A Multi-Site Field Study of Crowdsourced Contextual Help: Usage and Perspectives of End Users and Software Teams

Parmit K. Chilana[1], Amy J. Ko[1], Jacob O. Wobbrock[1], Tovi Grossman[2]

[1]The Information School | DUB Group
University of Washington
Seattle, WA 98195 USA
{pchilana, ajko, wobbrock}@uw.edu

[2]Autodesk Research
210 King St. E
Toronto, ON Canada M5A 1J7
tovi.grossman@autodesk.com

ABSTRACT

We present a multi-site field study to evaluate LemonAid, a crowdsourced contextual help approach that allows users to retrieve relevant questions and answers by making selections within the interface. We deployed LemonAid on 4 different web sites used by thousands of users and collected data over several weeks, gathering over 1,200 usage logs, 168 exit surveys, and 36 one-on-one interviews. Our results indicate that over 70% of users found LemonAid to be helpful, intuitive, and desirable for reuse. Software teams found LemonAid easy to integrate with their sites and found the analytics data aggregated by LemonAid a novel way of learning about users' popular questions. Our work provides the first holistic picture of the adoption and use of a crowdsourced contextual help system and offers several insights into the social and organizational dimensions of implementing such help systems for real-world applications.

Author Keywords

contextual help; crowdsourced help; field studies; field deployments; help systems; software support;

ACM Classification Keywords

H.5.2 [**Information interfaces and presentation**]: User Interfaces. *Graphical user interfaces*.

INTRODUCTION

Most software help is simply not helpful. One-on-one technical support is time-consuming, frustrating, and rarely puts users in touch with the experts who have answers [18]. Documentation, tutorials, and tooltips rarely provide the task-specific help that people need [19,29,31] because software designers cannot anticipate at design time the full range of *how*, *why*, *what*, and *where am I* questions that people ask [31]. Even crowdsourced solutions such as technical support discussion forums require users to browse lengthy, scattered, and often irrelevant conversations in search for even one helpful reply [22,32]. Worse yet, many users do not find these discussions at all, as the search queries that users write

often omit the key application terminology necessary for retrieving relevant help [11].

To address these problems of discoverability, recent approaches have explored integrating crowdsourced help within the application's user interface (UI). For example, the *CHIC* framework [34] for Eclipse adds links from each UI control to a wiki where users can author help. *TurboTax* help [37] and *IP-QAT* [24] display help discussions relevant to the current view in a sidebar within the application. The *LemonAid* help system [8] lets users retrieve Q&A at an even finer granularity by selecting a label, widget, link, image or another UI element relevant to a given help need.

Although these *crowdsourced contextual help* systems are quite promising, many questions remain about their effectiveness at providing help during real tasks. For example, while IP-QAT's initial evaluation [24] showed that it was more useful and easier to use than a basic discussion board, the study involved only one instrumented application and was carried out with paid volunteers who were *required* to contribute 3 items per day. LemonAid's initial evaluation [8] showed that LemonAid could retrieve relevant help in the top 2 results for over half of help seeking scenarios, but this finding was based only a simulated community of users and pre-defined tasks. As social systems often interact with the social and organizational contexts in which they are implemented [14], it is difficult to know if and when these lab study results apply to real users and their real situations and tasks [20,28].

To understand how crowdsourced contextual help is perceived in real settings and to increase the ecological validity of LemonAid's design [8], we partnered with multiple software teams to deploy and evaluate LemonAid in the field. After extending the original LemonAid prototype with community Q&A, answer notification, and back-end analytics features, we deployed it on a set of four diverse sites for periods ranging from 7 to 15 weeks, reaching user populations of 150 to over 40,000 users in size. From over 1,200 logs, 168 exit surveys, and 36 interviews with end users, we report on end users' perspectives on LemonAid's helpfulness, usability, and desirability for reuse, and how it compared to alternative forms of help seeking. We also present software teams' perspectives on issues related to LemonAid's adoption and integration and the utility of LemonAid's analytics data.

The result of these deployments is the first holistic picture of the adoption and use of a crowdsourced contextual help system, contributing:

- Extensions to LemonAid's crowdsourced contextual help [8] that support participation by end users and software teams, moderation of help activity, and improved discoverability of help content;

- Empirical findings that highlight the unique strengths of crowdsourced contextual help and selection-based help in particular, but also tradeoffs in integrating help into an application user interface;

- Results that illustrate the factors affecting software teams' adoption of a crowdsourced contextual help tool and teams' perceptions of the value of crowdsourced contextual help analytics data;

- Insights into the process of developing and integrating help systems into existing applications, and how help systems fit into larger social and organizational contexts.

RELATED WORK

Field Evaluations of Help Systems

Although "in-the-wild" field studies are increasingly common in HCI [4], field studies of software help have been rare. Although not directly in the domain of software help, Ackerman's [1] evaluation of *AnswerGarden,* a question-answering tool, comes closest to our study design. The study used multiple data collection procedures including questionnaires, usage data, and interviews and the findings provided rich insights into whether augmenting organizational memory was possible, and also contributed a number of lessons for designing organizational memory systems. Brandt et al. [3] evaluated their *Blueprint* tool for programmers and followed-up with a 3-month deployment at a large company, collecting usage data and interviews. The field study component helped them gain valuable insights into how participants integrated Blueprint into their everyday programming workflow. Although not a field evaluation, the study of IPQAT is one of the first to evaluate crowdsourced contextual help with users [24]. While the study ran for 2 weeks and there were 36 participants, all of these participants were using the same application and most of them were already users of product discussion forums. Furthermore, these participants were required to contribute 3 items per day and received additional monetary incentives to contribute more. In contrast to these works, our field study investigates the use of crowdsourced contextual help by hundreds of users for their real tasks across multiple sites.

