Lurkers in Online Communities

STS Research Paper
Presented to the Faculty of the
School of Engineering and Applied Science
University of Virginia

By

Grace Wu

May 1, 2020

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Signed: _____

Approved: _____ Date _____
Rider Foley, Department of Engineering and Society

Introduction

Online communities, defined as "any virtual social space where people can come together to get and give information or support, to learn or to find company", accommodate a wide demographic of users and serve a variety of purposes (Preece, 2001, p. 348). These communities can manifest in many forms including listservs, forums, instant messaging websites, and social networks. Online communities are formed and shaped by their particular interests which, in turn, affects how the user interface and the website itself are built to fulfill that purpose (Preece, Nonnecke, Andrew, 2004).

Research on online communities and user interfaces from this research will be implemented in the technical project, Social Networks and Archival Context (SNAC). SNAC is a free, online community that helps users discover information about people, families, and organizations that are documented in historical resources (primary source documents) and their connections to one another. SNAC, like the more famous website Wikipedia, is an international cooperative that includes (but is not limited to) archives, libraries, and museums. SNAC seeks to build a collection of reliable descriptions of people, families, and organizations that link to and provide a contextual understanding of historical records. The technical project focuses on creating an OpenRefine plugin to help compare and share data between SNAC and other archival organizations. OpenRefine is a Java-based tool that allows a user to upload data and then analyze, clean, reconcile, and augment it through an online interface (About SNAC, 2010).

Any online community's success is partly dependent on its levels of user participation. A significant reason why users are unable to participate in online communities is due to the usability of the software (Preece, 2001). The primarily focus of this thesis is to examine ways to encourage user participation. This will be done through two separate frameworks: Preece's

framework for usability and sociability as well as Habermas' theory of communicative action. Preece (2001) offers a framework for sociability and usability through improvements in user interface design while Habermas' theory suggests measuring the success of an online community through the meaningfulness of posts and discussion rather than the number of active users. These two frameworks will support the development team's ability to learn and design a user interface for the OpenRefine plugin that encourages more user participation and makes it easier for users to actively reconcile and import data into SNAC. This thesis will analyze online communities and focus on methods to improve user interfaces that will encourage increased, more meaningful participation.

Lurkers in Online Communities

Oftentimes, the majority of users do not participate in online discussions. In a study done by Van Mierlo (2014), four digital health social networks (DHSN) were monitored and 578,349 DHSN posts were observed for up to eleven years. Statistics revealed that less than 25% of the actors in each DHSN authored one or more posts. Users were sorted into three categories depending on how much they posted: 90% were "lurkers", 9% were "contributors", and the top 1% were "superusers". This study offers evidence of the 90-9-1 principle within online communities and shows that there is a large demographic of users (lurkers) that software developers can encourage to participate in online communities through better designed user interfaces. Nonnecke and Preece (2001) define lurkers as "one of the 'silent majority' in an electronic forum; one who posts occasionally or not at all but is known to read the group's postings regularly" (p. 294).

There are two main reasons why people lurk: lack of usability or personal reasons. In a survey done by Preece and Andrews (2004), 7.8% of respondents stated the lack of software usability as a reason for lurking. Despite the small percentage, software developers should strive towards equitable software usability for all users and consider utilizing features such as avatars, signatures, rankings, and point systems which enhances a person's identity and encourages participation in a quasi-anonymous online community (Liao, 2007).

Before discussion of individual features, it is important to define features within the larger context of a user interface. According to Myers (1995), a user interface is the "part that handles the output to the display and the input from the person using the program" (p. 66). Within the scope of this research, features such as the ability to post anonymously or create avatars are considered to be a part of the user interface as they serve a primary means of user interaction, by accepting information from the users and communicating output from the program as well as to other individuals in the online community. Therefore, it is important for software developers to design features of a user interface that decrease lurking due to software usability.

Anonymity in online communities has mixed results. Amichai-Hamburger (2013) argues anonymity can especially help those who are looking for others that share their own identity and help them to gain feedback and validation. On the other hand, anonymity can also lead to vitriol and hate comments. Therefore, total anonymity is not advised, as people are more likely to post negative content through trolling, lying, and harassing. Malinen (2015) suggests using pseudonyms, as it increases contribution of posts while significantly reducing negative postings, making it a happy medium between total anonymity and true self.

