Compass: Supporting Large Group Mentorship in a Chat-Based UI

KEHUA LEI*, UC Santa Cruz, USA
MINGRUI YU*, UC Santa Cruz, USA
MARISSA LEWELLEN*, UC Santa Cruz, USA
VENUS KU, UC Santa Cruz, USA
DAVID T. LEE, UC Santa Cruz, USA

While mentorship offers many benefits, student access to mentorship is often limited. In this paper, we introduce Compass, a chat platform where industry professionals mentor large cohorts of 30+ students with the support of novel features that enable full engagement without the typical chaos of group chat. Specifically, we conceptualize conversations as composed of not only individual messages, but also *multi-person conversational units* that collapse large numbers of small but related conversational exchanges into single conceptual units in the main dialogue. Doing so makes it possible to preserve a coherent linear flow of conversation while also supporting non-linear conversational exchanges that can be concisely summarized computationally and built on in the main conversation. We report on design lessons learned over a year of small real-world studies culminating in a final deployment in which 2 industry professionals successfully mentored 30+ students over a 10-week period. We find that both mentors and mentees find the chat UI effective and sometimes preferable, and discuss broader implications for the design of chat UI for large group conversations.

CCS Concepts: • **Human-centered computing** → **Interactive systems and tools**.

Additional Key Words and Phrases: large group mentorship, conversational UX for large groups, career mentorship

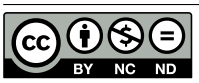
ACM Reference Format:
Kehua Lei, Mingrui Yu, Marissa Lewellen, Venus Ku, and David T. Lee. 2023. Compass: Supporting Large Group Mentorship in a Chat-Based UI. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 37 (April 2023), 25 pages. https://doi.org/10.1145/3579470

1 INTRODUCTION

Studies have shown that mentorship relationships help college students transform into elite performers, acquire specialized knowledge, improve GPAs and retention rates, and develop professional identities that allow for career success [11, 46]. Unfortunately, because it is hard to find mentors able to devote sufficient time, opportunities for student mentorship are limited, especially for those from underrepresented backgrounds or under-resourced institutions [2]. One way that colleges have sought to increase opportunities is through group mentorship [49], where a mentor supports 2-4 mentees at a time who each benefit from the expert mentor as well as from the

*The three authors contributed equally to this research.

Authors' addresses: Kehua Lei, klei4@ucsc.edu, UC Santa Cruz, USA; Mingrui Yu, UC Santa Cruz, USA, mingrayyu@gmail.com; Marissa Lewellen, UC Santa Cruz, USA, mlewelle@ucsc.edu; Venus Ku, UC Santa Cruz, USA, vwku@ucsc.edu; David T. Lee, UC Santa Cruz, USA, dlee105@ucsc.edu.



2573-0142/2023/4-ART37
https://doi.org/10.1145/3579470

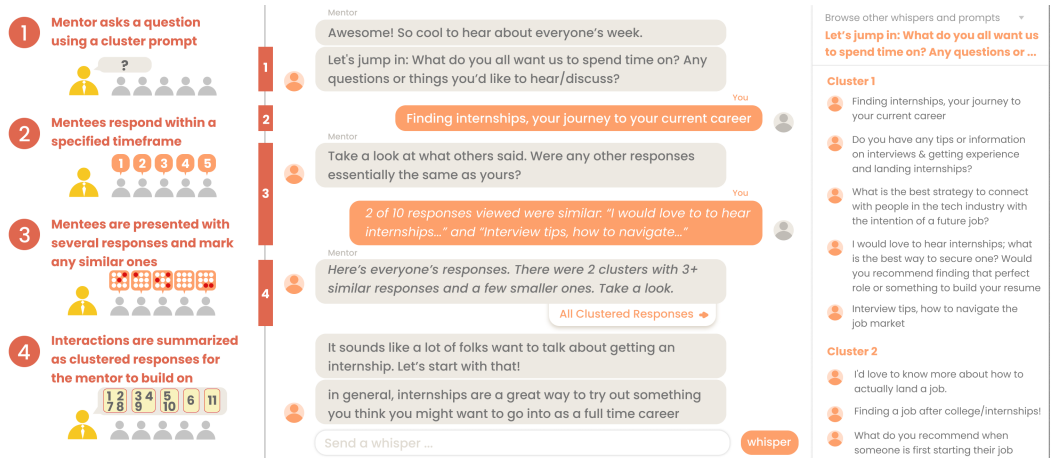


Fig. 1. Design illustrating how a mentor might find out what the entire group of mentees wants to discuss. Each mentee enters a response and specifies similar peer responses in a short conversational exchange that produces a synthesized view of the group intent. This multi-person conversational unit (one of several) allows everyone to participate while keeping a coherent flow of conversation.

added peer support. Unfortunately, despite the usefulness of group mentorship, this approach is not sufficient for addressing the limited opportunities for mentorship since groups are still small.

In this paper, we ask: *how might we design conversational interfaces that enable industry professionals to scale mentorship to large groups of 30+ mentees?* How might one do this while still preserving the level of support and connection found in small group mentorship, or while adding benefits uniquely possible in a large group chat-based setting?

The core design challenge we want to address is how one can enable large numbers of participants to be fully engaged in a chat-based mentorship session while still having a coherent conversation. Recent years have seen an explosion in the use of chat-based interfaces across diverse applications, with many describing chat as the next 'universal UI' due to its ease and familiarity [12, 27, 47, 65], and pointing to natural language as a time-tested 'interface' for receiving information and communicating intent. Yet, the linear nature of conversations–an attribute that helps facilitate listening and focus on a single topic at a time–also imposes significant limitations when groups are large, with either a large portion of participants unable to meaningfully express their thoughts, or a chaotic mess of disparate messages that prevent a coherent conversation. This challenge is perhaps the reason why much of the literature on conversational design has centered on supporting 1-on-1 interactions, leading to recent calls by researchers to more extensively explore 'beyond dyadic interactions' [58].

To support richer large group conversations, we conceptualize conversations as composed of not only individual messages, but also *multi-person conversational units* that collapse large numbers of small but related conversational exchanges into single conceptual units in the main dialogue. Doing so makes it possible to support non-linear, branched, or computationally-enhanced conversational exchanges (within each such unit) that can elicit participation from all, but in ways that can be concisely summarized and built on in the main conversation to preserve a coherent linear flow of conversation. We instantiate this within *Compass*, a platform we built for industry professionals to mentor large groups of mentees using two types of multi-person conversational units: *whispers* that enable mentees to initiate topics or questions without distracting from the main flow

of the conversation, and *prompt sequences* that enable mentors to: ask questions of the entire group (e.g. "What do people want to talk about today?"); facilitate responses, votes, clustering, or small group deliberation among the entire cohort of mentees; and have these conversational exchanges concisely summarized as group-wide intents that add naturally to the flow of the main dialogue (**Figure 1**).

This paper describes our exploration of multi-person conversational units through a research through design approach [24, 71]. We report on design patterns learned over a year of small real-world studies during which we evolved a starting set of conversational units and introduced new ones. This culminated in a final field study in which 2 industry professionals successfully mentored 30+ students in 3 mentorship sessions over a 10-week period. Both mentors and mentees found the chat-based UI effective and sometimes preferable to video-based mentorship. We describe the perceived advantages and disadvantages of chat-based mentorship, highlight design patterns key to a positive experience, and discuss broader implications for designing chat UI for large groups.

2 RELATED WORK

2.1 Career mentorship and design for mentorship

Many researchers in HCI have studied how one design technologies to support career development, e.g. for underserved job seekers [16] or crowd workers [53]. Within this literature, a growing body of work has considered how to support career mentorship, which recent research suggests consists of: psychological and emotional support, goal-setting and career path guidance, academic subject knowledge advancement, and role-modeling [10, 25, 40, 46, 49]. Much of this has traditionally been studied and implemented in the context of dyadic relationships between individual mentors and mentees [14]. However, a more limited set of studies have considered alternatives such as individual-team mentorship, in which a group of mentees are advised by a single mentor [52], and many-to-many mentorship (MTTM), in which two or more individuals serve as consistent mentors to a group of mentees [33, 54]. Studies show that group mentorship has better effectiveness than one-to-one format in some situations [17, 42], and that MTTM can result in increased collective engagement and growth in mentees, more focused mentorship sessions, and more mentor expertise as compared to other mentorship models [33, 54], motivating our focus on designing to support group mentorship.

There is also literature on career mentorship online. The earliest of these directions centers on *e-mentorship*, i.e. the use of email, chat, and video conferencing to support mentorship [4]. However, the focus in these studies is typically on the dynamics of mentorship online or on evaluation of new university mentorship programs [35, 60] rather than the development of new systems or system features. They also center on the traditional setting of one-on-one mentorship or small group mentorship. Others have studied the dynamics of *distributed mentorship* in communities of practice, e.g. through connecting crowd workers to provide peer coaching [9] or through mentorship within online forums [7, 21, 32, 39, 64]. However, since such dynamics are "ad hoc, i.e., they lack a previous history of interaction… they may fail to support the relational aspect of offline mentoring" [64]. *Our paper adds a new direction to existing literature through the design of novel chat-based interfaces for augmenting a mentor's ability to have an ongoing mentorship relationship with a large group of mentees.*

2.2 Designing UI for large group discussions in chat and other mediums

Several threads of research have considered how to better make sense of or better organize large group discussions that, absent intentional interventions, can be overwhelming due to the large

number of messages or interactions. These studies have occurred both in the context of group chat as well as in other mediums.

2.2.1 Orchestrating separate or distributed conversations. Outside of chat, this has included techniques for *orchestrating many separate but related conversations*, e.g. by using crowds and AI to extract shared structure from email replies for crafting personalized responses at scale [38], or using interactive visualizations to surface struggling students for scaling tutoring support [28]. Studies have also explored how to support *distributed conversations that surface salient opinions*, forgoing traditional dialogue for techniques that elicit, filter, and visualize opinions and preferences of a crowd. This could be in settings like online forums leveraging upvoting mechanisms to filter for top posts [19, 66] or tree-like structures to organize or synthesize ideas [62, 70]. It could also be in participatory democracy settings in which algorithms elicit and/or aggregate opinions and preferences using small micro-tasks (submitting a new idea, accepting or rejecting the opinions of others, or comparing between two ideas) [44] to identify and organize top ideas, polarizing ones, or opinions expressed by particular clusters of people [41, 55, 61].

While the settings described are useful in certain contexts, traditional conversations involving back-and-forth dialogue in a consistent group offer other benefits, such as allowing for deeper conversations grounded in and contributing to relational connection. The challenge in this context, however, is how to manage interactional incoherence [51] and minimize chat confusion [23] due to the potential for many people to be talking simultaneously and for multiple intertwined conversations, problems that appear even in small groups.

