Socializing Volunteers in an Online Community: A Field Experiment

Rosta Farzan, Robert Kraut
Carnegie Mellon University
rfarzan,robert.kraut@cs.cmu.edu

Aditya Pal, Joseph Konstan
University of Minnesota
apal,konstan@cs.umn.edu

ABSTRACT
Although many off-line organizations give their employees training, mentorship, a cohort and other socialization experiences that improve their retention and productivity, online production communities rarely do this. This paper describes the planning, execution and evaluation of a socialization regime for an online technical support community. In a two-phase project, we first automatically identified from participants' early behavior, those with high potential to become core members. We then designed, delivered and experimentally evaluated socialization experiences intended to build commitment and competence among these potential core members. We were able to identify potential core members with high accuracy from only two weeks of behavior. A year later, those classified as potential core members participated in the community ten times more actively than those not identified. In an evaluation experiment, some potential core members were randomly assigned to receive socialization experiences, while others were not. A year later, those who had participated in the socialization regime contributed more answers in the community compared to those in the control condition. The socialization experiences, however, undercut their sense of connection to the community and the quality of their contributions. We discuss what was effective and what could be improved in designing socialization experiences for online groups.

Author Keywords
Online volunteer communities, Socialization, Experiment

ACM Classification Keywords
H.5.3 Information Interfaces and Presentation: Group and Organization Interfaces: Evaluation/methodology,
Web-based interaction.

General Terms
Experimentation; Human Factors

INTRODUCTION
Volunteers are responsible for the user-generated content in most online communities, including videos on YouTube, code in open source communities, articles in Wikipedia and user support in both technical and health support groups. In these communities, a small fraction of the users generally provides a large proportion of the good content [16]. These members also do much of the administration and community building on the site [5]. Despite the importance of these core community members, neither site managers nor researchers understand how to select and develop a cadre of volunteers to take on these core roles.

Research on socialization in off-line organizations suggests that a well-designed regimen of socialization experiences can increase organizational members' commitment to the organization and their competence. Socialization is the process through which newcomers acquire the behaviors and attitudes essential to playing their roles in a group or an organization [17]. As organizational members become socialized, they move from the periphery of the organization towards the core [22]. Theories of organizational socialization describe institutionalized socialization practices that include socializing newcomers in a cohort, providing a clear sequence of stages through which members progress, providing training from experienced role models, and providing encouraging positive feedback, as well as constructive criticism [9]. The use of an institutional style of socialization by the organization is associated with more successful socialization outcomes: newcomers do their jobs better, are more satisfied with their jobs, become more committed to the organization, and leave the organization less often [20], [4].

Although evidence is strong that institutionalized socialization tactics are effective in developing commitment and appropriate behavior in conventional organizations, they are not common in online communities. Socialization processes in most online communities are informal and individualistic [19]. To thrive, online communities should incorporate processes for selecting and socializing committed members. One important step is to develop a process to identify newcomers who fit the community and will ultimately be valuable to it. Insuring a good fit between the newcomers and the community will lead to more positive outcomes for both newcomers and the community [11],[12],[10]. An efficient way to ensure good fit is to develop a screening mechanism. Often online communities employ a human evaluation process as the screen. For example, in many open-source software communities, volunteers must have offered code that meets quality standards before they gain committer status [13],[8]. However, human evaluation is by definition labor-intensive and time-consuming. Additionally, humans usually need long-term contribution data about a user to be able

CSCW 2012, February 11–15, 2012, Seattle, Washington, USA.

to make a reliable judgment. As a result, many potential core members may leave the community before being recognized.

This paper describes our experiences designing, implementing and evaluating socialization experiences in a question-and-answer community in which volunteers answer tax-related questions. The community recognizes users who contribute large numbers of high quality answers as "superusers". Superusers are selected through a manual human evaluation that takes into account individual users' contributions. The site contains more than 60,000 users and 83 superusers. Although anyone can answer questions on the site, 78% of answers judged as helpful by community readers are provided by the small group of superusers. Like many other online volunteer communities, this community faces the challenge of growing the number of committed users who contribute high quality responses. We use this community as an environment to explore practical ways that an online volunteer community can identify from their early behavior participants with potential to become core members, and to discover practical ways to socialize and train them so they are more competent and committed to the community. In a two-phase project, we first used machine learning to automatically identify from early behavior users with high potential to become core members. We then designed, delivered and experimentally validated socialization experiences to build commitment and competence among these potential core members. We discuss what was effective and what could be improved in designing socialization experiences for online groups.

IDENTIFYING POTENTIAL CORE MEMBERS

In the first phase of the project, we designed and implemented machine learning algorithms to automatically identify potential superusers. A complete description of the dataset we used, the classification algorithm, and its evaluation can be found in [18]. We used a dataset that included all of the observable behavior each user exhibited on the site from July 2006 to April 2009, including the text of their questions and answers, and meta-data such as time between posts and evaluations of their answers provided by other users. We reasoned that potential superusers should be highly motivated to help others, and they should have the required capability to answer questions correctly. We identified several possible indicators of motivation and ability. In this community users' motivation is reflected in quantity of contributions, frequency of contribution, and commitment towards the community. Ability of users is reflected in their domain knowledge, trustworthiness of their answers, and politeness and clarity in their responses. Table 1 below shows features we used to identify each indicator of motivation and ability.

We extracted motivation and ability features for the first two weeks of data per user to build the training dataset for machine learning. We used Support Vector Machine (SVM) and Decision Tree (DTree) learning models over these features to identify the potential superusers. The result of the evaluation of the algorithm shows that both motivation and ability factors have the same discriminating weight in classifying superusers. Information gain analysis shows that among all the features described in Table 1, number of answers, number of votes, number of best answers, frequency of login, average time elapsed between answers, and usage of pronoun I are the most informative features in distinguishing high potential users from low potential users.