Nonnecke and Preece (2001) offer several reasons for why people lurk that reflect users' different personalities: they want to learn more about the group first, they feel that there is no need to post, or they are still building an identity. Understanding a user's behavior and cognition will allow software developers to decide whether to include or exclude certain features when developing an application. It is equally important for software developers to consider the culture and values of their users. In a study by Malinen (2015), the author compares cultural differences between the US, Netherlands, and South Korea in a comparative case study. South Koreans tend to value collective activities, which means they are more likely to participate in online communities of interest and organizations within their local communities. Therefore, if the community did not pertain to their own personal interests and was in a low-context culture, South Koreans were not as likely to participate in the community as compared to the users in the US and the Netherlands. Software developers need to consider their user base and ensure that the user interface fits the users' cultural values.

There are other factors concerning individual users that could play a role in how a user interacts within an online community including educational experience, age, gender and employment status. In 2006, Nonnecke, Andrews, and Preece conducted a survey (n = 1188) examining these factors and found that 79.3% of Internet users surveyed had at least some college education and there were women users than men in the survey (50.5% and 49.5% respectively). They found that the population of lurkers and posters were very similar and there were no significant differences between age, gender, education, and employment (p. 10).

While some would argue that lurking is bad for an online community, Malinen (2015) argues that lurking is defined as a transformation from newcomer to regular member; it is a passive but non-negative way to enjoy an online community and can eventually allow users to

feel like they belong in the group, as "both forms of participation, reading and posting, have a positive influence on the development of a sense of community, and spending time in the online community and reading messages may actually lead to closer attachment to the group" (Malinen, 2015, p. 232).

Analyzing Discussion in Online Communities

The purpose of using two different frameworks is to measure the "success" of an online community and then analyze ways to improve user interface design tailored to a specific online community. The "success" of an online community can be considered through Habermas' theory of communicative action (Bohman and Rehg, 2014). In the context of online communities, when two or more users engage in a mutual deliberation and argumentation with the goal of reaching a common understanding of a topic, discourse occurs, eventually resulting in a consensus (Habermas, 1984). A consensus will occur when a user successfully convinces other members of the community of an idea and they take up an affirmative position towards the claim. Once a consensus is reached, the conversation in an online community will significantly decrease in terms of number of related posts and quality of meaningful discussion. Otherwise, discourse arises, where "claims…are tested for their rational justifiability as true, correct, or authentic" (Bohman and Rehg, 2014, para. 22). This cyclical process will occur many times throughout the lifespan of the online community. As long as there is recurring mutual deliberation in an online community, then users will be more incentivized to participate in the conversation, provided that software developers have created a user interface that promotes usability and sociability.

Preece (2001) offers a framework for analyzing the effectiveness of a user interface using usability and sociability. Usability is defined as "how intuitive and easy it is for individuals to

learn and interact with a product", where measures of usability include number of errors, productivity of the user, user satisfaction, etc. (Preece, 2001, p. 349). Sociability is defined as "developing software, policies, and practices to support social interaction online" (Preece, 2001, p. 349). Successful sociability is measured in terms of its purpose, people and policies. Table 1 gives a more comprehensive analysis of this framework and the criterion that can be used to measure the success of a community.

Malinen (2015) posits that the vitality and livelihood of an online community is not solely based on the number of contributions from its users, but rather the quality of participation from each user and the size of the online community. A large, well-established online community could benefit from more lurking, especially if lurking reduces the frequency of repeated questions or off-topic discussions. However, too much lurking could also be detrimental, especially in a relatively new community. A new community that does not have a "core, critical mass of active members for a self-sustaining interaction may need to encourage posting and discourage lurking" (Nonnecke et al., 2006, p. 18).

It is also important to consider the type of the community, as it can affect motivation for participation. In a mail survey done by Nonnecke and Preece (2000), results indicated that the average number of lurkers in a software support community was almost twice as high as one of a health support community, where lurking was defined as no posts in a 12-week data collection period. The difference in levels of lurking could be a result of health support communities being more focused on developing empathy amongst users or because health support communities are on average smaller than software support communities (Nonnecke and Preece, 2000). Regardless of the specific differences between various types of online communities, it is important for

software developers to consider what the users would use the site for and how exactly they would participate.

Therefore, it would be more meaningful to holistically examine the lifecycle of an online community through the meaningfulness of posts and discussion rather than measuring the "success" of an online community through the number of active users. Using Habermas' theory of communicative action, factors such as type of community, frequency and quality of posts, and size of community can be utilized to measure lurking levels to determine the vitality of an online community. From that point, it will be useful to use usability and sociability as a framework to compare and contrast the similarities and differences in features and user interface of various online communities (Preece, 2001).

