Visualizing Topics and Opinions Helps Students Interpret Large Collections of Peer Feedback for Creative Projects

PATRICK CRAIN and JAEWOOK LEE, University of Illinois, Urbana-Champaign
YU-CHUN YEN, University of California, San Diego
JOY KIM, Adobe Creative Intelligence Lab
ALYSSA AIELLO and BRIAN BAILEY, University of Illinois, Urbana-Champaign

We deployed a feedback visualization tool to learn how students used the tool for interpreting feedback from peers and teaching assistants. The tool visualizes the topic and opinion structure in a collection of feedback and provides interaction for reviewing providers' backgrounds. A total of 18 teams engaged with the tool to interpret feedback for course projects. We surveyed students (N = 69) to learn about their sensemaking goals, use of the tool to accomplish those goals, and perceptions of specific features. We interviewed students (N = 12) and TAs (N = 2) to assess the tool's impact on students' review processes and course instruction. Students discovered valuable feedback, assessed project quality, and justified design decisions to teammates by exploring specific icon patterns in the visualization. The interviews revealed that students mimicked strategies implemented in the tool when reviewing new feedback without the tool. Students found the benefits of the visualization outweighed the cost of labeling feedback.

CCS Concepts: • **Human-centered computing** → **Visualization systems and tools**; **Interactive systems and tools**; *Empirical studies in HCI*; *Visualization techniques;*

Additional Key Words and Phrases: Feedback sensemaking, feedback support, formative feedback, visualization design, learning

1 INTRODUCTION

Design-based learning is a powerful pedagogical approach that prepares students to develop solutions to real-world problems [17]. This learning style typically involves an iterative process where students create and refine prototypes based on feedback from stakeholders [6, 8, 15]. The feedback collected from different stakeholders such as peers, instructors, and external audiences is difficult

---

ACM Transactions on Computer-Human Interaction, Vol. 30, No. 3, Article 49. Publication date: June 2023.

49

to interpret because it often contains varying topics, opinions, and structure [1, 33, 42]. This challenge is exacerbated in learning contexts because students often skim formative feedback [14], lack concrete strategies for interpreting feedback [13, 44], and lack the opportunity or volition to implement those strategies [44]. Additionally, this challenge persists even when the individual pieces of feedback are of high quality [9, 29, 52].

Two threads of prior work most relevant to this article address the challenge of feedback interpretation. One thread explores using cognitive interventions such as reflection [7, 52] or paraphrasing [51] to make sense of feedback. These activities require individual, subjective effort that does not readily scale to large feedback collections or transfer between multiple recipients (e.g., teammates). A second thread explores aiding feedback interpretation through tool support by presenting interactive visual summaries to help users extract key themes from the content [20, 29, 30, 47]. However, existing tools typically assume feedback is generated with a known structure (e.g., using rubrics or fixed response options), limiting the utility of these tools for discovering patterns in the unstructured or semi-structured feedback common in design-based learning.

The present work intersects and extends these threads by deploying a new type of feedback visualization tool in a project-based user interface design course and studying how students used the tool to make sense of peer and instructor feedback. The tool structures free-form feedback by visualizing how it maps across feedback providers, topics, and opinion types (e.g., praise, suggestions, questions, and criticisms). The tool associates each feedback statement with an icon and arranges these icons in a grid indicating each statement's author (row), topic (column), and opinion (icon shape and color). We designed a field study to determine (1) what interpretation goals students pursue when reviewing formative feedback, (2) what patterns of tool use emerge in pursuit of those goals, (3) how the tool affects students' feedback review processes, and (4) how using students' own labels to generate the visualization compares to using a third party's labels. Though the tool was studied in a prior controlled experiment [53], that work did not study the tool's use in the context of users' own projects and feedback, patterns of tool interaction, or impact of labeling on users' review processes and visualizations.

In the course, 18 teams of 3–5 students each developed a 12-week user interface design project of their choice. Our study targeted three critical graded deliverables that would likely most benefit from revision in response to feedback: the project proposal (project weeks 1–2), low-fidelity prototype (project weeks 6–7), and functional prototype (project weeks 10–11). Students presented an initial version of each deliverable in an online studio session while peers and a **teaching assistant (TA)** wrote free-form feedback. Teams were given one week to review their feedback using the tool and revise each deliverable in response to the feedback. To study the tradeoffs of different labeling approaches, the research team labeled each student team's feedback for the project proposal. Students could revise the research team's labels for the low-fidelity prototype, and labeled the feedback themselves for the functional prototype. After each revision period, we surveyed students to learn about their goals for interpreting feedback, how they used the tool to accomplish these goals, and which aspects of the tool they found most valuable. At the end of the course, we interviewed 12 students about how reviewing feedback with the tool compared to their typical review processes, how these processes changed when using the tool, and how they felt about the different labeling approaches. To complement the students' perspectives, we interviewed the course's TAs to learn how using the tool affected their instruction, grading decisions, and perceptions of their own feedback.

Our investigation revealed that student teams used the tool to find the feedback that was valuable to them, assess the quality of their project deliverables (i.e., critical problems and key successes), prioritize and discuss which feedback to address, and justify their design decisions. These goals were primarily accomplished by exploring a multitude of patterns of opinion icons in the visualization and reviewing the providers' backgrounds. Students also used the icon patterns to

order their review of feedback details (e.g., begin with critical statements and end with praise). Students perceived that the benefit of the visualization outweighed the cost of labeling the feedback, and expressed that labeling the topics and opinions themselves prompted them to critically analyze each feedback statement. The distribution of labels assigned by the students was also consistent with the distribution of the research team's labels, suggesting comparable reliability between the two approaches. Several students reported adapting strategies they learned using the tool (such as color coding feedback by topic or highlighting repeated statements) when reviewing feedback for projects external to the course. The TAs reported leveraging patterns in the visualization to aid grading decisions and offer teams additional project guidance during office hours. The TAs also expressed considering how their feedback would be presented in the tool when writing it, such as not being overly critical to avoid an imbalance of red icons and not dissenting too far from majority opinions to avoid concerns over grading decisions.

This work makes three contributions to the human–computer interaction community. First, we demonstrate how interactive visualization techniques can be applied to help students discover, prioritize, and share critical issues in the feedback for their creative projects. The techniques implemented in the feedback visualization tool described in this article and the design implications resulting from the study of the tool should generalize to interpreting the unstructured or semi-structured feedback received in myriad contexts such as in online critique communities, academic peer review, and job performance evaluations. Second, we dispel concerns surrounding the costs of having students provide the metadata needed to generate a feedback visualization. Students found that the benefit of generating the visualization in the tool outweighed the cost of labeling the content (at most 60 minutes per feedback collection). Students also indicated a preference for labeling their own feedback because they found that labeling aided their comprehension of the feedback and they could appropriate the labels to match their own organizational style. Third, the collective data generated from students' tool interaction has the potential to aid instruction, e.g., by allowing instructors to determine whether students are writing feedback with an appropriate balance of opinions and topics and to curate examples of feedback that promote actions by the recipient. This article advances the idea that feedback *interpretation* is as important as feedback composition, and shows how interactive visualization tools can be leveraged to aid interpretation tasks and create new opportunities to teach specific feedback interpretation strategies.

2   RELATED WORK

Formative feedback is essential for reducing discrepancies between a student's understanding of a learning goal and their performance towards that goal [27, 35] or between a student's intentions and others' perceptions of their work [12]. However, feedback is not useful to a student unless they are able to interpret it and act upon it [22]. This interpretation, in turn, requires awareness of both the purpose of the feedback as well as tools and strategies for making sense of it [13, 44]. In the remainder of this section, we identify common barriers to interpreting feedback and situate our work with respect to existing strategies and tools for aiding feedback interpretation.

2.1   Challenges of Interpreting Feedback

Feedback may not be helpful to students if they are unable to interpret it, even if it is otherwise of high quality [9, 29, 52]. Difficulty interpreting feedback may arise from attributes of the feedback itself or from students' own cognitive barriers and behaviors related to reading the feedback. Winstone et al. found that students had difficulty interpreting and implementing individual pieces of feedback when they were unsure how it related to their personal goals or when the feedback contained complex language [44]. Cook et al. similarly demonstrated that while students who asked for specific, actionable suggestions for improvement made stronger revisions to their work than

students who asked for other kinds of feedback, students rarely asked such guiding questions when soliciting feedback [8]. Gibbs and Simpson [14] found that students often skim large amounts of feedback rather than thoroughly reviewing it in its entirety. Additionally, Foong et al. [13] showed that novices in particular may struggle to resolve feedback conflicting with their own frames of reference, and argue that online feedback engagement systems should provide additional structure and support for sensemaking.

Interpretation challenges may also stem from a student's own cognitive barriers and behaviors. Winstone et al. [44] identify four main cognitive barriers to feedback recipience: (1) awareness of feedback's meaning and purpose, (2) cognizance of strategies for interpreting feedback, (3) agency to implement those strategies, and (4) volition to engage with and implement the feedback. Several works reinforce these findings by demonstrating that gaps between a feedback provider and recipient in terms of domain knowledge [4, 16, 43], expertise [3, 10], and academic qualification [3, 54] may all inhibit interpretation. Other studies have demonstrated that a defensive mindset can also inhibit the reception and interpretation of feedback [40, 46, 49].

For the course involved in our study, student teams regularly received project feedback from the course staff and 20+ classmates, which was aggregated and presented to teams through an online spreadsheet tool in previous instances of the course. These students likely experienced many of the interpretation barriers previously described, resulting in minimal project revisions based on the instructor's observations. In the course instance involved in this article, students reviewed feedback using a new tool for interactively visualizing feedback's structure across providers, topics, and opinions, helping them better understand the meaning and purpose of the feedback. The tool also has the potential to reduce the cognitive burden required to learn and practice specific interpretation strategies (e.g., reconciling conflicting opinions by weighing reviewers' backgrounds), and gives students the agency to implement those strategies for any feedback set. Finally, our work explores different levels of involving students in the generation of the visualizations of the project feedback they receive, which has the potential to increase feedback engagement.

2.2 Strategies to Aid Feedback Interpretation

Research for facilitating feedback interpretation typically focuses on the representations of the feedback, cognitive strategies to help the feedback recipient extract themes from the feedback, and techniques for improving the feedback written by the provider. The first thread of research, which is most closely related to this article, leverages visual representations and interaction design to help users navigate large feedback collections. For example, the Voyant tool generates structured feedback for graphic designs from a non-expert crowd and presents the feedback using word clouds, interactive graphs, and annotations to help users extract high-level themes from the feedback [47]. A classroom study found the crowd feedback presented in Voyant helped students improve the quality of their graphic designs, and was perceived as more interpretative, diverse, and critical compared to free-form feedback generated for the same designs [48]. CrowdCrit [29] structures and aggregates crowdsourced design critiques through an interactive visualization showing the self-rated expertise of the providers, as well as quality ratings for the graphic design along several design principles. The authors found that users reported noticing more issues and producing better designs using aggregated crowd critiques than when using generic feedback. While these tools generate and visualize structured feedback from external crowds, the techniques used in these tools are less suitable for visualizing the topics and opinions in unstructured or semi-structured feedback, which is the focus of the tool studied in this article. The evaluations of Voyant and Crowdcrit also focused on evaluating the project outcomes that resulted from reviewing the feedback presented in the tool. By contrast, our work focuses on how students leverage the visualization and interaction design in the tool to review a collection of feedback, and explores how different approaches for

labeling the topics and opinions in the feedback (e.g., student-led vs. staff-led) impact the students' feedback review processes and the resulting visualization.

