts from the two most recent datasets available were used for this study: March to June 2016, and July to September 2016. As the entire available dataset exceeds the capacity that can be downloaded from the database server, this study examined a random sample of 50 posts collected from this 6-month time frame; these 50 posts had 2323 associated comments that constitute the dataset for the study. Table 1 provides descriptive statistics about the sample. Even though posts were collected from 2016, some were created earlier (the oldest was created in 2011), which can happen if users are still commenting or discussing the answers provided in the post.

Table 1: Descriptive statistics for the 50 posts

Oldest post creation date	03/15/2011
Latest post creation date	05/01/2016
Mean view count	216.12
Mean vote score = (upvote - downvote)	1.9
Mean no. of comments	46.16
Mean no. of answers	1.44
Total no. of comments	2323

## 4.3. Procedures

The nine comment categories developed in the classification development phase were applied to the 50 posts and associated 2323 comments in the larger study. Two coders each applied the categories to comments

associated with 25 posts; posts were randomly distributed between the two coders.

To test agreement between the two coders, a random sample of 10 posts (not included in the set of final 50 posts that were analyzed) were coded by both coders. The Krippendorff's alpha was found to be 0.762, establishing an acceptable amount of agreement between the two coders. Any differences in interpretation of the comment categories were discussed between the coders and resolved by arriving at a definition for each category that was mutually agreeable before the final set of 50 posts were analyzed.

Further, while coding their respective set of posts, if the first coder found a comment which they felt could not be classified using the derived categories, the second coder tried to classify the comment using the existing coding schema and discussed it with the first coder. If the second coder found an existing category could be applied, and if both coders agreed after discussion then no new comment category was added to the schema. In case both coders could not classify a commenter were not able to come a point of agreement, then a new category for comments would have been added. However, the coders did not find any such exceptions. Thus, the entire dataset could be classified using the coding schema as derived in the initial phase.

## 5. Results

All 2323 comments were classified into one of the nine categories, with 37% of the comments given in response to questions, and 63% in response to answers. This reflects the way most discussion in SO posts is around clarifications and modifications for answers. Table 2 presents a summary of the codes and their prevalence, and each code is described in more detail below with examples of comments assigned the code.

The picture of commenting that emerges provides indicators to the elements of code learning, including the kind of clarification needed for learning code, the right way to frame questions and provide answers, and nuances of giving improvements, alternatives and limitations. The nine categories of interaction apply to both learning code and to learning the norms of discourse in this environment, and thus support the practice of this community.

### 5.1 Comment Categories

**5.1.1. Comments with improvements (29%).** The largest percentage of comments address improving the way a question or answer is formulated; 35% given in response to questions, and 65% to answers. These comments help to refine unclear terms in questions or

shortcomings of answers, e.g., with text such as “what do you mean by ...”, or “could you explain what you mean ...”. An example is: “*What exactly is your question? You mean, when you replaced <button> tag with <a>, it's not working on click.*” This indicates that the question is not framed properly while also giving a clarification. Such a comment engages with the poster by creating a dialogue and models the way to ‘talk’ in this community, showing both that the question needs to be more specific, and how to be specific.

Table 2: Distribution of comment categories

Comments offering:	Percent of all comments	Proportion relating to Questions: Answers
Improvements	29	35 : 65
Code and explanation	14	52 : 48
Only explanation	11	38 : 62
Only code	10	61 : 39
Affect	10	10 : 90
Alternatives	9	26 : 74
Limitations	9	29 : 71
References/Links	6	59 : 41
Moderator comments	2	20 : 80

**5.1.2. Comments with code and explanation (14%).**

These comments include a bit of explanation along with a code example; 52% of such comments are given to questions and 48% to answers. An example is: “*No reason to extract the DOM node in the snippet used in the example, and then have to look it back up again. Better to just do: var \$button = \$('#btnUpdate'); And then in the If expressions just use \$button instead of \$(button). Has the advantage of caching the jQuery object.*” The code snippet provides an answer, but the comment goes further to explain the benefit of the change. This reflects an apprenticeship or peer-to-peer model of learning, with a more knowledgeable user providing information to another. It also demonstrates the site culture of taking time to explain rather than ‘fix’ the problem by presenting the answer. Moreover, the use of code fits with the purpose of the site, and thus presents answers in the right language for readers.

**5.1.3. Comments with only explanation (11%).**

These comments give a short description of the logic to use to write the associated code, but without giving code examples; 38% are given to questions and 62% to answers. An example given to an answer is: “*Form elements with type = “hidden” is just one case that can trigger: hidden. Elements with no height and width, elements with display = “none”, and elements with hidden ancestors will also qualify as :hidden*”. This

again reflects an apprenticeship or peer-to-peer model of learning, explaining how a concept works.