Help Seeking Behavior

Early studies of help seeking showed that when people experience breakdowns in using software, they begin a process of trial-and-error to find a resolution [5]. Carroll calls this the "paradox of the active user" [5], because even though it would be more efficient for the user to get help, the user opts to tackle the task on her own first [27]. In particular, users of early systems were known to avoid printed manuals [29,30] because such manuals were difficult to follow and were usually plagued with system-specific vocabulary [11]. The advent of searchable, online help documentation offered an improvement, but the static nature of the help content remained to be a problem [12]. Recently, online forums have become a new platform for help, with millions of users seeking help, and a smaller but still large population of people asking and answering questions [22,32]. Although forums offer a rich repository of application-specific help, they are often not easily navigable or searchable [32]. While crowdsourced contextual help approaches can bring more relevant help to users within the user interface, our study sheds light on several social, organizational, and technical factors that have to be aligned to deliver such help.

Software Support Field Studies

Another body of work evaluates software product support from the perspective of software teams. Many of these studies have focused on organizational, procedural, and management problems in the provision of support [18,26] and knowledge and management of support issue repositories [10,15,16]. Other efforts have focused on diagnosis and web-based one-on-one support practices [6]. A large body of work has also investigated peer-to-peer help in open source software (OSS) communities, where users can directly communicate with one another and with developers about support issues and bugs [21,22,32]. Still, we know little about how organizations integrate new provisions for help and support or how they may use data from these systems to inform design. To our knowledge, we are the first to contribute insights into how teams adopt a new crowdsourced contextual help tool and how teams perceive the value of the analytics data it generates.

EXTENDING LEMONAID FOR FIELD DEPLOYMENT

The key idea behind LemonAid [8] is that users enter a semi-transparent help mode overlaid onto the application's user interface and find help by selecting a label or image they believe is relevant to their problem (as seen in Figure 1.1). Upon a selection, LemonAid retrieves questions and answers related to the selection (Figure 1.2) based on an algorithm that retrieves help based on the text, tag type, and location of the selection within the user interface (described in detail in [8]). Users can then select a retrieved question to view its answers and indicate their helpfulness, or if they do not find a relevant question, they can submit one. Support staff and other users can then answer them. Users can also use keywords to search through existing questions, allowing LemonAid to behave like a site-specific Q&A search.

The prototype described in [8] included only a retrieval algorithm and a selection user interface. For deployment, we added several critical features:

Improving discovery of existing help content. One of our design goals was to facilitate users' navigation of the existing help content upon entry into the help mode. To facilitate this, LemonAid adds yellow question marks next to elements that have at least one question attached (Figure 1.1). These

Figure 1. Main components of the LemonAid interface in the help mode: (1) a user selects an on-screen label or image highlighted in yellow; (2) the user's selection triggers the retrieval of relevant questions; (3) the user can click on a question to see the answer(s) for that question and indicate whether the answer was helpful or not helpful.

questions marks are dynamically generated as soon as a user enters the help mode based on the selections of previously asked questions. To help users remember which UI elements they have already viewed, the system modifies the color of the visited question marks.

Encouraging user participation. Another goal was to encourage users to contribute questions and answers. When users add a new question in LemonAid, they can provide their email address and be notified of new answers without having to return to the site. LemonAid also allows users to report potential spam and offer "me too" votes on questions (Figure 1.2). When users view answers, they also contribute data on whether the answer was helpful or not helpful in order to dismiss the answer window (Figure 1.3).

Moderating help content. Many of the software teams that we approached were concerned about users' ability to deface their site with unwanted content. Therefore, we added basic a moderation feature, allowing support staff to receive e-mail notifications of new questions and answers so that they can both approve content, answer new questions, and improve user-contributed answers. The moderators also get notified when a user flags a post as inappropriate.

Analytics dashboard for monitoring help activity. We added a web-based administrator "dashboard" that shows an aggregate view of users' activities in LemonAid, specific to the host deployment site. Moderators can get a real-time snapshot of the elements where users are looking for help. They can also see rankings of popular questions based on views and votes, and the helpfulness of answers.

Implementation and setup: To set up LemonAid for each site, we needed: (1) a whitelist of the UI literals where Q&A would be attached in the interface, to prevent users from selecting privacy-sensitive content; and, (2) access to the team's source to include the LemonAid functionality on the site. We hosted all of the Q&A data, logs, and scripts related to LemonAid on our own servers. For (1), the previous version of LemonAid [8] offered a mechanism for extracting UI literals from the source code; teams desired more control over what part of the interface would actually become selectable on the screen, so we created a plug-in that would help us interactively choose selectable labels and images on specific pages. For (2), software developers only had to embed a link to the main LemonAid JavaScript file in their source code (as is required to use tools such as *Google*

Analytics, for example). One team decided to embed this link at the root level so that LemonAid would appear on all pages in the site; the other three teams placed LemonAid on their home page and selectively on other frequently accessed pages.

We also created a client-side browser plug-in that allowed us to test LemonAid on any site. Except for mobile devices, we were able to support LemonAid on most major browsers and platforms (although we did have problems with LemonAid on Internet Explorer versions 8.0 and lower for two of the deployments). We also had to resolve some minor JavaScript namespace conflicts on sites that used different versions of common JavaScript libraries, such as *jQuery*. The plug-in allowed us to sort out most of the compatibility issues independently of the host software teams, before deployment.

METHOD

Field Deployment Sites
The context for all of our deployment sites was a large university and its various software teams developing web applications to support teaching and research. To recruit field deployment sites, we met with software developers and site administrators across campus. Prior to making a commitment, the teams had a chance to see LemonAid demos, learn details about the field study, and discuss time and effort that they would be willing to commit. Six teams were enthusiastic about participating; the first site served as a pilot deployment and one of the remaining five had to leave the study due to conflicting schedules. The pilot site was a local intranet used by about 20 staff and students for internal project management. The pilot site allowed us to assess the usability of the interface, uncover implementation issues that had to be addressed for larger cross-platform deployments, and collect pilot data for improving our set-up and logging. We describe the four web sites augmented with LemonAid in Table 1.