A second thread of research emphasizes self-directed interventions to help the recipient interpret feedback. For example, Jackson et al. [18] demonstrated that having students draft action plans for the feedback they received on their coursework increased their subsequent feedback utilization and grades. Yen et al. [52] found that having designers perform a reflection activity after feedback review led to an increase in perceived quality of the revised designs compared to reviewing feedback without reflection. A related study found that having designers paraphrase feedback improved their comprehension of difficult words and concepts in the feedback, and led to more effective revisions than when no paraphrasing was performed [51]. Wu et al. [46] found that performing coping activities such as expressive writing or reflection after reviewing negative feedback increased resilience to harsh criticism. Cook et al. [7] found that reflection could help users better recall their design goals, question their choices, and prioritize revisions. While activities such as reflection and paraphrasing can aid feedback interpretation, these activities are subjective and do not directly help students discover themes or conflicting opinions that might be present in the feedback they receive. It is also not possible for others (e.g., teammates) to verify the outcomes without performing their own detailed review of the feedback. The visualization tool described in this article scaffolds a range of interpretation tasks (e.g., finding patterns of agreement across topics) identified in a prior study of feedback review practice [53]. These scaffolds promote discovering and sharing themes and patterns in the feedback and have the potential to enhance reflection activities.

A third research thread aims to help providers write effective feedback by bridging gaps in their expertise or domain knowledge. One approach to bridging this gap involves using expert-created rubrics to evaluate work, which can help novices write feedback rated nearly as helpful as expert feedback [55]. Shannon et al. [36] showed that live rubric-based peer review during in-class presentations helped students produce immediate, relevant, and diverse feedback, with over 80% of comments being rated as helpful by their peers. Another class of techniques leverages comparisons with exemplary works to bridge gaps in expertise. Kang et al. [23] showed that feedback providers wrote more specific, actionable, and novel feedback when asked to select visual examples of designs relevant to their feedback. Cambre et al. [5] demonstrated that students who composed feedback for one design while contrasting it with a second design wrote feedback that received higher expert ratings than feedback from students who did not review contrasting designs. These techniques aim to help students write quality feedback, whereas the focus of our work is deploying and studying the use of a new tool that aims to help students interpret feedback on their course projects. Also, even when individual pieces of feedback are of high quality, a feedback *collection* can still be difficult to interpret because it contains different topics, structure, and possibly conflicting opinions written by reviewers with different backgrounds. Our work examines how visualizing feedback's topic and opinion structure can aid interpretation by helping students extract themes and patterns from a feedback collection.

2.3 Feedback Visualization Tools

A feedback visualization tool helps users make sense of large amounts of written content, often by providing interactive visual summaries to help users explore patterns in the data. For example, CommunityPulse [19] helps civic leaders make sense of written feedback gathered from community members in response to civic design proposals. Among other tasks, the tool allows civic leaders to interactively explore the distribution of sentiment in comments written by the community members for each proposal. The authors found that the use of the tool reduced the time and expertise required for community input analysis. Review Spotlight [50] produces summaries of user-generated reviews using adjective–noun pairs, and allows users to explore the contexts of

these pairs in greater detail within the reviews. Users of the tool were able to form detailed impressions about restaurants and decide between two options faster than they could when using traditional review web pages. Crowdboard [2] projects real-time crowd feedback onto draggable sticky notes on a virtual whiteboard to help creators visualize and organize feedback on their ideas during brainstorming sessions. The authors found that creators valued the real-time crowd feedback the tool provided and incorporated it into their discussions, generating more creative ideas than creators who utilized a traditional whiteboard. Finally, Unakite [28] collects, organizes, and displays alternative solutions to programming problems in terms of their relative tradeoffs. A user study showed that the tool reduced the cost of capturing tradeoff-related information by nearly half, and that developers were able to understand these tradeoffs approximately three times faster.

The feedback visualization tool studied in this article shares the goal of helping users make sense out of a collection of written content. However, our work is original relative to this corpus of prior work because we are testing the use of the tool to help students determine how to best revise their open-ended design projects in response to feedback written by peers and instructors. This introduces sense-making tasks not directly supported by tools in prior work, such as finding topic patterns across multiple reviewers, reconciling conflicting opinions, and weighing comments based on the background of the comments' authors. We also study issues related to the tool that uniquely arise in this context, e.g., how the course staff feel about having their feedback juxtaposed with student peer feedback in the tool and comparing different approaches for generating the metadata needed for the visualization in the tool.

3 FEEDBACK VISUALIZATION TOOL STUDIED

We extended the design and implementation of a Web-based feedback visualization tool (Decipher [53]) for the classroom study. The tool adds structure to a collection of written feedback by visualizing how the feedback maps across providers, topics, and opinions. The goal of the tool is to help users discover useful patterns in the feedback that would be difficult to extract from the text alone—such as patterns of praise and criticism across topics or between providers—and to develop an action plan based on these patterns to improve the project.

The main page in the tool displays the text of the collection of feedback for the project on the left and presents the visualization of that collection of feedback on the right (see Figure 1). In the visualization, the feedback is organized in a grid by reviewer (rows) and by topic (columns). Topics are sorted by their frequency in the feedback, with the most frequent topic placed in the leftmost column. For each topic referenced by a provider, a graphical icon indicates if the piece of feedback consists of praise (thumbs up, green), criticism (thumbs down, red), a suggestion (bulb, yellow), a question (question mark, yellow), a neutral comment (filled, yellow), or a mixed opinion (half praise, half criticism). The cell is empty if a provider does not reference that topic in their feedback. Icon colors and shapes were used in tandem to improve visual accessibility.

The text and visualization are linked through interaction. Selecting a piece of feedback (e.g., a sentence) on the left highlights the graphical icon associated with that piece of feedback in the visualization. Likewise, placing the cursor over an icon in the visualization causes the tool to scroll to and highlight the corresponding feedback on the left. Clicking on an icon displays a window containing the associated piece of feedback and interaction for choosing an intended action (*must do*, *discuss it*, *disagree with it*, and *consider it*) for the piece of feedback and a checkbox for marking the intended action as completed. The user can search for keywords in the feedback and for pieces of feedback labeled with specific intentions. Search results are displayed by dimming the icons that do not match the search criteria.

The tool imports the collection of feedback and attributes of the providers from a CSV file. Once imported, the tool partitions the feedback into sentences. A user can then open the annotation page
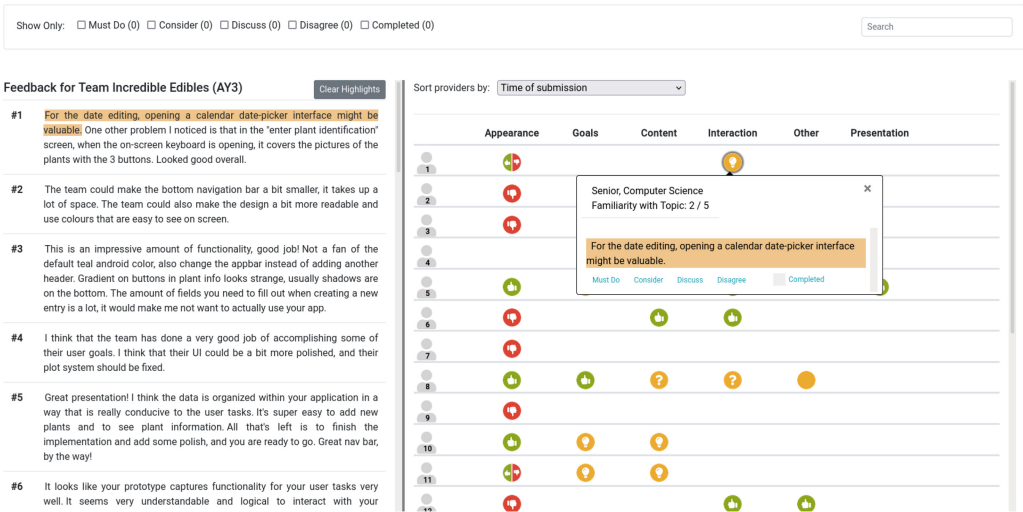
Fig. 1. The main user interface screen of the feedback visualization tool, currently showing a collection of feedback (left) and its visualization (right) for a team's functional prototype. The graphical icons on the right depict the opinions in the feedback by provider (the rows) and by topic (the columns). Hovering the cursor over an icon highlights the corresponding piece of feedback on the left, and vice versa. Clicking an icon displays a tooltip showing the feedback provider's background information, the piece of feedback on the left corresponding to the topic, and the intention labels for tracking feedback status. The top of the main user interface screen includes a row of checkboxes for filtering feedback by intention labels, a search bar to filter feedback by keywords, and an option to sort the feedback by attributes such as a provider's self-reported familiarity with the project topic, year of study, and major.

in the tool to label the topic and opinion of each sentence, and can recursively merge two or more adjacent sentences if desired. In our study, the users of the annotation page included the research team who labeled the feedback for the first and second project deliverables, as well as the students in the course who had the opportunity to modify the labels for the second project deliverable and label their own feedback for the final deliverable. The resulting labels are necessary for the tool to generate the visualization shown in Figure 1. Default topics are defined in the tool and, in our study, they were derived from the rubric associated with the corresponding project deliverable. For example, topics for the project proposal included project goals, solution alternatives, user audience, and feasibility. A user can add new topics in the tool, whereas the set of opinion categories was fixed and could not be revised by the user in the current implementation.

The attributes of a provider are displayed in a tooltip when the user places the cursor over the icon representing that provider in the visualization. For our study, the attributes were: familiarity with the project topic from Very Unfamiliar (1) to Very Familiar (5), major (computer science, social science, humanities, etc.), and year of study (junior, senior, grad student, or TA). The user can sort the visualization by these attributes (e.g., to find the feedback from the TA or from those most familiar with the project topic). A short document describing the visualization and interaction features of the tool was linked from a Help button placed on the main page of the tool.

4 RESEARCH QUESTIONS

We deployed the visualization tool in a course and had students in the course use it to interpret feedback for their course project. The field study was designed to answer four research questions:

— **RQ1**: What goals do students try to accomplish when using the feedback visualization tool and what patterns of use emerge to accomplish these goals?

— **RQ2**: What processes for reviewing feedback do students retain from using the tool, and what difficulties arise when reviewing feedback without the tool?

— **RQ3**: How does having students label their own feedback impact the resulting visualizations and their perceptions of the tool?

— **RQ4**: How can a feedback visualization tool be best appropriated for instructional purposes?

Answers to these questions will contribute to a base of knowledge and best practices for designing technology aimed at helping end users interpret a collection of feedback written by multiple providers.

5 METHODOLOGY

To answer the research questions, we deployed the feedback visualization tool (Section 3) in a project-based user interface design course taught at a public university in the U.S. A field study method was chosen so that we could study the use of the tool in authentic feedback review processes. The course was chosen because of its emphasis on having students gather and act on feedback as part of the iterative design process, and the specific tool was chosen because it visualizes the type of written feedback students already exchange as part of their project work in the course. In the course, students completed user interface design projects in teams and used the tool to review feedback written by classroom peers and the TAs for three project deliverables, each with an initial and revised version. Multiple uses allowed students to gain familiarity with the tool and how to best incorporate it into their feedback review processes. Using the tool for three project deliverables was the most that could be practically integrated into the course.

The research questions were addressed by analyzing data collected through surveys, interviews, and interaction logs. In addition, for the third research question, the research team progressively transitioned the labeling of the topics and opinions in the feedback from the research team (first use of the tool) to the students (final use of the tool). This scaffolding approach helped students understand why these labels were needed by interacting with the visualization first, and allowed the students to compare their experience with the tool when having to label their own feedback vs. having the feedback labeled by an external party.