**5.1.4. Comments with only code (10%).**

Some comments give code snippets only, usually a single line or a specific keyword; 61% of such comments were given in response to questions and 39% to answers. These are common when the question is asking for a precise code-related answer, and also for refining specific elements in answers, or correcting answers with code snippets. An example for this category is: “*Just try potato.include(“to”);*”, which suggests a short fix to a code snippet in an answer. Again, this fits with the language and genre of the site, reflecting community norms.

**5.1.5. Comments expressing affect (10%).**

These are comments expressing some standalone form of emotion related to the content in questions, answers or other comments; 10% of such comments were given to questions and 90% of such comments were given to answers. A common example is expressing gratitude to someone who answered the question or clarified an answer using a comment. Some evidence was found for both positive and negative sentiment. Positive sentiment was usually expressed by community members to thank those that helped them resolve their problems or to express their happiness about resolving something in a post. An example comment that expressed positive sentiment is: “*I thank you for taking the time to assist me*”. Negative sentiment was rare, and usually expressed in the form of frustration, anxiousness or sadness by community members when they did not get a satisfactory response to their question. An example of a comment expressing negative sentiment (frustration) is: “*please any one help! I am getting frustrated. I have tried this for the last 4 days!*”. In coding, positive and negative sentiment were not separated since most comments with affect were those expressing gratitude to other community members for their help and very few instances of negative sentiment were found. As sentiment or emotional expression can be quite nuanced, exploring this was not a goal of this research. However, future work will carry out further inspection of sentiment and affect on SO. Overall, the positive sentiment expressed can be expected to support the community, giving recognition to commenter, an aspect not otherwise rewarded in the point system.

**5.1.6. Comments giving alternatives (9%).**

These comments broaden the scope of questions and answers, by describing situations where the code logic might need to be extended or modified, or broadening the context of the question to make it more generalizable;

26% of such comments were given in response to questions and 74% to answers. An example from this category is; *“Have you changed the selector and tried like \$(“input.btnEliminar”). Sometimes, simple class selector doesn't work, we need to be more precise.”* The comment explains the shortcoming of a code snippet, i.e., that it would not work in certain web browsers, and suggests a suitable replacement so that the code can be used across multiple web domains. This exchange appears to build knowledge in a co-constructive way [41], with experts creating new knowledge that supports common understanding.

**5.1.7. Comments mentioning limitations (9%):** These comments express the limitations of answers, giving examples of where these would fail, sometimes with examples; 29% of such comments were given to questions and 71% to answers. An example comment of this category is: *“This doesn't work in the latest jquery. It's been deprecated.”* This alerts users to modifications (upgrades) made to the way jquery (a web programming language) works, while expressing a limitation of an answer. As for providing alternatives, this exchange adds to site knowledge.

**5.1.8. Comments with reference or links (6%):** These comments include links to other SO posts or refer to documentation for software or code; 59% are given to questions and 41% are given to answers. An example is: *“This is related although to a different question”*, providing the URL to another SO post which tackles similar issues mentioned in the post. This kind of comment demonstrates the internal knowledge held by members of the site, and the commitment this demonstrates in having members who have been present for sufficient time, and with sufficient attention to site content to know past questions and answers. This is the social capital held within the network, i.e., the knowledge held by members of the site that can be accessed and mobilized as needed [42].

**5.1.9. Moderator comments (2%):** These infrequent comments, usually posted by moderators, include information related to site management. This can mean arbitrating whether a question is relevant, marking a question as not relevant or as a repetition, or closing a discussion thread as not relevant to the platform; 20% of such comments were given to questions, and 80% to answers. An example is: *“It does answer the question's title. See the last paragraph. Also, feel free to edit the title to reflect the question.”* This indicates the answer is relevant, while also offering a way to conform to site norms by using a more relevant title. This example highlights the role that moderators play in controlling

the discussions and determining the relevance of content on SO. As such, they enforce local norms (of on-topic, non-repetitive discussion), while also providing expert guidance on how to follow local norms. This demonstrates the way informal learning about norms can happen, both through the direct comment to the poster, and to the visibility of this exchange to others.

