Tell Me More! Soliciting Reader Contributions to Software Tutorials

Patrick Dubois, Volodymyr Dziubak, Andrea Bunt

Department of Computer Science

University of Manitoba

ABSTRACT

Online software tutorials help a wide range of users acquire skills with complex software, but are not always easy to follow. For example, a tutorial might target users with a high skill level, or it might contain errors and omissions. Prior work has shown that user contributions, such as user comments, can add value to a tutorial. Building on this prior work, we investigate an approach to soliciting structured tutorial enhancements from tutorial readers. We illustrate this approach through a prototype called Antorial, and evaluate its impact on reader contributions through a multi-session study with 13 participants. Our findings suggest that scaffolding tutorial contributions has positive impacts on both the number and type of reader contributions. Our findings also point to design considerations for systems that aim to support community-based tutorial refinement, and suggest promising directions for future research.

Keywords: Online tutorials, user comments, categorizing.

Index Terms: H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces.

1 INTRODUCTION

When learning a new complex application, or seeking to improve their skills, users frequently consult online software tutorials [9]. While there is no shortage of such resources available, applying the workflows as described is not always straightforward. First, a given tutorial is typically written from a single perspective in terms of software knowledge and version, which is not always adequate for the tutorial's potentially wide and diverse user base. Users of varying levels of software expertise sometimes desire more explanation than the tutorial author has provided and can have difficulty adapting the instructions to their particular versions of the software [15]. Online tutorials also vary in quality [19], and can contain errors and omissions [15].

One way that the software tutorial community has attempted to address these challenges is through the comment section, where, in addition to expressing their appreciation to the author, tutorial readers post a range of tips, suggestions and corrections to the tutorial [15]. Prior work has also shown that promoting, tagging, and placing these user comments next to the tutorial improves their perceived utility [3]. How an interface might support users in *contributing* tutorial enhancements, on the other hand, remains largely unexplored.

In this paper we focus on *soliciting* tutorial enhancements from the tutorial user community. First, applying the idea of in-context, categorized tutorial comments proposed in prior work [3], we investigate a model for collaborative tutorial enhancement that 1) *elicits content-specific* tutorial reader contributions, 2) *guides* their contributions through *content-relevant note categories*, and 3) treats the contributions as first-class citizens within the tutorial by *integrating* them directly within the main tutorial.

We illustrate this approach through a prototype system called Antorial and explore its model of community tutorial enhancement from the perspective of an individual contributor. Specifically, we describe a multi-session study where 13 Photoshop experts completed and sought to improve Photoshop tutorials with both Antorial and a baseline commenting system.

Our results indicate that participants contributed more with Antorial than with the baseline system and that Antorial's scaffolding might have positive impacts on the type of notes contributed by tutorial readers. Our findings also highlight the importance of considering the sense of community associated with a tutorial and the perceived social costs of contributing.

In summary, our work contributes the following:

- We illustrate a model for eliciting community tutorial enhancements via structured scoped notes.
- We present results from a multi-session study demonstrating its impact on tutorial contributions.
- We derive important design considerations for systems supporting community-driven tutorial enhancement.

2 RELATED WORK

2.1 Improving Software Tutorials and Authoring

Tutorials play an integral role in supporting skill acquisition and development with complex software, hence there is a rich history of research on improving their utility. Examples include designing novel, engaging tutorial formats (e.g., [11,18]), and improving the degree of integration between a tutorial and the target application (e.g., [12,21]). Others have acknowledged the workload associated with authoring an effective tutorial, proposing a range of semi-automated or demonstration-based approaches (e.g., [7,11]) to help ease this burden.

Most directly relevant to our work are approaches to tutorial authoring and enhancement that promote or facilitate post-hoc contributions from paid crowd workers or a tutorial's user community. To enable tutorial users to more easily navigate a video tutorial's contents, prior work has used crowd workers to help segment it into individual steps [13]. Like our approach, others have explored capturing and/or leveraging the perspectives of other tutorial users. One example is allowing users to upload their own demonstrations of a tutorial as a complement to the original [16]. An evaluation showed that having multiple demonstrations available reduced users' frustration when completing a tutorial [16]. Others have enhanced tutorials via tagged user comments, "pinned" at the side of the tutorial [3]. Their evaluation demonstrated that having comments organized in this manner (in their case by the paper authors) improved users' subjective impressions of the utility of these comments [3]. We extend this body of work by focusing specifically on an approach to soliciting tutorial enhancements, and evaluate its impact on user contributions through a multi-session study.

{umdubo26, vdziubak, bunt}@cs.umanitoba.ca

16-19 May, Edmonton, Alberta, Canada

Figure 1: Antorial's interface. Notes for a specific step are categorized and displayed below the corresponding step upon demand.

2.2 Eliciting and Sharing Software Knowledge

Central to our approach is the tenet that users are willing to contribute their software knowledge, as a means of improving a software tutorial, an assumption supported by many prior studies. For example, studies have shown that some users author tutorials to showcase interesting workflows [22]. Users are already appropriating tutorial comment sections as a way of contributing additional material to complement the tutorial (e.g., corrections and tips) [15]. Studies of application- or web-page-specific Q&A systems (e.g., [5,20]) have shown that users are willing and able to answer others' application-specific questions.

We also build on findings demonstrating the value in structuring these community software-related contributions. For example, researchers have explored attaching Q&As directly to interface elements (e.g., [5,20]), categorizing tutorial comments according to their type and placing them next to the area of the tutorial that they reference [3], and projecting tweets about new tutorials onto the interface elements that they reference [8].