An author of this article was the instructor of the course. The instructor and TAs were blind to which students consented to allow their data to be used for the purpose of research. Members of the research team not affiliated with the course performed the interviews and collected all consent information. The study was approved by the Institutional Review Board at our university.

5.1 Design Course and Projects

The user interface design course targets upper-level undergraduate and beginning graduate students studying computer science. There were a total of 98 students (34 female, 64 male) in the course. About 10% of the students were from other departments on campus such as chemistry, health, psychology, and the arts. For most of the students, this was their first course on user interface design. The lecture topics included user research, prototyping, implementation, and evaluation techniques. The instructor organized students into teams of 4–6 students to balance skill sets. There were a total of 18 teams in the course. The teams applied the lecture topics to a 12-week user interface design project of their choice. Examples of these projects include a mobile app and hardware to remotely start a motorcycle, an app for scheduling the first and subsequent vaccine shots in one scheduling session, and an app for homeowners to request assistance with shoveling the snow from their sidewalks and for volunteers to respond to those requests.
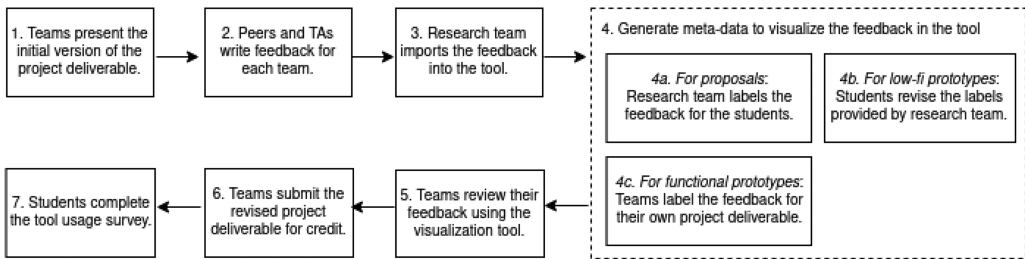
Fig. 2. An overview of the study procedure. The steps shown in the diagram were repeated three times, once for the project proposal, low-fidelity prototype, and functional prototype. Student teams presented the initial version of a project deliverable during their studio section. Peers and the TA wrote feedback for each team in an online form. The research team imported all the feedback into the tool. For the proposal (project weeks 1–2), the research team labeled the topics and opinions in the feedback and student teams could access the resulting visualization in the tool without additional work. For the low-fidelity prototype (project weeks 6–7), student teams could revise the labels assigned to the feedback by the research team to best fit their own interpretation of the content. For the functional prototype (project weeks 10–11), students labeled the feedback on their own and a default set of topics were available in the tool. Student teams then reviewed the feedback in the tool, submitted a revised deliverable for course credit, and completed a tool usage survey. After the project ended, members of the research team interviewed students and the two TAs in the course.

The projects were structured as a design process. There were a total of nine project deliverables, with one deliverable submitted each week for a grade. Teams presented the deliverables during a weekly design studio. Each studio section had 20–30 students, making up 5–6 teams. Teams were assigned using the CATME [32] software to balance team compositions based on students' skill sets. Presentations were performed via video conferencing technology due to the pandemic. Teams presented their project deliverables in the studio while classmates and the TA wrote feedback in an online form. The TA also facilitated five minutes of oral critique before proceeding to the presentation of the next project in that studio section.

Three project deliverables created by the student teams were targeted in this study: (1) the project proposal (project weeks 1–2), (2) the low-fidelity prototype (project weeks 6–7), and (3) the functional prototype (project weeks 10–11). The proposal was a document describing a project idea, including the user need, existing solutions, and user audience. The low-fidelity prototype consisted of sketches or mockups representing a first-cut solution for the project. The functional prototype was a programming implementation of a team's low-fidelity prototype. These three deliverables were selected for the study because they represented key milestones in the project timeline and incorporating a revision cycle for these deliverables would benefit teams the most. These deliverables were each split into an initial and revised submission. Teams presented the initial submission in studio, received a web link to access their feedback in the feedback visualization tool, revised the submission based on the feedback, and submitted the revision for course credit the following week. Teams could only access the feedback for the initial submissions in the tool. The revised submissions earned three times more credit than the initial submissions to incentivize attending to the feedback and making changes. Additional project deliverables such as user research reports were not included in the study design because the course timeline would not accommodate an initial and revised submission for each of the other deliverables. Teams did gather additional feedback from potential end users for other project deliverables. It was not practical for students to gather end user feedback in time to visualize it alongside the peer and TA feedback for the deliverables that were included in the study.

5.2 Students

At the onset of the course, students completed a survey about their experience receiving feedback for open-ended work (N = 76). In the survey, students described a recent project (not the project in the current course) which was revised based on the feedback received from multiple people, the format of the feedback, and the providers who wrote that feedback. Of the respondents, 89% reported having received feedback from multiple people for open-ended work two or more times while 11% reported no prior experience. Examples of open-ended work for which the students received feedback included written reports, programming projects, and senior design projects. For the instances described, 81% of the students reported that the collection of feedback was from 2–5 providers, most often classroom peers (74%) and TAs (14%). In the instances described, the format of the feedback was written (22%), verbal (25%), or a combination (53%). Students wrote that what made the collection of feedback difficult to understand was (1) knowing which statements in the feedback to prioritize for revising their work (e.g., to receive the highest grade), (2) translating vague and ambiguous statements in the feedback to actual revisions of their work, (3) remembering all the feedback received (particularly if received verbally) and what it meant in relation to their work, and (4) forgetting to address some feedback because it was received through multiple channels.

5.3 Tool Usage Surveys and Interaction Logs

Students completed three tool usage surveys, once after using the tool to review the feedback associated with the initial submission for each of the three project deliverables included in this study. Students completed the surveys individually, in part to allow each student to decide whether to give consent for their data to be used for the purpose of research. Students earned course credit for completing the surveys.

The tool usage survey associated with the project proposal had 20 questions related to our study. A student estimated the time spent reviewing the feedback in the tool and described how the team reviewed the feedback in the tool (e.g., as a group, or individually and then as a group). The student then responded to 14 Likert scale questions about the usefulness of the tool and each of its main user interface features (exploring topics, exploring opinions, seeing provider backgrounds, labeling intended actions, etc.), the quality of the topic and opinion labels provided by the research team, the quality of the feedback, and how difficult the feedback was to understand. The questions were phrased as a statement (e.g., "My team found the feedback visualization tool useful for understanding the feedback received.") and the responses were on a scale from 1 (Strongly Disagree) to 7 (Strongly Agree). In four open-ended questions, the student described the most useful insights the team discovered in the feedback, how the features of the tool were leveraged to find these insights, and the most significant benefits and problems the team experienced when using the visualization tool for reviewing the feedback.

The surveys corresponding to the usage of the tool for the low-fidelity and functional prototypes asked the same questions as described above. However, for the low-fidelity prototype, teams could revise the topic and opinion labels that were provided by the research team. The survey for this deliverable included two additional Likert questions about revising the labels ("My team revised the initial topic and opinion labels that were provided to us in the tool") and the importance of doing so ("My team found it important to revise the initial topic and opinion labels to best reflect our own understanding of the feedback statements").

For the functional prototype, the research team did not label the feedback. Each student team was instructed to use the tool to label the topics and opinions in the feedback they received on their own. The extent and appropriation of the labeling was up to them. This survey also asked about the perceived tradeoff between the effort required to perform the labeling and the benefits of being able to visually explore the feedback on a scale from 1 (cost outweighs benefit)

to 7 (benefit outweighs cost) and asked students about their preference for using a feedback visualization tool on a similar project in the future on a scale from 1 (prefer a traditional text viewing tool) to 7 (prefer the feedback visualization tool). Finally, the survey asked students to describe how the use of the feedback tool affected the process by which they review the feedback, compared to how they might have reviewed the same feedback if in text-only form.

The feedback visualization tool was instrumented to log students' interactions with the tool. This included the topic and opinion labels assigned to the feedback, the interactions with the graphical icons, the intention labels assigned, and the searches performed on the feedback.

5.4 Student and TA Interviews

After the project was complete, the research team conducted semi-structured interviews with 12 students and the 2 TAs in the course. Student interviewees were recruited through an open call placed in the final tool survey and sent to the course mailing list. Student interviewees consisted of four juniors, five seniors, and three graduate students. Seven were Computer Science or Computer Engineering majors, three were minoring in CS, and two were Health Technology majors. None of the students had previously used a similar feedback visualization tool. The TAs were Ph.D. students in Computer Science in the area of human–computer interaction. The interviews were conducted over Zoom, lasted 30–60 minutes, and were screen recorded with consent. Interviewees were each remunerated $40.

To begin the student interviews, we wanted to observe their process for reviewing feedback when that feedback was presented in a text viewing tool (Google Doc) and compare it to the processes observed and described when reviewing similar feedback in the visualization tool. We adapted a poster design and a collection of feedback for that poster that was developed in a prior study [53]. The feedback was written by users recruited from a micro-task platform. There were a total of six pieces of feedback for the poster and the word count was comparable to the word count of the feedback students typically received for a project deliverable in the course. We also confirmed the feedback collection varied in topics and opinions but was also scoped such that it could be reviewed during the first 15 minutes of the interview.

We asked the student to imagine helping the project's creator discover the most useful insights in the feedback and to show us their process for reviewing the feedback in a document editor (Google Docs), using any of the features in the editor desired. They were also asked about the strengths and weaknesses of reviewing the feedback in a document editor and to describe how they might have used the visualization tool to review this same collection of feedback.

For the next part of the interview, the student accessed the feedback for their project deliverables in the visualization tool. The student located what they recalled to be the most interesting insights in the feedback from any of the deliverables, and to show us how they discovered these insights with the tool. In addition, we asked about their first impressions of the tool, perceptions of labeling the feedback in the tool, and how the use of the tool helped them develop new processes (if any) for reviewing a collection of feedback written by multiple people.

For the TA interviews, we inquired about how the feedback they wrote was typically returned to students in other courses, their impressions of the tool, how the use of the tool helped (or hindered) their ability to grade, write feedback for, or mentor the projects, how they felt about their feedback being visualized alongside the student feedback in the tool, and how they believed that this might have affected the feedback they wrote.

5.5 Procedure

The instructor informed students that a feedback visualization tool would be used in the course and demonstrated the tool prior to its first use. When teams presented their initial project proposals

#	Content		Topic	Opinion
5-0	The prototype definitely does address your project goals.		Select from existing topics Goals Add a topic	praise
5-1	The only thing about the project may be feasibility.	Merge Above	Select from existing topics Goals Add a topic	criticism
5-2	It seems that there are quite a few features associated with each task.	Merge Above	Select from existing topics Interaction Add a topic	criticism
5-3	Maybe instead of listing palettes in the accessibility tool try to find something a bit more visually appealing.	Merge Above	Select from existing topics Appearance Add a topic	suggestion
5-4	Overall great job on the prototype.	Merge Above	Select from existing topics Presentation Add a topic	praise

Fig. 3. One page of the feedback labeling interface as seen by a student team for the low-fidelity prototype deliverable. Each page displays the feedback written by a single feedback provider and splits it into individual feedback statements. Students are able to assign an opinion and topic to each statement from predefined lists, define custom topics for the feedback statements, and merge related feedback statements together.
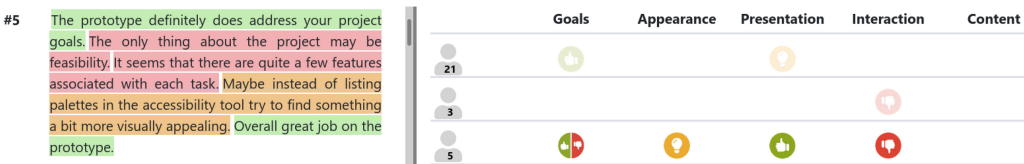

Fig. 4. How the labels assigned to the feedback in Figure 3 appear in the visualization's feedback review interface. Icons are arranged in a grid according to each feedback statement's assigned topic (column), feedback provider (row), and opinion (shape and color). Hovering over a feedback statement on the left highlights the associated icon on the right, while clicking the statement retains this highlighting, allowing students to highlight and compare multiple pieces of feedback at the same time.

in the studio, the other students and the TA in that studio section wrote feedback and answered the background questions (e.g., familiarity with the project topic) in an online form. The form instructed students to write one paragraph of detailed feedback for the project deliverable, including its strengths and weaknesses, and to explain their reasoning. The form also suggested topics to cover in the feedback based on the rubric associated with that project deliverable, although students were free to write whatever type of feedback they deemed appropriate.