## 6. Discussion

Applying the derived coding schema of nine comment categories to comments on a sample of 50 Q&A posts, provides insight into the role of comments in SO. Results show the ways in which comments engage community members in knowledge dissemination and co-construction when suggesting improvements, alternatives and or limitations, supporting community processes when expressing affect, moderating, and modeling local norms; supporting network outcomes of shared knowledge when offering comments; and demonstrating the social capital held within the network when providing recognition to others and referring to in-network resources. Table 3 summarizes the learning and community support provided by each category of comment.

This study finds commenting is applied to both questions and answers, but in different proportions. Comments offering improvements, explanation only, alternatives, limitations, and affect, as well as moderator comments, are tipped more to answers than questions, reflecting the site focus on answering questions. Only two categories, only code, and references or links, are tipped toward questions. Although further research is needed to evaluate usage, this is likely associated with questions that have a very quick solution or have been solved in other SO discussions and thus do not require a broader discussion. Thus, this also helps demonstrate knowledge management practices in SO: where questions have a simple or previous answer, the community does not need to spend time answering already answered questions; but, providing references and links depends on individuals being sufficiently familiar and altruistic to take on the referencing role. As for Reddit, the FAQ finder shows as an important role in managing question effort. It allows users to spend effort where questions are new, and elaborations are important for exploring and determining the best answer to the question.

One aspect of this study was to consider whether comments are a useful design feature for such learning sites. Given the results here, it appears that comments provide a strong supportive mechanism for understanding and expanding of questions and answers

and engaging in joint knowledge construction. Although more work is needed to assess this, particularly to compare with sites without commenting, our tentative conclusion is that comments are a useful feature, providing support for learning and community.

Table 3: Support for Learning and Community

Comments offer:	Learning and Community
Improvements	<ul style="list-style-type: none"> <li>Models community discourse</li> </ul>
Code and explanation	<ul style="list-style-type: none"> <li>Reflects apprenticeship and peer-to-peer learning</li> <li>Demonstrates site culture</li> <li>Models community discourse</li> </ul>
Only explanation	<ul style="list-style-type: none"> <li>Reflects apprenticeship and peer-to-peer learning</li> </ul>
Only code	<ul style="list-style-type: none"> <li>Models community discourse</li> </ul>
Affect	<ul style="list-style-type: none"> <li>Provides recognition, reputation and reward</li> <li>Provides community social support</li> </ul>
Alternatives	<ul style="list-style-type: none"> <li>Demonstrates co-construction of new knowledge</li> </ul>
Limitations	<ul style="list-style-type: none"> <li>Demonstrates co-construction of new knowledge</li> </ul>
References/ Links	<ul style="list-style-type: none"> <li>Demonstrates the internal knowledge structures of the site</li> <li>Demonstrates social capital held within the network</li> <li>Demonstrates emergent role of “FAQ finder” or “braider”</li> </ul>
Moderator comments	<ul style="list-style-type: none"> <li>Models and enforce local norms</li> <li>Demonstrates informal learning about norms</li> <li>Demonstrates community-defined role of Moderator</li> </ul>

## 7. Future Work

Our analysis provides insights into the way comments add value by supporting learning, knowledge dissemination and co-construction. Future work aims to examine SO interaction further to validate the nine categories of comments coded, by repeating the content analysis on a larger corpus of comments, with iterative coding rounds and multiple coders to achieve a good reliability score [18]. Part of that work could explore aspects of affect to better understand the different types of emotions conveyed in SO. A second line of future work will connect the categories of questions and answers from other research [24] with the categories of

comments we propose, and explore further how categories of questions, answers and comments align. Finally, future work will explore types of users (by reputation, skill level) in relation to comment use, and connect with the work of other researchers [11, 21].

## 8. Conclusion

In conclusion, the content analysis has revealed nine categories of comments that demonstrate support for learning, knowledge, and community. The categories and their use, suggest comments and carry valuable information that helps to improve the quality of discussions on SO. In particular, comments were found to support learning by offering explanation and code with explanation; support knowledge co-construction by providing improvements, alternatives, and noting limitations; support community through modeling language and genre; support community and knowledge management by referencing to in-network answers, keeping questions on-topic and non-redundant through moderator oversight, and using code-only shorthands for questions with quick answer. Further, the research highlights the importance of the unnamed community role of ‘past answer finder’ (‘FAQ finder’ in Reddit terms). This role appears to help streamline answering and allow others to put effort to addressing new questions and answers.

Thus, we find that comments are of significant value in these discussions, and that comments are a useful component of the SO community, and potentially valuable feature for other sites. These findings lay the groundwork for larger, more extensive study to validate further these content categories, compare to coding by other researchers [24, 2], and connect to the categories of users as derived in both other research works [21, 11]. Future work and further extensive analysis will also help to provide insights about how comments can support SO users, moderators, and community with more effective information curation and use of the SO platform, and how the comment features may add intrinsic value to other Q&A and knowledge exchange initiatives.