Framework	Design	How are they different?
Sociability	*Purpose*. A community's shared focus on an interest, need, information, service, or support, that provides a reason for individual members to belong to the community. *People*. The people who interact with each other in the community and who have individual, social and organization needs. Some of these people may take different roles in the community, such as leaders, protagonists, comedians, moderators, etc. *Policies*. The language and protocols that guide people's interactions and contribute to the development of folklore and rituals that bring a sense of history and accepted social norms. More formal policies may also be needed, such as registration policies, and codes of behavior for moderators. Informal and formal policies provide community governance.	Sociability focuses on human-human interaction supported by technology.
Usability	*Dialogue and social support.* The prompts and feedback that support interaction, the ease with which commands can be executed, the ease with which avatars can be moved, spatial relationships in the environment, etc. *Information design.* How easy to read, how understandable and how aesthetically pleasing information associated with the community is, etc. *Navigation*. The ease with which users can move around and find what they want in the community and associated websites. Many online community users have suffered from inconsistencies of data transfer and differences in interaction style and the website housing the community. *Access*. Requirements to download and run online community software must be clear. In addition, if high bandwidth and state of the art technology is needed to run the community there should be a low bandwidth text only versions and clear instructions about how to obtain it.	Usability focuses on how members of a community interact with each other via the supporting technology (human-computer interaction).

Table 1. Framework using sociability and usability to measure the success of an online community (Wu, adapted from Preece, 2001).

Research Questions and Methods

My research question is: Why do a small percentage of users contribute to discussions within online communities and how can software developers design user interfaces that encourage more participation? This mixed method study utilized both a survey and existing data and specifically focused on the forum-styled online community Piazza. Used by millions of students across thousands of campuses, Piazza allows students to ask questions and instructors to moderate and endorse answers with the goal of having "students teaching students" and "conversations…[that] continue long after office hours are over" (Piazza, 2020a; Piazza 2020b). While this study focuses on Piazza, the frameworks used can be applied to other online communities.

A survey was developed to measure levels of and the motivation behind lurking on Piazza. The survey was divided into three sections: lurking behavior, usability of certain features, and desired features. All questions were optional and individual responses were anonymized. The survey was conducted online over a period of two weeks and received 58 responses. Survey results indicated that all respondents were between 18-24 years old. All respondents had received at least a high school diploma or the equivalent, with 78.18% of respondents having received at least some college level education. As reported in a 2019 survey conducted by the Pew Research Center, 100% of adults between 18-29 years old have used the Internet, and 84% of high school graduates have used the Internet (Pew Research Center, 2019). Demographic data was collected in order to accurately represent the population of users who typically use Piazza.

Survey results were compared against existing Piazza data from four consecutive semesters (Spring 2018 – Fall 2019) of the same Computer Science course as evidence to observe patterns of lurking as originally offered by Van Mierlo (2014). A smaller subset of the Piazza data (Spring 2018 versus Spring 2019) was then used to support Habermas' theory of communicative action by examining the number of active unique users and the number of posts per day to observe whether an increase in posting significantly correlated to meaningful discussion in relation to course material and guidelines. This data was then corroborated with an interview with Wissem Gamra, a project manager at Piazza, who offered insights into how Piazza designs their user interface and iterates on features based on user feedback.

Piazza: A Case Study on Lurking in Online Communities

While participants reported a number of reasons as to why they do not post in online communities, this research identifies one specific reason: lack of usability. A lack of usability often stems from how software developers design the user interface and its respective features and tools. Implementing designs informed by user feedback and existing literature can increase usability, thus encouraging more participation. This process will be demonstrated through a case study using Piazza. Through this case study, we identify three main takeaways: a majority of users lurk, the effectiveness of specific features through the lens of usability and sociability, and finally, how all of this affects the lifecycle of an online community using Habermas' theory of communicative action.

In order to measure levels of lurking within Piazza, the total contributions of each student over the course of a semester was grouped in the five categories: "Never" (0 times), "Rarely" (1-

2 times), “Sometimes” (3-5 times), “Occasionally” (6-14 times), and “Frequently” (15 or more times). This was done to compare student perception of Piazza usage (survey) versus actual Piazza usage (existing Piazza data). Interestingly, the two are drastically different, as 52.73% of respondents in the survey stated that they rarely use Piazza, even though Piazza data indicated an average of 18.79% (Appendix A). In reality, a majority of students fell into the “Never” category.