At the conclusion of the studios that week, the research team aggregated all the feedback into a CSV file and imported it into the feedback tool. The research team and TAs then labeled the

Table 1. The Coding Schema Applied to the Interview and Open-Ended Survey Responses

Category	Label	# Students	Description	Example Idea Unit
Finding Valuable Feedback	Similarities	47	Focusing on insights regarding topics of interest repeated by multiple providers (e.g., by looking at clusters of icons)	"We attempted to look for patterns in feedback from our peers to ensure that we didn't skew our revisions based on one person's response."
	Ordering	46	Structuring feedback review in a certain order (e.g., criticism, then suggestions, then praise)	"We clicked on the arrows and first looked at red arrows which pointed what was wrong followed by the green good arrows for what is working on [our] prototype."
	Expertise	34	Sorting or otherwise locating feedback based on a peer's seniority or topic expertise (including finding TA feedback)	"Based on the seniority of the student, we could see that younger students were not able to offer as much feedback on feasibility but had opinions on design instead."
Assessing Quality	Proportions	41	Using proportions of opinion icons to identify assess a project's strengths and weaknesses	"The more useful aspects of the visualization tool are the culmination of comments and the icons under each category because we could quickly gauge how well our prototype was [received]."
	Strengths	5	Reviewing praise to identify strong areas to maintain in the revision	"Seeing the green thumbs up icon when you first open it up makes you feel good since it indicates that you are on the right track you just need some tweaks."
Team Coordination	Discuss	26	Insights from the tool sparked or reinforced discussions among team members	"We looked at all the critical feedback (using the right-hand side of the tool mostly) as a team and discussed/made a list of the changes we would make as we went."
	Debate	4	Features of the tool (e.g., assigning intention labels) prompted team to debate relative importance of feedback	"I found the 'must do/consider' buttons to be most useful. It forced us to debate the feedback we received."
Familiarization	Familiarize	17	Tool helped learner familiarize themselves with the feedback they received	"The biggest benefit was that we actually got very familiar with the feedback as we organized it…we really got to understand how our peers viewed our demo and which suggestions or critiques were the most important to discuss and work on."
Perception	Match	60	Tool's features matched the learner's goals (e.g., tool was intuitive or helpful)	"Having the labels of 'Must Do', 'Consider', 'Discuss', etc were helpful for understanding the tasks that were important and needed to be completed compared to extra or unnecessary tasks."
	Mismatch	33	Tool's features mismatched the learner's goals (e.g., tool overwhelmed the learner with information or features)	"Labeling the feedback sometimes seemed like a lot of extra effort compared to the benefits of visualizing the feedback."
Learning	Process	9	Tool helped learner develop a new process for reviewing or engaging with future feedback	"Since we can label fragments of each feedback paragraph, we can label different parts with different topics, and use it as a way of dividing work between team members."
Changes	Change	43	Learner wanted something added to, removed from, or changed about the tool (feature requests, QoL fixes, bug reports, etc.)	"If you could implement a 'comment counter' type of system and then sort the comments based on which comments match with the most commented stuff, that would be even more helpful."

Each idea unit in both the survey and interview data was assigned exactly one of the labels in the schema. Only the labels in the second column were directly used for the coding.

topics and opinions in the feedback. The topics were derived from the rubric used for grading the deliverable, while the set of opinion labels was fixed (see Section 3). The research team and TAs labeled the feedback together for one project deliverable for calibration, then divided the labeling of the feedback for the remaining projects. Each student team received a Web link to their project's feedback in the tool within two days of the completion of the recent studio. A student team received feedback from 20–25 peers and the TA. The student team reviewed the feedback in the tool, revised the project deliverable based on the feedback, and submitted the revision the next week for course credit. Students then completed a tool usage survey.

The above process was repeated for the low-fidelity and functional prototypes, but the strategies for labeling the feedback were adjusted to evaluate the tradeoffs between having an external party label the feedback and having the students label the feedback on their own. For the low-fidelity prototype, student teams were encouraged to revise the labels assigned by the research team to best fit their own interpretations of the content or needs. For the functional prototype, students were instructed to label the topic and opinions in the feedback on their own. A default list of topics was provided in the tool based on the rubric for the deliverable, but the teams could add new topics to this list. After this last use of the tool, the research team interviewed students and the TAs about their usage and perceptions of the tool.

After the completion of the interviews, the research team followed the generalized inductive approach [38] to develop and iteratively refine a coding schema. The interview and open-ended survey responses were partitioned into 1083 idea units. Two members of the research team coded the idea units according to the schema. To test inter-rater reliability, a sample of the data (80 units) was coded by both raters. Cohen's Kappa was 0.87, indicating strong agreement satisfactory for proceeding with the coding [26]. The remaining units were split between the two raters and coded independently. Table 1 summarizes the schema applied to the interview and open-ended survey responses.

6 RESULTS

A total of 18 teams received feedback for each of three project deliverables. For each deliverable, a project team typically received 20–25 pieces of peer feedback and one piece of feedback from the TA. On average, one team's collection of feedback contained 1447 words (SD = 397) and was split into 76 units of feedback (SD = 19) categorized into 5–9 topics (see Figure 1 for an example). Students rated the feedback they received from their peers helpful (mean = 5.9, SD = 0.9, 7-pt scale), and felt they were able to determine how best to revise their project deliverables using this feedback (mean = 5.6, SD = 1.0, 7-pt scale). Students reported using the tool for 20–60 minutes on average for reviewing the feedback for each project deliverable. About half the students reported using the tool individually to review feedback before meeting with their team to discuss it, while the other half reported reviewing the feedback in the tool as a team. This was accomplished through the use of screen sharing in a video conferencing tool, since all teamwork was conducted remotely. Table 2 shows a summary of responses to the tool usage survey for the final deliverable. Responses to the other two tool usage surveys have similar results and are omitted for brevity.

A total of 69 students (42 male, 27 female) consented to their data being used for the purpose of research. We answer our research questions by drawing from the student surveys and interviews, the data collected through the tool, and the TA interviews. In the next subsections, we use the notation S# to refer to student survey respondents, I# to refer to student interview participants, and TA# to refer to TA interview participants.

6.1 Goals When Using the Tool (RQ1)

We found from the interviews and open-ended survey questions that students aimed to accomplish three goals when reviewing a collection of feedback with the tool: find the valuable feedback, assess project quality, and facilitate communication and coordination amongst teammates.

6.1.1 Finding the Valuable Feedback. Students found the tool helpful for finding statements in the feedback that they perceived to be valuable for their project deliverable (Q5: mean = 5.64, SD = 0.87; Q7: mean = 5.51, SD = 1.08 in Table 2). Based on responses to the tool usage surveys and interviews, we identified three main patterns of use that students reported for finding the valuable feedback. One pattern reported by 46 students (67%) in the surveys and 5 students (42%) in the interviews was structuring their review of the feedback in a particular order based on the types of icons in the visualization. The typical order was to review the critical statements first, then the suggestions and questions, and end with the statements of praise.Students stated that using the critical statements as the initial entry point into the feedback collection helped them identify problematic areas of their projects, while reviewing suggestions afterwards directed them towards potential solutions.

> We first check all the red labels, then classify them, and extract the ones we think it is interesting and reasonable. Then check the yellow label, classify and discuss, and finally go through other checks to see if there is any missing information. [S38 (Team 16), Lo-Fi Prototype Survey]

> We always go to the red thumbs down, because like those are the negative things…the criticisms we can deal with, so we always, always go find where the red thumbs down are, and we try to take a look, and then we just see what it is saying and discuss it over our team call. and after that, we go to the insightful [suggestion] button here…we consider this neutral like, it's not really bad…they don't really say anything bad about the feature, they just say "oh maybe you can do this maybe you can do that"…we just put that into consideration…other than that we just skimmed through it. [I2 (Team 10)]

Table 2. The Mean and Standard Deviation for Students' Responses to the Rating Questions from the Last (Third) Tool Usage Survey

Question	Mean	SD
1. The feedback from peers was helpful for improving the project deliverable this past week (1 = strongly disagree, 7 = strongly agree).	5.76	0.94
2. The feedback from the TA was helpful for improving the project deliverable this past week (1 = strongly disagree, 7 = strongly agree).	6.37	0.76
3. My team found understanding the collection of feedback overwhelming (1 = strongly disagree, 7 = strongly agree).	3.18	1.43
4. My team found the feedback visualization tool helpful for understanding the feedback received for the project deliverable this past week (1 = strongly disagree, 7 = strongly agree).	5.27	1.12
5. My team was able to discover useful insights in the feedback when using the tool (1 = strongly disagree, 7 = strongly agree).	5.64	0.87
6. My team discussed the insights discovered in the feedback when using the tool (1 = strongly disagree, 7 = strongly agree).	5.77	0.98
7. My team was able to find the specific issues in the feedback that we were most interested in when using the tool (1 = strongly disagree, 7 = strongly agree).	5.51	1.08
8. My team found the feedback visualization tool helpful for determining how to best revise the project deliverable based on the feedback received (1 = strongly disagree, 7=strongly agree).	5.47	1.03
9. My team found it useful to assign the topic and opinion labels to the feedback statements (1 = strongly disagree, 7 = strongly agree).	4.64	1.50
10. My team would like to use the same tool to review feedback on a future project deliverable (1 = strongly disagree, 7 = strongly agree).	5.18	1.33
11. Please rate the usefulness of exploring the topic columns in the feedback (1 = not useful, 7 = very useful).	4.99	1.28
12. Please rate the usefulness of exploring the opinion icons in the feedback (1 = not useful, 7 = very useful).	5.04	1.28
13. Please rate the usefulness of reviewing the background of the feedback providers (1 = not useful, 7 = very useful).	3.60	1.69
14. Please rate the usefulness of labeling the intended actions (e.g., "must do") for each statement in the feedback (1 = not useful, 7 = very useful).	4.08	1.58
15. Please rate the usefulness of marking the intended actions as Complete (1 = not useful, 7 = very useful).	3.86	1.60
16. Please rate the usefulness of labeling the topics and opinions in the feedback (1 = not useful, 7 = very useful).	4.90	1.30
17. Please rate the usefulness of reviewing the feedback in order on the left side of the screen (1 = not useful, 7 = very useful).	5.45	1.20
18. If you needed to review feedback for a similar project in the future, please rate your preference for using a text viewing tool (e.g., a Google Doc or Web page with the feedback text) or a feedback visualization tool like the one used in the course (1 = prefer text tool, 7 = prefer feedback visualization tool).	5.37	1.31
19. Please rate how you feel about the tradeoff between the effort required to perform the labeling and the benefits of being able to visually explore the feedback (1 = cost outweighs benefit, 7 = benefit outweighs cost).	4.69	1.35
20. Estimate the total degree of change that you believe the team made from the initial to the revised project deliverable by addressing issues in the feedback (1 = no change, 7 = significant change).	5.25	1.15

The data from the first two tool usage surveys showed similar ratings and is omitted for brevity. Responses were structured as 7-point Likert items, with endpoints shown in each question. Q11–17 were presented in the form of a grid in the actual survey.