Before making LemonAid live, we asked each software team to seed the help database with FAQs or other help content relevant to each page. In two cases, additional staff and the first author were also involved in seeding the help content.

Mixed-Method Data Collection
We treated each deployment as an individual case [36], but used the same data collection procedure for each site. We used a mixed-method approach to better capture the plurality of experiences and perspectives of users [13,36]. We collected data from three sources: (1) usage logs, (2)

Site	Description	User Population
LIBRARY	Provided access to a wide range of print and electronic resources available within the university library system and on web. One of the largest academic library systems in the US with millions of cataloged books, journals, and other materials.	Over **40,000** university students, faculty, and staff, visiting scholars, and the general public.
DEPT	A large academic department site that offered information about courses, upcoming events, admission, faculty, collaboration opportunities, and research facilities.	Over **900** undergraduate and graduate students, faculty, staff, prospective students, and the public.
EDC	An electronic data capture (EDC) system used for collecting data from clinical trials in an electronic format. Offered features for data entry, creating reports, analyzing data, exporting data, and sharing data with multiple users.	Over **500** staff personnel within the medical school and affiliated research center (restricted access).
RDB	A records database management system for storing and organizing data needed to maintain internal records of staff researchers. Used periodically by staff to review and update personal information, publications, grants, and to collaborate.	Over **150** staff and researchers in a unit within School of Medicine (restricted access).

Table 1. Summary of the four deployment sites.

surveys, and, (3) interviews. The data collection occurred during the spring and summer of 2012.

Usage Log Data
We instrumented LemonAid to gather a time-stamped log including: entry into help mode, exit from help mode (by clicking on the exit help button or leaving the current page), selections of elements on the screen, selections of questions, helpfulness of answers viewed, votes on questions, flags used to mark questions as spam, content of submitted questions and/or answers, content of search keywords, and lists of users subscribed to receive answers to questions. LemonAid gathered all of the foregoing data anonymously, meaning that we could not ascribe activity in the help mode to individual users, only users in aggregate.

Exit Survey
When users exited the help mode for the first time (detected by LemonAid's browser cookie), they were presented with a brief exit survey. Users had the option of being included in a $50 gift card drawing as an incentive to participate and could win additional prizes if they wanted to participate in a follow-up interview. Survey responses and identities of the respondents were not linked.

The survey asked users three questions about their experience, each scored on a 7-point Likert scale, ranging from "strongly disagree" to "strongly agree." These questions were: (1) "I found something helpful;" (2) "I would use this help feature again," and (3) "the help feature was intuitive." We also asked users closed demographic questions about their role (e.g., staff, student, faculty, etc.), their frequency of website use (ranging from daily, few times a week, few times during the quarter, or few times a year), and preferred method of getting help for web applications ("using the site's help", "searching online", "asking someone", "calling a support line", or "trying on my own"). We required users to provide a response for each of these questions to maintain consistency in the responses.

One-on-One Interviews with Users and Software Teams
Using the list of survey respondents who wished to be interviewed (~56% of all respondents), we recruited interviewees from each deployment site. We conducted all interviews on campus, except four were on the phone.

The interviews were semi-structured and lasted around 30 minutes on average. Our goal was to understand users' help seeking behavior in the context of modern help tools and how users saw LemonAid fit in with these tools. Interviewees had access to LemonAid during the interview to facilitate recall. We also probed into interviewees' perceptions of LemonAid's usability and usefulness, particularly compared to other help formats. We also probed into aspects of LemonAid that the interviewees believed were confusing or could be improved.

Near the end of each deployment period, we also interviewed the software team members involved in the deployments. We began the interviews by probing into teams' perspectives on the experience of integrating LemonAid compared to other forms of help. We also asked interviewees to describe how they currently managed user feedback and whether the feedback was used to inform design decisions. We also showed them usage data captured by LemonAid's Administrative Dashboard and asked them to explore and comment on the utility of the aggregated data, if any. These interviews lasted about 45 minutes.

Analysis
To assess how LemonAid was used in practice and whether LemonAid was helpful, usable, and desirable for reuse, we used the concept of data triangulation [17] and looked for convergence in the logs, survey data, and interview data. We began the analysis by parsing the usage logs to get an overview of usage patterns and help activity in LemonAid (the main variables are shown in Table 2). For the survey responses, we investigated associations between our key outcome variables and user demographics.

To analyze the qualitative data, we audiotaped and transcribed each interview, then organizing, coding, and analyzing with qualitative data analysis software. In the first pass, we coded for data related to questions about LemonAid's usage, and end users' perceptions of utility and usability. In the next pass, we examined interviewees' descriptions of help seeking strategies using an inductive analysis approach [25,35]. This inductive analysis approach was useful for seeing patterns in the data and in identifying recurring themes across the different deployments. A similar inductive approach was also used for analyzing interviews with software team members.

LEMONAID USAGE AND USERS
The first set of results we present concern who used LemonAid and what they did with it, based on the usage data and demographic survey responses.

	LIBRARY	DEPT	EDC	RDB
Length of deployment	15 weeks	6 weeks	8 weeks	8 weeks
# of help sessions	972	185	65	40
# of element selections	2057	509	141	91
# of answer views	562	194	61	36
% of viewed answers marked as helpful	77.7%	71.0%	75.1%	69.2%
# of new questions posted	16	7	3	5
# of answer views on new questions	121	27	30	13
% of answers to new questions marked helpful	74.3%	55.6%	67.7%	57.1%
# of search queries	43	23	4	4

Table 2. Summary of usage activity across the 4 LemonAid deployments.

Role	LIBRARY (n=121)	DEPT (n=33)	EDC (n=9)	RDB (n=5)
Grad Student	40.0%	29.0%	--	--
Undergrad	39.1%	35.5%	--	--
Staff	8.2%	16.1%	100%	100%
Faculty	6.6%	3.2%		
Other	8.0%	12.9%	--	--

Table 3. Distribution of survey respondents' roles. EDC and RDB were restricted access sites accessed only by staff.