Table 1. Features indicating users' motivation and ability

Quality	Indicator	Features
Motivation	Quantity	#of answers(+) # of questions(+)
	Frequncy	Average time elapsed between two answers(+)
	Commitment	# of logins(+) Time spent in the community(+)
Ability	Knowledge	# of best answers(+)
	Trustworthiness	# of votes on answers(+) # of positively voted answers(+) Ratio of answers with negative votes to positively voted answers(-)
	Politeness and Clarity	Language dimensions such as presence of typos(-), use of singular pronouns(-), use of negative terms(-)

† (+) and (-) signs represent the increasing or decreasing effect of each indicator on motivation or ability.

Since all the features in the algorithm are related to users' responses, including users with few posts or none does not affect the classification result. Therefore, we excluded users with fewer than ten responses from the dataset. That left 1,974 users in the dataset. Then we used our binary classifier on those 1,974 users to select potential superusers. The classifier classified all the users into one of two classes: potential superusers or not-superusers. Out of 1,974, 75 were classified as potential superusers based on data from their first two weeks in the community. Out of these 75, 40 were already recognized as superusers by community. That means, 53% (40/75) precision and 48% recall (40/83) for the algorithm. However, in terms of precision of the algorithm, the remaining 35 participants could have been misclassified as superusers by the algorithm (i.e., false positives) or could really have been potential superusers who were missed during the manual selection process. We conducted a formal human evaluation to further assess the effectiveness of the automatic classifier. We hypothesized that some of these 35 users were users with high potentials to become superusers. Human evaluation allowed us to examine this hypothesis. Employees whose job responsibilities included promoting people to superuser status evaluated a sample of users selected by the algorithm to be either high-potentials, medium-potentials, and low-potentials. For each category, we modified our classifier to return a ranked list of all 1,974 users instead of binary classifications. We then selected the top 35, middle 35 and bottom 35 to represent high, medium, and low potential superuser respectively. Twenty-five judges rated these 105 users. Each evaluator was presented with a list of 9-12 users from three categories in a random order. Each user was evaluated by two judges. The human judges had access to the full profile of each of those users on the site, which included all their responses in addition to statistics about their tenure, total number of responses provided, total number of positive and negative votes they received, and number of helpful responses provided. Human judges had access to similar information as the algorithm but the information was not limited to only two weeks of data.

Table 2 shows the confusion matrix, showing the agreement between the algorithm's assessments and the human judgments. The algorithm's extreme assessments in the high and low categories matched the human judgments very well (77% and 57% accuracy, where 33% is expected by chance). Human judged 57% of the algorithm-classified medium category as having high potential which suggests the algorithm is more conservative than the humans. Those in the medium category can potentially be trained to match superuser status.

Table 2. Agreement between the algorithm and the human judgments

		Human judgments		
		low	medium	High
Algorithm	Low	20	10	5
	Medium	9	8	18
	High	3	5	27

The results suggest that the identification of potential core members in this community from their early behavior can be done successfully with reasonably high accuracy. However, there is no ground truth for which of these people really had superuser potential. Therefore, it is impossible to assess the validity of the model without observing them in the subsequent tax season. Even though human judges might still be needed to vet the candidates, a preliminary sorting by the algorithm decreases the manual work significantly and can guard against judges failing to give serious consideration to high quality candidates.

LEADERSHIP CHALLENGE

The second phase of the project focused on the design and implementation of socialization practices to help the high-potential candidates build commitment and competence. Before we intervened, the site had no formal socialization process for new users. Newcomers were not aware of other newcomers and did not have methods by which they could communicate with others for support. Newcomers had to figure out on their own how to get started and how to make progress. They received very little feedback from the community, especially when they made mistakes. Additionally, identifying experts and communicating with them was a cumbersome and unnatural process on the site. Our goal was to discover practical ways of addressing these shortcomings to increase participants' motivation and abilities. As described earlier, prior research in organizational socialization, education and social psychology indicates that formal socialization is associated with positive outcomes. We developed Leadership Challenge incorporating elements of a formal socialization process that were compatible with an online production community in a low-cost and self-sustaining way – cohort-based socialization in which the newcomers were isolated from the surrounding community, sequenced training materials, and providing feedback. We did not include a formal mentoring program because of its likely expense.

Cohort based socialization

Newcomers to organizations are more likely to know what they should do, become more productive, are more satisfied and stay longer if they join the organization in a group with other newcomers and go through early socialization experiences with others in their cohort [4], [20]. Being part of a cohort helps newcomers share their experiences, collectively figure out the organization's values and work procedures, receive social support from their peers and build a longer term social support network. If training and other socialization experiences occur off the job, in an environment where consequential performance demands are minimized and fear of failure is reduced, they can practice job-related skills in a non-threatening way. This sequestered style of socialization improves their self-confidence and their job-related skills [4].

We implemented off-the-job socialization by conducting the Leadership Challenge between tax seasons, when newcomers could train on questions culled from the previous season. Answering these questions incorrectly would not put taxpayers at risk. We implemented cohort-based socialization by providing a discussion forum as a place for participants to interact with each other, compare their answers, and engage in group discussions. As described below, we attempted to stimulate discussion by requiring participants to grade others' answers in their training exercises. We populated the discussion forum with participants' answers and comments on these training exercises to stimulate interaction.