## 9. References

- [1] J. Ahn, and I. Erickson, “Revealing mutually constitutive ties between the information and learning sciences,” *The Information Society*, vol. 32, no. 2, pp. 81–84, 2016.
- [2] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, “Discovering value from community activity on focused question answering sites,” in *Proceedings of the 18th ACM SIGKDD International conference on Knowledge discovery and data mining*, 2012.

- [3] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, "Steering user behavior with badges," in *Proceedings of 22nd international conference on World Wide Web*, 2013.
- [4] M. Asaduzzaman, A. S. Mashiyat, C. K. Roy, and K. A. Schneider, "Answering questions about unanswered questions of stack overflow," in *Proceedings of 10th Working Conf. on Mining Software Repositories*, pp. 97-100, IEEE press, 2013.
- [5] A. Barua, S.W. Thomas, and A. E. Hassan, "What are developers talking about? an analysis of topics and trends in stack overflow," *Empirical Software Engineering*, vol. 19, no. 3, pp. 619-654, 2014.
- [6] A. Bregman, and C. Haythornthwaite, "Radicals of presentation in persistent conversation", in *Proceedings of the 34<sup>th</sup> Hawaii International Conf. on System Sciences*, 2001.
- [7] A. Bregman, and C. Haythornthwaite, "Radicals of presentation: Visibility, relation, and copresence in persistent conversation", *New Media and Society*, vol. 5, no. 1, pp. 117-140, 2003.
- [8] N.R. Budhathoki, and C. Haythornthwaite, "Motivation for open collaboration: Crowd and community models and the case of OpenStreetMap", *American Behavioral Scientist*, vol. 57, no. 5, pp. 548-575, 2013.
- [9] F. Calefato, F. Lanubile, M.C. Marasciulo, and N. Novielli, "Mining successful answers in stack overflow", in *Proceedings of the 12th Working Conference on Mining Software Repositories*, pp. 430-433, IEEE Press, 2015.
- [10] A. Forte, N. Kittur, V. Larco, V.H. Zhu, A. Bruckman, and R.E. Kraut, "Coordination and beyond: social functions of groups in open content production", in *Proceedings of the ACM 2012 conference on CSCW*, pp. 417-426, ACM, 2012.
- [11] A. Furtado, N. Oliveira, and N. Andrade, "A case study of contributor behavior in Q&A site and tags: the importance of prominent profiles in community productivity", *Journal of the Brazilian Computer Society*, vol 20, no. 1, p. 5, 2014.
- [12] S. Gilbert, "Motivations for participating in online initiatives: exploring motivations across initiative types", Unpublished doctoral dissertation, University of British Columbia, 2018. Available online at <https://open.library.ubc.ca/cIRcle/collections/ubctheses/24/items/1.0372890>
- [13] B. Hartmann, D. MacDougall, J. Brandt, and S.R. Klemmer, "What would other programmers do: suggesting solutions to error messages", in *Proceedings of SIGCHI Conf. Human Factors in Computing Systems*, pp. 1019-1028, ACM, 2010.
- [14] C. Haythornthwaite, "Facilitating collaboration in online learning." *Journal of Asynchronous Learning Networks*, vol. 10, no. 1, pp. 7-24, 2006.
- [15] C. Haythornthwaite, "Online knowledge crowds and communities", in *J. Echeverria, A. Alonso & P.J. Oiarzabal (Eds.), Knowledge Communities*, pp. 193-209, 2011.
- [16] C. Haythornthwaite, "Crowds and communities: Light and heavyweight models of peer production", in *Proceedings of 42<sup>nd</sup> Hawaii International. Conference on System Sciences*, 2009.
- [17] C. Haythornthwaite, M.M. Kazmer, J. Robins and S. Shoemaker, "Community development among distance learners: Temporal and technological dimensions", *Journal of Computer-Mediated Communication*, vol 6, no. 1, p. JCMC615, 2000.
- [18] C. Haythornthwaite, P. Kumar, A. Gruzd, S. Gilbert, M. Esteve del Valle, and D. Paulin, "Learning in the wild: coding for learning and practice on Reddit", *Learning, Media and Technology*, vol. 43, no. 3, pp. 219-235, 2018.
- [19] M.M. Kazmer, "Beyond C U L8R: disengaging from online social worlds", *New Media & Society*, vol. 9, no. 1, pp. 111-138, 2007.
- [20] P. Kumar, A. Gruzd, C. Haythornthwaite, S. Gilbert, M. Esteve del Valle, and D. Paulin, "Learning in the wild: coding Reddit for learning and practice," in *Proceedings of 51<sup>st</sup> Hawaii International Conference on System Sciences*, 2018.
- [21] L. Mamykina, B. Manoim, M. Mittal, G. Hripcsak, and B. Hartmann, "Design lessons from the fastest Q&A site in the west", in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 2857-2866, ACM, 2011.
- [22] J.E. McGrath, *Groups, interaction and performance*, Prentice-Hall, Englewood Cliffs, NJ, 1984.
- [23] J.E. McGrath, and A.B. Hollingshead, *Groups interacting with technology*, Sage, Beverly Hills, CA, 1994.
- [24] S.M. Naschi, J. Sillito, F. Maurer, and C. Burns, "What makes a good code example? A study of programming Q&A in Stack Overflow", in *28th IEEE International Conference on Software Maintenance (ICSM)*, pp. 25-34, IEEE, 2012.
- [25] M.S. Poole, and A.B. Hollingshead, *Theories of small groups: interdisciplinary perspectives*. Sage Publications, 2004.
- [26] C.B. Smith, M.L. McLaughlin, and K.K. Osborne, "Conduct control on Usenet", *Journal of Computer-Mediated communication*, vol. 2, no. 4, p. JCMC2410, 1997.
- [27] J. Lave & E. Wenger, *Situated learning: Legitimate peripheral participation*. Cambridge University Press, 1991.
- [28] C. Preston, "Braided learning: An emerging process observed in e-communities of practice." *International Journal of Web Based Communities*, vol. 4, no. 2, pp. 220-243, 2008.