LemonAid Usage Activity

To describe the usage activity in LemonAid, we use an individual *help session* as the unit of analysis since we were unable to track individual users. We define a help session as an episode consisting of all the activities that occurred when a user entered the help mode and pressed exit or left the page. In our analysis, we ignored all of the help sessions where entry into the help mode was immediately followed by an exit out of the help mode or the current page. About 20% of the logs followed this pattern.

LemonAid's deployments across the four different sites resulted in over 1,200 help sessions. Table 2 summarizes this activity for each deployment site. Since the LIBRARY site was the longest deployment and had the largest user base, it yielded the highest usage of LemonAid (972 help sessions). The LIBRARY logs showed that 16 new questions were added during the deployment, constituting about 1.6% of the total help sessions. We also found that no end users answered a question; library staff answered all new questions. We did find that the 16 new questions asked by users received 121 views, accounting for about 21.5% of all question views and 74.3% of the corresponding answers were marked as helpful. Prior work has shown that this level of activity is typical of what occurs in technical forums [21,22,24] and more broadly on the Internet (e.g., "*the 1% rule*" [2]) where most users are consumers of online content rather than contributors.

Demographics of Survey and Interview Participants

We received 168 responses to exit surveys across the different deployments (with the highest number of responses from the LIBRARY site (*n*=121, response rate=12%), followed by the DEPT (*n*=33, response rate=18%), EDC (*n*=9, response rate=12%), and RDB (*n*=5, response rate=13%). A breakdown of the survey respondents based on their role is shown in Table 3. Note that the EDC and RDB sites required authentication and had respondents who were exclusively staff members at the university or an affiliated institution. The LIBRARY and the DEPT sites were largely public sites (with only some features requiring authentication) and garnered responses more broadly from the university's graduate and undergraduate students, staff, faculty, and external visitors. We interviewed 36 users across the four deployments, with the majority of interviewees being users of the LIBRARY site (18), followed by DEPT (8), EDC (6), and RDB (4). Our interviewees represented a similar proportion of roles as the

survey respondents, with graduate students and staff constituting the majority. Table 4 shows the breakdown of the interviewees based on their role. The graduate and undergraduate students came from a range of departments including Psychology, Drama, Communications, Biostatistics, Political Science, Computer Science, Information Science, among others. The staff were primarily from the School of Medicine and affiliated clinical research centers.

Grad Student	31.4%
Staff	25.7%
Undergrad	22.9%
Admin Staff	17.1%
Faculty	2.9%

Table 4. Interviewee roles.

Our survey data for the LIBRARY site showed that the majority of respondents were regular site users, with 70% reporting visiting site the daily or few times a week (Figure 2a). To get a sense of how users normally found help on web sites, we asked respondents to indicate their preferred method of finding help. For the LIBRARY deployment, users' preferred form of help (Figure 2b) differed significantly (Pearson $\chi^2_{(4,N=121)}$=36.15, *p*<.0001), with "trying on my own" and "using the site help" accounting for most responses (35.5% and 28.9%, respectively).

END USERS' PERSPECTIVES ON LEMONAID

We present our main findings about end users' perceptions of LemonAid by combining quantitative analyses from logs and surveys and qualitative findings from interviews.

Helpfulness of LemonAid

To assess the helpfulness of LemonAid, we sought convergence in our three sources of data: (1) helpfulness of Q&A selections captured during actual use (Figure 1.3); (2) data collected from an exit survey which explicitly asked whether users found something helpful; and, (3) interview data that shed light on why or why not something was helpful.
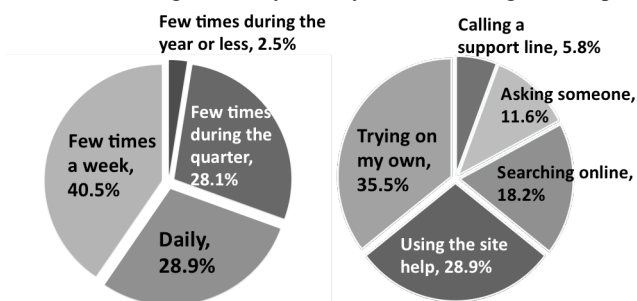

Figure 2. (a) Frequency of site use among LIBRARY users; (b) Preferred method of help seeking among LIBRARY users.

(a) "I found something helpful"　　(b) "I found the help feature to be intuitive"　　(c) "I would use the help feature again"
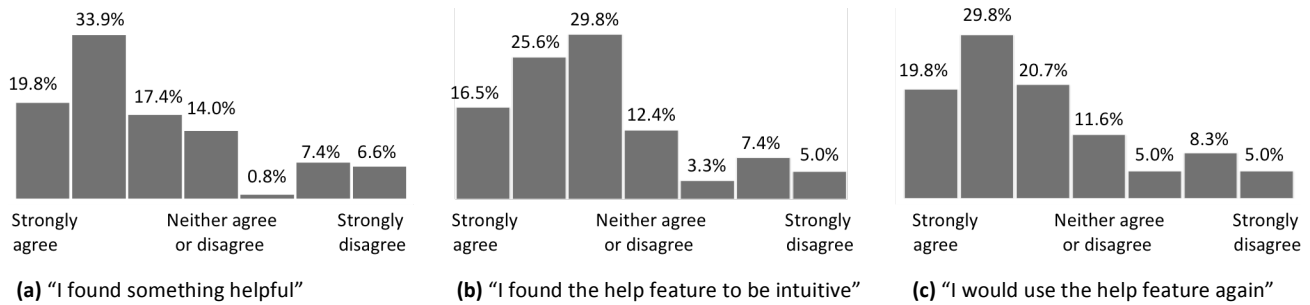
Figure 3. Distribution of LIBRARY users' survey responses.