Goal setting and clear sequence of stages

Participants in the Leadership Challenge were given tax questions to answer and the answers of others to grade. Research on both goal-setting and socialization in organization suggests that providing people with clear goals and a defined sequence of stages to achieve as they move from the periphery to the core of an organization or community improves both role clarity and self-efficacy [4]. Providing clear goals can assist newcomers in figuring out what is expected from them, and achieving these goals can increase their sense of self-efficacy [9],[3]. Research in psychology and organizational behavior indicates that goals and goal-setting strongly motivate people, especially when the goals are specific and challenging [15]. A fixed time-table for advancing in the organization improves newcomers' role clarity [4].

We translated these ideas into the Leadership Challenge by giving targets for the number of questions they should answer and answers they should evaluate to advance through a sequence of pages in each of five functional areas of tax expertise (e.g., basic income, deductions and education). Participants could receive silver badges in each area for com-

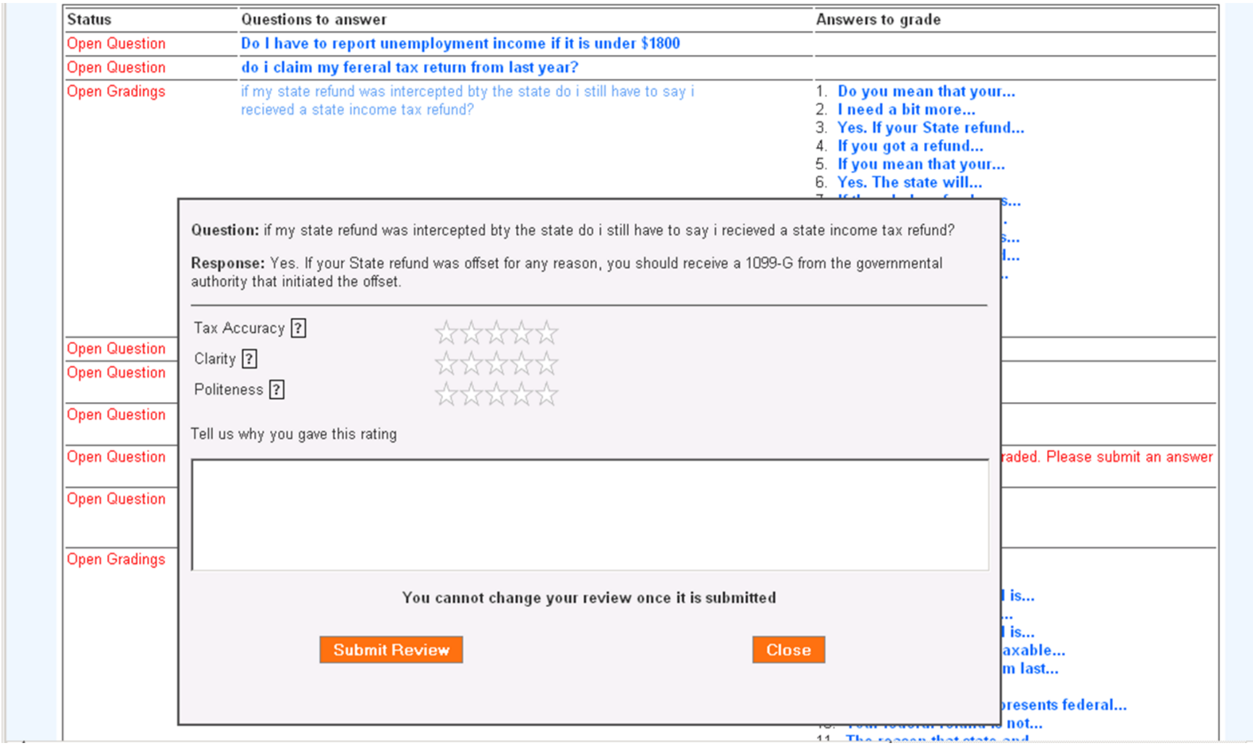


Figure 1. Interface for answering questions and grading

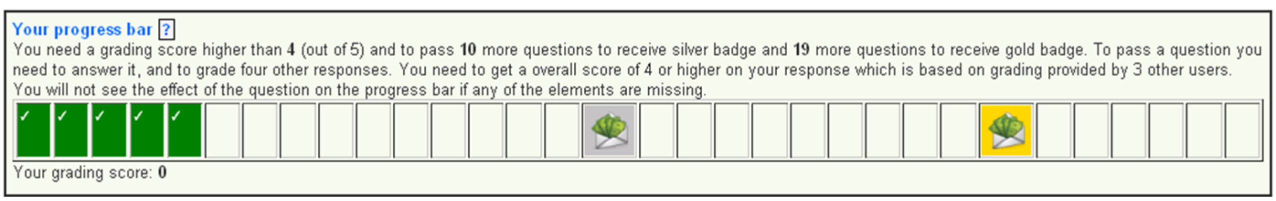
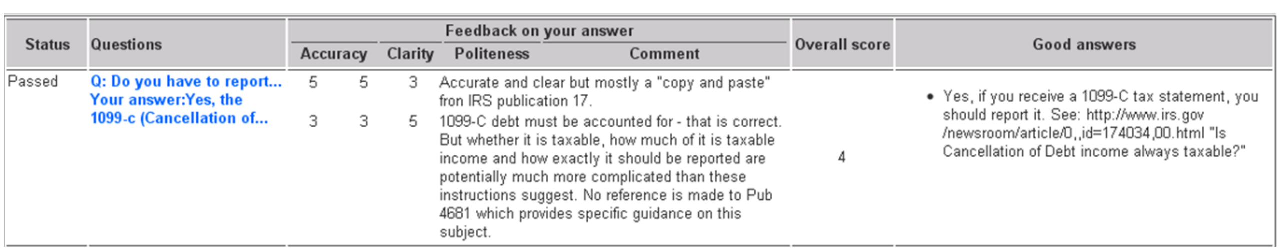


