

"At Least I Know Why!": Contextualizing User Reactions to Negative Feedback on StackExchange

WALEED AHMED, University of Illinois Urbana-Champaign, USA

SEAN FARHAT, University of Illinois Urbana-Champaign, USA

ADVAI PODDUTURI, University of Illinois Urbana-Champaign, USA

ZACHARY ROBERTSON, University of Illinois Urbana-Champaign, USA

With the fast evolution of the online community and crowd-sourcing platforms, there has been growing focus on newcomer experience and the multiple factors that influence it. In this paper, we aim to answer some integral questions about newcomer experiences relating to community guidelines and content removal. In particular, we collect and analyze behavioral data from the StackOverflow community to investigate contextual factors that influence user receptiveness to criticism. We construct a statistical model to measure the effect of informativeness versus other contextual factors. To this end, we are able to successfully establish the importance of guidelines in predicting user response. We show a 4.15% increase in the number of users who ask another question upon the closure of their first question, if they had read the community guidelines.

Additional Key Words and Phrases: datasets, user retention, online communities

ACM Reference Format:

Waleed Ahmed, Sean Farhat, Advai Podduturi, and Zachary Robertson. 2022. "At Least I Know Why!": Contextualizing User Reactions to Negative Feedback on StackExchange. *ACM Trans. Graph.* 37, 4, Article 111 (August 2022), 8 pages. <https://doi.org/XXXXXXXX.XXXXXXX>

1 INTRODUCTION

Content removal is a common experience for new users in online communities. How does this impact their experience? Will they stay on the platform? In this paper, we want to investigate the role of guidelines on user retention. Do users who engage with official guidance demonstrate greater resilience to having their questions closed?

Broadly we structure our investigation as a data analysis of users on the StackExchange network. We focus on the newcomer experience and are interested in how users orient, or disorient, themselves in the community. In particular, we focus on users who may have broken community norms such as when they post a question that eventually is closed. Our investigation is structured by the following research questions:

- RQ1: How has the community attempted to address the user retention problem so far?
- RQ2: How do contextual factors affect the likelihood of posting another question in the future?

To address RQ1 we investigate a recent response by the community to help users create first questions. To answer RQ2 we collect user behavior data from the StackOverflow community and

Authors' addresses: Waleed Ahmed, waleed3@illinois.edu, University of Illinois Urbana-Champaign, USA; Sean Farhat, seanf2@illinois.edu, University of Illinois Urbana-Champaign, USA; Advai Podduturi, advairp2@illinois.edu, University of Illinois Urbana-Champaign, USA; Zachary Robertson, zwr3@illinois.edu, University of Illinois Urbana-Champaign, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Association for Computing Machinery.

0730-0301/2022/8-ART111 \$15.00

<https://doi.org/XXXXXXXX.XXXXXXX>

perform statistical analysis to establish which contextual factors are most important for predicting user reactions to question closure. Our main contribution is to establish the importance of guidelines over specific assistance with question writing in predicting user response to question closure. This is in line with research on user responses to post removal in the Reddit community and could be seen as a replication of [Jhaver et al. 2019].

2 STUDY CONTEXT: STACKOVERFLOW

Stack Overflow has had several discussions on how to improve the platform for newcomers; the overall conclusion from experienced users was that newcomers needed a guide to orient themselves and get accustomed to the community norms before contributing [MechMK1 2020]. Otherwise, newcomers would have their posts deleted or closed, mostly because they were not following best practices. Stack Overflow responded to this feedback by offering two newcomer-oriented components: a site tour and an AskQuestion wizard.

2.1 Site Tour

The site tour takes the form of a readable webpage [MechMK1 2022] that covers the important basics: the point of the site, tips on how to ask good questions/contribute, and short introductions to features such as Tags, Reputation, and Commenting. The tour is easily accessible, especially for newcomers. When making a new account, the very first message in your inbox is a message directing you towards the tour. In addition, there is also a large suggestion to do the tour on every users account page. An excerpt of the tour can be found in Figure 1.