2.2.2 Structured annotations for summarizing conversations. There are at least three different approaches people have taken. First, researchers have explored using *structured annotations to summarize past texts*, supporting members in making sense of and catching up on chats involving multiple interleaved conversations. For example, in Zhang and Cranshaw [69], users tag/annotate messages which are then synthesized in structured summaries containing labels for discourse acts (e.g. Question, Answer, Update, Resource) [57]. Similarly, in crowd-powered conversational support systems, crowd workers update a running summary of important aspects of the conversation for future crowd workers [31, 43].

2.2.3 Backchanneling/multichanneling to separate side conversations. A second approach that has been explored is to use *backchanneling/multichanneling as a means for participants to fully engage while separating out side conversations* from the primary dialogue (which could be in chat or in another medium such as video). This allows for large-scale engagement without taking away from the main conversation. These side channels can take the form of standard text chats [15], chats overlayed onto a Zoom video or livestream [3, 48], threaded conversations [22], topical channels [1], or multiple chat lanes with different "velocities" to separate out messages that should be longer lasting [50].

2.2.4 Active moderation or facilitation. Finally, a third thread of work considers ways to *provide structure for actively moderating the conversation* so that engagement is more organized. This could look like a pre-defined script with timed scenes and prompts describing topics for conversation [20], a chatbot providing topic prompts and eliciting participation from those who have been silent [36, 37], interfaces that augment human moderators through structured scripts and AI-generated prompt suggestions [45], or widgets for voting and polling [68]. The work that most directly inspired ours was the idea of a real-time large-scale synchronous dialogue process (RLSDP) [5], where a moderator could pose open-ended questions that participants respond to and vote on to create a summarized view of the top-ranked responses.

2.2.5 Our approach and contribution. Like Bilich et al., we are interested in chat UI that can engage a large group in a synchronous conversation. However, we sought to go beyond a feeling of just filling out a series of polls to deeper back-and-forth dialogue fostering relational connection. In existing work, engaging with many participants simultaneously means living with and managing significant levels of incoherence (e.g. in livestreaming). Our paper resolves this through contributing: (1) a new conceptualization of conversations as built on multi-person conversational units, (2) a generalizable and extensible framework for using this conceptualization to engage large groups of people simultaneously in a coherent conversation, (3) an implementation of this idea in a system with several illustrative examples of multi-person conversational units deployed and evaluated "in the wild" that enabled two mentors to hold mentorship sessions with 30+ simultaneously engaged mentees, and (4) design lessons and implications for building on this framework that came out of a year-long design based research process for developing the platform.

3 COMPASS AND MULTI-PERSON CONVERSATIONAL UNITS

3.1 Conceptualizing conversations as built on multi-person conversational units

A multi-person conversational unit is a set of one or more (but typically many) conversational exchanges that can be grouped together as playing a single role within the larger conversation (see **Figure 2**). For example, in a conversation about trip-planning, a discussion on where to go could be considered a single unit in the larger conversation, with that unit playing the "role" of making a decision or narrowing down a set of options to be discussed further. In a typical in-person classroom discussion, a single unit might encompass an instructor's prompt, multiple student discussions in small groups, and a summary to the class from representatives within each group. In this case, the unit is playing the "role" of facilitating reflection on the prompt or eliciting the different ways that students are thinking about the prompt. In an extreme case, one could consider a series of side conversations tangential to the topic at hand as all represented as a single unit without any role (or a "null" role) relative to the larger conversation. **Figure 2a** shows how a mentorship session might be decomposed into sections that reveal the structure of the conversation. Some of these such as "Decide the topics of the session" and "Understand mentees' goals" correspond to a single multi-person conversational unit as they involve a series of conversational exchanges that all support a single role (determining the topics to address during the session and understanding the goals of each mentee, respectively). Other sections correspond to several multi-person conversational units.

If one were to make an analogy using the language of intent-based conversational design, a multi-person conversational unit might be thought of as a *group intent that has been addressed, at least relative to "what matters" in the larger conversation.* Of course, "what matters" is highly relative since "the group" does not have a single intent in reality and there could also be multiple ways of viewing a conversation as composed as multi-person conversational units depending on how "zoomed in or out" one wants to view the conversation. We will not concern ourselves with these in our first (admittedly imperfect) introduction of the concept. What is important to us is that one can begin to think about units of conversation that involve many people, but might be represented compactly within the larger conversation.

3.2 A framework for large group chat UI using multi-person conversational units

Our conceptualization allows one to define a general framework for large group chat UI. In our framework, every "type" of multi-person conversational unit would need to be designed separately, and would consist of: (1) algorithms and/or interface widgets that coordinate the entire set of conversational exchanges for the given unit (see **Figure 2b**), (2) a compact summarized view

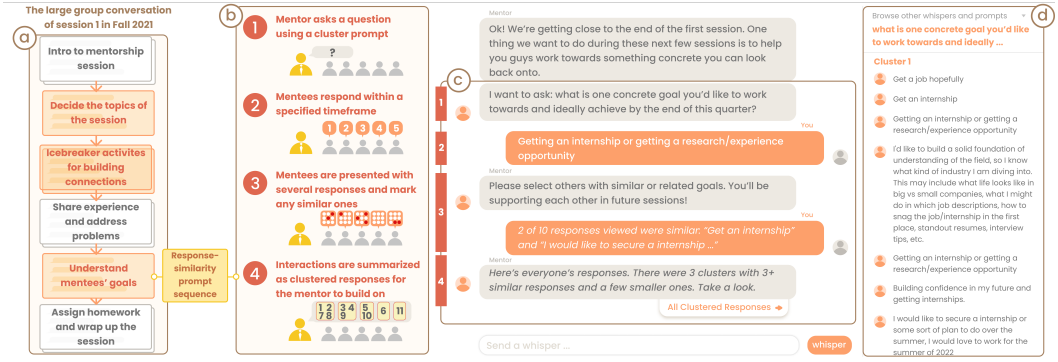


Fig. 2. An illustration of how a large group conversation might play out with multi-person conversational units. The first column on the left (a) shows how a mentorship session might be broken down in several multi-person conversational units, each playing a single role in the larger conversation. Orange boxes refer to those for which we designed custom UI to support. The remaining columns show how we support one of these units ("Understand mentees' goals") with: (b) algorithms and/or interface widgets that coordinate conversational exchanges, (c) a compact summarized view of the unit in the main chat dialogue, and (d) an expanded, but still summarized, view of the unit to be viewed and interacted with in a side panel.

of that unit within the main chat dialogue (see **Figure 2c**), and (3) an expanded, but still summarized, view of important information for that unit to be viewed and interacted with in a side panel (see **Figure 2d**). These three aspects build on many existing approaches to supporting large group conversations, and integrate them in a general framework. Techniques for moderation and orchestration of conversation (**Sections 2.2.1** and **2.2.4**) are at the core of defining algorithms and interface widgets to coordinate the conversational exchanges. Techniques for structured summarization (**Sections 2.2.2**) are relevant to the summaries in both the main chat and the side panel, but also benefit from the structured orchestration of the conversational exchanges to be summarized. Finally, techniques for backchanneling/multichanneling (**Section 2.2.3**) appear in side panel interactions as well as the way conversational exchanges are orchestrated for certain units.

3.3 How large group mentorship unfolds within Compass

Compass is a platform for large group mentorship developed using Angular, NgRx, Firestore, and Cloud Functions. Mentorship in Compass unfolds through a primarily mentor-facilitated conversation integrating smaller conversational exchanges and interactions using UI for multi-person conversational units. At the beginning of a new cohort, users can view and join the chat room for the upcoming cohort. Within the navbar, users can see a list of participant profiles. Hovering over participant profiles allows users to view a brief introduction written by the participant.

Conversations take place in an interface similar to most chat applications. The key difference is that some conversational exchanges are orchestrated using algorithms and input widgets, and only presented as a single summarized message within the main chat (see **Figure 1**). To view these in more detail, users click on the compact version to browse details in a side panel. The intention is for the main chat UI to present a coherent, mostly linear view of the conversation so that it is easy for users to have a mental model of the conversation. Less important exchanges are collapsed to provide awareness and social translucence [18], but without taking up inordinate space.

Our final version had five multi-person conversational units. Besides a trivial "whispering" unit, most of our conversational units were mentor-initiated "prompt sequences", predefined or algorithmically-determined sequences of prompts and responses. To initiate a prompt sequence,

mentors specify the initial prompt text, transition text for later steps (if applicable), and the amount of time for each step. While a prompt sequence is active, the input bar changes for each step of the sequence and indicators show the amount of time remaining and the number of responses received by other mentees. Prompt sequences are the primary feature that enable Compass to move the experience from a monotonous Q&A session to a more intimate conversation with engaged participants. By the end of our iterations, we had defined 4 different prompt sequences (**Figure 3**): the *response-vote*, *response-similarity*, *response-deliberation*, and *choose-deliberate* prompt sequences, named to reflect the stages in the sequence.

The specific set of 5 units were *not* chosen at the outset, but were added one-by-one as we conducted pilot studies and identified conversational needs within our mentorship context. These goals increased in ambition over time, starting from simply providing useful advice to providing relational connection, so as to test the framework's applicability to diverse intents and conversational needs.

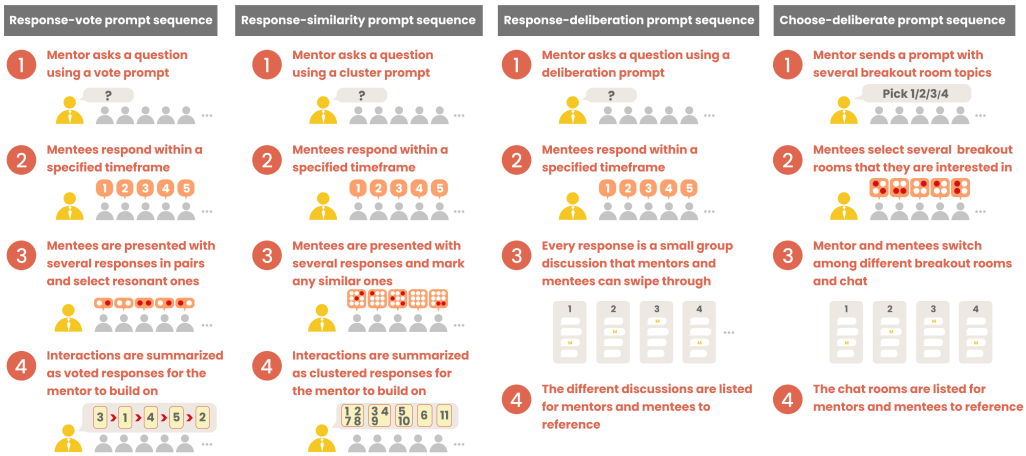


Fig. 3. The main steps for our four prompt sequences, each of which enable mentors to engage and coordinate the group towards particular goals. All participant interactions for a given prompt sequence are displayed compactly within the main chat UI.

3.4 The five multi-person conversational units implemented in Compass

In the descriptions that follow, we include figures of the input widgets for prompt sequences, but do not depict how messages appear in the main chat UI since these are similar to **Figure 1**. We note that the figures in this paper depict more compact versions of our final hi-fidelity designs. Our actual implementation was less polished due to prioritizing functional iterations over completing interface details that were not functionally critical.