Across all four deployment logs, we found that users selected "helpful" for over 73.2% of the viewed answers (LIBRARY=77.7%, DEPT=71.0%, EDC=75.1%, RDB=69.2%, Table 2). There is some noise in this log data—some of our interviewees indicated that they were simply browsing questions and not looking for help on a particular topic, but were forced to assess the helpfulness of the answer set anyway. Still, our exit survey data largely corroborates the log data: 70.6% of respondents on average indicated having found something helpful during the use of LemonAid (LIBRARY=71.1%, DEPT=65.6%, EDC=77.8%, RDB= 80.0%). The distribution of the survey responses from the LIBRARY site is shown in Figure 3a. A majority of users felt LemonAid was helpful (Pearson $\chi^2_{(6,N=121)}$=60.25, $p<.0001$). Only 14.9% of respondents disagreed at some level that LemonAid was helpful, whereas 71.1% agreed at some level.

We also analyzed whether certain types of users were more likely to find LemonAid helpful. For the LIBRARY data, we found that respondents' frequency of site use was correlated significantly and negatively with whether they felt LemonAid was helpful (Spearman ρ=-.23, N=121, $p<.01$), meaning users who indicated visiting the site more frequently were less likely to find something helpful through LemonAid. Our interview data corroborates this finding as the more frequent users indicated that they were already familiar with the site's main features and many of the existing help questions were "too basic" for them. In contrast, the interviews consistently revealed that new or less frequent site users found LemonAid to be helpful for learning about unfamiliar parts of the interface:

Well, what I liked about it is that…you might find a question that you didn't even know you had, so I like that…I was just hovering and looking at stuff and somebody asked a question about audio books, and it reminded me, "oh yeah, I need to get an audio book." So I like that about it, so I think it's good in that when you don't have a specific question or you are just looking at what other people are asking, then it's really helpful. (p5)

We analyzed the relationship between users' preferred method for finding help and helpfulness of LemonAid for the LIBRARY data (shown in Table 5). Although this association was not significant (Spearman ρ=.06, N=121, p=.54), we did learn from the interviews that users who normally preferred to find help using the site's existing help through FAQs or documentation found LemonAid's contextual Q&A approach to be more helpful. Although Table 5 shows that users who normally preferred to find help by asking someone were less

likely to find LemonAid helpful, this accounted for only about 12% of users overall who asked someone for help (Figure 2b). Many of our interviewees felt that there was often a social cost to asking someone for help. LemonAid, in contrast, allowed users to learn from other people's experiences without having to bother anyone or lose face:

I think that students nowadays like to help themselves a lot more. In terms of figuring out a website, it can be kind of embarrassing to not know how to get around a website. So I think the nice thing about this [LemonAid] is that it's 24/7 and it doesn't require that stepping out of what maybe your comfort zone, you know. If English is your second language, or you're really new…and you're nervous walking up to the librarian at the library, or someone's just not available when you want a question answered, I think it [LemonAid] can be a real positive experience…(p16)

Still, many of our interviewees noted that in-person explanations were richer and allowed opportunity for instant back-and-forth than textual, asynchronous Q&A. Some interviewees also revealed that even if they did not ask others for help, they were sometimes on the *other* side answering technical questions for less tech-savvy family members or friends. In these cases, the interviewees felt that hands-on help was the most preferred form of help (regardless of the available alternatives that improved help retrieval):

They really just want somebody to do it for them. They really don't want to learn, it they just want it done so they'd rather just call someone and have it talked through or done for them even though there's a help function that'll explain it step by step... [p19]

Perhaps more visual, interactive, and synchronous forms of Q&A in future versions of LemonAid could make the content appear more "hands on" and be more helpful for such users.

Usability of LemonAid

Survey responses across *all* deployments indicated that on average 72.5% of users agreed to some extent that the LemonAid help feature was intuitive (LIBRARY=71.9%, DEPT=68.8%, EDC=75.0%, RDB=100.0%). The distribution of the responses from the library deployment is shown in Figure 3b. Users expressed significant agreement as to

	Agree	Neither Agree/Disagree	Disagree
Asking someone	64.3%	14.3%	21.4%
Calling a support line	85.7%	0.0%	14.3%
Searching online	77.3%	4.5%	18.2%
Trying on my own	58.1%	23.3%	18.6%
Using the site help	82.9%	11.4%	5.7%

Table 5. Cross tabulation of LIBRARY users' responses to "preferred method of help" vs. "*I found something helpful*"

LemonAid's intuitiveness (Pearson $\chi^2_{(6,N=121)}$=44.74, p<.0001), with a majority of respondents indicating that they felt LemonAid was intuitive (71.9%).

Our interviews also revealed that most users found the help interface to be easy to understand even though it was their first time using this crowdsourced contextual help tool. One consistent theme was that users appreciated having direct access to help content without having to visit a separate interface, unlike most other help tools:

I liked the visual overlay. I think for a second I thought, "Whoa!" I'm not searching through a separate site to get to the help screen like [default help] where a different window pops up and you scroll through topics…I think that [LemonAid] is extremely intuitive. Having the help topics spread out by the areas in which I would need them as opposed to having a separate window open... (p6)

This finding is consistent with other studies that show that switching modes between an application and help often causes users to forget why help was requested in the first place and it takes them longer to return to their task [9]. Users also liked that they could simply turn off the help mode when finished with their help-seeking task and could return back to the application. But, some users mentioned that if they were working on a complex, multi-step task, going back and forth in the help mode could be distracting and having keyboard short-cuts perhaps would be useful.

Potential Reuse of LemonAid

As our survey and interviews primarily probed the first use of LemonAid, we also wanted to know if users were likely to use it again. Our survey results (Figure 3c) showed that 72.3% of respondents indicated they would use LemonAid again (Pearson $\chi^2_{(6,N=121)}$=44.74, p<.0001): (LIBRARY=70.3%, DEPT=81.3%, EDC=66.7%, RDB=100.0%). Whether users found LemonAid helpful correlated significantly and positively with whether they felt they would reuse LemonAid again (Spearman ρ=.63, N=121, p<.001).