The value of structured and/or integrated community contributions has also been demonstrated in domains outside of learning complex software. For example, allowing learners to anchor their comments to an online video's timeline led to a greater sense of social presence and likelihood of contribution [17]. Social annotation has also been explored in the context of educational documents, where systems let students annotate documents for themselves, their peers, or the instructor (e.g., [25,27]). In these examples, students gained a greater appreciation of the material and there were instances of students helping other students, mirroring our hope that tutorial readers will help other readers. Structuring comments through categorization has been shown to increase engagement with online discussions [10] and productivity in collaborative visual analytics tasks [23].

2.3 Models of Collaborative Document Refinement

The approach to tutorial enhancement that we investigate in this work can be viewed as a form of collaborative document refinement that harnesses the wisdom of the community. Many models of collaborative document refinement have been proposed and/or extensively studied in the literature. One common approach is the Wiki model, where a group of authors collectively refines a document, a popular instance being Wikipedia. The Wiki model generally assumes a single "best" document that the community works to achieve collectively. Within this model, community members take on many (evolving) roles to help ensure that articles are held to community standards (e.g., [2,14]). Edits are also tracked to enable documents to revert to prior states in the case this collective action is not seen to produce improvements.

Other models of collaborative authoring rely on a lead author, or small set of authors, to evolve a document, with collaborators providing feedback on ways to enhance it. Prior work in this space has explored a number of supporting technologies, including obtaining quick access to feedback and/or suggestions using paid crowd workers [1] and structuring collaborator comments to enable authors to process them more effectively [26].

In this paper, we explore user contributions within a model of collaborative refinement where the objective is not to move towards a single "best" document but rather to elicit and make accessible a range of community enhancements.

3 Antorial: Eliciting Tutorial Contributions

In this section, we describe an approach to eliciting structured tutorial enhancements and illustrate it through a prototype system called Antorial. In designing Antorial, our primary goals were:

1. To *elicit* tutorial enhancements that are *content-specific*
2. To *guide* contributors towards posting *beneficial enhancements*
3. To *integrate* community notes within the tutorial as *first-class citizens*.

We describe Antorial's features in light of these design goals.

3.1 Eliciting Content-Specific Enhancements

Antorial allows users to attach *notes* to individual steps of the tutorial (see Figure 1). Antorial uses the term *note* rather than *comment* to shift the emphasis from general tutorial-relevant discussions found in comment sections (e.g., words of appreciation, requests for help on other tasks [15]) to contributions that pertain to the content of the tutorial in question.

We selected a tutorial step as the scope for notes based on prior work indicating that the vast majority of text- and image-based tutorials are step-based, most of which are clearly labelled [19]. Prior work has also explored ways to segment tutorials according to steps, for example, using crowd workers [13]. Collectively, this suggests that a step is both a conceptually meaningful unit within the tutorial community and that a tutorial could be segmented into steps either programmatically or via crowd annotation.

We chose to enable users to attach notes to *individual* steps for two reasons. A tutorial step typically contains a single instruction or related series of small instructions that produce a single effect. Second, when prototyping more flexible models (e.g., enabling users to attach notes to collections of steps) we found that our note elicitation and display interfaces were more complicated. We return to the potential for more flexible models in our Discussion.

Figure 2 shows the mechanism of contributing a note for a particular step. When the user hovers over the step they would like to contribute to, a "submit note" icon appears (Figure 2, B). Clicking this icon will then automatically select this step for the note, and start the note submission process.

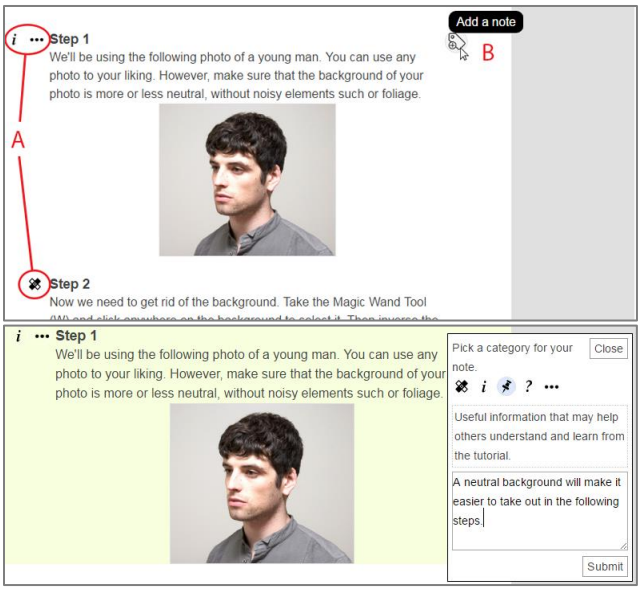

Figure 2: (Top) Interaction with Antorial: (A) clicking on icons next to a step reveals notes under that step (see Figure 1); (B) hovering over a step reveals an icon for adding a new note. (Bottom) Submitting a new note for Step 1.

3.2 Guiding User Contributions

Beyond requiring users to associate notes with a step, Antorial guides their contributions through its inclusion of note categories.

Creating Antorial's categories was an iterative process. We started with software tutorial comment categories identified in previous work [3,15]. However, we found that these categories reflected existing commenting behaviour (in free-form systems) and were not intended to guide contributors towards enhancing the tutorial's content. With this different perspective in mind, we reviewed comments from over 100 Photoshop tutorials, for a sense of content enhancements that were both common and potentially useful. For each comment, we noted potential reasons for posting it (e.g., "I can't find the command," "the tutorial seems to be missing a step"). We then applied open coding on these reasons. From these codes, we selected an initial set of categories (and created icons), which we refined through a small laboratory evaluation with five Photoshop experts. The categories and icons used in our study are found in Table 1. We emphasize that they are starting points for investigating the utility of this approach to note elicitation and that developing a robust set will likely require a complete iterative design and evaluation cycle.