3.4.1 Whispers to seamlessly integrate mentee-initiated thoughts. Mentors can change the 'state' of the text input bar so that mentee chats are normal messages, whispers, or determined by the mentee. When mentees whisper, their message is represented in the main chat as a small profile picture placed in a single line along with any other whispers sent since the last message. Participants view the whispers by clicking on the line of whispers to see them in a side panel. The side panel UI looks just like a chat, except that it only contains the whispers. Mentors can write messages

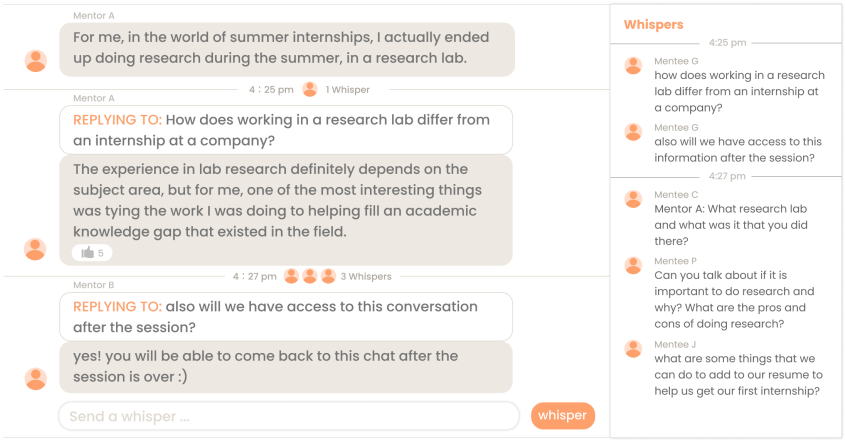


Fig. 4. An illustration of how the whisper feature allows mentees to interject with follow-up questions, thoughts, or other comments without distracting from the flow of conversation. Whispers are represented in the main chat as a single compact line and shown in detail within the side panel. If mentors reply to whispers, then they are shown in the main chat UI.

replying to whispers, which would display them within the main chat UI (**Figure 4**). Whispers are a trivial conversational unit that allow mentees to interject with follow-up questions, thoughts, clarifications, or even unrelated comments without distracting from the main conversational flow. Whispers are essentially an implementation of backchanneling within our framework, plus a compact indication of the backchanneling activity within the main chat.

3.4.2 Response-vote prompt sequence: eliciting ranked responses across the group. The response-vote prompt sequence elicited ranked responses from all mentees. In this prompt sequence, mentors ask a question for all participants to respond to within a set amount of time, e.g. 90 seconds (**Figure 5-1**). Once the time runs out, a follow-up prompt is sent asking mentees to make a series of votes comparing algorithmically chosen pairs of responses from other mentees. During this time period (e.g. 30 seconds), the input text bar is replaced with an interactive widget displaying each pair of responses (**Figure 5-2**). Once voting has completed, a compact algorithmically produced summary is added to the main chat. Clicking on this allows one to see the ranked responses in the side panel. This prompt sequences brings comparison-based preference elicitation methods used in orchestrating participatory democracy initiatives into the chat context to orchestrate

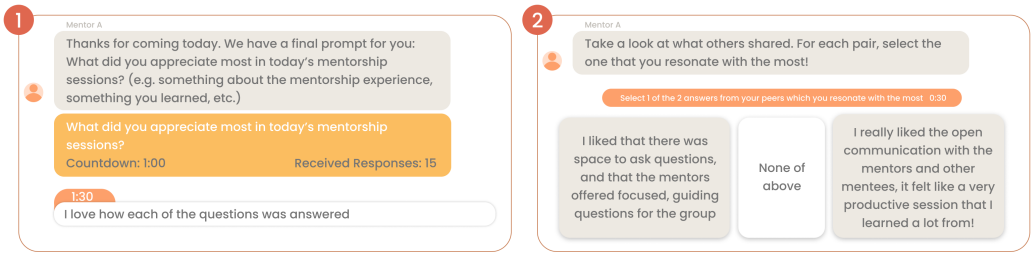


Fig. 5. The interface of the response-vote prompt sequence.

mini-interactions and produce structured summaries, making it possible to ask questions and hear thoughts from all mentees without overwhelming participants.

3.4.3 Response-similarity prompt sequence: eliciting response clusters across the group. The response-similarity prompt sequence that elicited clustered responses from all mentees was developed to enable mentors to more easily determine the set of unique topics among the cohort responses and to remove the feeling of having a low-ranked response. It also starts with a question from the mentor and responses from all participants (**Figure 6-1**). However, this time, the follow-up prompt displays an algorithmically chosen set of 10 responses from other participants and asks mentees to select whether any are similar to their own and could be addressed together (**Figure 6-2**). It then produces a grouping of the responses into clusters. The clustering algorithm for this was a novel one we developed and required several iterations due to differences in how people perceived similarity, but the details are out of scope for this paper.

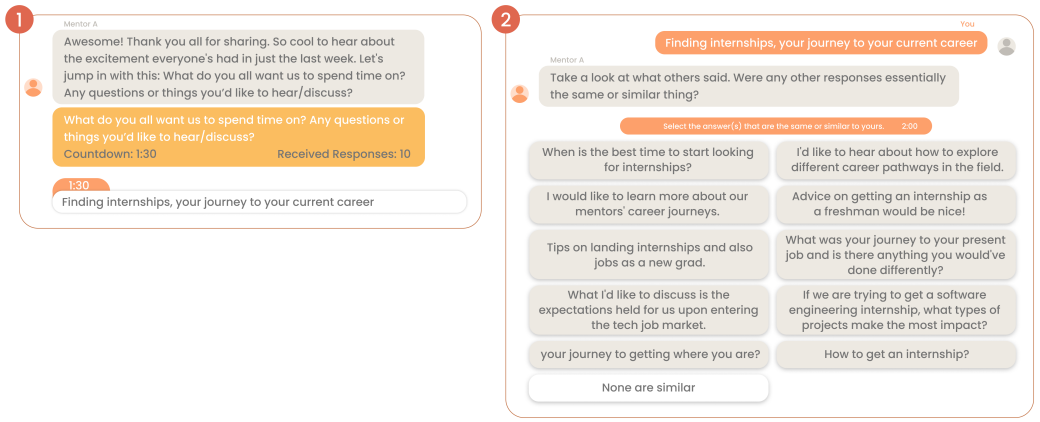


Fig. 6. The interface of the response-similarity prompt sequence.

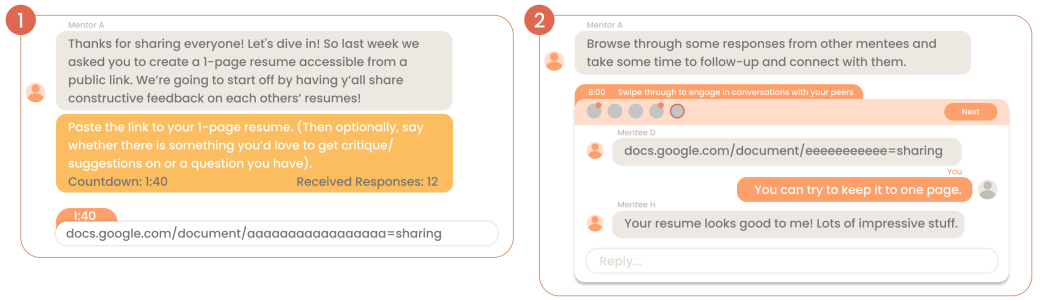


Fig. 7. The interface of the response-deliberation prompt sequence.

3.4.4 Response-deliberation prompt sequence: facilitating small group conversations. The response-deliberation prompt sequence was designed to facilitate conversations between mentees to foster relationships and peer support. It also starts with a question from the mentor, e.g. "How are you guys feeling when you think about careers/life after college?" or "Share a link to your 1-page

resume, and whether there is something you'd like feedback on" (**Figure 7-1**). But this time, after participants submit their responses, each mentee response results in a new mini chat room within a follow-up prompt asking mentees to browse through the chat rooms to engage with peers, e.g. through sharing their own stories, giving peer critique, etc. (**Figure 7-2**). The aggregated view in the side panel is simply the list of mini chat rooms that participants can drill into to view.

3.4.5 Choose-deliberate prompt sequence: facilitating targeted small group conversations. The choose-deliberate prompt sequence was also designed to facilitate conversations between mentees, but sought to encourage more depth through a smaller number of rooms centered on shared mentee interests or goals. It starts from a list of mentor-specified topics that mentees can choose from (**Figure 8-1**). Upon selecting one or more of these topics, mentees are placed into mini chat rooms corresponding to each of these topics and asked to engage with each other based on a mentor-specified follow-up prompt (**Figure 8-2**). Mentees can change topics (and thus, switch rooms) at any time.

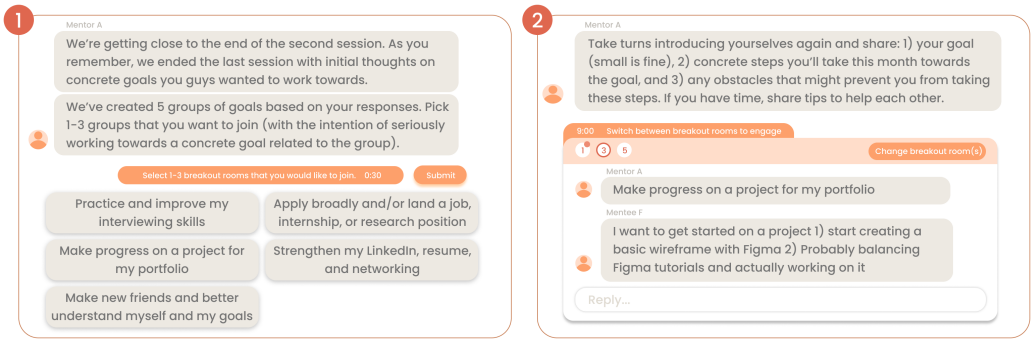


Fig. 8. The interface of the choose-deliberate prompt sequence.

3.5 Prototyped features for supporting mentors

While we focused on building UI for mentees, we used Google Docs to prototype features that mentors needed, which mostly centered on dealing with limitations in typing speed. This included supporting mentors in pre-writing scripts for introductions, prompts, and conclusions; in curating and reusing responses to past cohorts; and in using predefined peer connection prompts to provide value to mentees while mentors are reading messages and crafting responses. We also prototyped a modified word cloud generator to help mentors quickly identify the topic in response clusters produced from the response-similarity prompt sequence (details are out of scope for this paper).