Figure 2. Progress Bar



Status	Questions	Feedback on your answer				Overall score	Good answers
		Accuracy	Clarity	Politeness	Comment		
Passed	Q: Do you have to report... Your answer:Yes, the 1099-c (Cancellation of...	5	5	3	Accurate and clear but mostly a "copy and paste" fron IRS publication 17.	4	• Yes, if you receive a 1099-C tax statement, you should report it. See: http://www.irs.gov /newsroom/article/0,,id=174034,00.html "Is Cancellation of Debt income always taxable?"
		3	3	5	1099-C debt must be accounted for - that is correct. But whether it is taxable, how much of it is taxable income and how exactly it should be reported are potentially much more complicated than these instructions suggest. No reference is made to Pub 4681 which provides specific guidance on this subject.		

Figure 3. Interface for reviewing feedback from others

pleting 50% of questions correctly and gold badges for completing 80% of questions correctly. A question was considered complete if it received an overall score of greater than four out of five and if the participant graded at least two answers to the same question from others. The score was calculated as a weighted sum of peer reviewed scores on the three dimensions of accuracy, clarity, and politeness. A progress bar made the goal and their progress towards the goal salient (Figure 2).

To make the goals and the steps to achieve them more salient, we sent participants in the Challenge a personalized email reminder at the end of each week. Email reminders have been shown to be effective in encouraging contribution in online communities, especially when the messages are personalized [14], [7]. The reminders emphasized the amount of work the recipients had already done and highlighted what more they could do. A sample reminder is shown below.

```
Thanks for participating in the Leadership
Challenge.  We would like to let you know new
questions are available on the site now.

Here is a summary of your activity over last week:

You responded to 8 questions and graded 2
responses of your peers.  You have 12 more
questions to reply to and 18 more responses
to grade.  According to scientific research,
answering questions, getting feedback from others
and evaluating others answers all help you become
more expert, by giving practice and offering
multiple points of view.

Login now ...
```

Training through worked examples

Educational research shows that providing learners with worked examples (i.e. the steps in a correct problem solution) fosters learning and stimulates deep understanding [2]. The paradigm of problem solving with worked examples originated in mathematics and physics. Worked examples typically consist of a statement describing the problem and an explanation of how to solve it, both meant to guide learners in solving similar problems [23]. Learning from worked examples is especially effective for beginners and for initial skill acquisition. According to the ACT-R cognitive theoretical framework, in the first stage of skill acquisition, learners refer to known examples and try to relate them to problems they are trying to solve [1]. This research from the educational domain suggests that learning from worked examples is a promising technique to help newcomers improve their domain knowledge.

We implemented these ideas in the Leadership Challenge by constructing a set of tax-related questions from an archive of questions asked in a prior tax season. The questions were vetted by a tax expert to be informative, potentially teaching those who answered them something new, interesting or non-obvious in a particular area. Every week ten new questions were offered for participants to answer. Once they answered a question, they could view all of the good answers from the archive (See Figure 1). A sample question and answer on regulations related to basic income is shown below.

My federal refund went DOWN when I entered in my second W2, why is that? My federal refund amount is decreasing the more W2s I enter, is that supposed to happen? If I don't owe taxes can I just enter in one and not the others to get a bigger refund?

```
No.  You must enter all income received in 2009.
The reason your refund goes down after you enter
the 2nd W-2, is that your standard deduction and
exemption is subtraced from your 1st W-2 before
the tax is calculated...
```

Providing feedback

Research in peer-evaluation has shown that students learn from providing feedback to their peers and benefit from receiving feedback from others with similar experiences, especially when the review process is scaffolded, anonymous, and reciprocal [6]. Additionally, peer-reviewing can increase interaction among the newcomers and therefore improve their sense of community. After participants in the Leadership Challenge answered a question, they were required to grade answers submitted by other newcomers on three dimensions (accuracy, clarity, and politeness) and to complete a free-format comment justifying their evaluations (Figure 3).

FIELD EXPERIMENT

Often only a small percentage of people invited to an experiment actually participate in it. Since we required at least one control group for comparison, inviting only 75 high-potential users would have provided too few people to form the community. Additionally, the result of our human evaluation showed that more than half of users with medium potentials did exhibit superuser capabilities. Therefore, to increase the number of invitees, first, we relaxed our criteria of the minimum number of posts from ten to five and, second, we used our ranking algorithm to select all the users with medium or high potential to our experiment. This resulted in inviting 283 people.

Based on these criteria, we invited the 283 users to join the Leadership Challenge. The Challenge was run for six weeks in the off season, isolated from the live Tax Support Community site. All the users received the same invitation. To prevent selection biases, participants were randomly assigned to the control or experimental group only after they responded to the invitation. To create a critical mass of participants to interact with each other in the experimental condition, we assigned twice as many participants to the Leadership Challenge experimental condition as to the control condition. We used a randomized wait-list control experimental design, in which we promised the control group they could participate in the Leadership Challenge in the future. The experimental group received the treatment immediately after signing up and control group did not receive any treatment. The participants in both control and experimental group received a free copy of a tax software as compensation.