2.2 AskQuestion Wizard

The AskQuestion wizard is a newer feature that is under active development, having an initial version in March 2019. Originally, it took the form of an interactive, guided mode of question-asking for all users under 100 Reputation. The next version was converted to a more fleshed out webpage for the very first question users asked. This was refined until the current version, which began rolling out in March of 2022. The first time that a user asks a question, they are taken to a page where each component of the question is broken down and accompanied by a commentary bubble suggesting best practices. This was built upon years of iterative discussion between the Stack Overflow team and the community. Notably, this mode checks for duplicate posts and even suggests existing posts that it believes are semantically similar. This is probably in response to the largest cause of closed posts: duplicate questions. An excerpt of the AskQuestion Wizard can be found in Figure 2.

3 RELATED WORK

Attempts to measure and quantify the user experience are useful because they inform us of which aspects may be most relevant or impactful to improve. There are many different metrics for user retention that can be used based on the length and number of users in the measurement. We present 3 retention metrics that we short-listed from our background research. The strongest metric was to use a ratio of recency / inter-action-time where a ratio larger than 1 is an indication of a retention problem [Ascarza et al. 2018]. As an example, a user who typically uses the site every 1 month with a recency of 6 months would have a ratio of $6:1 = 6$ which shows a retention problem. Similarly, a user who goes on the site once a year with the most recent interaction being 6 months ago does not indicate a retention problem since the ratio is $6:12 = 0.5$. However, this metric requires a lot of data on the user so that the ratio calculated is not skewed by a few extra interactions. So, we explored another metric that measured retention based on the number of content created by the user, the number of active months, and the range of active months [Dev 2021]. Here, an active month refers

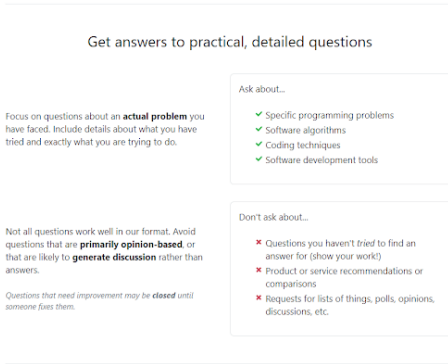


Fig. 1. Tour Page

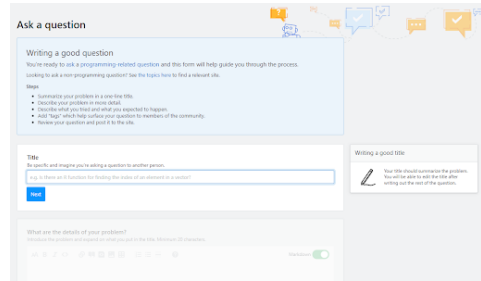


Fig. 2. AskQuestion Wizard (2022)

to when the user creates at least one piece of content. The range of active months is the period between the user's first active and last active month. This metric allows us to measure the slope of activity, especially for newcomers, and focus only on content generation (posts, comments, and likes) rather than just visits to the site. The last metric we considered is a simple 0/1 indicator, such as the user upvoting a post in the past X months [Ascarza et al. 2018]. Then, our retention metric is simply the ratio of users who hit the indicator over total users. We still need to refine what user action is best to use for the 0/1 indicator.

Socialization tactics like welcome messages, assistance, and feedback are known to help improve user experience [Choi et al. 2010]. More integrated approaches such as encouraging teamwork and collaboration are also effective socialization tactics [Li et al. 2020, 2015]. Specific guidelines about the norms are a common and effective strategy for improving user experience [Li et al. 2020; Myers West 2018]. In particular, one paper finds that reading user guidelines can help improve user reception to criticism and content removal on Reddit [Jhaver et al. 2019]. However, some work suggests that user guidelines can also stifle user expression and creativity [Fiesler et al. 2018]. Positive or negative, this work gives warrant to the hypothesis that the communication of user guidelines may have an outsized impact on the user experience.

4 METHODOLOGY

In this section we overview the investigation methods used to formalize our research questions quantitatively. The key technical aspect of the work is progress on interacting with data StackExchange provides on its users.

4.1 Data Explorer

Ultimately, we are interested in measuring engagement with these remediation attempts. Stack Overflow provides a measurable metric if a user has read the tour: the Informed badge. Somewhat related, there used to be an Analytical badge that would be awarded if a user read every section of the FAQ. However, this badge was retired at least seven years ago. Measuring interaction with the AskQuestion wizard is a little less straightforward. There is no badge, as every new user is taken through this process. We believe taking the sample of all new users after March 2019 as the "question guided" group is a reasonable metric.