4 DESIGN ITERATIONS AND LESSONS ACROSS A YEAR OF PROTOTYPING AND USER EVALUATION

We carried out design-based research from Fall 2020 to Fall 2021 in which we prototyped aspects of a large group conversational experience within Compass, evaluated it in increasingly real-world settings, and made iterations based on observations and participant surveys. As we observed unmet needs, we constructed dialogue excerpts depicting how those needs might be met in traditional in-person mentorship and then brainstormed how we might support the same benefits or feelings in a large group chat environment, either by finding ways to directly simulate the experience or by leveraging the unique possibilities of chat or large groups.

4.1 An overview of our three rounds of prototyping

Our iterations started with a series of 3 chat-based mentorship sessions on applying to graduate school in Fall 2020 and a series of 3 Zoom-based program kickoffs in which Compass was used in a supporting role to facilitate peer connections at the beginning of Spring, Summer, Fall 2021. These experience prototypes [6] were not meant to probe all aspects of the large group mentorship experience (or even any mentorship experience in the case of the program kickoffs), but to evaluate the experience of participating in or facilitating a large group conversation using our prototyped features, and in a real-world context where we could elicit authentic conversations. These were followed by two field deployments in Summer 2021 and in Fall 2021 in which Compass was used to run a "Tech Industry Mentorship Pilot Program" open to all students who were mentored by two industry professionals in 3 mentorship sessions.

All studies took place in synchronous contexts, so as to fully test the extent to which multi-person conversational units could handle the challenge of orchestrating a large volume of simultaneous interaction, and so that we could begin authentic evaluations without needing to implement a mobile app with push notifications. In what follows, we briefly describe the three prototyping rounds and motivations for major changes for each round. A large number of smaller improvements were also made throughout this process. We do not detail these since they are tangential to the core concept of multi-person conversational units, but a few are depicted in **Figure 9**.

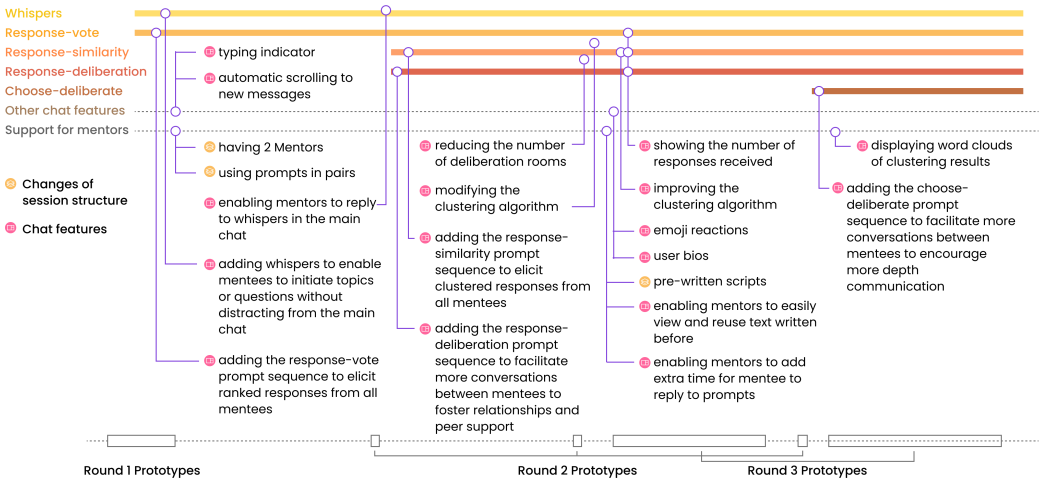


Fig. 9. Illustration of the main design iterations we made from Fall 2020 to Fall 2021.

4.1.1 Round 1 prototypes: mentorship sessions on applying to graduate school. The first set of studies were 3 mentorship sessions in Fall 2020 on applying to graduate school, all mentored by members of the research team, and each 45 minutes long and involving 6-10 students. This prototype only had two multi-person conversational units: *whispers* that provided ways for mentees to express thoughts at any time (see **Section 3.4.1**) and a *response-vote prompt sequence* that produced ranked responses from mentees, e.g. to questions like "What questions do you have or what topics would you like to discuss?" (see **Section 3.4.2**) These sessions validated that the concept could work, with participants feeling that their questions and needs were being addressed, and that it was helpful when others asked questions they hadn't thought of. However, we also observed that some people didn't like the voting when their responses didn't get many votes. People also did not

feel personally connected and felt that the sessions were very slow, causing them to be unengaged or even anxious. Mentors felt a lot of pressure to respond quickly due to typing speed limitations. They also wanted to get a sense of the responses as a whole rather than just top responses, and did not feel that they had time to really "listen" to and absorb mentee questions, or to share longer stories.

4.1.2 Round 2 prototypes: peer conversations during an online program launch. Motivated by what we learned, we introduced two new prompt sequences: a *response-similarity prompt sequence* (see **Section 3.4.3**) that clustered mentee responses to help mentors determine topics more easily and a *response-deliberation prompt sequence* (see **Section 3.4.4**) aimed at fostering peer connection. We also supported mentors by having multiple mentors and introducing the idea of asking questions in pairs. Specifically, one could ask an "elicitation" question (focused on learning what mentees want to talk about) immediately followed by a "peer connection" question (focused on helping mentees get to know each other) to give the mentor time to look through mentee questions and begin writing responses.

A major goal in our second round of prototypes was to go beyond just answering questions in a Q&A-like experience to fostering feelings of connection, and to also evaluate Compass with a larger number of participants by using it within Zoom-based program kickoffs of around 50 participants. For this set of prototypes, we invited students to use Compass while the informational talk was going on or during short periods of time dedicated to engaging with other peers.

The *response-similarity prompt sequence* did not work due to flaws in our algorithm, but the *response-deliberation prompt sequence* was successful (using the question, "Propose a question for connecting with others! Something you'd like to learn about others or a fun topic for people to discuss / take a stance on!"). Many participants felt it made the session more engaging than the past and some preferred it over Zoom breakout rooms. They liked chatting about many topics at once and meeting many people they don't usually talk to. They expressed feeling personally connected through finding similarities with others, understanding more about others, and engaging in silly conversations.

Most of the negative feedback related to usability issues and bugs. However, some felt intimidated by the number of topics or preferred breakout rooms for connecting on a more personal level. Since the mentor only used Compass to facilitate prompt sequences (and conveyed core content through the Zoom talk), we did not observe the same frustration with typing speed. Of course, this also meant that this prototype was limited to evaluating the deliberation prompt rather than the possibility of facilitating an engaging and coherent large group conversation over chat.

4.1.3 Round 3 prototypes: field deployments of a career mentorship pilot program. Since people expressed feeling connected, we decided that we were ready to move to running a real-world career mentorship program (the "Tech Industry Mentors Pilot Program") with 2 industry professionals jointly mentoring a large group of students across three one-hour long mentorship sessions spaced roughly 1 month apart. Our main changes involved addressing numerous usability improvements, bugs, and performance issues, as well as iterating on the clustering algorithm.

We ran this with two different cohorts, once in Summer 2021 with 22 mentees and then our final deployment in Fall 2021 with 36 different mentees. Mentees were recruited through campus mailing lists and social media. Participation in the mentorship program was open to all students regardless of whether they wanted to participate in filling out research surveys, with research participants entered into a lottery for a $25 Amazon gift card. While this meant that we were not able to assess the experience of the entire cohort, it meant that the responses of those who did participate reflected an authentic large group mentorship experience.

The 2 industry mentors were the same across both cohorts. Both were recent graduates (around 3 years since graduation) recruited through personal connections, with one working as a software engineer and the other working as a customer experience analyst, both at large well-known tech companies. To minimize challenges specific to learning the interface for the first time or figuring out how to best structure a chat-based mentorship session, we held one-hour mentorship prep sessions before each mentorship session. These prep sessions involved: 1) explaining/demonstrating new platform features that we added since the last session, and 2) discussing/reviewing the agenda and partial scripts for the mentorship session (which got more detailed over time, as will be discussed).

As you may recall, the biggest weaknesses of our early prototypes were: 1) mentees feeling unengaged and mentors feeling stressed due to slow sessions constrained by typing speed, and 2) mentees not feeling personally connected to mentors and each other. Overall, despite our iterations, these two weaknesses still dominated the summer experience.

With regards to engagement and typing speed, having multiple mentors helped, but it also meant they had to spend time discussing and coordinating. This was mostly addressed by Fall through: 1) finally getting the clustering algorithm to produce decent results, 2) showing the number of responses received so far so that people who finished submitting early could see that something was happening, 3) creating a question bank of predefined peer connection questions that a mentor could use as needed, 4) working with mentors to prepare a detailed session plan including pre-written text for parts they could predict (e.g. introductory text, prompts, and concluding text), and 5) enabling mentors to view and reuse text they had written for the summer cohort. Several participants mentioned being surprised at having their minds changed between the pre-study survey when they expressed preferring Zoom and the post-session one survey when they expressed preferring chat. Mentors responded similarly, describing how natural and comfortable they felt as compared to summer and how exciting it was to engage with the large group of mentees.

With regards to feelings of connection, the deliberation prompts that were used (for discussing career goals, their feelings about their career, and their progress in practice interviews) were not as successful as the "Propose a question for connecting..." prompt of our Round 2 prorotypes, as they required a lot more initiative to engage in. To improve feelings of connection, we: 1) added emoji reactions to support lightweight interactions, 2) made it so that every participant could write short intros displayed as tooltips when hovering over their profile pictures at the top of the chatroom or next to their messages, 3) used deliberation prompts where mentee responses involved clear calls for conversation, and 4) created a new *choose-deliberate prompt sequence* that made it possible to have conversations around common topics of interest. We trialed our new deliberation prompts in the 2nd and 3rd sessions of Fall, during which we observed significantly more engagement in the mini-chatrooms and at a deeper level of conversation. In the handful of survey responses we did receive, mentees expressed liking the more personalized conversations, feedback, and engagement from mentors. Unfortunately, we were not able to collect larger amounts of survey data to more strongly assess whether there was improved feelings of connection because we saw a significant dropoff across the sessions (from 36 to 14 to 9 participants) due to midterms and a technical difficulty we ran into during the second session which required us to shift the 2nd and 3rd session dates.

4.2 Lessons learned on designing chat-based interfaces for large group mentorship

These design iterations capture several lessons about designing for large group mentorship in chat-based UI. We briefly highlight a few of these, focusing in on insights related to designing chat using multi-person conversational units.

4.2.1 Tailor multi-person conversational units to specific intents and human factors. We found that multi-person conversational units can engage large groups of people in coherent conversations, but they need to be designed for specific dialogue contexts and tuned accordingly, including the design of the widgets, algorithms eliciting, and summary views. We saw this in several ways. First, while both voting and clustering were functionally able to elicit desired topics of discussion from mentees, voting wrongly assumed that mentors would feel comfortable with only addressing the top ranked responses, and also had the unintended effect of making people feel like their response was not valued by others. Second, the design of the clustering algorithm had to account for the different ways humans respond to requests to "select similar responses" (some chose to not participate, others only selected the single closest response, and still others had very encompassing interpretations of similarity). Third, the first deliberation prompt we designed was found to only be suitable with prompts that elicited clear calls to conversation. Future work will need to explore the space of dialogue intents that need to be supported in common large group mentorship contexts.