Data

At the end of the socialization period we conducted a survey to collect data about participants' evaluation of the Leadership Challenge along three dimensions - the extent to which participation in the Challenge influenced their tax knowledge, connection to the community and enjoyment. These questions were asked only of those in the Leadership Challenge (i.e. the experimental condition). The survey also included a scale measuring participants' future commitment to the Tax Support Community. This part of the survey was sent to both the control and the experimental group. Additionally, we collected participants' behavioral data[1] over the 2011 tax season to analyze the effect of the Leadership Challenge on participants' subsequent participation in the Tax Support Community. The behavioral data included a measure of quantity of participation (the number of questions that participants answers) and two on quality of participation (the total quality points they received for their answers, and the number of appreciations they received from other members of the community). Users can provide appreciations for answers they like by giving a "thank you" to the person answering a question. Users can accumulate points based on the feedback that others provide to their answers. Points are awarded such that each "thumbs up" awarded one point, each "helpful" tag awarded five points and each "solved" tag awarded 10 points. The points a member of the community earned were public information and were displayed whenever the member's moniker was displayed.

Of the 52 users who responded to the invitation, 16 were assigned to control group, and 36 to the experimental group. Thirteen of the 16 people in the control group filled out the survey, as did 24 of the 36 people in the experimental group.

In the next sections, we divide the results into three segments: accuracy of the automatic classification, a description of participants' behavior during the experiment and an evaluation of the effectiveness of the socialization intervention. We present the evaluation results in two subsections, because effectiveness changed from immediately following the interventions (when the interventions reduced participants' reported commitment to the community compared to those in the control condition) to a longer term follow-up (when the interventions increased the amount of participants' actual participation in the community compared to those in the control condition).

Accuracy of automatic selection

We invited 283 users to the Leadership Challenge based on the result of their leadership potential score from the automatic classifier, which measured how similar their early behavior on the site was to the early behavior of existing superusers. To test the predictive validity of the classification algorithm, we compared the 283 potential superusers' participation rates in the Tax Support Community in the 2011 tax season to participation of those not classified as potential superusers. The sample of non-selected users included 66,000 people. To address the imbalance in sample sizes, we compared the 283 potential superusers both to all 66,000 non-potential leaders and to a series of 1% random samples from the 66,000 population. The results of random sampling and complete datasets were similar. Here we report the results from comparing random sample with experimental data.

[1]The data does not include any personal information

Because participation rates are count data truncated at zero, we used negative binomial regression models to analyze them. The model included number of days since joining the community ("days"), number of answers in previous tax season (tax2010), and experimental condition as the independent variables. Experimental condition was coded as zero for those who were not selected by our algorithm, one for those who were selected but did not respond to our invitation, and two for those who were selected and responded to our Leadership Challenge invitation. The model predicted number of answers provided in the 2011 tax season. Results presented in Table 3 show that 283 users selected by the algorithm, whether they participated in the Leadership Challenge or not, participated at least 100 times more compared to those who were not selected. This result confirms the utility of the automatic classification in selecting those who demonstrated more potential. In addition, users who were selected and accepted the invitation to the Leadership Challenge contributed significantly more than those not selected. The algorithm was an accurate filter weeding out people with less motivation to contribute to the community.

Table 3. Estimated means of number of questions answered

condition		N	Mean	SE
Not selected		606	.01	.005
Selected		283	2.42	.181
	declined	231	1.13	.112
	accepted	52	8.67	1.440

† Covariates appearing in the model are fixed at the following values: tax2010=6.06 and days=934.50

Short-term evaluation of Leadership Challenge

General information about participants' behavior during the Challenge period is shown in Table 4. Their participation varied substantially. Some took the Challenge very seriously, logged in more than twice a week during the Challenge, answered more than 70% of available questions and provided good feedback to others. In contrast, others logged in only once to explore the Challenge. Similarly the quality of feedback on the Challenges varied significantly from very informative comments with references to relevant materials such as, "Wrong answer. The IRS Publication 17 states C. ... . If it is, you cannot deduct rental expenses that are more than your rental income for the unit. So, if you use your house for more than 14 days during the year and you decide to rent your house, you will need to check the yes box that used the house or dwelling unit more than 14 days. The conversion date is not the start of the tax year." In contrast, other feedback was perfunctory, consisting of only one word comments such as "ok", "clear", "correct."

Self-sustained community

We incorporated a peer-review process in the socialization practice for its learning benefit and its advantage in empowering the community as self-sustainable. However, as mentioned earlier, since the experiment was conducted in a small, isolated community, the number of participants was insufficient to make the community self-sustained. The problem may have been intensified by the amount of work required in a short period of time. Overall, participants provided 420

Table 4. Leadership Challenge Participation Descriptive Statistics

	Mean	Median	Min	Max	SD
Number of distinct days logged into the Challenge	5.42	3	1	17	5.2
Number of distinct sessions logged into the Challenge	8.83	3.5	1	40	11.9
Number of questions answered	16.21	7.5	0	107	25.1
Number of answered reviewed	16.67	1	0	116	29.5
Length of feedback in number of words	27.79	17.5	1	336	37.53

answers, but only 141 of them (34%) received at least two reviews from the community (Two reviews were the minimum needed to assure reasonable fairness in a peer evaluation process. In fact, the research on educational peer evaluation suggests a minimum of three reviews.) Of the 141 answers that were graded, 110 received a "correct" grade and counted towards winning a badge, but none of the participants received enough feedback to win any badges. We conducted a survival analysis to assess the effect of receiving feedback on users' survival. We calculated a feedback score for each participant as follows and classified participants into high- or low-feedback groups by a median split on the feedback score.

$$Feedback\ \ Score = \frac{Number\ \ of\ \ reviewed\ \ responses}{Total\ \ number\ \ of\ \ responses}$$

The result of a survival analysis shows that participants were significantly more likely to continue to answer questions when they received more feedback (See Figure 4). However, feedback did not seem to affect the number of days they logged into the Challenge (See Figure 5). This suggests that participants in the low-feedback group continued to log into the site to see if they received feedback on their submitted answers.