For statistical modeling, our sampling population consisted of all StackOverflow users whose first question was closed. We used a random sampling approach to gather 50k users from this population. We created SQL queries on the DataExplorer website to collect our initial data. After

this, the data was cleaned and prepared for statistical modeling. We also filtered the population so that account creation must be after the introduction of the Informed badge.

4.2 Variables

The main dependent variable of interest is whether or not the user ever posts again, after having their first question closed. For each user, we collected features that we hypothesized would be relevant predictors for the dependent variables. Our control variables were the question's score, the time between account creation and first question, and the number of posts made before a first question. Our independent variables were categorical variables indicating the user had read the tour page before/after posting the question and the reason for closure.

4.3 Data Analysis

We used a Logit regression model for their ease of interpretation. We standardized all data so that we could approximate effect-size with the relative size of the coefficients.

5 QUALITATIVE FINDINGS

Our choice of dependent variables for the statistical model was motivated by qualitative findings regarding new users and their early posting behavior. Primarily, we were interested in whether the presence of the Informed badge or AskQuestion Wizard correlated to any significant patterns of behavior.

To address the efficacy of these tools, we looked to answer two questions:

- (1) Did its presence lead to better *quality* first questions?
- (2) Did its presence encourage newcomers to post again?

To answer both questions, we first grouped all the first posts by users into 4 different quality categories described in the following section, plus an "accepted" category indicating whether the question received an Accepted answer.

5.1 Measuring Quality of Posts

When looking at users' posted questions, we naturally needed a metric to quantify the quality of their posts. Adopting the metric that StackOverflow uses when conducting their internal studies, we extracted several variables from the DataExplorer to classify every post. The breakdown can be found in Table 1.

Table 1. Measures for Question Quality

Quality	Closed	Score	AnswerCount
Good	No	> 0	N/A
	No	0	> 0
Neutral	No	0	0
Bad	Yes	N/A	N/A
	N/A	< 0	N/A
Closed	Yes	N/A	N/A

5.2 The AskQuestion Wizard

Our findings for the AskQuestion Wizard can be found in Figure 3.

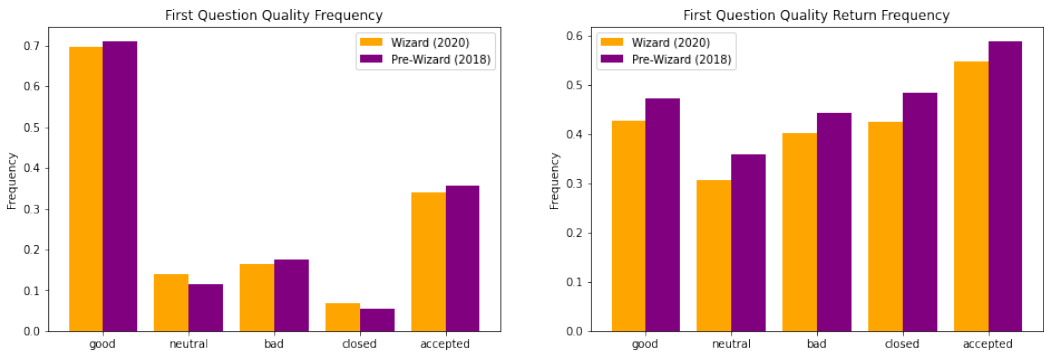


Fig. 3. The effect of the AskQuestion Wizard. To account for the unbalanced period of time before/after the introduction of the wizard, we sampled all new users in the year before it was introduced (2018) and the year after (2020) as representative groups. (Note: not all new users in the Wizard group experience the wizard.)

In the first chart, we measured the question quality of all first posts by new users in a Pre-Wizard and Wizard era. In the second chart, we measure how many users, respective to the quality of their first question, posted again.