4.2.2 Leverage multi-person conversational units and the persistence of text to augment mentors' capacity to engage. While text has the disadvantage of being slower to type in real-time, its persistent nature makes it possible to create pre-written text or to enable re-use of text. We found that this was critical to the experience, and that multi-person conversational units were particularly suited to leveraging this. For example, the predefined bank of "peer connection" prompt sequences allowed mentors to go beyond just engaging mentees through pre-written text introductions, prompts, and concluding text to engaging mentees through *pre-defined orchestration of peer interaction* that significantly amplified of mentors' ability to engage. A second example was how multi-person conversational units made it possible for mentors to reuse text from past cohorts. In a small group mentorship context, it would not have been as easy to predict that text could be reused from prior cohorts. But in a large group context, mentors can be fairly confident that a response-similarity prompt sequence asking for topics of discussion in the first meeting will likely result in questions about landing an internship, interviewing techniques, etc., enabling mentors to share previous content much more quickly and to have more time for personalized responses or participation in mini-conversations.

We note that while limitations in typing speed were particularly salient in synchronous settings, the lessons learned about how to augment mentors' capacity to engage apply even to asynchronous settings. While the synchronous setting allowed us to stress test multi-person conversational units, we believe that the experience would be further improved in an asynchronous setting, e.g. where participants have hours or even days to respond to steps in a prompt sequence.

4.2.3 Orchestrate multiple simultaneous mini-conversations to enhance personal connection. While the response-vote and response-similarity prompt sequences were enough to achieve engagement, we saw that mini-conversations within deliberation prompt sequences were critical for creating a more personalized experience and for fostering feelings of connection. Although mentors were not able to participate in all rooms, they were able to find conversations where they could add value and could summarize select insights at the end of the prompt sequence. One can also imagine these features allowing for a large group of "guest mentors" to join in just for a single deliberation prompt.

We see many interesting directions for exploring deliberation prompts further. Our response-vote algorithm was able to leverage existing algorithms for eliciting and aggregating pairwise comparisons. Our response-similarity algorithm had to be designed ourselves, but we could still draw from existing ideas in networks and clustering. But preference elicitation does not typically deal with unstructured building blocks like mini-discussions (we are only aware of one paper on algorithms for orchestrating multiple rounds of discussions towards larger objectives [26]).

One starting point for defining what makes for a good mini-conversation "unit" is that they should be built on clear calls to conversation that don't require significant mentee initiative. For example, successful prompts in our studies included ones where: 1) mentees suggested questions for peers to answer, 2) mentees shared resumes for others to critique, or 3) mentees joined rooms corresponding to goals to share and discuss progress. In contrast, asking mentees to share how they were feeling or discuss career goals did not result in the deep conversations we had hoped for.

5 "IN-THE-WILD" EVALUATION OF THE LARGE GROUP CHAT-BASED MENTORSHIP EXPERIENCE

In this section, we give a deeper view into how mentees and mentors experienced large group chat-based mentorship in our final fall field deployment. Unlike **Section 4**, where our focus was on designs iterations and what we learned about designing interfaces for large group mentorship, here we focus on detailing the experience, with a view towards discussing the possibilities and limitations of chat-based UI for supporting large scale conversations in **Section 6**.

5.1 Survey and platform data

The results in this section are mostly based on data from a pre-study survey and a post-session study after each session. We also use platform analytics to describe the different levels of engagement in the mentorship session.

5.1.1 Pre-study survey. Our pre-study survey asked participants about their motivations for joining the mentorship program and their ideal experience. We also added questions asking them to rank three options for a career center mentorship program: 1) Zoom-based mentorship in small groups (e.g. 3 mentees), but only for 10% of students who apply; 2) Zoom-based mentorship in large groups (e.g. 30 mentees), available to all students who apply; and 3) Chat-based mentorship in large groups (e.g. 30 mentees), available to all students who apply. Of the 36 mentees in Fall, 17 filled out the pre-study survey and gave consent to its use for research.

5.1.2 Post-session surveys. Our post-session surveys asked both mentors and mentees what they liked, disliked, and felt throughout the mentorship session. We also asked them how strongly they agreed with feeling engaged, feeling like an active participant, feeling like their questions or needs were addressed, and feeling personally connected (on a 5-point Likert scale with options Strongly disagree, Somewhat disagree, Neutral, Somewhat agree, and Strongly agree), and we asked them to rank the same three options in the pre-study survey (Zoom in small groups, Zoom in large groups, chat in large groups). Of the 36, 14, and 9 mentees participating in the three fall sessions, only 14, 4, and 3 respectively filled out the post-session surveys and gave consent for research use.

5.1.3 Limitations. We had limited data for the 2nd and 3rd sessions because of the significant dropoff in participation coupled with the fact that mentees were not required to participate in filling out the surveys. As a result, our description of the experience is largely based on mentee's views of the first sessions. Certain aspects of the experience such as deliberation prompts only occurred in the later sessions, so our description of those aspects is more limited. This means that while our evaluations showed that we were able to successfully demonstrate our framework's ability to engage a large and consistent group of people simultaneously in a coherent back-and-forth conversation, we did not yet achieve our second more ambitious goal of relational connection within a large group. Nevertheless, even for this, our Round 2 prototypes and the few data points we did get in Fall indicate that relational connection in large groups is feasible. The insights we discovered in the process provide a starting point for future exploration (**Section 4.2.3**).

Another point of discussion in limitations is whether the drop-off issue would have been resolved with a different choice of setting, e.g. if we didn't choose to explore mentorship which comes with expectations of relational connection, or if we studied an asynchronous or hybrid setting where typing speed wouldn't be an issue. We think that these could very well have made a difference and that they are interesting future directions. We discuss possibilities of asynchronous chat in **Section 4.2.2** and we touched on hybrid experiences in our Round 2 prototypes described in **Section 4.1.2**. It would be interesting to fully explore multi-person conversational units in that context through video-based prompt sequences as well as text-based ones. With all this said, we think that the setting we chose was appropriate for this first paper on multi-person conversational units as it allowed us to stress test the idea.

We also note that the dropoff in participation in Fall was not due to the quality of the experience, but due to the busyness of the academic quarter and due to technical difficulties which caused us to need to change the 2nd and 3rd session dates. In a survey we conducted of those who did not continue (7 respondents), 71.4% said the reason was that they got busy or stressed with school and 28.6% said they couldn't make the time after it got moved.

5.2 Session dynamics: how time was spent during the mentorship sessions

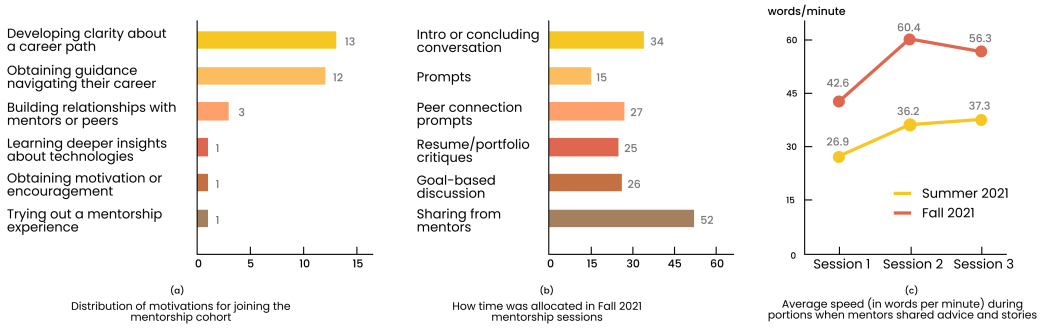


Fig. 10. On the left, the chart shows the ranking result of 6 motivations that we identified from the pre-study survey. In the middle, the diagram displays how time was spent during roughly 3 hours in the fall sessions. On the right, The line chart displays mentors' average speed in each session in the Summer and Fall prototyping studies.

The fall mentorship sessions were one-hour long each, with the roughly 3 hours of total time spent on (see **Figure 10b**): **34 minutes** of introductory or concluding conversation (including intros, group check-ins, and tasks for the next session), **15 minutes** of prompts (to learn what questions mentees had and what goals they had), **27 minutes** of peer connection prompts that occurred while mentors were writing responses to questions (voting, clustering, and deliberation prompts asking about people's countries of origin, best career advice received, memory from childhood, campus career resources they'd wish for, question they'd like to ask others), **25 minutes** of resume/portfolio critiques in mini-chatrooms (that mentors also participated in), **26 minutes** of goal-based discussion in groups (that mentors also participated in), and **52 minutes** of mentors sharing advice and stories in response to mentee questions or discussions.

The 52 minutes of mentors sharing advice and stories consisted of 2559 words with an average speed for sessions 1 to 3 of 42.6 wpm, 60.4 wpm, and 56.3 wpm respectively. Since they were able to reuse some prior text, this was significantly faster than summer when they took 80 minutes to share 2658 words worth of content with an average speed for sessions 1 to 3 of 26.9 wpm, 36.2

wpm, and 37.3 wpm respectively (see **Figure 10c**). This could have been even faster as the mentors purposely paused between sentences to prevent participants from being overwhelmed.

5.3 Mentee experiences improved their views of chat-based versus Zoom-based mentorship

In our Fall pre-study survey (with 17 respondents), most participants expressed skepticism about chat as a medium. When mentees ranked options for a career mentorship program, choosing between 1) Zoom in small groups limited to 10% of students, 2) Zoom in large groups with no participation limits, and 3) chat in large groups with no participation limits, the chat-based option was ranked very low by most participants. If large groups were assumed (options 2 and 3), only 11.8% preferred the chat-based option in summer compared to 35.7% in Fall (see **Figure 11**).

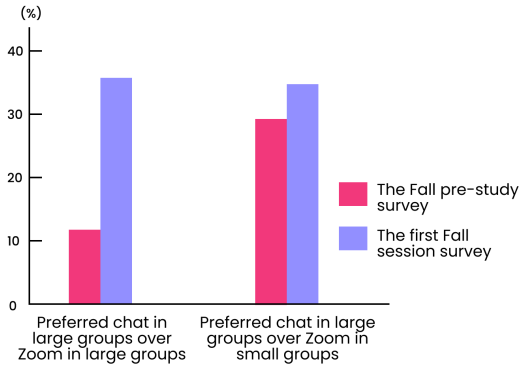


Fig. 11. The diagram shows 1) mentees' preferences between chat and zoom in large groups, and 2) mentees' preferences between chat in large groups and zoom in small groups before and after the first session in Fall.