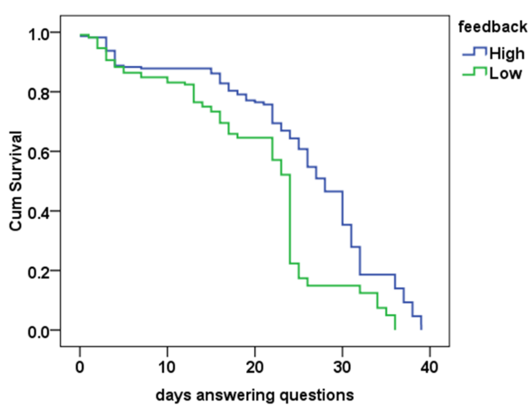


Figure 4. Effect of feedback on question-answering survival

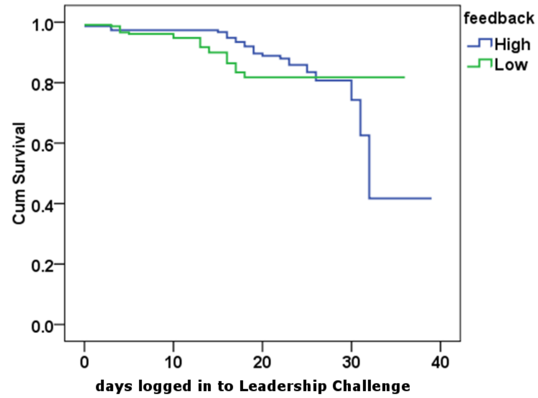


Figure 5. Effect of feedback on overall login survival

Effect of personalized reminders

As described earlier, we sent weekly personalized email reminders to participants who signed up for the Leadership Challenge. The data, presented in Figure 6, shows that each reminder caused a burst in the number of answers that participants produced and the size of this effect did not diminish with repeated reminders, at least over this six-week experiment. These results suggests the robust value of sending reminders to increase participation.

Subjective evaluations of the Leadership Challenge

We asked participants to evaluate the effectiveness of the Leadership Challenge in terms of how much it improved their tax knowledge, how much it helped them to learn about the community, and how satisfying the experiment was. Sample questions in each category are shown below. Answers were provided on 5-point Likert scales ranging from "Strongly disagree" to "Strongly agree."

We averaged the questions assessing each aspect of the Challenge into three reliable scales (Cronbach's $\alpha s > .67$, Table 5). While the overall opinion for all three categories was positive (Table 5), participants rated the value of the Challenge for increasing tax knowledge significantly higher than its value for learning about the community (Non-parametric test of related samples, $Z = -2.3, 2 - tailed\ p = .021$). Moreover, the correlation between the number of sessions they logged into the challenge and their evaluation of the challenge was negative ($\rho = -.404, p = .05$). This pattern suggests that the lack of activity in the site and the inadequate feedback participants provided each other led to dissatisfaction with the community.

```
Q1.  Participating in the challenge sharpened my
tax knowledge

Q2.  Participating in the challenge helped me
become more familiar with other users on the
community

Q3.  Participating in the challenge was fun
```

Behaviorial evidence reveals that very little interaction hap-

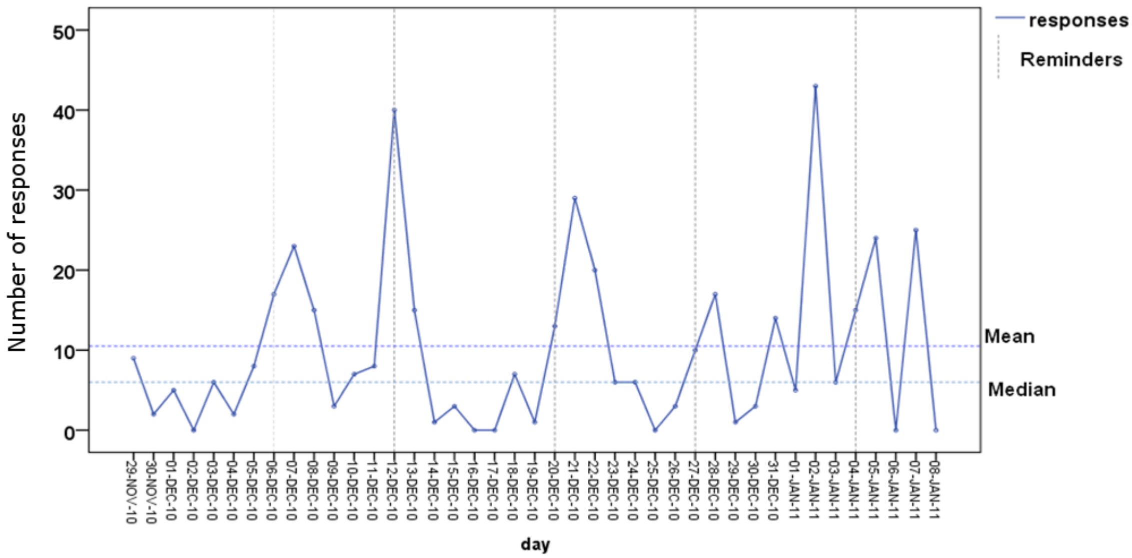


Figure 6. Effect of personalized reminders on number of users responding questions and total number of responses

pened among the participants. In particular, the discussion forum that was designed to support the cohort interaction was rarely used. Thus both attitude scales and behavioral evidence suggest that the experiment was not effective in promoting the feeling of the community. Participants found the training process helpful in learning about tax knowledge but not about the community.