These results are a strong contrast to prior work that has shown that users are frustrated with help systems and fear clicking on help [29]. In fact, about two-thirds of our interviewees said that they would like to see LemonAid even on other sites and desktop applications. For example, one interviewee recounted frustration with a site and its internal help tools she used recently, wishing that she could instead access help via LemonAid on that site:

You search [on the site] and it gives you three very different things. The only two hits have nothing to do with what I want. If there was a LemonAid type help for each section it would help… I could type in the keywords into LemonAid and see how many other administrators across campus have the same question or other questions…that would be helpful! (p21)

Utility of LemonAid Compared to Other Forms of Help

In addition to getting insights about LemonAid's usage, our interviews also focused on understanding how LemonAid compared with modern help tools such as built-in help, FAQs, online searching, and discussion forums.

LemonAid Provides More Relevant Help Content

The majority of interviewees expressed frustration in using current forms of software help. Although a third of the users said they consulted built-in help, the vast majority of users completely avoided consulting help because they feared seeing "long pages of text." Although this behavior has been known for years [29], countless applications still offer only textual built-in help. Users also felt that the built-in help content was often outdated, even though more updated content and tips were available through online channels:

Sometimes it feels that they [tech support] don't go through and update…even if they do set up 5 standard categories, it feels like they're not looking at other message boards to see the huge need for this other thing…they're not adapting and updating that (p23)

In contrast, users felt that the help content in LemonAid was more current and relevant since it could be added and edited at any point and could represent a wide range of actual problems rather than those anticipated by software designers. Furthermore, several users mentioned that the questions and answers in LemonAid were easier to understand than official built-in FAQs that often used a lot of system-specific jargon and were presented in a long list:

FAQ's are a pain because, first of all, they don't use the same language that a normal user would most of the time but rather an internal jargon. I tried LemonAid and I could relate to the questions if I was a student or whoever. This pushes the popular questions out. An FAQ is more like an artificial taxonomy giving a lot of useless information. (p29)

LemonAid Reduces the Burden on Choosing Keywords

Although about 30% of our interviewees said they prefer to search when seeking help, they identified a number of problems with keyword-based searching. One common reason that users were not able to find relevant content was due to the mismatch between how users described their questions versus how other users or system designers may have answered the same questions (the classic "vocabulary problem" [11]). The issue of choosing relevant keywords was particularly acute for non-Native English speakers.

As seen in Table 2, only a small percentage of users ended up using the built-in LemonAid search feature (e.g., there were 43 search queries in the LIBRARY deployment or about 4.4% of all help sessions). Although it is possible that some users may not have noticed this feature, many interviewees pointed out that LemonAid reduced the need to search altogether because they could find relevant questions and answers by clicking around in the interface:

When you're looking at something in the forums, you have to know what it's called...people can call it different things. With the kind of thing you're talking about, you have to know the terms then you may not find what you're looking for…With this [LemonAid], you're like, 'Oh, I don't know what's happening here!', click the Help button, click on what's confusing you and you don't have to worry about inputting any terms, or what it's called; I like that.. (p9)

This supports the key premise behind LemonAid's selection-based retrieval of relevant results, that users will be more

able to choose relevant keywords and terms in the UI than to come up with keywords themselves to find help [8].

LemonAid Facilitates In-Context Social Learning

Another recurring theme in our interviews was that users appreciated the social aspects of retrieving help through LemonAid—for example, being able to browse through other users' questions, answers, and votes:

It's sometimes nice to hear what other people's experience is. Sometimes it can be kind of validating. If I'm confused about something, it's nice to know that other people have been similarly confused. (p6)

Users also felt that the discovery of new information or tips through LemonAid was valuable because the content came from other users or staff and not a predefined document. For example, one user of the LIBRARY site explained:

I think [LemonAid] may save time if you are doing research or something like that so you have the ability to get input from other sources very quickly at your fingertips. To me, it's almost like a gold mine in that sense because it leads me into a direction that I may have not been able to get just going through traditional methods. (p14)

Despite the advantages of social learning, some users also raised concerns about the authority and quality of answers posted on forums (and in LemonAid in the future):

You never know the technical level of people in the forum. So sometimes, there's great help…other times, there are people who barely figure out how to turn the thing on, and they're looking for very basic help on the topic, and it's very difficult to differentiate until you just read sometimes pages of content… (p3)

Another user described the frustration of wading through repetitive questions and comments on forums:

…you have to sort through a lot of stuff to get there [answer in a fourm]…and there's repetition, so part of that depends on how well the forum is moderated and how well it's cleaned up so that repetitive answers or answers that are just slightly different but still the same aren't just clogging up the numbers of response... [p11]

Although few users contributed questions and answers in our deployments, it is possible that problems that impede social learning on forums could affect LemonAid as well. However, as some users pointed out, because users invoked LemonAid from within the application rather than from a separate forum, people were perhaps more cautious about posting content because the help seemed "official," as it was overlaid on the application itself. LemonAid perhaps also succeeds in limiting the number of repetitive "I have the same question" comments because users tended to find relevant questions and could express the same sentiment by clicking on *me-too*, rather than posting a new redundant question. In fact, none of the new 16 questions posted on the LIBRARY site were duplicate questions and users voted "me-too" 63 times.

SOFTWARE TEAMS' PERSPECTIVES ON LEMONAID

In addition to understanding end users' perspectives on LemonAid, we also wanted to gauge the reactions of the teams who deployed LemonAid. For each of the deployments, we interviewed the lead developer who handled the integration of LemonAid, and for the LIBRARY and DEPT deployments, we also interviewed one site administrator and two support staff (7 people in total).

Motivation for Integrating LemonAid

Although every software team that we worked with had different motivations for integrating LemonAid, a common theme was that with small teams, it was difficult to offer one-on-one support. Although the sites already offered some forms of built-in help, their developers felt that LemonAid could improve the discoverability of existing help, especially at locations in the interface where users were more likely to need it. Another motivation was curiosity about what kind of questions users would *want* to know answers to.