In the scope of this case study, contributions are “posts, responses, edits, follow-ups and comments to follow-ups” in Piazza (Piazza, 2020b). Therefore, lurkers were broadly defined as users who fell into the “Never” (never contributed on Piazza) or “Rarely” (contributed 1-2 times) categories. These criteria fit Nonnecke and Preece’s (2001) definition of a lurker, that who is “one of the ‘silent majority’…who posts occasionally or not at all but is known to read the group’s postings regularly” (p. 294). As expected, a majority of users were classified as lurkers (Figure 1).

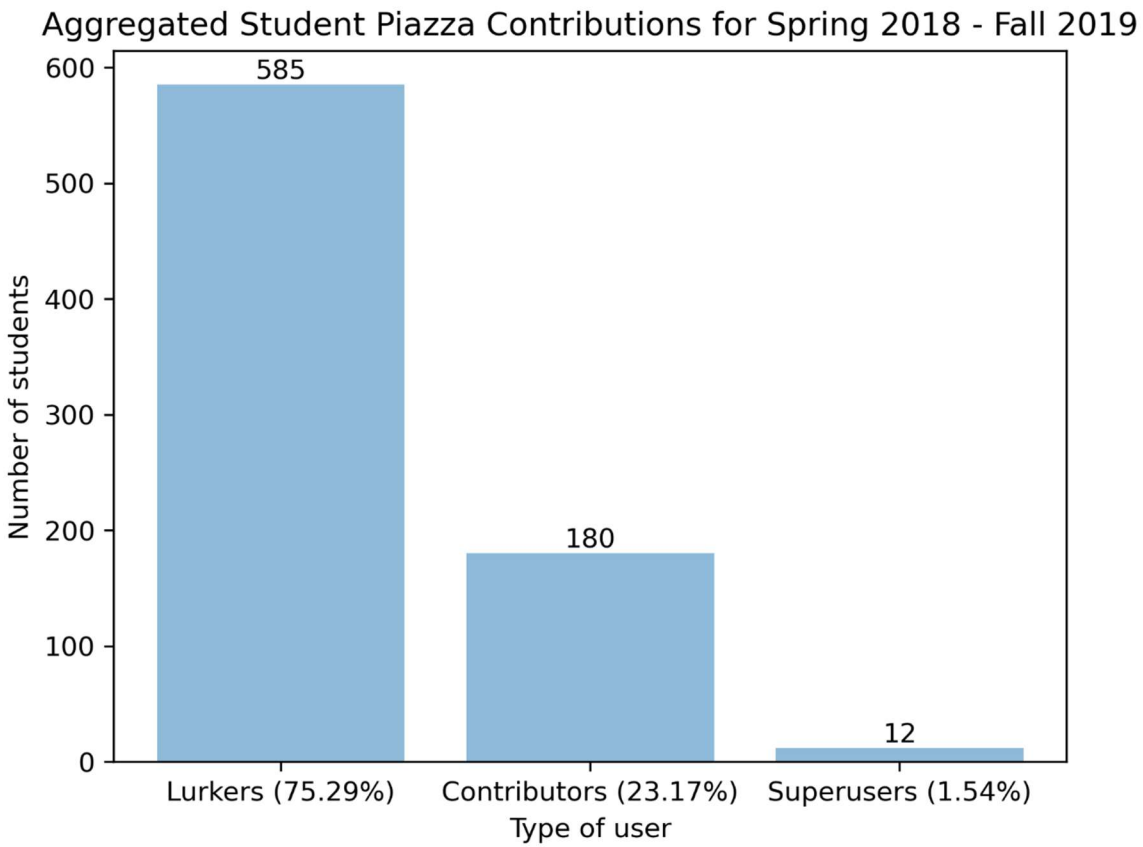


Figure 1. Aggregated data across four semesters where lurkers were defined as users who "Never" or "Rarely" contributed, Contributors were defined as users who "Sometimes" or "Occasionally" contributed, and Superusers were defined as users who "Frequently" contributed.