- [29] A. Bandura, *Social learning theory*, Prentice-Hall Englewood Cliffs, NJ, 1977.
- [30] E. Wenger. "Communities of practice and social learning systems: the career of a concept", in *Social Learning Systems and Communities of Practice*, pp. 179-198, Springer, 2010.
- [31] S. Buckingham Shum and R. Ferguson, "Social Learning Analytics", *Journal of educational technology & society*, vol. 15, no. 3, pp. 3-26, 2012.
- [32] C. Haythornthwaite, M. de Laat and B. Schreurs, "A social network analytic perspective on e-learning", *Handbook of E-Learning Research*, pp. 251-269, Sage, 2016.
- [33] C. Haythornthwaite, "Learning, connectivity and networks", *Information and Learning Sciences*, vol. 120, no. 1, pp. 19-38, 2019.
- [34] E. Wenger, "Communities of practice: Learning, meaning, and identity," Cambridge Univ. Press, Cambridge, UK, 1998.
- [35] R. Ferguson, Z. Wei, Y. He and S. Buckingham Shum, "An evaluation of learning analytics to identify exploratory dialogue in online discussions", in *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, pp. 85-93, ACM Press, NY, 2013.
- [36] G.N. Gunawardena, C. Lowe, and T. Anderson, "Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing", *Journal of Educational Computing Research*, vol. 17, no. 4, pp. 397-431, 1997.
- [37] N. Mercer, "Sociocultural discourse analysis: analysing classroom talk as a social mode of thinking", *Journal of Applied Linguistics*, vol. 1, no. 2, pp. 137-168, 2004.
- [38] M.M. Kazmer, "The process of disengaging from online learning community revealed through examination of threaded discussions", *International Journal of Web-Based Communities*, vol. 8, no. 4, pp. 521-536, 2012.
- [39] K. Krippendorff, *Content Analysis; An introduction to its methodology*, Sage publications, 2018.
- [40] H-F. Hsieh and S.E. Shannon, "Three Approaches to Qualitative Content Analysis", *Qualitative Health Research*, vol. 15, no. 9, pp. 1277-1288, 2005.
- [41] M. Scardamalia and C. Bereiter, "Knowledge building: Theory, pedagogy and technology", *The Cambridge Handbook of the Learning sciences*, pp. 97-118, Cambridge University Press, NY, 2006.
- [42] N. Lin, *Social capital: A theory of social structure and action*, Vol. 19, Cambridge University Press, Cambridge, UK, 2001.
- [43] R. Boyatzis, *Transforming qualitative information: Thematic analysis and code development*, Sage, 1998.
- [44] K.K. Cetina, *Epistemic Cultures: How the sciences make knowledge*. Harvard University Press, 2009.
- [45] C. Miller, "Rhetorical community: The cultural basis of genre", *Genre and the New Rhetoric*, p. 57, 2003.