As part of guiding readers towards content refinement, some categories require users to specify additional information. For example, the '*What is this?*' note category tries to guide the contributor towards notes specific to features of the software and, therefore, contributors are encouraged to provide the software or software version to which the note applies.

3.3 Integrating Notes within the Tutorial

In displaying the notes, we faced a tension between our goal of promoting notes as first-class citizens and the possibility of introducing too much clutter in the event of high note volume. We ultimately decided on a relatively conservative solution for this initial prototype that indicates the presence of notes by displaying the icons for their categories next to the tutorial step (e.g., see Figure 2, A), but keeps their content hidden until requested.

Table 1: Antorial's pre-defined note categories

Icon	Category	Definition
	Corrections	Point out an error in the tutorial, or provide a correction.
	What is this?	Clarification for something that is crucial in completing the tutorial.
	Tell me more!	Information that is not crucial for completing the tutorial, but may help someone learn something new.
	Questions	Asking a question for the tutorial. Other users may provide a reply.
	Other	Label your note with your own category.

Clicking on the category icons displays the notes directly below the step (see Figure 1). By displaying notes immediately after a step (upon demand) Antorial aims to reinforce the importance of contributor notes as complements to the originally-authored steps.

4 Multi-Session Study

We conducted a multi-session study to investigate how Antorial's model of collaborative tutorial refinement impacts contributions as well as to gather subjective impressions of the strengths and weaknesses of the general approach. We chose Photoshop as our target application owing to its popularity and for access to a wide enough pool of knowledgeable contributors.

4.1 Participants

Fourteen experienced Photoshop users participated in our study, recruited from a local university, Reddit, as well as word-of-mouth. We pre-screened participants to ensure that they had enough knowledge of Photoshop to complete and to contribute to tutorials. Given some of the challenges of eliciting self-assessments of expertise with a complex and diverse application like Photoshop (e.g., [3]), our pre-screening process consisted of asking participants to share some prior work and answer questions about some of Photoshop's features. Our call for participation also sought individuals who liked contributing their software knowledge, as our goal was to test the impact of Antorial on users who are likely to be contributors in the first place. Participants were 18 to 40 years old, and four were female. Participants received a $100 gift card.

4.2 Conditions

Our study had a within-subjects design with two conditions; participants were asked to complete and improve two tutorials with each of the following two interfaces:

1. **Antorial:** Participants used the prototype as described above.
2. **Baseline.** The baseline condition represented a conventional threaded tutorial commenting interface, where users can enter free-form contributions. As shown in Figure 3, we made one modification to the status quo by placing comments beside the tutorial rather than below. We did so to avoid unfairly biasing the results against a baseline with contributions completely deemphasized. We chose to list comments in chronological order (newest at the bottom) to make it easier for readers to follow the temporal order of discussions. Comparing Antorial to baseline systems that differ in their positioning and ordering of submissions is an area of future work.

4.3 Tutorials

As mentioned above, participants completed and worked on improving four Photoshop tutorials (two per condition).

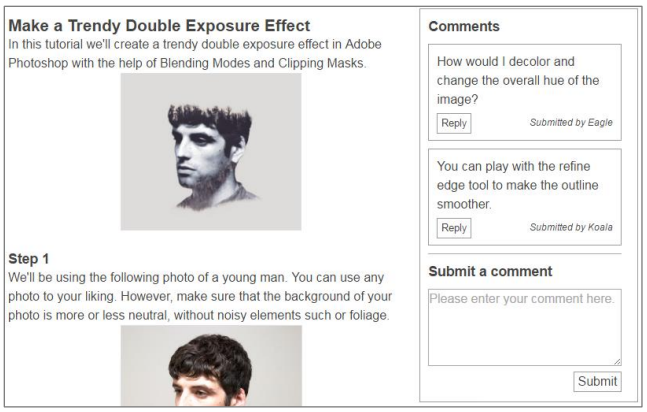

Figure 3: The baseline condition

A professional Photoshop user (a designer) assisted us in hand-picking four tutorials from publicly available online sources. To cover a range of Photoshop skills, we selected two tutorials that use parametrized commands, such as layer styles (creating metal text, adding rain to a photo), and two that use free-form tools, such as the paintbrush (making a double exposure effect, creating floral text). Being mindful of participant time, we compressed the tutorials by removing some extraneous steps; in the end, each of the tutorials had a single goal described by its title.

We paired the tutorials so that participants completed one parameterized tutorial and one free-form tutorial per condition. Our Photoshop expert judged the parameterized tutorials to be easier, so participants always started each pair with a parameterized tutorial to enable a logical progression of difficulty.

We pre-seeded each tutorial with three notes to give participants some initial sense of potential contributions (one question and two elaborations). To ensure that all participants, regardless of when they entered the study, had the same opportunities to enhance the tutorials, notes from other participants were not visible during the study.

4.4 Procedure

We began by asking participants to fill in a demographics questionnaire. To motivate participants to contribute, we introduced a small amount of deception (approved by our institution's research ethics board). Specifically, we told participants that we were considering using tutorials as part of a course and requested their help in improving them. We also indicated that we were experimenting with different interfaces.