The survey also helps answer methods in which software developers design user interfaces that encourage more participation. Despite Piazza's boasts that its main feature is a fast response time (mean=14-minutes), the most frequently requested feature was a method of reminding instructors about unanswered questions (Appendix B) (Rusli, 2011). Interestingly enough, respondents also indicated that "Instructor answer" and "Anonymity" features were the top two "Extremely Useful" features (Appendix C). The second most requested feature was a less cluttered user interface. One respondent stated, "I wish it was visually easier on the eyes. It's really hard to navigate through past posts when I am actually looking for a specific post that I read about earlier…". Survey responses implied that the usability, specifically information design and navigation, of the user interface inhibited users' ability to effectively accomplish tasks. As a result, usability should be measured through other user feedback metrics such as user satisfaction and productivity.

Without user feedback, developing methods to encourage participation is extremely difficult for software developers. Online communities with clear channels of communication between the users and the developers facilitates community growth and allows developers to continually improve upon features. In an interview with Wissem Gamra, he noted that user feedback was mainly gathered from trial and error and testing. Gamra states the CEO of Piazza, Pooja Sankar, believed in the concept of product-market fit. This pushed Piazza to have the end users, students, directly test the product at Stanford career fairs. In conjunction with beta testing, Wissem stated that they also wrote "half-finished functionalities" where developers would pitch an idea for a feature, "throw a button on the website, and see how many people click on it. Meanwhile, we'd display something like: 'Not yet available' or something. Based on demand, we would see if we keep developing the feature or not."

So long as there exists a conversation between the users and the developers, it is possible to measure the meaningfulness of conversation within an online community using Habermas' theory of communicative action. In order to do this, two parallel semesters (Spring 2018 and Spring 2019) were observed to see if there were any patterns of increased or decreased activity following important course "milestones" such as tests, project and assignment deadlines. During both semesters, the number of posts and active users increased in the beginning of February, the same time the first big class assignment was due. Additionally, smaller sharp increases in the number of active users and posts corresponded with biweekly project checkups. A significant decrease in activity occurred in the earlier half of March during spring break when most students tend to step away from school work. Thus, the relationship between sociability (what is deemed valuable within the online community, in this case, "milestones" within a course) and increased, "meaningful" activity is dependent on the values of the user and what they deem to be important. As long as usability and the needs of the user are successfully met, there will be a constant ebb and flow of activity within an online community.

Generalizations of Online Communities

While Piazza is a specific case of an online community whose success depends on both the usability of its software and the lifecycle of its online community, there are multitudes of other online communities that have been growing just as (if not more) rapidly than Piazza, such as Reddit. Within 20 years, Reddit has managed to cultivate "loyal communities", which have a strong network of user interactions within each niche community populated by "loyal users" who contribute by bringing in new, esoteric content and discussion, allowing the online community to thrive (Hamilton et al., 2017).

How is Reddit so successful at cultivating "loyal communities"? This comes from understanding the relationship between the purpose and values of the online community and the definition of meaningful activity described by Habermas' theory of communicative action. A community's purpose and values are driven by a mutual focus on an interest, information, service, or support that draws users in (Preece, 2001). However, meaningful activity is based on the values of the individual user and what they deem to be important, which may be different from others. As a result, two or more users are able to engage in mutual deliberation and argumentation with the goal of reaching a common understanding of a topic, allowing for meaningful activity (Habermas, 1984). Yet, this cannot be possible without addressing usability first and making sure that all users who want to contribute are able to. Software developers should strive to design a user interface whose features can be accessible and usable by all. Once this has been successfully established, only then can discourse occur.

This study sought to examine why a small percentage of users contribute to online discussion and how software developers can design user interfaces that encourage more participation through surveys and analysis of existing data. It is important to recognize that because of the

limited scope of data collection, the data collected is not necessarily representative of all online users or online communities. In the future, it would be useful to analyze other online communities in a similar manner to this case study but on a larger scale in order to gain a more holistic perspective. Furthermore, I would redesign the survey in order to collect more data on demographics and try to elicit more responses. Additionally, I would conduct more interviews with software developers and gain insight on how companies utilize user feedback in order to improve their user interfaces.

Examining the relationship between users, software developers, and the user interface in which they indirectly communicate through frameworks such as usability and sociability as well as Habermas' theory of communicative action has helped me gain insight on improved methods of designing applications. It is the software developer's goal to translate user requirements and feedback into a tangible product that is not inhibited by lack of usability.