A second pattern for finding the valuable feedback was observing vertical and horizontal clusters of the same category of opinion icons in the visualization. These clusters indicated redundant critiques of a topic or a reviewer who was particularly critical or supportive. This pattern was mentioned by 47 students (68%) in the comments they wrote in the tool usage survey and by 3 students (25%) in the interviews. Students also gave positive ratings for exploring the feedback by topic (Q11: mean = 4.99, SD = 1.28) and by opinion (Q12: mean = 5.04, SD = 1.28) in the tool.

> ...we would first look out for the red thumbs down...the column or the row with the most, those were obviously the ones we looked at first...the columns with the most feedback on it are usually the areas you need to pay some attention to, unless it's all thumbs up. So we definitely knew that...we had the backend and everything running, but we hadn't so much focused on the UI part of it...that's basically how we approached it...first look for all the negatives, and then also look for which column has the most amount of feedback given for it. [I1 (Team 18)]

> When we met up as a group, the first thing we would do is look at the visualization side of it, and see which columns had the most notes. So appearance and interaction...seems like where we spend the most time, and appearance has a lot of critiques on that....we would then look at that and say "okay, let's make note of all of these things...let's read through them." [I6 (Team 5)]

A third pattern for finding valuable feedback reported by 34 students (49%) in the surveys and 4 students (33%) in the interviews was prioritizing feedback based on reviewers' self-reported familiarity with the project topic and year of study.

> One of the interesting patterns we discovered is that people who are less familiar with the topic don't seem to care about existing solutions as much as those who are more familiar with it. We found out this by using the sorting feature of the tool. [S56 (Team 11), Project Proposal Survey]

> We found interesting differences in the feedback from different levels of students. Based on the seniority of the student, we could see that younger students were not able to offer as much feedback on feasibility but had opinions on design instead. [S37 (Team 9), Project Proposal Survey]

Among those who prioritized feedback review by year of study, several students reported reviewing critical statements from the TA first, and then reviewing peer feedback on those same topics for further elaboration and to find suggestions for improvement. Other students explored these relationships in the feedback in the opposite order, using the feedback from the TA to confirm or elaborate issues mentioned in the peer feedback. In both cases, the organization of the feedback by topic and opinion in the tool further enabled these explorations.

> We found that the feedback from the TA was a very strong benchmark and very whole at identifying a broad range of issues and concerns with our project, both good and bad. All the reviews from my peers touched upon at least one concern raised by the TA and expanded upon their viewpoint. These patterns were apparent because of the identifying likes, dislikes, and neutral hand icons which helped us quickly sort through the reviews and understand the sentiment across the board for a particular issue based on the hand icon colors. And clicking on the hand icons helped us highlight the review which causes that hand color. This helped us a lot. [S9 (Team 10), Project Proposal Survey]

> So I went through all the students first and the TA was the last one. and it either confirmed something...if all the students were saying this one thing and then the TA

> agreed with them, it almost put more weight on that...versus if a couple students said something and the TA was like "actually I think this is good because x y z"...I think this is good then that'd be a point of discussion with my group mates. [I6 (Team 5)]

In sum, students found the valuable feedback by reviewing criticisms and suggestions before other icon types, observing criticisms that were repeated by multiple providers, and prioritizing feedback from peers who reported high familiarity with the project topic and from the TAs. However, exploring the reviewer background in the tool was rated as neutral (Q13: mean = 3.60, SD = 1.69), possibly indicating only some students used the reviewer background as a way to find the feedback that was valuable to them. The usage patterns described in this section may not be exhaustive, as students may not have recalled all the different uses of the tool and new patterns might develop with additional experience with the tool. Also, individual students likely performed different subsets of these patterns based on the feedback content, the project deliverable, and their own preferences.

6.1.2 Using Clusters of Icons for Assessing Project Quality. A total of 41 students (59%) in the surveys and 5 students (42%) in the interviews reported using the visualization to evaluate the quality of specific aspects of their project deliverable or its overall quality. Many students mentioned observing the proportion of praise, criticism, questions, and suggestions to evaluate the quality of their project deliverable and estimate the amount of work that would be required for the revision. Students reported visually scanning the criticism icons in each topic column to quickly determine which areas of their projects were in most need of attention. Similarly, five students (7%) in the surveys and one student (8%) in the interviews reported observing clusters of praise icons for assessing which aspects of their project deliverable they should avoid changing while revising other aspects of their work.

> [The tool] helped me quantize the amount of praise comments we had, and the amount of improvement comments we had. We were able to obtain a better summary gauge without reading every single feedback, which is something we couldn't have done on an all text feedback form. [S24 (Team 5), Functional Prototype Survey]

> In terms of the praise, we would all just go through alone and look through general things. I think in our app, we had a good sense of people really liked our ui and obviously the app looks great, so what we knew was that any changes we make, we can make them, but let's not do it at the cost of changing our colors, our layout, the way it just looks when people enter the app, because clearly people really like that. [I5 (Team 16)]

Students often reported that they reviewed opinion icons to identify the strengths and weaknesses of their projects before reading the associated text. In this sense, the distribution of the icons themselves functioned as a high-level form of feedback summarizing the detailed text feedback.

6.1.3 Communicating and Coordinating Revisions. The third main goal that 26 students (38%) in the surveys and 5 students (42%) in the interviews reported pursuing with the tool was facilitating communication and coordination amongst team members. Students reported using the tool's feedback organization features to structure their team discussions, as well as using feedback trends to develop their arguments during these discussions. For example, when reviewing peer feedback with which they disagreed, students were prompted to justify their decisions not to address certain issues raised in the feedback.

> After the annotations were all done, we would sort of discuss based on what feedback we got. So if there were any questions or suggestions on what we should do...during

our group meeting...we would discuss the feedback and use that as a basis of any design changes we would make...we'd include our own opinions on top of that, like, if someone gave feedback and we all very strongly disagreed with that, then we might not follow it as much. But we mainly used it as a way to discuss how we could do changes. [I11 (Team 7)]. We tried to address all suggestions and criticism in the feedback, either by actually implementing changes, or by explaining the reasoning behind our design choices for the things we did not change in our presentation. [S55 (Team 1), Functional Prototype Survey]

Four survey respondents and one interview participant stated that assigning intention labels to their feedback forced their teams to debate the relative importance of each piece of feedback, which helped them prioritize revisions more effectively and make sure they addressed the most important feedback.

I found the "must do/consider" buttons to be most useful. It forced us to debate the feedback we received. [S59 (Team 12), Project Proposal Survey]

...whenever we take a look at [our feedback]...we'd click on "disagree" because of the project scope and how we only have like two weeks to develop it, then we just have to disagree with this current feedback. So I really think that these "must do", "consider", and "discuss" [labels were] really helpful for our team, because we could just say "hey what should we do with this specific feedback?" [I2 (Team 10)]

Students leveraged the above mechanisms to communicate with their teammates and coordinate revisions for their project deliverables. Although the tool was designed primarily for individual use, the behaviors some students demonstrated in this study highlight several opportunities for the tool to facilitate team-based feedback review.

6.2 Comparing Feedback Review with and without the Tool (RQ2)

As part of the post-project interviews, we performed a short deprivation protocol to study which strategies (if any) performed with the tool continued when reviewing feedback in a text-only form. The interviewees used a document editor (Google Docs) to identify what they believed to be the most important insights in a collection of six pieces of feedback for a design project unrelated to their own course projects. Interviewees were encouraged to use any of the tool's features they felt necessary to mark up the text or take notes. Interviewees were given 20 minutes to review the feedback, extract what they felt were the most important insights, and think aloud while reviewing the feedback.

We observed that most students performed strategies similar to those implemented in the visualization tool and created a visible trace of their processing of the feedback. For highlighting, two students applied color-coded highlighting to identify different topics in the feedback, and two other students used color to differentiate criticism from suggestions. Five students used a single color to highlight important insights, while the remaining three students did not use color to highlight at all. Regarding additional strategies, two students copied the most critical statements from the feedback to the end of each piece of feedback, and in one case a student added topic labels (e.g., "Content:" or "Presentation:") and notes to the copied statements. Similarly, a different student selected feedback statements and added comments in the format "[topic] [opinion]". Three students used boldfacing to mark the feedback statements they found most valuable. Two students reviewed the feedback without marking up the feedback text in any way. While students mimicked some of the same techniques implemented in the visualization tool, they weren't able to apply these techniques as effectively without the tool's consistent visual structure. For example, students who

used highlighting typically developed arbitrary mappings of the colors to the topics or opinions in the feedback, and did not consistently apply these mappings.

Following the deprivation protocol, interviewees were asked about the strengths and weaknesses of reviewing text-only feedback, as well as the strengths and weaknesses of the visualization tool. Interviewees cited the difficulty of remembering (N = 7) and organizing (N = 4) the feedback as the biggest weakness of reviewing feedback in text-only form. By contrast, interviewees cited the ability to aggregate and categorize data (N = 5) and identify the most common and important criticisms (N = 5) as the biggest strengths of the visualization tool. These responses reflect the value of the visualization tool in its facilitation of different strategies and visualization of topic and opinion structure in a consistent manner. Interviewees also cited the ability to develop their own interpretations of the feedback without possible biases imposed upon the feedback's presentation as the biggest strength of reviewing feedback in a text viewing tool over the visualization tool (N = 6).

When asked in the second part of the interviews whether their strategies for reviewing feedback changed after using the feedback visualization tool for the course projects, six interviewees reported developing new processes for engaging with feedback, many of which they felt would carry beyond the use of the tool. One student mentioned they began highlighting plain-text feedback they received with different colors to help them review and organize the feedback more effectively.

> I really liked the idea of color coding. Usually whenever I go in and look at anyone's feedback and it's just in text form, I would probably like make notes on a separate document for it but I wouldn't visually go in and change the way the feedback looks, but I think that helped me process things a lot easier when I went back into the feedback after the first time I looked at it. So I know that's something I've kind of even implemented in general, like actually going and working and annotating with the text instead of like on a separate document. [I10 (Team 2)]

A different student reported that after working with the tool, they began taking notes on recurring themes in plain-text feedback they received to reduce the burden of reviewing and reprocessing it later. Another student who was in the habit of skimming plain-text feedback stated that after working with the tool, they learned to appreciate the value of finding and marking the commonalities in the feedback they received, noting they wanted to ascertain that they were making full use of the feedback.