Participants were asked to complete tutorials remotely. To provide participants with sufficient time and flexibility to complete and improve each tutorial, they were allotted three days per tutorial. As a guideline for expectations, we told participants to spend less than two hours per tutorial. When participants were finished with a tutorial, they contacted the researcher to receive access to the next one. To account for potential order effects, we counterbalanced the order of conditions across the participants. We assigned tutorial pairs to conditions in a round-robin fashion.

Prior to each condition, participants received a summary of the features of the system in lieu of a live interface demonstration. After completing the condition, participants completed a Likert-scale questionnaire (Table 2). To conclude the study, participants took part in a semi-structured interview and debriefing session, where we explained the nature of the deception.

For each participant, the study lasted from 2 to 22 days depending on how quickly they completed each tutorial, with an average length of 11 days. We estimate the total time commitment per participant to be 7-8 hours.

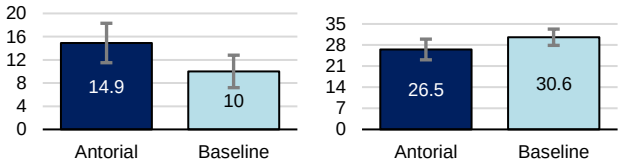

Figure 4: (Left) Mean number of notes per participant; (Right) mean number of words per note. Error bars represent standard error.

4.5 Data Collection and Analysis

Our study data consisted of participants' contributions, their responses to the post-condition questionnaires (Table 2) and their perspectives on the strengths and weaknesses of Antorial's approach collected via the semi-structured exit interview.

When analyzing the data, we coded participant contributions using an open coding scheme that we refined iteratively. We analyzed the interviews, which were transcribed in full, thematically through joint sessions among the paper authors, with continual revisits to the raw transcripts to help ensure the validity of our themes. We compared quantitative means using an RM-ANOVA with *Interface Type* (Antorial vs. Baseline) as the within-subjects factor. We also included *Interface Order* (Antorial First vs. Baseline First) as a between-subjects factor to check for potential order effects. Finally, we compared distributions for categorical variables using Pearson's CHI-squared test. We report results as significant if $p < 0.05$.

The post-session interview revealed that one participant became frustrated by the study's workload in the first condition (the baseline condition), and decided to stop contributing. We removed this participant's data from our analysis since they did not experience both conditions, leaving us with 13 full participants.

4.6 Results: Contribution Volume and Content

Over the course of the study, the 13 participants that we included in our analysis made 319 contributions: 192 with Antorial compared to 127 with Baseline. Figure 4 (Left) depicts the results on a per-participant basis. On average, participants made 14.9 contributions with Antorial (*SE*. = 3.4) compared to 10.0 with Baseline (*SE*. = 2,8), a difference found to be significant ($F_{1,11}$ = 7.32, p = 0.02, η^2 = 0.4). This represents an average increase of 49% when contributing with Antorial over Baseline. The Interface Type × Interface Order interaction effect was not significant ($F_{1,11}$ = 0.59, p = 0.48, η^2 = 0.05).

To get a sense of whether Interface Type impacted the verbosity of participants' contributions, we also analyzed the number of words per contribution. Figure 4 (Right) shows that participants did write slightly less with Antorial on a per-contribution basis (M = 26.53, SE = 3.43) than with Baseline (M = 30.59, SE = 2.7), but this difference did not reach significance ($F_{1,11}$ = 2.61, p = 0.14, η^2 = 0.192). The Interface Type × Interface Order interaction effect was also not significant ($F_{1,11}$ = 1.48, p = 0.29, η^2 = 0.119).

In examining contributions, we found that one source of additional verbosity in the baseline (where, on average, participants wrote four extra words per contribution) was that participants often had to add extra context. For example, 47% of Baseline contributions identified some step number, as the following illustrates:

> *Step 3 - If you want to hide […] layers, you can click on the eye icons (on the left side of each layer) to hide them. (P6)*

Conversely, step information is built into Antorial's notes.

In further analyzing participant contributions, we found that they differed along the following primary dimensions: type, intended audience, scope, and whether the contribution attempted

to provide potentially helpful supplemental information or simply critiqued the tutorial. We discuss each dimension in turn below.

4.6.1 Contribution Type

We identified seven types of contributions that participants aimed to make: to present an alternate workflow or method of command access, to express appreciation, to pose a question or mention a difficulty, to correct a portion of the workflow, to correct the wording or grammar, to expand an explanation, and to extend the workflow to improve the final result. Figure 5 illustrates the distribution of contribution type across the two conditions, which were significantly different according to a Pearson's Chi-Square test ($\chi^2 = 17.772$, $p = 0.007$).

We felt that all the notes submitted by our participants, apart from those identified as appreciation, could potentially be useful to at least some reader. We opted not to further quantify their quality as we found doing so was highly subjective and audience dependent. Whether an individual note ultimately adds value from a reader's perspective will depend on factors such as their expertise, interest in skill acquisition vs. task completion, etc. We therefore leave assessing the utility of these types of user-supplied notes as an area of future work.

The most common contribution type was to expand upon the explanation provided in the tutorial, which accounted for just over a third of the contributions in each condition. In examining differences in the two distributions for contribution type, there appeared to be a tradeoff between workflow corrections and grammar/wording corrections. Specifically, more Antorial contributions offered corrections to the workflow (Antorial: 13%, Baseline: 3.7%), while Baseline's corrections were more focused on the writing quality (Antorial: 1.9%, Baseline: 9.6%). The remaining contribution types were roughly equally represented across the two conditions.