Mentees who changed their responses to rank chat as their top choice said that *"it was really smooth to use compared to Zoom which often has lag" (P12)* and that *"most of the questions I had were addressed better through chat than any zoom mentorship has" (P11)*. One explained:

> *"You've changed my opinion on this! I found the different features and filtering of responses to be really helpful, and in Zoom people often don't have their cameras on anyway, so even though I like the idea of being able to express oneself with body language and facial expressions, etc., this chat experience has a lot of advantages that outweigh missing out on the visual info." (P14)*

We note that session one did not yet utilize the deliberation prompts which personalized the experience even more. If participation had not dropped so much due to technical issues and the busyness of the academic quarter, we believe we would have seen further increases as we saw this for the 3 survey respondents to both session one and two.

5.3.1 What people who didn't prefer chat said. Even among those who ranked chat-based mentorship as their lowest choice, many expressed that they *"still [felt] like the chat session was very helpful (P16)"*:

> *I still believe zoom is a better medium then chat, however I think chat has a lot of potential... this was way more useful than I thought the chat medium could possibly be. (P4)*

That being said, it was also easy to identify common themes in why people preferred Zoom-based small groups that are unlikely to change in the foreseeable future. Among those whose preferences for chat did not increase and who gave a reason for it, every single one described some variant of Zoom-based mentorship or small groups being *"more engaging"*, *"more personal"*, *"more personalized"*, or *"more intimate"*. Based on our experience, we think that it's possible to significantly improve how engaging and personalized the chat experience is, but we do not see chat creating a personal and intimate experience close to that of small groups. Thus, while large group chat-based mentorship may be valuable in providing guidance for career decisions and clarity about career paths (especially in enabling greater access and peer learning), small groups will likely continue to be stronger for developing deeper, more intimate relationships.

5.4 Most mentees felt that their questions and needs were addressed

The strongest value provided in the session was the practical value of getting answers to questions. Of 14 respondents to the post-session one survey, 71.4% expressed somewhat agreeing with or strongly agreeing with the statement "*I felt like my questions or needs were addressed*" (**Figure 12**). Mentees felt it was wonderful to "*get answers to my questions*" (P21) and to "*hear their experiences and advice*" (P11). They felt the advice/information was "*really helpful and informative!*" (P12) and "*important/relevant to me*" (P15). They liked it when others "*ask[ed] questions that I was going to ask*" (P23) and felt "*relieved that other students have similar concerns to mine*" (P14). But they also enjoyed it when "*the questions didn't apply to my current situation [as] it was still interesting to see what the mentors thought*" (P16). They appreciated "*how [mentors] changed their topics with what people wanted to hear from them*" (P9).

Of course, mentors still weren't able to address all the questions, which some mentees disliked, saying "*a lot of questions went largely unanswered*" (P17), and "*they weren't able to respond to [my questions] simply because of the limited time*" (P12). For others, this didn't bother them, saying "*they said they would go deeper into it next time*" (P15).

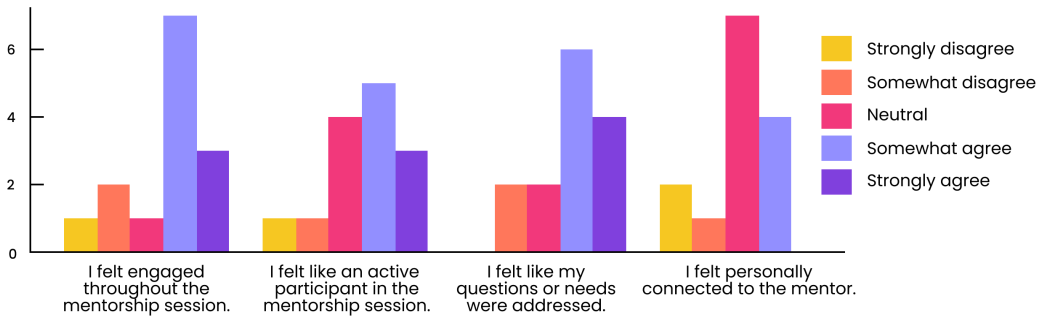


Fig. 12. This diagram shows mentees' responses on whether they felt engaged, felt like active participants, felt their questions were addressed, and felt connected after the 1st Fall 2021 session.

5.4.1 Mentees liked being able to go back to reference the chat history. An additional value that chat-based interfaces provided in this area was the ability to *"go back and reference the chat history" (P16)* and *"read the responses as we saw fit"* (P18). One mentee described using this affordance during the session: "*when there was downtime..., I could browse back to previous things people had typed that I may have not read yet*" (P14).

Our platform analytics also supported this, showing that some participants attending the session but came back to the platform later to review the text (9, 1, and 1 mentee(s) in sessions 1, 2, and 3

respectively). Other participants chose to skip the sessions altogether (or perhaps couldn't make that time), and chose to just read the text after the session had concluded (4, 5, and 1 mentee(s) in sessions 1, 2, and 3 respectively). These participants are not included in the previous counts, but were still able to gain some value from the sessions.

5.5 Most mentees felt engaged, but some still felt that the sessions were slow

Of 14 respondents to the post-session one survey, 71.4% expressed somewhat agreeing with or strongly agreeing with the statement "*I felt engaged throughout the mentorship session*" (**Figure 12**). The strong feeling of engagement people experienced was partly due to the value they derived from getting their questions and needs addressed, with one mentee saying, "*I engaged because I was learning important pieces of advice!*" (P17) Our platform analytics also reflected engaged participation (as opposed to lurking [63]). In the first session, of the 36 mentees, 24 sent normal messages, 30 sent whispers, 29 responded to at least one prompt, 23 voted on at least one prompt, and 25 used emojis.

5.5.1 Chat helped shy mentees feel more comfortable engaging. One strength of the impersonal nature of chat is that it was more comfortable for shy mentees. One said, "*I liked how easy it was to ask questions. Since it was anonymous, it felt a lot less pressure.*" (P12) Another said "*It was actually a lot more comforting and less intimidating compared to Zoom*" (P7).

5.5.2 Multi-person conversational units supported engagement. Though we did not specifically ask about the new features, some mentees expressed that the interface supported feelings of engagement. They liked "*how clusters were formed and [how] students could chat in the side*" (P17) and "*having side conversations through whispering [and] grouping ideas together through clusters*" (P16). Mentees said:

> "*I liked how dynamic it was despite it being all through text. Loved the whispering feature, the main chat, the reaction emojis, and just the overall interface... I felt excited, curious, validated, and optimistic.*" (P11)
>
> "*I found the different features and filtering of responses to be really helpful... the clustering of responses helped enhance that feeling of similarity to other students*" (P14)

5.5.3 Typing still made sessions feel slow. However, despite people feeling engaged overall, waiting on mentors to type made the session feel "*somewhat slow and laggy*" (P4) and made them feel "*a little bit bored at times*" (P7). They explained that "*it was really hard to pay attention*" (P22) "*while waiting for messages to pop up*" (P9) This explains another reason why some mentees preferred Zoom for which they believed "*response times would be faster*" (P16).

The feelings of slowness may be the reason why mentees expressed lower feelings of being active participants as compared to feeling engaged, with 57.1% of mentees expressing somewhat agreeing with or strongly agreeing with the statement "*I felt like an active participant in the mentorship session*" after the first session (**Figure 12**). Note however, that only 14.3% of mentees expressed somewhat disagreeing with or strongly disagreeing with feeling like an active participant, so this was still overall positive.

It is an interesting that some mentees still felt this way after we had doubled the speed compared to summer. It is unclear whether this feeling would have gone away with further increased speed. If so, it would not be hard to do because the mentors purposely added delays when reusing/copy-pasting text to keep mentees from feeling overwhelmed. Another possibility is that video allows for intermediate levels of communication (through fillers like "hmmmm" and "let me think" or with body language) that keeps mentees engaged even as mentors are processing. In contrast, in our chat UI, mentors were either silent or sharing organized insights.

5.6 Most mentees did not feel personally connected, but deliberation prompts likely improved this

In contrast to their feelings of engagement, active participation, and having their needs and questions addressed, mentees overall did not feel personally connected. Of 14 respondents after session 1, only 28.5% expressed somewhat agreeing with or strongly agreeing with "*I felt personally connected to the mentor*" with an almost equal number of 21.4% expressing somewhat disagreeing with or strongly disagreeing with that statement (**Figure 12**). Mentees expressed that they "*did not feel very close to my peers or the mentors*" (P15) and that "*it was also a lot less personal*" (P22). They felt that small groups "*tend to be more intimate with sharing their personal experiences*" (P23): "*the fewer the participants there are, the closer we are able to get to the mentors*" (P18).

Mentees also expressed that seeing participants "*usually makes it feel more real*" (P18) and that "*there is something more personal about speaking in person with mentors*" (P15). Being in-person comes with other benefits like getting a "*sense of the room*" and allowing one to "*have small side conversations before or after the session in private*" (P14). Some mentees felt like the information conveyed through chat could be "*replaced with a forum channel, where mentors post answers to questions*" (P21) or "*done on Zoom as well*" (P16) through features like chat or polling.

5.6.1 Deliberation prompts helped support feelings of connection and motivated goal-setting. We only had 3 survey respondents attend both sessions two and three when deliberation prompts were used. While we cannot make definitive conclusions, all three expressed improved connection with either the mentors or peers, saying that engaging with mentors and peers in the mini-chatrooms "*felt more connected*" (P15), "*more personalized*" (P12), and "*very relevant and collaborative*" (P14).

We also saw indications that the use of deliberation prompts for critiques and goal setting helped expand the value mentees obtained beyond just getting their questions answered. All three survey respondents liked the resume critiques, saying that "*you're receiving direct feedback [with] actionable steps that you can take*" (P12), that "*it felt very relevant and collaborative*" (P14), and that the "*smaller breakout groups [enabled] more personalized conversations... I was taking notes :)*" (P15). One mentee described its impact on them:

> "*it encouraged me to get a resume together and actually get started on a project... something that I probably wouldn't have had the motivation to do without the career mentorship program* (P12)".

5.7 The platform supported mentors in having a positive mentorship experience

Mentors felt that they successfully shared valuable experiences and provided personal and individualized advice to support mentees in both summer and fall sessions. With the new prompt sequences and the improved interfaces, the experience of fall sessions was much better than summer sessions, which exceeded their expectations. In the summer sessions, both mentors felt stressed due to the limitations of typing speed. They felt "*pressure to respond to as many conversations as possible*" (M2), and "*it just takes a lot longer to type out thoughts compared to speaking*" (M2). After our new prototyped features for mentors in Fall, they expressed feeling more confident "*being able to communicate effectively on the platform*" (M1) and described it as helping them "*focus more on the topics*" (M2).

Mentors preferred our chat-based platform over zoom for large group mentorship. They said that the multi-person conversational units allowed them to "*quickly ascertain what a large group of people want to talk about*" (M1) and "*provide more value in a more impactful way than just picking random topics*" (M2). The deliberation and breakout rooms also helped them to build a deeper connection with mentees, especially in mini-chatrooms for resume/portfolio critiques, where they

felt the discussions were "*more substantial*" (M1) and allowed them to "*give individualized feedback*" (M1).