> When going through feedback I would honestly skim through multiple pieces of feedback and try to see if there were any common elements. If I did see common elements then I would kind of jot them down, and every time that I saw something similar to what I read before I would just kind of increase the value of that comment, just so I could remember that more people were arguing for this approach. That I feel like is a better way of going through feedback than just trying to remember it all in your head and trying to proceed from there because that just gets very confusing very fast [I1 (Team 18)]

> It definitely changed a little bit. Before, like I mentioned I just go through a couple of them and then also, I probably only see big themes...like big commonalities between them. But I think now if I were just given like a bunch of text based feedback...I kind of go through line by line and then really think about what smaller commonalities each of these reviews have. And then focus on those a lot more and highlight them so I don't [forget] them, which is something I never used to do...and just make sure I'm fully utilizing everything that people have written for me. [I4 (Team 11)]

Table 3. Distribution of Opinion Categories by Project Deliverable

	Praise	Criticism	Neutral	Suggestion	Question
Proposal	62.33%	33.41%	4.26%	N/A	N/A
Low-fidelity Prototype	48.99%	17.75%	0.00%	26.52%	6.74%
Functional Prototype	50.80%	13.93%	0.00%	30.26%	5.01%

The research team labeled all the feedback for the first project deliverable (Proposal), teams could revise the labels assigned by the research team for the feedback for the second deliverable (low-fidelity prototype), and student teams labeled their own feedback for the third deliverable (functional prototype). The "Suggestion" and "Question" labels were not yet implemented in the tool at the time of the first project deliverable.

A fourth student who had previously considered praise feedback unhelpful reported that they more thoroughly considered the praise they received in the context of consciously maintaining stronger aspects of their design.

> ...a lot of time when people are given feedback, they're so worried about fixing the bad things...that sometimes they do it at the cost of changing something that was good in their app or just good in their work that they did. So I think something that I've really learned is that looking at good feedback and understanding what someone says about something that you did well is super important and can actually help your weaknesses and help you not make changes that you shouldn't be making. [I5 (Team 16)]

With the feedback visualization tool, students found the feedback that was valuable to them by observing clusters of the same category of opinion icon, among other strategies. Without the tool, we found that some interviewees (N = 10) also attempted to process the topics and opinions in the feedback, but used ad-hoc color codes, inline comments, or free-form notes, while others (N = 2) processed the feedback in memory. These results suggest that user strategies for processing feedback might have limited effectiveness without the support of a tool designed to foreground the topics and opinions in the feedback with a consistent visual presentation.

6.3 Impact of Labeling Approach on Visualization and Feedback Perceptions (RQ3)

The feedback tool requires metadata in the form of topic and opinion labels in order to generate the visualization for a feedback collection. An important consideration is how this metadata should be created and by whom. We compared three approaches: the research team labels the feedback, student teams revise the labels assigned by the research team, and students label the feedback on their own. We begin the results with the student-led labeling.

Student teams labeled their own feedback for the third project deliverable. There were a total of 1096 statements in the peer and TA feedback, or about 61 statements for each team. The interaction data showed that the student teams fully labeled all the statements in their respective collections of feedback. We also found that the distribution of labels assigned by the student teams was similar to the distribution of the labels assigned by the research team for the second project deliverable (see Table 3). Although different collections of feedback were labeled by the students and by the research team, the similar distributions suggest that students did not interpret the feedback for their own projects with a particular bias (e.g., with overly positive or negative interpretations).

On the tool usage survey associated with the third project deliverable, 60% of respondents estimated their team spent at most 30 minutes labeling the feedback, and 30% estimated their team spent between 30 and 60 minutes. When asked how their team labeled the feedback, 58% of respondents indicated that one member of the team labeled the feedback, 25% indicated that the team labeled the data together, 10% indicated that the team divided the work, and 7% did not answer or reported other strategies. Students agreed with the statement that labeling the topics

and opinions in the feedback was useful (Q16: mean = 4.9, SD = 1.3) and perceived the benefit of seeing the resulting visualization to outweigh the cost of labeling the feedback (Q19: mean = 4.7, SD = 1.4).

When asked about the tradeoffs between the three approaches experienced for labeling the feedback, interviewees (N = 12) mentioned two benefits of labeling the feedback on their own. First, nine interviewees mentioned that labeling the feedback forced them to think more carefully about each piece of feedback they received and to become familiar with the feedback more quickly than in the prior project deliverables. The remaining three interviewees did not cite any particular advantages to labeling the feedback themselves over having the research team label it for them. In response to the open-ended question about the tool's main strengths, survey respondents (N = 17, 25%) also noted the benefits of familiarizing themselves with the feedback through the labeling process. Students noted that assigning the labels prompted them to begin thinking about the most critical pieces of feedback before they saw the resulting visualization.

> ...for the third iteration where we were like in charge of labeling our own feedback that was like...we were like oh it actually you know it's like not tedious but like a process you kinda have to go through each sentence and really think about "oh this is for appearance, this is for content, I think this is positive, I think this is negative, this is an idea, this a confusion point"...I think that was like where we probably got the most value out of 'cause we were thinking about it a lot more and actually working with this a lot more and categorizing it for ourselves rather than just agreeing with what was already there. [I4 (Team 11)]. The biggest benefit was that we actually got very familiar with the feedback as we organized it. When it was organized for us, we didn't pay attention to every piece of feedback, but this time we really got to understand how our peers viewed our demo and which suggestions or critiques were the most important to discuss and work on. [S25 (Team 2), Functional Prototype Survey]

A second benefit was being able to appropriate the labels to fit their own needs. For example, two students reported in the tool usage survey for the third deliverable that their respective teams developed a new schema for the topic labels in the tool, appropriating the topic labels for use as an action list.

> The opportunity to categorize feedback to help us prioritize the issues was helpful. Also, we split up the team into frontend and backend using a divide and conquer approach. Therefore, the categories allowed us to easily delegate work. [S5 (Team 16), Functional Prototype Survey]

For disadvantages of labeling the feedback on their own, eight interviewees cited issues with the labeling process, including the difficulty deciding upon the appropriate label for a piece of feedback, difficulty tracking where they left off when splitting the labeling into multiple sessions, and the lack of an undo button. The other four interviewees cited the time required to label the feedback in the tool as the primary disadvantage. However, three of these interviewees also stated the value of the resulting visualization outweighed the labeling effort.

For the second project deliverable, students were given the option to revise the labels assigned by the research team. The interaction logs indicated that none of the project teams revised these labels. The lack of revision was likely due to the fact that students agreed with the statement on the tool usage survey that the labels assigned by the research team matched what they would have assigned themselves (mean = 5.16, SD = 1.06, from the second tool usage survey).

> I don't think we changed any of the labels, to be honest. I think we did look through them. But we agreed with most of them. So we're like, "okay, cool." [I4 (Team 11)]

> And even in the second time, where we had the ability to go back and edit, we didn't re-
> ally use...that opportunity, because it was already kind of done for us. And we honestly
> got a little lazy, because it was already done. [I10 (Team 2)]

Despite not revising the research team's labels in practice, all interviewees preferred having the
option to revise the pre-labeled feedback. When asked which of the three labeling approaches they
preferred (pre-labeled without revision, pre-labeled with revision, and self-labeled), six intervie-
wees stated they preferred labeling the feedback themselves, three preferred revising the research
team's labels, and one felt both of these approaches were equally valuable. The remaining two
interviewees had more nuanced opinions, stating pre-labeled feedback might be most effective
for the first time they used the tool, or when they were working with a group they were less
familiar with.

> If you are comfortable with your team, if you work together often, I would say that
> doing it manually...is gonna be very helpful because you have the manpower to do it.
> and then you're enjoying your time anyways on the team call. it's not like a burden for
> you. but maybe some people who [don't] click with their teammates and just want to
> get over with everything quickly...if I were to put myself in their shoes, I would rather
> have the automatically annotated feedback just so I don't have to spend too much time
> with these random people...it really depends on a case by case basis. [I2 (Team 10)]

> I would probably pick [self-annotation], but if you were just talking in general, I think
> it's a good idea to have [editable staff annotations] as an introductory approach. if it's
> not for me or for other students, it's nice to have an introduction to it and then have
> the ability to annotate, because I think you understand the tool a lot better that way.
> [I10 (Team 2)]

In sum, from these different approaches for labeling the feedback, we found that students are
willing to fully label the topics and opinions in the feedback for their projects in order to produce
the resulting visualization, assign labels without inflating or deflating the criticism they receive,
and report processing the feedback more carefully than if the labels are provided to them.

6.4 Appropriating the Tool for Instructional Purposes (RQ4)

During the interviews, the TAs referenced several different uses of the tool for instructional pur-
poses. TA1 mentioned using the tool to review student feedback to help synthesize her own feed-
back. She also reported writing her feedback precisely such that the topics in the feedback would
be easy to recognize and label in the tool.

> There were a couple of times somebody would say something...and I could kind of call
> that out in my feedback like "oh so and so mentioned I think this is a good idea" or
> I may say something like "oh I saw a lot of the feedback said something about this
> particular feature, maybe you want to consider doing something with that." So I sort
> of used it to inform the feedback I was writing. [TA1]

> One thing is it sort of made me more aware of the categories... thinking more explicitly
> about the topics of the feedback and where it might fall on the visualization. So I might
> tweak the wording a little bit to make it more clear like "oh this is about the idea and
> this is about the implementation". So I guess it made me be a little bit more precise
> with my language. [TA1]

TA1 also mentioned that she used the tool in part to justify her grading decisions, stating she
was more inclined to deduct points if a lot of criticism icons were present in the visualization.

> I don't think it hindered me at all. I think there may have been a couple times when we were using the tool where I sort of looked at the overall like color of the of the comments that were given here to see, is there a lot of green? I guess it helped with the severity of deducting points, if that makes sense. So I might be more inclined to deduct more points if I saw a lot of red and or yellow in the chart. [TA1]

TA2 mentioned revisiting the visual structure of her feedback in the tool (e.g., the critical points) to guide her later discussions with the project team during office hours. The same TA also reported being conscious of the need to balance the opinions in her critique of the project deliverable because she did not want her feedback to appear inconsistent with the peer feedback adjacent in the tool.

> Whenever a student comes to my office hour [to] discuss about their project, I can't remember all 10 teams' feedback at that point sometimes, so I go to this tool, look at my previous feedback, and this tool has these nice visualizations that I can look into the negative feedback I gave to that team, so I just quickly look through what I said negatively about their project, and then remember what I said and give advice. [TA2]

> ...my impression of reading the comments from students was that they were mostly positive. They were very positive. compared to my feedback, because I have to have some negative points if I need to deduct points. I felt since everyone was so positive about the project, I felt kind of guilty to say negative things in my own feedback, so I tried to have a more positive tone and I also tried to make positive statements when I'm making feedback. [TA2]

These uses of the tool for instruction should be considered preliminary because our sample only included two TAs. However, these preliminary findings do indicate an opportunity for future work exploring how a feedback visualization tool could aid course instruction.

7 DISCUSSION

We deployed a feedback visualization tool in a project-based course to learn how students leverage this type of tool for interpreting feedback received from multiple providers—in this case, classroom peers and the TAs. Here, we summarize the main findings for each of the research questions and discuss these findings in light of the tool's interaction design, the learning context of the study, and the broader literature.

7.1 RQ1: Pursuing Feedback Goals

For RQ1, we found that students wanted to accomplish three goals when using the tool. One of these goals was finding the feedback that they believed was valuable for improving their project. This was accomplished by using the tool to explore feedback by its opinion structure (critical first), observing clusters of the same category of icon within the topic columns, and reviewing the background of the providers to weight their specific feedback statements. Students stated they typically read critical comments first to determine the problems they needed to address in their project deliverable, and sometimes ignored positive comments altogether.

The student's preference to attend to the critical statements first conflicts with how providers are typically taught to write formative feedback. Providers are typically taught to write design feedback in the form of a "feedback sandwich", offering praise before criticism and ending on a positive note [11]. The expectation is that the recipient will read the feedback in the order it was written. Prior work has found that reading the feedback from a positive to negative valence order improves the recipient's perception of the feedback collection relative to placing the negative

feedback in other positions [45]. Future work could experiment with different constraints to affect how recipients are able to review the feedback (e.g., requiring recipients to open all praise comments before critical comments or suggestions are revealed). The color scheme used for the opinion icons also made it easy for students to attend to the critical statements (red icons) before the praise and suggestions (green and yellow icons), as red hues are known to have a strong pop-out effect [25]. Future work could allow users to configure different color schemes for the icons. Presentational choices such as icon color and ordering constraints demonstrate the potential impact of separating feedback's representation from its content, and encourage further exploration of how different means of presenting feedback can facilitate feedback review.