Table 5. Participants' opinion about Leadership Challenge

	Mean	SD	Cronbach's α
Knowledge	3.62^{a}	.94	.87
Community	3.36^{b}	.81	.67
Satisfaction	$3.45^{a,b}$	1.09	.96

† Different superscript (a,b) indicate significant differences between values ($p<0.05$)

The survey also included four questions intended to measure participants' intentions to contribute to the community in the upcoming tax season (e.g. "I intend to participate in the community this tax season"). Answers to all four questions were interrelated (Cronbach's $\alpha = .91$). We used the average answer to all four questions per users as a measure of commitment. Comparison of the participants' subjective opinion in the control and the experimental groups shows that the control group expressed a significantly greater desire to participate in the future (Mean commitment measure: 3.74 versus 2.92, independent samples t-test, T=3.04, p=.005), although behavioral data discussed below shows they participated less.

Long terms effects of participation in the Challenge

To objectively assess the effects of the Leadership Challenge, we examined behavioral data to compare the quantity and quality of participation in 2011 tax season for those in the experimental and control conditions. We used the number of answers provided as a measure of quantity and number of appreciations and number of points they received as measures of quality of answers. As shown in table 6, participants in the experimental condition answered more questions in the 2011 tax season than did those in the control condition, controlling for the number of answers provided in the previous year and number of days since they joined the community.

Table 6. Estimated Marginal Mean - Effect of socialization practice on quantity of contribution

condition	N	Mean	SE
Control	16	5.29	1.57
Experimental	36	9.99	1.76
p-value	.04		

† Covariates appearing in the model are fixed at the following values: tax2010=6.06 and days=934.50

Additionally, for the experimental group, we assessed how their participation during the Leadership Challenge itself predicted their follow-up participation during the 2011 tax season. Independent variables in the model are the number of questions users answered during the Challenge, the number of answers they reviewed, and the proportion of their answers that received reviews from others, controlling for participation in the previous tax season. Since the distribution of number of questions they answered and answers they reviewed was highly skewed, we converted these independent variables into binary variables using a median split. Again we used negative binomial regression to predict the number of questions they answered during the 2011 tax season.

The result shows that receiving feedback during the Challenge motivated more contribution in future. Those who had higher proportion of their answers reviewed in the Challenge, answered more questions during the 2011 tax season. A 10% increase in the proportion of reviewed answers resulted in a 45% increase in answering questions in 2011 tax season. (IRR=4.5, SE=.53, p=.004). Participants who were more active during the Challenge both in terms of number of questions they answered, and number of answers they reviewed, answered significantly more questions during the 2011 tax season. This is an expected result which can suggest that those were just more active users.

To assess the effect of the socialization intervention on the quality of contribution, we normalized the number of points and appreciations by dividing them by the number of answers for each participant. Since users who have zero answers cannot receive any points or appreciations, the distribution of our response variables are zero-inflated. To account for that, we used a regression model with Tweedie distribution with log link function [21]. In terms of number of appreciations, the difference is not significant but in terms of number of points during the 2011 tax season the control group received a higher number of points per answers as shown in Table 7.

Table 7. Estimated Marginal Mean - Effect of socialization practice on quality of contribution

		Control	Experimental
# of appreciations	Mean (SE)	.28 (.16)	.58 (.21)
		p-value.	.31
# of points	Mean (SE)	1.53 (.61)	.45 (.17)
		p-value.	.032

† Covariates appearing in the model are fixed at the following values: # of points per answer in 2010=1.09, # of appreciations per answer in 2010=.14

In summary our analysis shows that the socialization practices offered to the participants encouraged more contribution, especially when the participants received feedback during the training period. However, the socialization did not improve the quality of responses compared with the control group who did not receive any training.

DISCUSSION AND FUTURE WORK

This research shows that one can identify potential core members in an online community with high accuracy from only two weeks' worth of early behavior in the community. A year later, those categorized by our algorithms as potential core members participated in the community ten times more actively than those not identified. Approximately 20% of these potential core members who were offered socialization experiences agreed to participate in a six-week long program. Those assigned to participate in the socialization experiences increased their contributions to the community compared to comparable volunteers in the control group. A subset of them participated very actively, logging in to the Challenge site more than twice a week and answering more than 70% of the questions posed to them. These people expressed their satisfaction explicitly in the final survey: "Fun program. Wish I were able to spend more time with it but other priorities made that difficult". These factors suggest that a formal training program is possible in an online community like the one studied here.

However, the small number of participants in the experimental group meant that most of them got insufficient feedback when they answered questions or posted comments to the forum. As a result, participants probably became frustrated with the lack of interaction, especially those who were active on the Challenge and expected reciprocal feedback from the community; they were disappointed by both the amount and the quality of the feedback they received. Their frustration was reflected in the comments they provided in the exit survey. "I felt like I was the only one participating, so it was not very worthwhile. I stopped answering questions because no one else was offering feedback or answering other questions. ..." or "If you allow 'laymen' to grade the answers of tax professionals (and give [them] a failing grade), you should allow a dialog [so that the professional can] explain why the grader is wrong or misguided." Additionally, some participants complained that the amount of work expected of them in the Challenge period was too great.