A more practical motivation was that the developers did not have to change the underlying code of their website:

What was appealing [about LemonAid] was that it was an overlay on top of it [the site] and since you're not requiring me to write any code, I think that's what sold it to me, I didn't have to write any code. It's really easy to do…(T02)

Deploying and Moderating LemonAid

Two teams were initially concerned about the possibility of spam, especially since the tool did not require a login. Despite the initial concerns, we found that spam was not a major issue during the deployments—only 5 questions were marked as spam across the four deployments over several weeks. Upon inspection, we found that 4 of these 5 questions were originally put in by staff and were not actually spam; 1 other question was an incomplete sentence that was marked by a staff member to flag it for removal.

Similar to the concerns of end users, administrators were also concerned about maintaining the quality and accuracy of answers. Furthermore, administrators also had concerns about the "public defacing" that was possible with LemonAid's content overlaid on the site:

[Help through LemonAid] is more on the site…and I really like that. The only thing is that…the information needs to be 100% correct and the people who are able to put in answers to the questions need to really make [sure] that it is [correct], you know, because of liability and lawsuits. (T05)

Another administrator mentioned that he once had to intervene in a discussion on a user mailing list when he sensed that users' answers were projecting incorrect information. He would consider his job in administering LemonAid to be no different:

It definitely takes the load off of the administrators to have users helping each other, right? But you do have to monitor and make sure that they are giving the correct information. I've already seen a few occasions where I had to clarify something on an email [in a mailing list] that some user sent out. (T03)

Utility of LemonAid Analytics Data

Teams had few formal methods for learning about users after deployment. The common theme was gathering basic usage analytics data through services such as Google Analytics. Although teams collected support questions through emails or

phone calls, there was no systematic process that teams used to inform designs based on this data. This is consistent with studies that show the disconnect between software support and usability engineering [7].

After asking teams about these existing practices, we showed them the LemonAid data for their site in the form of a basic dashboard, like the one found in web-based analytics tools. Most of the interviewees felt that they were not able to obtain this type of data from anywhere else and LemonAid would be a useful way to augment existing usage-based analytics services. For example, one software developer pointed out that unlike LemonAid, data from tools such as Google Analytics did not reveal users' intentions in terms of why users were clicking on certain areas or spending more time on a particular page:

I think a harder thing to get at in regular analytics is what people actually struggle with…you can get the stuff that they like, you know the stuff that's popular or works well…but stuff that people are confused about, you don't get through regular analytics…and this LemonAid is useful because it's not just about numbers…it's showing the actual questions that people have…and phrased in a way that they [users] would post these questions, in their language…(T04)

Another administrator was particularly excited to see data about what users considered to be relevant questions based on where they clicked and what they voted on:

I think from the useful perspective, just seeing what questions people have on what elements and what are being asked. Because that kind of informs us where we might have some problems, lack of understanding, or where we can do some PR. And that's a lot of where we need some help…how do we make [the site] more accessible? (T07)

Although some team members were enthusiastic about using this data to argue for design and content changes, they also felt that it would be most useful over a longer period of deployment and with greater user participation.

DISCUSSION

The primary goal of this field study was to investigate how users perceive the usability, helpfulness, and reuse value of crowdsourced contextual help in real tasks. Although we found that LemonAid was helpful, intuitive, and desirable for reuse for over 70% of end users across all the deployment sites, our findings point to several social aspects of generating and maintaining help content that should be addressed in the design of crowdsourced contextual help systems.

The majority of concerns raised by end users were about the timeliness, quality, and authority of answers in the long run. As mentioned in our results, frequent site users were less likely to find LemonAid to be helpful than new users. The frequent users often had more advanced questions but were not sure if and when their questions would be answered and who would be providing the answers. Even though staff moderators were ready to promptly answer users' questions as they emerged, the availability and role of moderators in the help interaction could perhaps be conveyed more clearly to users in the interface.

In the current set of deployments, the host teams were able to devote staff time and resources to monitor the questions and provide answers. It may be that to sustain the same level of quality in answers, a long-term commitment from the host teams would be necessary. Since many modern organizations have already opted to create peer-to-peer support forums [22,33], perhaps engaging with users through crowdsourced contextual help will be a natural extension. In particular, the teams felt that investing in one-to-many support is more efficient and provides greater cost-savings in the long run compared to supporting users one-on-one.

Still, in many cases, the true value of crowdsourced help may be from users helping each other, particularly when staff moderators are not the power users of the application. Future work in crowdsourced contextual help should explore strategies for encouraging the contributions of these power users, reducing the burden on staff while maintaining the authority and quality of answers. For example, the badges, awards and leaderboards that help make sites such as *Stack Overflow* successful [23] might translate well to crowdsourced contextual help systems.

Although the web offers a plethora of resources for software help, as seen in our results, about two-thirds of users still preferred to find help by (1) trying on their own, or, (2) using the site's help. This finding suggests that users overall are reluctant to leave the interface when seeking help. Our results further suggest that the majority of users who benefit from modern forms of help (e.g., searching, forums) are more likely to be tech-savvy users. Users who are less tech-savvy are more likely to need help, but they are also less likely to search for help on the web or on forums. Thus, we believe that there is potential in further understanding users' software help-seeking behavior in different contexts and in exploring strategies for delivering more relevant help in the context of the users' tasks. LemonAid and other crowdsourced contextual help systems offer only one step in this direction.

Study Limitations

Our study has several limitations that should be taken into account when interpreting our findings. First, the field study method inherently lacks the control and precision that could be attained in a controlled setting [36]. There is also a sampling bias given that all of our participants had some level of university education and our deployment sites were all hosted within one U.S. university. Some of the initial usage at each site was possibly influenced by a novelty effect as announcements were sent out to users to advertise the feature. Administrators for each site did point out, however, that the number of LemonAid-based help sessions overall were similar to the number of help requests they receive on average. The way the interview data was collected and coded affects its interpretation and there may be alternative explanations.