We can interpret these results from 2 perspectives. The first is a within-group analysis, which can be generalized to a general trend since the results were similar, regardless of group. We see that, for both groups, users mostly post good first questions, with a somewhat evenly-split low rate of neutral or bad ones. A very low number are closed, while a moderate number have accepted answers. With regards to posting again, we see that, again regardless of group, new users are the most likely to post again if they get an accepted answer, followed by posting a good quality question, a closed one, bad quality, and lastly a neutral one. This final result is the most surprising and lends to an important insight: *the presence of feedback, supportive or critical, is more encouraging to newcomers than no feedback at all.*

The second perspective looks at the results to directly answer our initial questions: did the Wizard help? From the data, the answer seems to be *no, the AskQuestion Wizard did not make a big difference.* For the most part, in terms of question quality and rates of posting again, users in the era before the AskQuestion Wizard actually posted higher quality questions in slightly higher rates and returned to post again in moderately higher rates.

So, our results led us to not pursue building a model around the AskQuestion Wizard, though we did gain a notable insight into the importance of the presence of feedback on encouraging new users to engage with the platform again. The questions involving what steps StackOverflow has taken to direct feedback towards newcomers and those effects remains for future work.

5.3 The Informed Badge

Our findings for the Informed badge can be found in Figure 4. In the first chart, we measured the question quality of all first posts by users who had the Informed badge vs. those who did not. In the second chart, we measure how many users, respective to the quality of their first question, posted again.

In general, the results for question quality are similar than that of the Wizard. Most questions are good, a similarly low amount of neutral and bad, a very low amount of closed, and a moderate amount of accepted. The most notable takeaway is the much higher rate of accepted first questions from Informed users. However, the query for return frequency was computationally intractable

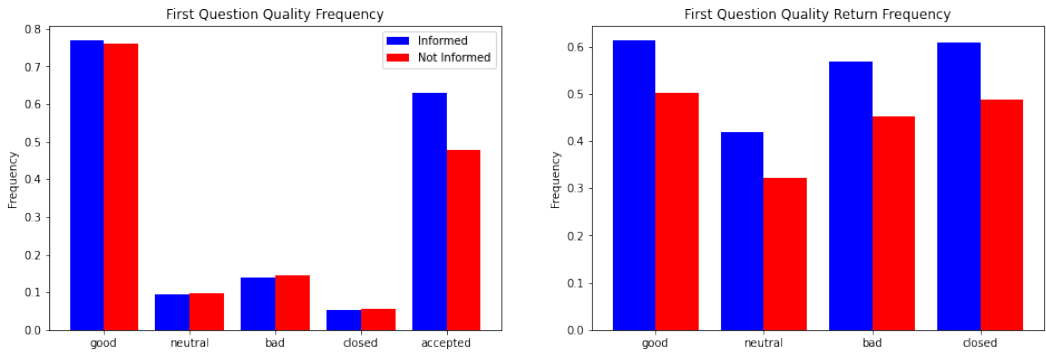


Fig. 4. The effect of the Informed badge. Due to the complex query constraints, the return frequency for the "accepted" group was intractable and timed out, so it is omitted.

when applied to the accepted group, so we could not analyze it further. The other trends were generally the same across both groups.

The results for return frequency are where the Informed badge shines. Across the board, regardless of the quality of the question, Informed users were more likely to post again. Each category had at least a 10% higher rate of return, while the largest effect occurred in the closed group: Informed users who had their first question closed were around 25% more likely to post again than non-Informed users.

This final result motivated our choice of variables when building a predictive model: look at Informed/non-Informed users who had their first question closed. To be more thorough, in our model, we broke down the Informed users into those who obtained the badge before they asked their first question, and those who got it after.

6 QUANTITATIVE FINDINGS

There are some descriptive statistics for the Informed users on the platform. In the population at large, we find that roughly 17% of users have the Informed badge. Among users whose first question was closed roughly 25% have the Informed badge. This is partially explained by the fact that 22% of first questions come from users with the Informed badge. Overall, closure rate on first question from Informed users is slightly lower than non-Informed (5.38% vs. 5.59%).

For statistical modeling we first calculated descriptive statistics for the dependent variable PostAgain. Overall, we found that 48.57% of users would post again. Users who have the obtained the Informed badge before asking a first question that gets closed end up asking another question 51.84% of the time. For users who have not obtained the Informed badge this number is 47.69%. Following this, we explored how this outcome is related to various contextual factors of the initial question posting experience. An overview of our results for the statistical modeling can be found in Table 2.

6.1 Posting Context

We find that contextual factors such as the post's score and the amount of time between account creation and first question have relatively high effect sizes. On the other hand, the number of posts before the first question has a relatively small effect size. The effect of having a question closed as a duplicate seems to be insignificantly negative.