6 DISCUSSION AND FUTURE DIRECTIONS IN CONVERSATIONAL DESIGN FOR LARGE GROUPS

We have already discussed lessons on designing chat-based interfaces to support large group mentorship in Section 4.2. We found that multi-person conversational units all need to be designed for specific dialogue intents and that their corresponding widgets, algorithms, and summary views need to be tailored to those unique contexts and human factors (see **Section 4.2.1**). We discussed how well-designed conversational units and the persistence of text can augment mentors' capacity to engage participants (see **Section 4.2.2**). Finally, we reflected on the use of simultaneous mini-conversations for improving connection among participants (see **Section 4.2.3**). In this section, we discuss our reflections on the design of conversational interfaces for large groups beyond mentorship.

6.1 Exploring large group conversational settings beyond mentorship

We see many opportunities for exploring large group conversations beyond mentorship. For example, how might chat-based interfaces be used for deliberative democracies [29] to mediate conversations on issues relevant to a community or society, and potentially in areas that are polarizing? How might they support more conversational dynamics in large-scale ideation [8, 59]? How might they support a crowd-based form of focus groups that can obtain both the depth of qualitative research and the rigor of quantitative research, along the lines of Wiki Surveys [56]? Exploring any such setting will require designing a larger ecosystem of multi-person conversational units attuned to diverse conversational intents (see **Section 4.2.1**).

Additionally, our approach to large group conversations thus far relies heavily on having a small number of facilitators who have control over the flow of conversation. While this seems reasonable, at least as a starting point, for the above examples (societal deliberation, ideation, and focus groups), there may be scenarios in which it is worth exploring dynamics that do not have this assumption. One fruitful direction may be to build on studies of coordination and control in collaborative tabletop interfaces [67] to explore dynamics of coordination and control for how facilitator roles can be claimed, relinquished, or determined collectively by a group.

6.2 Algorithms and AI for large group mentorship

We also see interesting directions for algorithms and AI. New multi-person conversational units may motivate new algorithms or even new forms of elicitation such as ones built on mini-conversations (see **Section 4.2.3**). Another really interesting direction is to build on the growing interest in exploring beyond dyadic chatbot interactions [58] to consider the role of bots in large group conversations. One thing we observed is that large numbers of participants make it easier to predict group intents due to the law of large numbers (see **Section 4.2.2**). We see many ways that chatbots can allow mentors to focus their energy on personalizing content for mentees. For example, chatbots could help further clarify participant intents in responses for better clustering, facilitate the peer connection questions, or help augment the mentor if the same topic comes up in many mini-conversations.

6.3 Beyond being there: chat as a medium for large group conversations

The overarching argument of this paper is that chat holds interesting possibilities for large group conversation that are currently unexplored. In the early days of CSCW, researchers sought to

design video-based computer-mediated communication to get as close to the in-person experience [13], but then realized that one would always fall short and should instead leverage the unique affordances of technology to go "*beyond being there*" [30]. We argue that a similar shift in mindset applied to chat-based interfaces could help expand the design space for supporting large group conversations.

As we experienced in our design iterations (see **Section 4**), chat will always be slower and less intimate when compared with video or in-person communication. It does not allow for seeing other's faces or other forms of non-verbal communication can help people stay engaged in a conversation and facilitate the exchange of affective information [34]. However, the root bottleneck in large group conversations (for both chat and video) is that it's impossible for mentors to communicate with single mentees for as long or in as personal ways as in small groups. This is where looking at the unique advantages of chat are interesting.

First, chat supports non-linear communication. Multiple people can type at the same time without interfering with each other. People can browse past messages, which means it is possible to participate in multiple side interactions at once without taking one away from a primary conversation. Second, chat is easier to process computationally. This makes it possible to design algorithms and/or interfaces that can elicit and summarize a large number of intents from a group. Finally, chat is flexible. Since people can review past dialogue, they can engage fully in the conversation, connect ideas more easily, participate while multi-tasking in doing other work, or engage asynchronously over a larger span of time. Our work shows one path for using chat to support large group conversations. There may be other ways to leverage the unique affordances of chat to enable large group conversations.

7 CONCLUSION

In this paper, we presented our exploration of multi-person conversational units, which collapse large numbers of small but related conversational exchanges into single conceptual units for supporting large group online communication. In doing so, multi-person conversational units make it possible to fully engage all participants through non-linear, branched, or computationally-enhanced conversations while still maintaining a coherent linear flow of conversation across the entire group. We instantiated this concept within a chat platform, Compass, and evolved it through real-world studies across the span of a year, culminating in our final mentorship pilot program in which 2 industry professionals successfully mentored 30+ students in 3 mentorship sessions over the span of 10 weeks. We described the design lessons we learned through our design process, the large group mentorship experience, and strengths and weaknesses of chat-based interfaces for facilitating large group conversations. Our hope is that this will open up future work in the design of multi-person conversational units for applications other than mentorship, and ultimately, make it possible for large groups or crowds of people to engage in rich and meaningful conversations.

ACKNOWLEDGMENTS

We would like to thank Amy Liu and Nathaniel Thompson for being mentors in the real-world career mentorship program. We thank the anonymous CSCW reviewers for their generous feedback, which helped us significantly improve this paper. Thank you to Maryam Siddiqui, Rebecca Noordeen, Dina Chen, Daniel Felipe Serrano Mora, Aproop Kamat, Sai Kambampati, Gia Alexis Garado, Ryan Anderson, Calum Chen, Jennifer Dutra, Yuqing Yan, Kristin Jai, and Jeremy Tan for contributing to the exploration of early ideas and designs, development of Compass, and the teaser video. Thank you to the mentees for their participation in the mentorship sessions. Finally, we thank Veronica Rivera and Dustin Palea for their feedback on early drafts.

REFERENCES

[1] Nazanin Andalibi, Frank Bentley, and Katie Quehl. 2017. Multi-channel topic-based mobile messaging in romantic relationships. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–18.

[2] Jeffrey E Barnett. 2008. Mentoring, boundaries, and multiple relationships: Opportunities and challenges. *Mentoring & Tutoring: Partnership in Learning* 16, 1 (2008), 3–16.

[3] Tom Bartindale, Peter Chen, Harrison Marshall, Stanislav Pozdniakov, and Dan Richardson. 2021. ZoomSense: A Scalable Infrastructure for Augmenting Zoom. In *Proceedings of the 29th ACM Int'l Conf. on Multimedia*. 3771–3774.

[4] Laura L Bierema and Sharan B Merriam. 2002. E-mentoring: Using computer mediated communication to enhance the mentoring process. *Innov. High. Educ.* 26, 3 (2002), 211–227.

[5] Jordan Bilich, Michael Varga, Daanish Masood, and Andrew Konya. 2019. Faster Peace via Inclusivity: An Efficient Paradigm to Understand Populations in Conflict Zones. *AI for Social Good Workshop at NeurIPS* (2019).

[6] Marion Buchenau and Jane Fulton Suri. 2000. Experience prototyping. In *Proceedings of the 3rd Conf. on Designing interactive systems: processes, practices, methods, and techniques* (New York City, NY, USA) *(DIS '00)*. Association for Computing Machinery, New York, NY, USA, 424–433.

[7] Julie Campbell, Cecilia Aragon, Katie Davis, Sarah Evans, Abigail Evans, and David Randall. 2016. Thousands of Positive Reviews: Distributed Mentoring in Online Fan Communities. In *Proceedings of the 19th ACM Conf. on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) *(CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 691–704.

[8] Joel Chan, Steven Dang, and Steven P Dow. 2016. Comparing Different Sensemaking Approaches for Large-Scale Ideation. In *Proceedings of the 2016 CHI Conf. on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2717–2728.

[9] Chun-Wei Chiang, Anna Kasunic, and Saiph Savage. 2018. Crowd Coach: Peer Coaching for Crowd Workers' Skill Growth. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (Nov. 2018), 1–17.

[10] Norman H Cohen. 1995. The Principles of Adult Mentoring Scale. *New directions for adult and continuing education* 66 (1995), 15–32.

[11] Janet W Colvin and Marinda Ashman. 2010. Roles, risks, and benefits of peer mentoring relationships in higher education. *Mentoring & Tutoring: Partnership in Learning* 18, 2 (2010), 121–134.

[12] Alistair Croll. 2016. On chat as interface - Alistair Croll - Medium. https://acroll.medium.com/on-chat-as-interface-92a68d2bf854. Accessed: 2021-10-15.

[13] Richard L Daft and Robert H Lengel. 1986. Organizational Information Requirements, Media Richness and Structural Design. *Manage. Sci.* 32, 5 (May 1986), 554–571.

[14] Kathryn H Dansky. 1996. The Effect of Group Mentoring on Career Outcomes. *Group & Organization Management* 21, 1 (March 1996), 5–21.

[15] Alan R Dennis, Julie A Rennecker, and Sean Hansen. 2010. Invisible whispering: Restructuring collaborative decision making with instant messaging. *Decision Sciences* 41, 4 (2010), 845–886.

[16] Tawanna R Dillahunt, Jason Lam, Alex Lu, and Earnest Wheeler. 2018. Designing Future Employment Applications for Underserved Job Seekers: A Speed Dating Study. In *Proceedings of the 2018 Designing Interactive Systems Conf.* (Hong Kong, China) *(DIS '18)*. Association for Computing Machinery, New York, NY, USA, 33–44.

[17] David L DuBois, Nelson Portillo, Jean E Rhodes, Naida Silverthorn, and Jeffrey C Valentine. 2011. How effective are mentoring programs for youth? A systematic assessment of the evidence. *Psychological Science in the Public Interest* 12, 2 (2011), 57–91.

[18] Thomas Erickson and Wendy A Kellogg. 2000. Social translucence: an approach to designing systems that support social processes. *ACM Trans. Comput.-Hum. Interact.* 7, 1 (March 2000), 59–83.

[19] Baldo Faieta, Bernardo Huberman, and Paul Verhaeghe. 2006. Scalable online discussions as listening technology. In *Proceedings of the 39th Annual Hawaii Int'l Conf. on System Sciences (HICSS'06)*, Vol. 1. IEEE, 15c–15c.

[20] Shelly Farnham, Harry R Chesley, Debbie E McGhee, Reena Kawal, and Jennifer Landau. 2000. Structured online interactions: improving the decision-making of small discussion groups. In *Proceedings of the 2000 ACM Conf. on Computer supported cooperative work*. 299–308.

[21] Umer Farooq, Patricia Schank, Alexandra Harris, Judith Fusco, and Mark Schlager. 2009. Sustaining a Community Computing Infrastructure for Online Teacher Professional Development: A Case Study of Designing Tapped In. In *Learning in Communities: Interdisciplinary Perspectives on Human Centered Information Technology*, John M Carroll (Ed.). Springer London, London, 111–138.

[22] David Fono and Ron Baecker. 2006. Structuring and supporting persistent chat conversations. In *Proceedings of the 2006 20th anniversary Conf. on Computer supported cooperative work*. 455–458.