The Future of Online Communities

Online communities present a new and innovative way for humans to interact. While this research was able to consider an online community through the frameworks of usability and sociability as well as Habermas' theory of communicative action, it would be useful to research if levels of lurking change throughout the lifecycle of an online community. Future online communities will need to continually shift and evolve to fit the needs of their users. Subsequently, the nature of online communities means that there will always be lurkers. However, as more research is done on the "silent majority", software developers will hopefully be able to design user interfaces to better fit the needs of all users.

In closing the interview, Gamra stated, “I personally believe that you always have to get your products out there in the market [as soon as possible], because no matter how much you think you know your clients and their needs, you actually do not know until they start telling you what they like.” Regardless of the type of application, whether it be SNAC or a niche online community, software developers who fail to understand the needs of the user will ultimately be unable to develop successful applications.

References

About SNAC. (2010). Retrieved October 28, 2019, from https://portal.snaccooperative.org/about.

Amichai-Hamburger, Y. (Ed.). (2013). The social net: Understanding our online behavior. OUP Oxford.

Bohman, J. and Rehg, W. (2014). Jürgen Habermas. *The Stanford Encyclopedia of Philosophy (Fall 2014 Edition).* Retrieved from https://plato.stanford.edu/archives/fall2014/entries/habermas/

Habermas, J. (1984). The theory of communicative action, Volume 1. Boston: Beacon.

Hamilton, W. L., Zhang, J., Danescu-Niculescu-Mizil, C., Jurafsky, D., & Leskovec, J. (2017). Loyalty in online communities. *Eleventh International AAAI Conference on Web and Social Media*.

Liao, Y. Y. (2007). Promoting online discussion participation by integrating identity-enhancing feature from digital games (Doctoral dissertation, Ohio University), 27-39.

Malinen, S. (2015). Understanding user participation in online communities: A systematic literature review of empirical studies. *Computers in Human Behavior*, 46, 228-238.

Myers, B. A. (1995). User interface software tools. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *2*(1), 64-103.

Nonnecke, B., Andrews, D., & Preece, J. (2006). Non-public and public online community participation: Needs, attitudes and behavior. *Electronic Commerce Research*, *6*(1), 7-20.

Nonnecke, B., & Preece, J. (2001). Why lurkers lurk. AMCIS 2001 proceedings, 294.

Pew Research Center. (2019, June 12). Demographics of Internet and Home Broadband Usage in the United States. Retrieved from https://www.pewresearch.org/internet/fact-sheet/internet-broadband/

Piazza. (2020a). Our Story. Retrieved from https://piazza.com/about/story

Piazza. (2020b). Why Piazza Works. Retrieved from https://piazza.com/product/overview

Preece, J. (2001). Sociability and usability in online communities: Determining and measuring success. *Behavior & Information Technology*, 20(5), 347-356.

Preece, J., Nonnecke, B., & Andrews, D. (2004). The top five reasons for lurking: improving community experiences for everyone. *Computers in Human Behavior*, 20(2), 201-223.

Nonnecke, B., & Preece, J. (2000). Lurker demographics: Counting the silent. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 73-80).

Rusli, E. (2011, July 3). Homework Help Site Has a Social Networking Twist. Retrieved from https://www.nytimes.com/2011/07/04/technology/04piazza.html?pagewanted=1&ref=technology

Tsikerdekis, M. (2013). The effects of perceived anonymity and anonymity states on conformity and groupthink in online communities: A Wikipedia study. *Journal of the American*

Society for Information Science and Technology, *64*(5), 1001-1015.

Van Mierlo, T. (2014). The 1% rule in four digital health social networks: an observational study. *Journal of Medical Internet Research*, *16*(2), e33.