4.6.2 Target Audience

We also found that contributions differed in their intended target audience. Some were clearly directly at the tutorial author, whereas others were more directed at potential tutorial followers. For example, the note below appears to be addressed to the author. It suggests an area within the tutorial where the author can expand, and goes on to describe some specific suggestions (omitted for space reasons).

> *Maybe add some pointers afterwards to what can be done to take the effect further […] (P8)*

This note, on the other hand, addresses potential readers directly, and even qualifies the level of expertise targeted:

> *For advanced users: If you convert your layer to a Smart Object before adding your filters they become "Smart Filters" that you at any time can double-click to adjust the settings. (P8)*

Most of the contributions in both interfaces were aimed at the tutorial author (Antorial: 78%, Baseline: 78.9%). This is perhaps not surprising considering the scenario we described to participants: that we were considering using these tutorials in a course and wanted their help to improve them. Additionally, there was no interaction between contributors, which may have reduced the incentive to offer advice directly to other tutorial followers. Despite these aspects of the study scenario, some participants did continue to view their contributions as a way to communicate with other tutorial followers (roughly 9% of contributions in each condition), or worded their contributions in a way that they could target either audience (e.g., 13% in Antorial). Overall, Interface Type did not appear to impact the distribution of comments across target audience ($\chi^2 = 0.236$, $p = 0.889$).

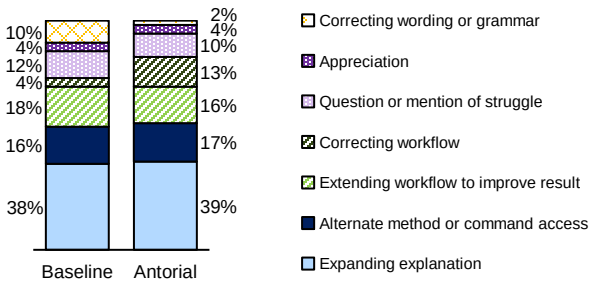

Figure 5: The distribution of comments' contribution type

4.6.3 Contribution Scope

We found the participants' contributions had the following scopes: the entire tutorial, multiple steps, a single step, a specific word(s), or a new step. Despite Antorial's explicit step-based elicitation method, the difference in scope distribution across the two conditions did not reach significance ($\chi^2 = 6.263$, $p = 0.180$), with the most common scope in both conditions being a single step (79% with Antorial and 69% with Baseline). Interestingly, participants did write notes with wider scope even with Antorial. For example, to address multiple steps, participants would list additional affected steps, as they would in Baseline. To recommend an additional step, they would submit a note on the step following the location they would place the new step. We return to the possibility of explicitly supporting these additional scopes in our Discussion.

4.6.4 Supplement vs. Critique

We found that contributions differed in the degree to which they simply pointed out issues or actually helped to work towards solutions. A *supplemental* note provided specific directions on ways to improve or correct the workflow, or the tutorial's explanations. For example:

> *It is better to explain at this point why you are making some colour and focus adjustments to the photo. In this case, the photo has warm tones which is associated to bright and sunny, therefore colour adjustment is necessary. (P10)*

The second type, a *critique*, did not specify exact ways to enhance the tutorial, but pointed out places where the improvement could/should take place:

> *You might want to clarify in greater detail how you masked out the dark beige portions […]. (P12)*

Most contributions were in fact supplemental (74% with Antorial and 79% with Baseline). The difference in distribution across the two conditions, however, was not significant ($\chi^2 = 1.023$, $p = 0.312$).

When examining these contributions according to their target audience we found that all those targeted at tutorial readers offered potential advice as opposed to simply critiquing. Whereas those targeted at tutorial authors contained a mix of critiques (28%) and supplemental notes (72%). This suggests that there may be value in further emphasizing the role of notes in communicating with other readers. We return to the potential role of audience and how systems might further promote reader-centric enhancements in our Discussion.

4.6.5 Antorial Categories

Antorial supplies its own note types (i.e., its categories) in an attempt to guide the tutorial notes. Using a strict definition of correctness, we coded an Antorial note as correctly categorized if and only if the participant selected the *most appropriate* category

Table 2: Post-condition Likert-scale questionnaire results (1 = Strongly Disagree, 5 = Strongly Agree)

Statement	Antorial Mean (*SE*)	Baseline Mean (*SE*)	$F_{1,11}$	*p*	η^2
I felt encouraged or motivated to post comments	4.45 (0.19)	4.38 (0.26)	0.31	0.59	0.027
I felt like it took a lot of work to post a new comment	2.16 (0.33)	1.83 (0.33)	1.19	0.30	0.098
I felt comfortable posting comments	4.93 (0.08)	4.85 (0.11)	1.19	0.30	0.097
I felt confident that the comments I posted will be useful	4.79 (0.11)	4.85 (0.11)	0.22	0.65	0.020

given our definitions in Table 1. Using this strict criteria, we found that most notes (55%) where not, in fact, categorized as we would have envisioned. As we describe in our next section, participants had highly individualistic responses towards the categories, with some not wanting to expend the effort to categorize their notes.

4.7 Results: Subjective Response

The post-condition questionnaires show that participants were generally comfortable and motivated to post comments with both interfaces (Table 2), with little difference between conditions.

In terms of overall preference, participants' responses were split: five preferred Antorial, five preferred Baseline, and three participants were neutral. In what follows, we discuss some important considerations that participants raised during the interviews for systems that aim to support collaborative tutorial enhancement. We also describe how these considerations influenced participants' preferences for the systems investigated.