Students used their peers' self-reported familiarity with the project topic and year of study to weigh feedback statements, but did not report using other attributes such as major of study. The right amount of background information to collect about the feedback providers and project creators and present in a feedback visualization tool is an open question. For example, including provider background data such as gender or ethnicity could facilitate unwanted biases from the project creators during feedback review [34], whereas presenting too little information could inhibit creators' interpretation of the feedback. Conversely, disclosing demographic information about the project creators to the providers could cause creators to question the legitimacy of critical feedback they receive [31, 39, 41]. The most appropriate background information to present might also differ by project topic and stage, all of which argue for giving instructors the option to toggle the collection and visibility of various background information for the feedback providers and project creators in the tool. The dataset generated through the use of the tool (e.g., labels indicating how students intend to act on specific feedback statements) might also enable analysis that sheds light on how students with different backgrounds exchange feedback and build a network of trusted peers. Future work could extend the implementation of the tool to allow project creators to solicit the most appropriate feedback for them from within these networks.

A second goal students pursued with the tool was assessing the quality of their projects by comparing the relative proportions of different opinion icons across topics or within a particular subset of the topics. One way to further enhance the tool in support of this goal would be to implement visual summaries of the icon categories for each topic column and overall. Observing the ratios of icon categories in the tool might be difficult for large feedback collections, e.g., if using the tool to visualize the course evaluation comments written by hundreds of students. To scale to larger feedback sets without overwhelming the user with information, the tool could incorporate hierarchy into the visualization by organizing reviewers and topics into a taxonomy and allowing creators to "drill down" [21] to progressively explore the details. Finally, a third goal students pursued with the tool was facilitating team discussion and coordination, e.g., by assigning intention labels to the feedback statements and later searching and filtering the feedback by these labels. Team coordination could be further supported by allowing students to enter an estimated amount of time needed for making revisions prompted by the feedback statement, assigning the revisions to a team member, and generating an action list for each team member. The time estimates could be used to ensure balanced workloads between team members and to help prioritize the revisions in the project timeline.

7.2 RQ2: Processes for Reviewing Feedback

When reviewing text feedback during the interviews, students demonstrated techniques that mirrored the tools' features, such as color coding praise and critical comments and marking intentions to act on statements by boldfacing those statements. To determine whether these techniques were indeed inspired by the tool, we performed the same interview protocol with five new students enrolled in a later instance of the same course which did not incorporate the visualization tool.

These students were asked to think aloud while reviewing a given collection of feedback using a text viewing tool. While students in both sets of interviews attempted to find the most valuable insights by looking for redundancy, students who did not use the visualization tool leveraged alternative techniques that required extra steps to process the feedback and identify these insights. Specifically, interviewees who did not use the visualization tool highlighted with arbitrary color mappings (N = 2/5) or not at all (N = 3/5), did not attempt to organize the feedback (N = 4/5), and backtracked when reviewing feedback (N = 4/5). By contrast, students who used the visualization tool typically highlighted the feedback as they were reading it (N = 10/12), corroborating their reports of familiarizing themselves with feedback through the annotation process. These students also used color-coded opinion highlighting that mirrored how they looked for opinion clusters in the tool (N = 8/12), and marked important insights with either boldface (N = 2/12) or inline comments (N = 3/12) similar to their use of the tool's intention labels. Our findings suggest the visualization tool helped students develop general strategies for reviewing feedback which can be applied with or without tool support. These findings also highlight the opportunity to teach effective feedback review practices through tool design.

7.3 RQ3: Impact of Labeling Approach on Feedback Engagement

We found that students were willing to fully label the topics and opinions in their feedback to produce the visualization, assigned opinion labels without particular bias (such as preferring labeling feedback as praise or suggestions rather than as criticisms), and reported processing the feedback more carefully than if the labels were provided to them. Additionally, survey responses indicated that students believed the benefits of generating the visualization and familiarizing themselves with the feedback for their project outweighed the costs of assigning the labels. These findings support having students generate the metadata for their own feedback when deploying similar feedback visualization tools in course contexts. However, to scaffold student use of the tool [24, 37], course staff may want to initially label some feedback themselves to demonstrate the resulting visualization so students understand why the labels are necessary and can calibrate their labeling decisions. The course staff might choose to label all the feedback in situations where the instructors' perspectives and consistency of labels between teams are most desirable.

The labeling interface in the tool provides a list of the feedback statements and pull-down menus for selecting the topic and opinion categories for each statement. The user interface is straightforward, but the statements are shown in the form of a list rather than in the form of a narrative. An alternative labeling interface might allow the user to label the data while reading the feedback content, similar to how the interviewees annotated the text in the deprivation protocol. For example, the tool could provide a specific highlighter for each opinion category and the user could select text and drag representations or copies of that text into topic categories. The resulting visualization would be the same, but the labels could be captured in way that aligns with what was observed during the deprivation protocol in the interviews.

There are additional options for generating the metadata that were not tested in this study. One option is to ask the providers to label the feedback as they write it, which could help them communicate their ideas and alleviate work for the recipients. However, this approach could also increase the provider's cognitive burden when writing the feedback and lead to inconsistent labels across multiple providers. Another option is to use machine learning models to assign topic and opinion labels that the students could revise. We conducted an exploratory test of this approach with our dataset. Using the students' and research team's labels as a baseline, our exploration found that an off-the-shelf model achieved a 59.4% accuracy on assigning opinion labels across all feedback collected in the study for each team and deliverable. This accuracy is likely a lower bound, as a training dataset could be expanded over time as more feedback statements are labeled

in the tool. Note that we did not have sufficient data to build and test statistical models for the topic labels because the topics were different for each project deliverable. A future study should compare how delegating the labeling task to different sources (provider, recipient, instructor, or automation) affects not only accuracy and effort, but also the recipient's comprehension of the feedback.

7.4 RQ4: Instructional Impact of the Tool and Future Improvements

Both TAs used each team's visualization to help determine a grade for the associated project deliverable and write meta-level feedback explaining that grade. They also used the visualization of the feedback to discuss strengths and weaknesses of a project with a student team and provide additional guidance. One assistant also noted that she thought about how her feedback would appear in the tool and tried not to write comments that would appear inconsistent with the peer feedback for that same project. Future work should explore how seeing a visualization of the topic and opinion structure in their own feedback affects the feedback an instructor later writes.

Our interviews with the TAs indicate there are additional opportunities to extend the tool to enhance course instruction. For instance, topic columns in the tool that contain many criticism icons across many projects might indicate gaps in project knowledge for the recipients or unrealistic expectations of the providers. An instructor might use the tool to identify such topic columns and prioritize the coverage of these issues in course materials or manage students' expectations of their peers' projects. Instructors might also use praise labels to curate examples of good project deliverables and best practices that could be incorporated into lecture materials or linked as examples in the tool. Finally, instructors could measure progress on open-ended assignments by linking students' revisions back to planned changes identified in the intention labels assigned to the feedback.

To improve feedback review, the most commonly requested feature was the ability to attach notes to individual pieces of feedback, usually for explaining how a piece of feedback was addressed or why it wasn't addressed. Since feedback statements labeled with the same topic in the tool did not always address the same issue, students requested the ability to add keywords for each statement in the tool, allowing them to highlight subsets of icons that share keywords. Finally, the TAs wanted students to be able to explain in the tool when the students disagreed with statements in their feedback or chose not to implement a specific suggestion. These suggested features indicate simple but powerful opportunities for improving the tool's usability for instructors and students alike.

8   LIMITATIONS AND FUTURE WORK

The results reported in this article were derived from a field study methodology. Though this method allowed us to answer questions about the use of a feedback visualization tool in an authentic classroom setting, we are unable to compare these results to students reviewing feedback with existing tools. Future controlled experiments are needed to determine how the choice of tool for feedback review affects feedback comprehension, interpretation strategies, and quality of revisions made to a project. We also studied the use of the feedback visualization tool in the context of a course where students earn grades for their course performance. Additional work is needed to test the use and benefits of a feedback visualization in other contexts, such as when performing design projects for enjoyment, for entrepreneurship, or as part of one's job responsibilities. Future work should also investigate whether students adapt their labeling of the feedback over multiple uses of the tool (e.g., do students adapt the labels to try to affect grading decisions?) and study the feedback situations for which students choose to use this type of tool on their own.

Students worked on the course projects in teams, and each team may have appropriated the tool in different ways (e.g., reviewing feedback as a team vs. reviewing feedback individually and discussing it later as a team). Our dataset was not large enough to determine how the uses of the

tool reported by students relate to different teamwork styles. Because it was a research prototype, the tool deployed in this study did not yet implement some collaborative features, such as assigning "must do" actions to specific members of a team or allowing feedback annotation by distributed team members simultaneously. How these and other possible collaborative features affect the use of a feedback visualization tool remains an open question. Because we deployed a specific feedback visualization tool in a single course, it is also possible students practiced feedback review processes that were not observed in our study. Moreover, the design of the feedback sessions in the course produced the type of feedback that is best visualized in the tool: feedback by multiple providers with different backgrounds, foci, and opinions. The use of the tool might not be as beneficial if used in courses where the feedback is generated from a few peers with similar backgrounds. The results reported in this article should be expanded by testing the use of feedback visualization tools with different user populations, project types, and course cultures.

We see several additional directions for future work. One direction is to extend the tool to import and visualize the feedback written in different review contexts, such as comments posted in online critique communities (e.g., Reddit and Behance), in PDF documents, and in academic peer review platforms. A second direction is to extend the tool to support the *composition* of feedback, in addition to the interpretation of the resulting feedback collection. For example, future work could extend the tool to visualize the feedback as it is submitted and evaluate how showing providers the in-progress visualization affects the feedback they write (e.g., do they direct their comments to empty cells in the visualization?). Likewise, the tool could be extended to allow the content creator to visually mark the topics for which they are most interested in receiving feedback, and make these markings available to feedback providers at the onset of the composition process. A third direction is to leverage the tool to build an open dataset of feedback statements annotated with topic, opinion, and intention labels. The dataset could, for example, be used by instructors to find examples of feedback statements that promote revision actions for students and contrast these with examples of feedback statements that are not acted upon. Finally, the tool could be extended to support additional metadata such as the estimated time to implement a feedback statement. This metadata could be used to generate action plans based on how much time the content creator is willing to invest in revising the work.

9 CONCLUSION

Feedback for creative projects can be difficult to understand, especially when it involves resolving conflicting opinions, judging the credibility of suggestions, and prioritizing the issues raised across a multitude of topics. In this article, we reported findings for how student teams engaged with an interactive tool that visualizes the topic and opinion structure within a collection of feedback written by peers and TAs for their design projects. We found that the tool was useful for scaffolding students' processes for reviewing the feedback. Students leveraged the structure of the visualization in the tool to explore the critical statements in the feedback first (thumbs down icons), to assess the quality of their projects (ratio of praise to criticism icons across topics), and to facilitate team discussion about how to best revise their projects in response to the feedback. We also found that students were willing to create the metadata needed to generate the visualization in the tool. Students indicated that the effort required to label the feedback was outweighed by the benefits of familiarizing themselves with the feedback and having access to the resulting visualization. Though the goal of the tool is to help students learn and apply skills for feedback interpretation, the results of our study also revealed opportunities to use the tool to benefit instruction. These opportunities include curating examples of feedback with an appropriate balance of praise, criticism, and suggestions, identifying student projects that are succeeding (e.g., disproportionate praise) or in need of additional mentorship (e.g., disproportionate criticism), and identifying topics that need further

explanation (e.g., seldom referenced in the feedback). The tool and findings reported in this article contribute to a future in which students are taught not only how to write good feedback, but are also taught skills for effectively interpreting and acting on that feedback for their creative projects.