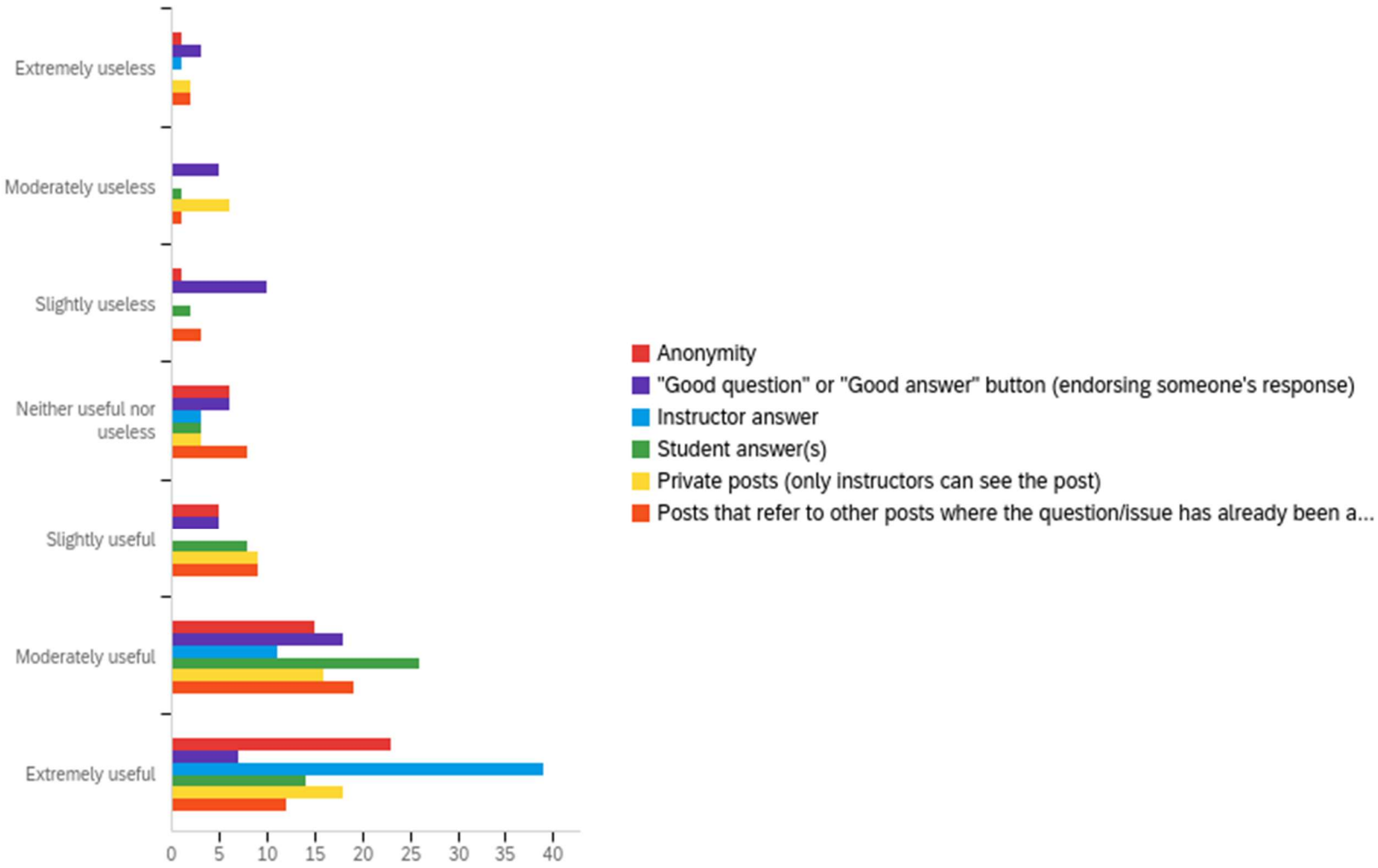
Appendix

Question	Options
How often do you use Piazza?	• Never (7.27%) • Rarely (once or twice a semester/class) (21.82%) • Sometimes (once every couple months) (20.00%) • Occasionally (once every week) (29.09%) • Frequently (more than once a week) (21.82%)
How often do you POST on Piazza?	• Never (36.36%) • Rarely (once or twice a semester/class) (52.73%) • Sometimes (once every couple months) (7.27%) • Occasionally (once every week) (1.82%) • Frequently (more than once a week) (1.82%)

Appendix A. Survey questions on lurking (N = 58).

"What feature do you wish Piazza had that would make your experience better?"
• A feature to remind instructors of unanswered questions (6) • Better user interface (too cluttered) (6) • A voting feature for questions or answers (4) • FAQ page for each class (2) • Default notification for emails should be less frequent (2) • Chat feature that includes private messaging (2) • Increased (or complete) anonymity (2)

Appendix B. Features that users would like in Piazza, aggregated and ordered by frequency from survey results. The value in parenthesis listed is the actual number of individuals who answered.



Appendix C. Survey respondents rated how useful they found each feature to be. Note that the "Instructor answer" and "Anonymity" were chosen by 72.22% and 45.10% of respondents to be "Extremely Useful", respectively.