4.7.1 Supporting Specificity and Context through Step-Based Notes

Most participants felt that Antorial's per-step commenting model was more tailored towards eliciting enhancements than the free-form comments in the baseline system. They found that the model helped them focus on specific enhancements and reduced the amount of additional context they had to provide. The following two quotes are representative of these sentiments:

> *Then I can focus on the step I'm talking about, if you noticed for the other interface whenever I made a comment I would say something like "In Step 7, this this this." But for this one I don't have to do that. Everyone knows this is the step we're talking about. (P3)*

> *[Antorial] is more specific than [the Baseline] […] The [baseline] is just for regular stuff, [Antorial] is much more specific. (P1)*

4.7.2 Guiding Contributions through Categories

Participants' views on structuring tutorial contributions with pre-defined categories diverged. Some participants did not mind specifying categories, as long as it makes it more convenient for other people, particularly other contributors:

> *For the [contributor], it keeps suggestions and information [more] organized […]. In the eyes of the reader, […] it will be easier for the people answering [comments and questions]. People have asked and you'll see if they have been answered already. You won't have to answer twice because you can find things. (P10)*

Several participants also said that specifying a category helped them craft their notes and understand which contributions would be considered valuable:

> *[The] icon guided me to make my comments. […] I would go here and say 'maybe I need to look for something I don't understand, or something to correct' (P3)*

Four out of the five participants who preferred the baseline system, however, specifically complained about the overhead that the categories introduced. These participants felt that the categories were not completely clear, that selecting one required too much thought, and/or that they restricted contributions in way that was not beneficial:

> *I knew what my comment was but I really had a hard time figuring out what the category should be. I actually looked at the other comments to see what other people were saying. Maybe if I could get an idea [from the comments] maybe I'd learn, but that wasn't the case. (P11)*

> *I prefer it when something is more open because it doesn't make a person feel like they're restricted by what they should or shouldn't put down […] sometimes those [categories] would be in the same comment I was trying to make and I think "ok, should I split this up into three?" and I really didn't like doing that and would [rather just] put everything there. (P2)*

In relation to P11's quote above, our study method made it difficult for participants to leverage community categorization practices, in that we pre-seeded each tutorial with only three notes. It is possible that more notes would help contributors better establish and understand norms.

Out of the people who saw benefit in specifying categories, not all of them felt that they understood the categories provided in Antorial, but were content with their workarounds, such as selecting "Other" when they did not feel that their note fit into a particular category. For example, P6 used only the '*Tell me more!*' category for all his 10 notes.

Overall, our sense from the interviews is that the categories could certainly be refined to reduce confusion, but the extent to which this would alleviate all of participants' concerns is an open question. Some participants who preferred free-form note entry, simply wanted to express their tutorial input and move on.

4.7.3 Sense of Community vs. Social Cost

Our interviews also revealed some potentially interesting tradeoffs with having contributions permanently visible. In the baseline system comments were always displayed next to the tutorial (see Figure 3). With Antorial, users had to click the category icons to reveal the notes below a step (see Figure 2, top). On one hand, several participants indicated that the always-visible notes in the baseline system provided them with a better sense of the community. For example, one participant indicated that the baseline system suggests an active discussion with other artists:

> *More discussion with [Baseline]. It feels more like the commenting section - it feels like you'd have more discussion with the other artists than on [Antorial]. Um, the replies back and forth with the artists that you'd normally get with Reddit with the tossup of ideas. (P9)*

Another participant said that this sense of community gives them confidence that their questions might get answered quickly:

> *If I'm going to be asking questions, and I have people respond also, most tutorials you watch them or follow them without having the community to support you if you don't understand something. I would feel very happy to have such [support]. (P3)*

The notion of promoted user comments providing a sense of community surrounding a tutorial has been identified in prior work [3].

On the other hand, the visibility of contributions also seemed to

make some participants concerned with the social cost of contributing. For example, a participant indicated that they were more conscious of their posting volume in the baseline system (where they posted only 5 notes) than in Antorial (where they posted 22 notes). With the baseline system, the participant said:

> *I didn't want to annoy other people and post a lot. So I combined all the things I had to say step by step into one tutorial, by the end of it. (P11)*

This is supported by this participant's contributions in Baseline, which tended to group together multiple brief points, as opposed to in Antorial, where they were more descriptive.

Exploring how properties of the systems (in addition to contribution visibility) might impact these more social aspects of tutorial contribution is an interesting avenue of future research.

4.7.4 Providing Multiple Perspectives as Motivation to Contribute

A motivation for Antorial's model of collaborative tutorial enhancement is that there is often no "correct" version of a given workflow. In our interviews, two participants said that a desire to provide an alternative perspective motivated them to contribute:

> *[…] there is not a right or wrong way to paint the sunset, you know, you see it, you make your visual translation, and that's your interpretation. And just like Photoshop, if you want to achieve an objective there are 100s of ways to do it, so one is not necessarily the right way. There are more efficient ways, but nothing is really right or wrong. (P6)*

> *[The tutorial] may be doing it one way but there might be another way. So I don't like giving one method and saying that is the only way, there [are] 1000 different ways […]. (P9)*

These quotes lend some initial support for Antorial's approach to enabling a community of tutorial users to contribute multiple perspectives.