REFERENCES

[1] Claire Aitchison. 2014. Learning from multiple voices: Feedback and authority in doctoral writing groups. In *Writing Groups for Doctoral Education and Beyond.* Routledge, 67–80.

[2] Salvatore Andolina, Hendrik Schneider, Joel Chan, Khalil Klouche, Giulio Jacucci, and Steven Dow. 2017. Crowdboard: Augmenting in-person idea generation with real-time crowds. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition.* 106–118.

[3] Robert G. Bing-You, Jay Paterson, and Mark A. Levine. 1997. Feedback falling on deaf ears: Residents' receptivity to feedback tempered by sender credibility. *Medical Teacher* 19, 1 (1997), 40–44.

[4] Kirsten R. Butcher and Tamara Sumner. 2011. Self-directed learning and the sensemaking paradox. *Human–Computer Interaction* 26, 1–2 (2011), 123–159.

[5] Julia Cambre, Scott Klemmer, and Chinmay Kulkarni. 2018. Juxtapeer: Comparative peer review yields higher quality feedback and promotes deeper reflection. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* 1–13.

[6] Kwangsu Cho and Charles MacArthur. 2010. Student revision with peer and expert reviewing. *Learning and Instruction* 20, 4 (2010), 328–338.

[7] Amy Cook, Steven Dow, and Jessica Hammer. 2020. Designing interactive scaffolds to encourage reflection on peer feedback. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference.* 1143–1153.

[8] Amy Cook, Jessica Hammer, Salma Elsayed-Ali, and Steven Dow. 2019. How guiding questions facilitate feedback exchange in project-based learning. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* 1–12.

[9] Patrick A. Crain and Brian P. Bailey. 2017. Share once or share often? Exploring how designers approach iteration in a large online community. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition.* 80–92.

[10] Monique A. Dijks, Leonie Brummer, and Danny Kostons. 2018. The anonymous reviewer: The relationship between perceived expertise and the perceptions of peer feedback in higher education. *Assessment & Evaluation in Higher Education* 43, 8 (2018), 1258–1271.

[11] Anne Dohrenwend. 2002. Serving up the feedback sandwich. *Family Practice Management* 9, 10 (2002), 43.

[12] James Elkins. 2012. *Art Critiques: A Guide.* New Academia Publishing.

[13] Eureka Foong, Darren Gergle, and Elizabeth M. Gerber. 2017. Novice and expert sensemaking of crowdsourced design feedback. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–18.

[14] Graham Gibbs and Claire Simpson. 2005. Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education* 1 (2005), 3–31.

[15] John Hattie and Helen Timperley. 2007. The power of feedback. *Review of Educational Research* 77, 1 (2007), 81–112.

[16] Pamela J. Hinds. 1999. The curse of expertise: The effects of expertise and debiasing methods on prediction of novice performance. *Journal of Experimental Psychology: Applied* 5, 2 (1999), 205.

[17] Karen Holtzblatt and Hugh Beyer. 1997. *Contextual Design: Defining Customer-Centered Systems.* Elsevier.

[18] Maria Jackson and Leah Marks. 2016. Improving the effectiveness of feedback by use of assessed reflections and withholding of grades. *Assessment & Evaluation in Higher Education* 41, 4 (2016), 532–547.

[19] Mahmood Jasim, Enamul Hoque, Ali Sarvghad, and Narges Mahyar. 2021. CommunityPulse: Facilitating community input analysis by surfacing hidden insights, reflections, and priorities. In *Proceedings of the Designing Interactive Systems Conference 2021.* 846–863.

[20] Mahmood Jasim, Pooya Khaloo, Somin Wadhwa, Amy X. Zhang, Ali Sarvghad, and Narges Mahyar. 2021. CommunityClick: Capturing and reporting community feedback from town halls to improve inclusivity. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–32.

[21] Mikael Jern. 1997. Information drill-down using web tools. In *Visualization in Scientific Computing'97.* W. Lefer and M. Grave (Eds.), Springer, 9–20.

[22] Anders Jonsson. 2013. Facilitating productive use of feedback in higher education. *Active Learning in Higher Education* 14, 1 (2013), 63–76.

[23] Hyeonsu B. Kang, Gabriel Amoako, Neil Sengupta, and Steven P. Dow. 2018. Paragon: An online gallery for enhancing design feedback with visual examples. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* 1–13.

[24] Duenpen Kochakornjarupong, Paul Brna, and Paul Vickers. 2005. Who helps the helper? A situated scaffolding system for supporting less experienced feedback givers. In *Proceedings of the 2005 Conference on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology.* 851–853.

[25] Michał Kuniecki, Joanna Pilarczyk, and Szymon Wichary. 2015. The color red attracts attention in an emotional context. An ERP study. *Frontiers in Human Neuroscience* 9 (2015), 212.

[26] J. Richard Landis and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 1 (1977), 159–174.

[27] Janet Lefroy, Chris Watling, Pim W. Teunissen, and Paul Brand. 2015. Guidelines: The do's, don'ts and don't knows of feedback for clinical education. *Perspectives on Medical Education* 4, 6 (2015), 284–299.

[28] Michael Xieyang Liu, Jane Hsieh, Nathan Hahn, Angelina Zhou, Emily Deng, Shaun Burley, Cynthia Taylor, Aniket Kittur, and Brad A. Myers. 2019. Unakite: Scaffolding developers' decision-making using the web. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 67–80.

[29] Kurt Luther, Jari-Lee Tolentino, Wei Wu, Amy Pavel, Brian P. Bailey, Maneesh Agrawala, Björn Hartmann, and Steven P. Dow. 2015. Structuring, aggregating, and evaluating crowdsourced design critique. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 473–485.

[30] Narges Mahyar, Michael R. James, Michelle M. Ng, Reginald A. Wu, and Steven P. Dow. 2018. CommunityCrit: Inviting the public to improve and evaluate urban design ideas through micro-activities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.

[31] Rodolfo Mendoza-Denton, Michelle Goldman-Flythe, Janina Pietrzak, Geraldine Downey, and Mario J. Aceves. 2010. Group-value ambiguity: Understanding the effects of academic feedback on minority students' self-esteem. *Social Psychological and Personality Science* 1, 2 (2010), 127–135.

[32] Daniel Meulbroek, Daniel Ferguson, Mathew Ohland, and Frederick Berry. 2019. Forming more effective teams using CATME teammaker and the Gale-Shapley algorithm. In *Proceedings of the 2019 IEEE Frontiers in Education Conference*. IEEE, 1–5.

[33] Kamila Misiejuk, Barbara Wasson, and Kjetil Egelandsdal. 2021. Using learning analytics to understand student perceptions of peer feedback. *Computers in Human Behavior* 117 (2021), 106658.

[34] Gabriela Morales-Martinez, Paul Latreille, and Paul Denny. 2020. Nationality and gender biases in multicultural online learning environments: The effects of anonymity. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[35] D. Royce Sadler. 1989. Formative assessment and the design of instructional systems. *Instructional Science* 18, 2 (1989), 119–144.

[36] Amy Shannon, Jessica Hammer, Hassler Thurston, Natalie Diehl, and Steven Dow. 2016. PeerPresents: A web-based system for in-class peer feedback during student presentations. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. 447–458.

[37] Lorrie A. Shepard. 2005. Linking formative assessment to scaffolding. *Educational Leadership* 63, 3 (2005), 66–70.

[38] David R. Thomas. 2006. A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation* 27, 2 (2006), 237–246.

[39] Gladman Thondhlana and Dina Zoe Belluigi. 2017. Students' reception of peer assessment of group-work contributions: Problematics in terms of race and gender emerging from a South African case study. *Assessment & Evaluation in Higher Education* 42, 7 (2017), 1118–1131.

[40] Yaacov Trope, Ben Gervey, and Niall Bolger. 2003. The role of perceived control in overcoming defensive self-evaluation. *Journal of Experimental Social Psychology* 39, 5 (2003), 407–419.

[41] Tom R. Tyler. 2001. Why do people rely on others? Social identity and social aspects of trust. In *Trust in Society*. K. S. Cook (Ed.), Russell Sage Foundation, 285–306.

[42] Iris Vardi. 2008. The relationship between feedback and change in tertiary student writing in the disciplines. *International Journal of Teaching and Learning in Higher Education* 20, 3 (2008), 350–361.

[43] Karl E. Weick, Kathleen M. Sutcliffe, and David Obstfeld. 2005. Organizing and the process of sensemaking. *Organization Science* 16, 4 (2005), 409–421.

[44] Naomi E. Winstone, Robert A. Nash, James Rowntree, and Michael Parker. 2017. 'It'd be useful, but I wouldn't use it': Barriers to university students' feedback seeking and recipience. *Studies in Higher Education* 42, 11 (2017), 2026–2041.

[45] Y. Wayne Wu and Brian P. Bailey. 2017. Bitter sweet or sweet bitter? How valence order and source identity influence feedback acceptance. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*. 137–147.

[46] Y. Wayne Wu and Brian P. Bailey. 2018. Soften the pain, increase the gain: Enhancing users' resilience to negative valence feedback. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–20.

[47] Anbang Xu, Shih-Wen Huang, and Brian Bailey. 2014. Voyant: Generating structured feedback on visual designs using a crowd of non-experts. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 1433–1444.

[48] Anbang Xu, Huaming Rao, Steven P. Dow, and Brian P. Bailey. 2015. A classroom study of using crowd feedback in the iterative design process. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 1637–1648.

[49] Yueting Xu and David Carless. 2017. 'Only true friends could be cruelly honest': Cognitive scaffolding and social-affective support in teacher feedback literacy. *Assessment & Evaluation in Higher Education* 42, 7 (2017), 1082–1094.

[50] Koji Yatani, Michael Novati, Andrew Trusty, and Khai N. Truong. 2011. Review spotlight: A user interface for summarizing user-generated reviews using adjective-noun word pairs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1541–1550.

[51] Yu-Chun Yen. 2021. *Scaffolding Feedback Interpretation Process for Creative Work Through Reflection, Paraphrasing, and Information Visualization*. Ph.D. Dissertation. University of Illinois at Urbana-Champaign.

[52] Yu-Chun Grace Yen, Steven P. Dow, Elizabeth Gerber, and Brian P. Bailey. 2017. Listen to others, listen to yourself: Combining feedback review and reflection to improve iterative design. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*. 158–170.

[53] Yu-Chun Grace Yen, Joy O. Kim, and Brian P. Bailey. 2020. Decipher: An interactive visualization tool for interpreting unstructured design feedback from multiple providers. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[54] Ling Ying-Leh, Abdul Ghani Kanesan Abdullah, and Aziah Ismail. 2015. Feedback environment and coaching communication in Malaysia education organizations. *Asian Journal of Social Sciences & Humanities* 4, 1 (2015), 66–73.

[55] Alvin Yuan, Kurt Luther, Markus Krause, Sophie Isabel Vennix, Steven P. Dow, and Bjorn Hartmann. 2016. Almost an expert: The effects of rubrics and expertise on perceived value of crowdsourced design critiques. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 1005–1017.