CONCLUSION

Our field study of LemonAid contributes the first field evaluation of crowdsourced contextual help, showing that LemonAid and similar help systems have a strong potential to

both simplify how users find software help, but also provide novel data to software designers about an applications' usability and utility. This study shows that, while difficult, field evaluations can be quite valuable in demonstrating not only that an innovation is valuable, but why it is valuable in the context of social and organizational factors.

ACKNOWLEDGEMENTS

REFERENCES

1. Ackerman, M.S. Augmenting organizational memory: a field study of answer garden. *ACM Trans. Inf. Syst. 16*, 3 (1998), 203–224.

2. Arthur, C. What is the 1% rule? *The Guardian*, 2006.

3. Brandt, J., Dontcheva, M., Weskamp, M., and Klemmer, S.R. Example-centric programming: integrating web search into the development environment. *ACM CHI* (2010), 513–522.

4. Brown, B., Reeves, S., and Sherwood, S. Into the wild: challenges and opportunities for field trial methods. *ACM CHI* (2011), 1657–1666.

5. Carroll, J.M. and Rosson, M.B. *Paradox of the active user.* The MIT Press, 1987.

6. Chilana, P.K., Grossman, T., and Fitzmaurice, G. Modern software product support processes and the usage of multimedia formats. *ACM CHI* (2011), 3093–3102.

7. Chilana, P.K., Ko, A.J., Wobbrock, J.O., Grossman, T., and Fitzmaurice, G. Post-deployment usability: a survey of current practices. *ACM CHI* (2011), 2243–2246.

8. Chilana, P.K., Ko, A.J., and Wobbrock, J.O. LemonAid: selection-based crowdsourced contextual help for web applications. *ACM CHI* (2012), 1549–1558.

9. Clark, I.A. Software simulation as a tool for usable product design. *IBM Systems Journal 20*, 3 (1981), 272–293.

10. Das, A. Knowledge and productivity in technical support work. *Management Science 49*, 4 (2003), 416–431.

11. Furnas, G.W., Landauer, T.K., Gomez, L.M., and Dumais, S.T. The vocabulary problem in human-system communication. *CACM 30,* 11 (1987), 964–971.

12. Goodall, S.D. Online help: a part of documentation. *ACM SIGDOC* (1992), 169–174.

13. Greene, J.C. and Caracelli, V.J. Defining and describing the paradigm issue in mixed-method evaluation. *New Directions for Evaluation 1997*, 74 (1997), 5–17.

14. Grudin, J. Why CSCW applications fail: problems in the design and evaluationof organizational interfaces. *ACM CSCW* (1988), 85–93.

15. Halverson, C.A., Erickson, T., and Ackerman, M.S. Behind the help desk: evolution of a knowledge management system in a large organization. *ACM CSCW* (2004), 304–313.

16. Jiang, W., Hu, C., Pasupathy, S., Kanevsky, A., Li, Z., and Zhou, Y. Understanding customer problem troubleshooting from storage system logs. *USENIX* (2009), 43–56.

17. Jick, T.D. Mixing qualitative and quantitative methods: Triangulation in action. *Admin Sci Quat 24*, 4 (1979), 602–11.

18. Kajko-Mattsson, M. Problems within front-end support. *J Software Maintenance & Evolution 16*, 4-5 (2004), 309–29.

19. Kearsley, G. *Online help systems: design and implementation*. Ablex Publishing Corp., 1988.

20. Klein, H.K. and Myers, M.D. A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Q. 23*, 1 (1999), 67–93.

21. Ko, A.J. and Chilana, P.K. How power users help and hinder open bug reporting. *ACM CHI* (2010), 1665–1674.

22. Lakhani, K.R. and Von Hippel, E. How open source software works:"free" user-to-user assistance. *Research policy 32*, 6 (2003), 923–943.

23. Mamykina, L., Manoim, B., Mittal, M., Hripcsak, G., and Hartmann, B. Design lessons from the fastest q&a site in the west. *ACM CHI* (2011), 2857–2866.

24. Matejka, J., Grossman, T., and Fitzmaurice, G. IP-QAT: In-product questions, answers, & tips. *ACM UIST* (2011), 175-84.

25. Miles, M.B. and Huberman, A.M. *Qualitative Data Analysis: An Expanded Sourcebook*. Sage Pub, 1994.

26. Negash, S., Ryan, T., and Igbaria, M. Quality and effectiveness in web-based customer support systems. *Information & Management 40*, 8 (2003), 757–768.

27. Novick, D.G., Elizalde, E., and Bean, N. Toward a more accurate view of when and how people seek help with computer applications. *ACM SIGDOC* (2007), 95–102.

28. Olsen Jr, Dan R. Evaluating user interface systems research. *ACM UIST* (2007), 251-258.

29. Rettig, M. Nobody reads documentation. *Commun. ACM 34*, 7 (1991), 19–24.

30. Rieman, J. A field study of exploratory learning strategies. *ACM Trans. Comput.-Hum. Interact. 3*, 3 (1996), 189–218.

31. Sellen, A. and Nicol, A. Building user-centered on-line help. *Human-computer interaction*, Morgan Kaufmann (1995), 718–723.

32. Singh, V. and Twidale, M.B. The confusion of crowds: non-dyadic help interactions. *ACM CSCW* (2008), 699–702.

33. Steehouder, M.F. Beyond technical documentation: users helping each other. *IEEE IPCC* (2002), 489– 499.

34. Stevens, G. and Wiedenhöfer, T. CHIC - a pluggable solution for community help in context. *ACM NordiCHI* (2006), 212-21.

35. Strauss, A. and Corbin, J. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage Publications Inc, 1998.

36. Yin, R.K. The case study method as a tool for doing evaluation. *Current Sociology 40*, 1 (1992), 121–137.

37. TurboTax Support. http://turbotax.intuit.com/support/.