4.7.5 Enabling General Discussions

Participants suggested that future iterations of Antorial should also explore general discussions as a complement to scoped notes. For example, a participant indicated that general comments might be more review-style, giving an overview of the tutorial:

> *I did notice other individuals who are commenting just commented about the tutorial overall which I thought was interesting, I thought they were commenting almost as a review, like oh this is an interesting way to put rain on the photo. (P11)*

Given that each tutorial was pre-seeded with the same style of note (regardless of condition), the fact that this participant viewed some notes in the baseline system as more review-oriented is interesting. This is perhaps reflective of the view that commenting sections are generally for expressing opinions. Further investigation, however, on how users might perceive comments posted in different areas is needed, as well as how providing multiple scopes might influence contribution types and volume.

5 Discussion and Future Work

Our study results indicate that participants made close to 50% more contributions with Antorial's structured approach than with a baseline commenting system. There is also some evidence that Antorial impacted the types of contribution, with a shift from wording corrections to workflow corrections. The latter type of contribution is arguably more beneficial for tutorial readers – adapting to grammar issues is likely easier than adapting to errors in the workflow. From the standpoint of community tutorial enhancement, it is encouraging that most contributions (both with Antorial and the baseline) were aimed at enhancing the tutorial as opposed to merely critiquing without suggesting a path forward.

While our evaluation is an important first step in demonstrating potential strengths and weaknesses of this type of approach to community tutorial enhancement, there are several areas of future study. First, our findings are based on 13 participants. Future studies should explore the generalizability of the approach to a broader sample of contributors, including contributors with less expertise in the target application. Following suggestions that the HCI community consider resources as a factor in assessing the suitability of a sample size [4], we note that this was a resource intensive study particularly in its time commitment. There are also aspects of our method that lack external validity, such as our use of deception to motivate contributions and that participants' contributions were not visible to each other. A longer-term public deployment is needed to assess the impact of our approach on contributions in a less constrained setting and to provide deeper insight into how different approaches to comment elicitation and display might influence both the sense of community and perceived social costs of posting. Finally, a longer-term deployment would enable us to determine whether the increased contributions we saw with Antorial were due to a novelty effect.

Our results also point to additional open research questions, including those described below.

5.1 Generalizability and Scalability of Approach

In this work, we focused on soliciting enhancements to text- and image-based software tutorials. While our study targeted Photoshop tutorials, there is hope that the approach would generalize well to a range of other types of step-based software tutorials. In terms of generalizing to video tutorials, there has been some work on eliciting corrections to learning videos [6]. Soliciting other community enhancements (beyond text-based comments) that could be readily integrated within a video represents an interesting design challenge for future work.

As our first investigation of structured tutorial contributions, we built a relatively simple system that accepts only simple text input. There are many ways to enrich this input. For example, contributors could submit rich text, images, videos, [16] voice comments [25] and gestures [25]. The way we present these could also be enhanced, such as having a showcase for images submitted as results of following the tutorial [3].

There is also the issue of scalability as the number of contributors and notes increases. In the face of a large volume of contributions, moderation strategies, both author- and community-centric ones, may need to be investigated and included.

5.2 Exploring More Flexible Postings Models

While Antorial's extra posting overhead did not appear to have a negative impact on contribution rates or content, participants' preferences for Antorial vs. the baseline system were mixed. A central issue appeared to be with Antorial's categories. Some participants felt that the categories served their intended purpose, which was to guide contributions. Others found them confusing, and unnecessary. At a minimum, our categories need refinement to make them clearer to a wider range of potential contributors. Our findings also suggest that it would be worthwhile to explore more flexible contribution models. For example, there may be value in letting contributors easily "opt out" of providing extra structure for their notes. It might also be possible to shift the cost vs. benefit tradeoff by incentivizing the process, by only integrating categorized notes into the tutorial as a "reward", with other notes relegated to a less prominent "general" pool. Moving forward, it will be important to understand how differing types

and levels of contribution structure impact both contributions and the consumption of these contributions by tutorial readers.

Antorial uses a tutorial's step as the scope for a note, and our results lend support for this decision: the most common scope across Antorial and the baseline system was an individual step and users found workarounds in Antorial for notes with a larger scope. The range of scopes that we saw in participants' notes, however, indicates that further exploration of flexible scopes is also warranted. The challenge will be to balance the expressive power of such flexibility with the extra overhead and complexity that it might introduce. For example, permitting varying degrees of note scope would likely lead to contributors having more decisions to make at posting time, and conveying varying levels of note scopes within the tutorial might also introduce more visual complexity for those seeking to complete the tutorial.

5.3 Considering the Learner and Author Perspectives

Ideally, community tutorial enhancement mechanisms should provide a way for contributors to communicate with the tutorial author and other learners. Our results indicate a mix in intended audience within our participants' notes. Our results also indicate that when contributors addressed learners, they always provided concrete attempts at tutorial enhancement rather than critiques. The question then becomes how to reinforce the idea of other learners as an audience for contributions. An approach could be to acknowledge contributors, making it clear that learners play a part in this acknowledgement. For example, tutorials could include a contributor recognition section where learners mark enhancements that they found useful. This could also be addressed through category descriptions, with categories explained in a way that emphasizes posting notes to tutorial readers.

A second concern is how a tutorial author might respond to users contributing directly to their tutorial, particularly if contributions are prominently recognized. There might be tensions between incorporating these enhancements, thereby creating a sense of community, and maintaining control and recognition for the final product. We would be interested in exploring how different points in the design space impact this tension.

6 Summary

We presented an approach to eliciting software tutorial enhancements through structured, content-specific community contributions. Through a multi-session study, we demonstrated several potential strengths and weaknesses of this approach as compared to a default free-form commenting system, in terms of its impact on user contributions and subjective response. Our results also point to several potential directions for future work, including exploring the impact of more flexible elicitation schemes, and understanding the tutorial author's perspective.