Table 2. Logit model for whether or not a user posts again after having their first question closed. This model includes posting context as control variables in addition to independent variables.

	Coef.	Std.Err.	z	P> z
Score	0.1534	0.0219	6.9995	0.0000
Hours Before Question	0.1618	0.0103	15.6410	0.0000
Duplicate	-0.0260	0.0104	-2.4954	0.0126
Posts Before Question	0.0315	0.0197	1.5975	0.1102
Informed Before	0.0717	0.0104	6.9267	0.0000
Informed After	0.2875	0.0111	25.9387	0.0000

6.2 Community Guidelines

We find that the effect of reading the tour page (Informed badge) varies by when the reading was done. If the badge is obtained before posting a first question, we have a moderate positive effect size on predicting if the user will post again. In contrast, we find that reading the tour page after having their first question closed has a large positive effect size on predicting posting again.

7 LIMITATIONS AND FUTURE WORK

For the statistical modeling, there are a few limitations worth discussing. Foremost, we only find a small effect size for users who obtain the Informed badge before posting their first question. It seems plausible that measuring whether a user obtains the Informed badge after posting could confound our analysis of user retention.

We try to address this problem with establishing causality by looking at the AskQuestion wizard as an instrumental variable predictor for whether a user posts again. However, we do not find a strong effect here. While it is possible the AskQuestion wizard was not successful, it is also possible that guidelines do not have as strong of a causal effect as originally hypothesized.

The natural followup to this work is to investigate the neutral feedback effect: having any form of feedback is more encouraging than getting none. We believe that this result is not too surprising. The neutral response puts their question in limbo. They technically have not done anything wrong since it was not closed, but it was not good enough to get an answer. This middle ground provides no direction as to how they can improve. Future work should investigate how often this occurs, what groups it affects, and whether StackOverflow does anything to mitigate this problem.

8 CONCLUSION

In this work we introduce the hypothesis that reading community guidelines has an outsized effect in predicting user response to criticism. Moreover, we are able to establish significant statistical evidence in favor of this hypothesis. However, our approach is limited by confounding variables, and our attempt to establish a viable instrument for measuring causal effects was not successful. Despite this, we find that initial reception and orientation on the platform is an important predictor of whether or not a user will be able to receive criticism.

REFERENCES

- Eva Ascarza, Scott A Neslin, Oded Netzer, Zachery Anderson, Peter S Fader, Sunil Gupta, Bruce GS Hardie, Aurélie Lemmens, Barak Libai, David Neal, et al. 2018. In pursuit of enhanced customer retention management: Review, key issues, and future directions. *Customer Needs and Solutions* 5, 1 (2018), 65–81.
- Boreum Choi, Kira Alexander, Robert E Kraut, and John M Levine. 2010. Socialization tactics in wikipedia and their effects. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. 107–116.
- Himel Dev. 2021. *Modeling the successes and failures of content-based platforms*. Ph. D. Dissertation. University of Illinois at Urbana-Champaign.
- Casey Fiesler, Joshua McCann, Kyle Frye, Jed R Brubaker, et al. 2018. Reddit rules! characterizing an ecosystem of governance. In *Twelfth International AAAI Conference on Web and Social Media*.
- Shagun Jhaver, Darren Scott Appling, Eric Gilbert, and Amy Bruckman. 2019. "Did you suspect the post would be removed?" Understanding user reactions to content removals on Reddit. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–33.
- Ang Li, Zheng Yao, Diyi Yang, Chinmay Kulkarni, Rosta Farzan, and Robert E Kraut. 2020. Successful online socialization: lessons from the Wikipedia education program. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–24.
- Guo Li, Haiyi Zhu, Tun Lu, Xianghua Ding, and Ning Gu. 2015. Is it good to be like Wikipedia? Exploring the trade-offs of introducing collaborative editing model to Q&A sites. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. 1080–1091.
- MechMK1. 2020. What about the community is "toxic" to new users? <https://meta.stackexchange.com/questions/342779/what-about-the-community-is-toxic-to-new-users>
- MechMK1. 2022. Tour Page. <https://stackoverflow.com/tour>
- Sarah Myers West. 2018. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society* 20, 11 (2018), 4366–4383.