The small size of the group receiving socialization, the design that isolated them from the community and our desire for the socialization experiences to be self-sustaining undoubtedly undercut their effectiveness. These limitations may explain why participants receiving the socialization experiences reported less commitment to the group the more they participated, although they participated more and the quality of their answers was worse than those from the control group. Although prior research on socialization suggests that socialization in a cohort isolated from the larger community should be beneficial, our experiment identified an important challenge. Without a critical mass of others actively participating, socialization in a cohort can undercut social experience. Incorporating the new cohort into the community and explicitly encouraging existing core members to mentor them might be a solution, providing more chances for newcomers to receive the feedback necessary to improve their domain and community knowledge. Bringing them into the live community might be especially important if too few newcomers join in any defined period.

The research presented here provides insight to designers of online communities on the effectiveness of identifying high potential contributors and coaching them to become more active members. Although formal socialization programs are frequent offline, they are rarely used in online communities. Our research shows that designers and managers of online communities can be more proactive in identifying and nurturing future core members. Even though not all aspects of the socialization experiences were successful, the research shows the promise of developing successful and self-sustaining socialization programs for growing the leadership in an online community. Future research should focus on additional ways to integrate more sustainable and effective socialization practices into online communities.

ACKNOWLEDGEMENTS

This work was funded by NSF grants 0812148 and 0808692. We would like to thank Howard Chen and Igor Podgorny from the Intuit Live Community for their assistance with this research.

REFERENCES

1. ANDERSON, J. R., FINCHA, J. M., AND S. DOUGLASS. The role of examples and rules in the acquisition of a cognitive skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition 23* (1997), 932–945.

2. ATKINSON, R. K., DERRY, S. J., RENKL, A., AND WORTHAM, D. Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research 70*, 2 (2000), 181–214.

3. BANDURA, A. Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist 28*, 2 (1993), 117–148.

4. BAUER, T. N., BODNER, T., ERDOGAN, B., TRUXILLO, D. M., AND TUCKER, J. S. Newcomer adjustment during organizational socialization: A meta-analytic review of antecedents, outcomes, and methods. *Journal of Applied Psychology 92*, 3 (2007), 2707–2721.

5. BUTLER, B., SPROULL, L., KIESLER, S., AND KRAUT, R. Community effort in online groups: Who does the work and why? *Leadership at a distance 11* (2007), 171–194.

6. CHO, K., AND SCHUNN, C. D. Scaffolded writing and rewriting in the disipline: A web-based reciprocal peer review system. *Computers and Education 48*, 3 (2007), 409–426.

7. CHOI, B., ALEXANDER, K., KRAUT, R. E., AND LEVINE, J. M. Socialization tactics in wikipedia and their effects. In *Computer Supported Collaborative Work* (2010).

8. DUCHENEAUT, N. Socialization in an open source software community: A socio-technical analysis. In *Computer Supported Cooperative Work* (2005), vol. 14, pp. 323–368.

9. JONES, G. R. Socialization tactics, self-efficacy, and newcomers' adjustments to organizations. *Academy of Management 29*, 2 (1986), 262–279.

10. KRAUT, R. E., BURKE, M., AND RIEDL, J. Dealing with newcomers. In *In Evidence-based social design: Mining the social sciences to build online communities.*, R. E. Kraut and P. Resnick, Eds. MIT Press., Under contract.

11. KRISTOF, A. L. Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel Psychology 49*, 1 (1996), 1–49.

12. KRISTOF-BROWN, A. L., ZIMMERMAN, R. D., AND JOHNSON, E. C. Consequences of individuals' fit at work: A meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. *Personnel Psychology 58*, 2 (2005), 281–320.

13. KROGH, G. V., SPAETH, S., AND LAKHANI, K. R. Community, joining, and specialization in open source software innovation: A case study. *Research Policy Special Issue On Open Source Software Development 32*, 7 (2003), 1217–1241.

14. LING, K., BEENEN, G., LUDFORD, P. J., WANG, X., CHANG, K., LI, X., COSLEY, D., FRANKOWSKI, D., TERVEEN, L., RASHID, A. M., RESNICK, P., AND KRAUT, R. E. Using social psychology to motivate contributions to online communities. *Journal of Computer Mediated Communication 10*, 4 (2005).

15. LOCKE, E. A., AND LATHAM, G. P. Building a practically useful theory of goal setting and task motivation. *American Psychologist 57* (2002), 705–717.

16. MOCKUS, A., FIELDING, R. T., AND HERBSLEB, J. D. Two case studies of open source software development: Apache and mozilla. *ACM Transactions on Software Engineering and Methodology 11*, 3 (2002), 309–346.

17. MORRISON, E. W. Newcomers' relationships: The role of social network ties during socialization. *Academy of Management Journal 45*, 6 (2002), 1149–1160.

18. PAL, A., FARZAN, R., KONSTAN, J. A., AND KRAUT, R. E. Early detection of potential epxerts in question answering communities. In *User Modeling, Adaptation, and Personalization (UMAP)* (2011), Springer.

19. PREECE, J., AND SHNEIDERMAN, B. The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction 1*, 1 (2009), 13–23.

20. SAKS, A. M., UGGERSLEV, K. L., AND FASSINA, N. E. Socialization tactics and newcomer adjustment: A meta-analytic review and test of a model. *Journal of Vocational Behavior 70* (2007), 413–446.

21. TWEEDIE, M. C. K. An index which distinguishes between some important exponential families. *Statistics: applications and new directions* (1984), 579–604.

22. VAN MAANEN, J., AND SCHEIN, E. H. Toward of theory of organizational socialization. *Research in Organizational Behavior 1* (1979), 209–264.

23. VANLEHN, K. Cognitive skill acquisition. *Annual Review of Psychology 47* (1996), 513–539.