Acknowledgements

We thank Mackenzie Plowman for her input, and the National Sciences and Engineering Research Council and the University of Manitoba's GETS program for funding this research.

References

[1] M. S. Bernstein, G. Little, R. C. Miller, B. Hartmann, M. S. Ackerman, D. R. Karger, D. Crowell, K. Panovich. 2010. Soylent: A Word Processor with a Crowd Inside. Proc. UIST: 313–322.

[2] S. L. Bryant, A. Forte, A. Bruckman. 2005. Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia. Proc. GROUP: 1–10.

[3] A. Bunt, P. Dubois, B. Lafreniere, M. Terry, D. Cormack. 2014. TaggedComments: Promoting and Integrating User Comments in Online Application Tutorials. Proc. CHI: 4037–4046.

[4] K. Caine. 2016. Local Standards for Sample Size at CHI. Proc. CHI: 981–992.

[5] P. K. Chilana, A. J. Ko, J. O. Wobbrock. 2012. LemonAid: Selection-Based Crowdsourced Contextual Help for Web Applications. Proc. CHI: 1549–1558.

[6] A. Cross, M. Bayyapunedi, D. Ravindran, E. Cutrell, W. Thies. 2014. VidWiki : Enabling the Crowd to Improve the Legibility of Online Educational Videos. Proc. CSCW: 1167–1175.

[7] T. Dong, M. Dontcheva, D. Joseph, K. Karahalios, M. W. Newman, M. S. Ackerman. 2012. Discovery-based Games for Learning Software. Proc. CHI: 2083–2086.

[8] V. Dziubak, P. Dubois, A. Bunt, M. Terry. 2016. Switter: Supporting Exploration of Software Learning Materials on Social Media. Proc. DIS: 1209–1220.

[9] M. Ekstrand, W. Li, T. Grossman, J. Matejka, G. Fitzmaurice. 2011. Searching for Software Learning Resources Using Application Context. Proc. UIST: 195–204.

[10] S. Faridani, E. Bitton, K. Ryokai, K. Goldberg. 2010. Opinion Space: A Scalable Tool for Browsing Online Comments. Proc. CHI: 1175–1184.

[11] T. Grossman, J. Matejka, G. Fitzmaurice. 2010. Chronicle: Capture, Exploration, and Playback of Document Workflow Histories. Proc. UIST: 143–152.

[12] C. Kelleher, R. Pausch. 2005. Stencils-Based Tutorials: Design and Evaluation. Proc. CHI: 541–550.

[13] J. Kim, P. T. Nguyen, S. Weir, P. J. Guo, R. C. Miller, K. Z. Gajos. 2014. Crowdsourcing Step-by-Step Information Extraction to Enhance Existing How-to Videos. Proc. CHI: 4017–4026.

[14] M. Krieger, E. M. Stark, S. R. Klemmer. 2009. Coordinating Tasks on the Commons: Designing for Personal Goals, Expertise and Serendipity. Proc. CHI: 1485–1494.

[15] B. Lafreniere, A. Bunt, M. Lount, M. Terry. 2013. Understanding the Roles and Uses of Web Tutorials. Proc. ICWSM: 303–310.

[16] B. Lafreniere, T. Grossman, G. Fitzmaurice. 2013. Community Enhanced Tutorials: Improving Tutorials with Multiple Demonstrations. Proc. CHI: 1779–1788.

[17] Y. Lee, W. Lin, F. Cherng, H. Wang, C. Sung, J. King. 2015. Using Time-Anchored Peer Comments to Enhance Social Interaction in Online Educational Videos. Proc. CHI: 689–698.

[18] N. Li, M. B. Rosson. 2014. Using Annotations in Online Group Chats. Proc. CHI: 863–866.

[19] M. Lount, A. Bunt. 2014. Characterizing Web-Based Tutorials: Exploring Quality, Community, and Showcasing Strategies. Proc. SIGDOC: 6:1–6:10.

[20] J. Matejka, T. Grossman, G. Fitzmaurice. 2011. IP-QAT: In-Product Questions, Answers & Tips. Proc. UIST: 175–184.

[21] C. Nguyen, F. Liu. 2015. Making Software Tutorial Video Responsive. Proc. CHI: 1565–1568.

[22] D. Perkel, B. Herr-Stephenson. 2008. Peer pedagogy in an interest-driven community: the practices and problems of online tutorials. Proc. Media, Communication and Humanity: 1–30.

[23] W. Willett, J. Heer, J. M. Hellerstein, M. Agrawala. 2011. CommentSpace: Structured Support for Collaborative Visual Analysis. Proc. CHI: 3131–3140.

[24] H. Yang, C. Lai. 2010. Motivations of Wikipedia content contributors. Computers in Human Behavior 26, 6: 1377–1383.

[25] D. Yoon, N. Chen, B. Randles, A Cheatle, C. E. Löckenhoff, S. J. Jackson, A. Sellen, F. Guimbretière. 2016. RichReview++: Deployment of a Collaborative Multi-modal Annotation System for Instructor Feedback and Peer Discussion. Proc. CSCW: 195–205.

[26] Q. Zheng, K. Booth, J. McGrenere. 2006. Co-Authoring with Structured Annotations. Proc. CHI: 131–140.

[27] S. Zyto, D. R. Karger, M. S. Ackerman, S. Mahajan. 2012. Successful Classroom Deployment of a Social Document Annotation System. Proc. CHI: 1883–1892.