[23] Hugo Fuks, Mariano Pimentel, and Carlos Jose Pereira de Lucena. 2006. RU-Typing-2-Me? Evolving a chat tool to increase understanding in learning activities. *Int'l Journal of Computer-Supported Collaborative Learning* 1, 1 (2006), 117–142.

[24] William Gaver. 2012. What should we expect from research through design?. In *Proceedings of the SIGCHI Conf. on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 937–946.
[25] Jean E Girves, Yolanda Zepeda, and Judith K Gwathmey. 2005. Mentoring in a post-affirmative action world. *Journal of Social Issues* 61, 3 (2005), 449–479.
[26] Ashish Goel and David T Lee. 2016. Towards Large-Scale Deliberative Decision-Making: Small Groups and the Importance of Triads. In *Proceedings of the 2016 ACM Conf. on Economics and Computation* (Maastricht, The Netherlands) *(EC '16)*. ACM, New York, NY, USA, 287–303.
[27] Dan Grover. 2021. Chinese Mobile App UI Trends. http://dangrover.com/blog/2014/12/01/chinese-mobile-app-ui-trends.html. Accessed: 2021-10-15.
[28] Philip J Guo. 2015. Codeopticon: Real-time, one-to-many human tutoring for computer programming. In *Proceedings of the 28th Annual ACM Symp on User Interface Software & Technology*. 599–608.
[29] Amy Gutmann and Dennis F Thompson. 2009. *Why Deliberative Democracy?* Princeton University Press.
[30] Jim Hollan and Scott Stornetta. 1992. Beyond being there. In *Proceedings of the SIGCHI Conf. on Human Factors in Computing Systems* (Monterey, California, USA) *(CHI '92)*. Association for Computing Machinery, New York, NY, USA, 119–125.
[31] Ting-Hao Huang, Joseph Chee Chang, and Jeffrey P Bigham. 2018. Evorus: A crowd-powered conversational assistant built to automate itself over time. In *Proceedings of the 2018 CHI Conf. on human factors in computing systems*. 1–13.
[32] Julie S Hui, Matthew W Easterday, and Elizabeth M Gerber. 2019. Distributed Apprenticeship in Online Communities. *Human–Computer Interaction* 34, 4 (July 2019), 328–378.
[33] Russell L Huizing. 2012. Mentoring Together: A Literature Review of Group Mentoring. *Mentoring & Tutoring: Partnership in Learning* 20, 1 (Feb. 2012), 27–55.
[34] Wijnand IJsselsteijn, Joy van Baren, and Froukje van Lanen. 2003. Staying in touch: Social presence and connectedness through synchronous and asynchronous communication media. *Human-Computer Interaction: Theory and Practice (Part II)* 2, 924 (2003), e928.
[35] Christina Algiere Kasprisin, Peg Boyle Single, Richard M Single, and Carol B Muller. 2003. Building a Better Bridge: Testing e-training to improve e-mentoring programmes in higher education. *Mentoring & Tutoring: Partnership in Learning* 11, 1 (April 2003), 67–78.
[36] Soomin Kim, Jinsu Eun, Changhoon Oh, Bongwon Suh, and Joonhwan Lee. 2020. Bot in the bunch: Facilitating group chat discussion by improving efficiency and participation with a chatbot. In *Proceedings of the 2020 CHI Conf. on Human Factors in Computing Systems*. 1–13.
[37] Soomin Kim, Jinsu Eun, Joseph Seering, and Joonhwan Lee. 2021. Moderator chatbot for deliberative discussion: Effects of discussion structure and discussant facilitation. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–26.
[38] Nicolas Kokkalis, Chengdiao Fan, Johannes Roith, Michael S Bernstein, and Scott Klemmer. 2017. Myriadhub: Efficiently scaling personalized email conversations with valet crowdsourcing. In *Proceedings of the 2017 CHI Conf. on Human Factors in Computing Systems*. 73–84.
[39] Yubo Kou and Colin M Gray. 2018. "What do you recommend a complete beginner like me to practice?": Professional Self-Disclosure in an Online Community. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (Nov. 2018), 1–24.
[40] Kathy E Kram. 1988. *Mentoring at work: Developmental relationships in organizational life*. University Press of America.
[41] Travis Kriplean, Jonathan Morgan, Deen Freelon, Alan Borning, and Lance Bennett. 2012. Supporting reflective public thought with considerit. In *Proceedings of the ACM 2012 Conf. on Computer Supported Cooperative Work*. 265–274.
[42] Gabriel P Kuperminc and Jessica D Thomason. 2013. Group mentoring. *Handbook of youth mentoring* 2 (2013).
[43] Walter S Lasecki, Rachel Wesley, Jeffrey Nichols, Anand Kulkarni, James F Allen, and Jeffrey P Bigham. 2013. Chorus: a crowd-powered conversational assistant. In *Proceedings of the 26th annual ACM Symp on User interface software and technology*. 151–162.
[44] David Timothy Lee, Ashish Goel, Tanja Aitamurto, and Helene Landemore. 2014. Crowdsourcing for participatory democracies: Efficient elicitation of social choice functions. In *Second AAAI Conf. on Human Computation and Crowdsourcing*. aaai.org.
[45] Sung-Chul Lee, Jaeyoon Song, Eun-Young Ko, Seongho Park, Jihee Kim, and Juho Kim. 2020. Solutionchat: Real-time moderator support for chat-based structured discussion. In *Proceedings of the 2020 CHI Conf. on Human Factors in Computing Systems*. 1–12.
[46] Daniel Jacob Levinson. 1978. *The seasons of a man's life*. Random House Digital, Inc.
[47] Jonathan Libov. 2015. Futures of text. https://whoo.ps/2015/02/23/futures-of-text. Accessed: 2021-10-15.
[48] Matthew K Miller, John C Tang, Gina Venolia, Gerard Wilkinson, and Kori Inkpen. 2017. Conversational chat circles: Being all here without having to hear it all. In *Proceedings of the 2017 CHI Conf. on Human Factors in Computing Systems*. 2394–2404.

[49] Amaury Nora and Gloria Crisp. 2007. Mentoring students: Conceptualizing and validating the multi-dimensions of a support system. *Journal of College Student Retention: Research, Theory & Practice* 9, 3 (2007), 337–356.
[50] Kanayo Ogura, Yoko Matsumoto, Yoshiyuki Yamauchi, and Kazushi Nishimoto. 2010. Kairos Chat: a novel text-based chat system that has multiple streams of time. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*. 3721–3726.
[51] Jacki O'Neill and David Martin. 2003. Text chat in action. In *Proceedings of the 2003 Int'l ACM SIGGROUP Conf. on Supporting group work*. 40–49.
[52] Kate Philip and Leo B Hendry. 2000. Making sense of mentoring or mentoring making sense? Reflections on the mentoring process by adult mentors with young people. *Journal of Community & Applied Social Psychology* 10, 3 (2000), 211–223.
[53] Veronica A Rivera and David T Lee. 2021. I Want to, but First I Need to: Understanding Crowdworkers' Career Goals, Challenges, and Tensions. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1 (April 2021), 1–22.
[54] Marjatta Saarnivaara and Anneli Sarja. 2007. From university to working life: mentoring as a pedagogical challenge. *Journal of Workplace Learning* 19, 1 (Jan. 2007), 5–16.
[55] Matthew J Salganik and Karen EC Levy. 2015. Wiki surveys: Open and quantifiable social data collection. *PloS one* 10, 5 (2015), e0123483.
[56] Matthew J Salganik and Karen E C Levy. 2015. Wiki surveys: open and quantifiable social data collection. *PLoS One* 10, 5 (May 2015), e0123483.
[57] John R Searle and John Rogers Searle. 1969. *Speech acts: An essay in the philosophy of language*. Vol. 626. Cambridge university press.
[58] Joseph Seering, Michal Luria, Geoff Kaufman, and Jessica Hammer. 2019. Beyond Dyadic Interactions: Considering Chatbots as Community Members. In *Proceedings of the 2019 CHI Conf. on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19, Paper 450)*. Association for Computing Machinery, New York, NY, USA, 1–13.
[59] Pao Siangliulue, Joel Chan, Steven P Dow, and Krzysztof Z Gajos. 2016. IdeaHound: Improving Large-scale Collaborative Ideation with Crowd-Powered Real-time Semantic Modeling. In *Proceedings of the 29th Annual Symp on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 609–624.
[60] P B Single and C B Muller. 2001. When Email and Mentoring Unite: The Implementation of a Nationwide Electronic Mentoring Program–MentorNet, the National Electronic Industrial Mentoring Network for Women in Engineering and Science. https://eric.ed.gov/?id=ED472832. Accessed: 2022-1-11.
[61] Christopher Small, Michael Bjorkegren, Timo Erkkilä, Lynette Shaw, and Colin Megill. 2021. Polis: Scaling Deliberation by Mapping High Dimensional Opinion Spaces. *Recerca: Revista de Pensament i Anàlisi* 26, 2 (2021).
[62] Marc Smith, Jonathan J Cadiz, and Byron Burkhalter. 2000. Conversation trees and threaded chats. In *Proceedings of the 2000 ACM Conf. on Computer supported cooperative work*. 97–105.
[63] Na Sun, Patrick Pei-Luen Rau, and Liang Ma. 2014. Understanding lurkers in online communities: A literature review. *Comput. Human Behav.* 38 (Sept. 2014), 110–117.
[64] Maria Tomprou, Laura Dabbish, Robert E Kraut, and Fannie Liu. 2019. Career Mentoring in Online Communities: Seeking and Receiving Advice from an Online Community. In *Proceedings of the 2019 CHI Conf. on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–12.
[65] Kyle VanHemert. 2015. The Future of UI Design? Old-School Text Messages. *Wired* (June 2015).
[66] Bogdan Vasilescu, Alexander Serebrenik, Prem Devanbu, and Vladimir Filkov. 2014. How social Q&A sites are changing knowledge sharing in open source software communities. In *Proceedings of the 17th ACM Conf. on Computer supported cooperative work & social computing*. 342–354.
[67] J O Wobbrock, M R Morris, and A D Wilson. 2009. User-defined gestures for surface computing. *GROUP ACM SIGCHI Int. Conf. Support. Group Work* (2009).
[68] Min Wu, Arin Bhowmick, and Joseph Goldberg. 2012. Adding structured data in unstructured web chat conversation. In *Proceedings of the 25th annual ACM Symp on User interface software and technology*. 75–82.
[69] Amy X Zhang and Justin Cranshaw. 2018. Making sense of group chat through collaborative tagging and summarization. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–27.
[70] Amy X Zhang, Lea Verou, and David Karger. 2017. Wikum: Bridging discussion forums and wikis using recursive summarization. In *Proceedings of the 2017 ACM Conf. on Computer Supported Cooperative Work and Social Computing*. 2082–2096.
[71] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI Conf. on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. Association for Computing Machinery, New York, NY, USA, 493–502.

Received January 2022; revised July 2022; accepted November 2022