MixTAPE: Mixed-initiative Team Action Plan Creation Through Semi-structured Notes, Automatic Task Generation, and Task Classification

SAJJADUR RAHMAN*, Megagon Labs, USA
PAO SIANGLIULUE, B12, USA
ADAM MARCUS, B12, USA

Checklists and action plans are a proven mechanism for project-based collaboration. Synthesizing project-specific plans is challenging, as project managers must consider multiple sources of information, from structured surveys to semi-structured conversations with stakeholders. In a needfinding study with project managers, we identified challenges in creating action plans for teams. We built *MixTAPE*, a mixed-initiative system that addressed these challenges with three components: a semi-structured note-taking interface for capturing stakeholder conversations, a plan generator for automatically combining multi-source information into action plans, and classification models for assigning and prioritizing action items. We evaluated *MixTAPE* in an observational study of 32 website design projects. Compared to a previously unstructured process, *MixTAPE* generated 1.45*X* as many tasks that are more consistent, while reducing the plan creation time by 33.70%. Through interviews and surveys, we found that participants rate *MixTAPE* highly across several measures. Based on our findings, we discuss the implications and opportunities for mixed-initiative action plan creation.

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

Additional Key Words and Phrases: Action Plan Synthesis; Collaboration; Mixed-initiative

ACM Reference Format:
Sajjadur Rahman, Pao Siangliulue, and Adam Marcus. 2020. *MixTAPE*: Mixed-initiative Team Action Plan Creation Through Semi-structured Notes, Automatic Task Generation, and Task Classification. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2, Article 169 (October 2020), 26 pages. https://doi.org/10.1145/3415240

1 INTRODUCTION

Synthesizing project plans for teams is challenging. Project managers have to consider multiple sources of information and requirements, from structured survey input to semi-structured interviews and conversations with stakeholders. Given the multiple sources of information, project managers have to create plans with clear and succinct tasks and descriptions, and consider details on each task like ownership and priority. Sources of information are at times in conflict with one-another, and at times present redundancies that should not make their way into the final plan. The quality of the synthesized plans have far-reaching impact on factors like project completion [3]

*This work began when the author was part of the University of Illinois (UIUC).

Authors' addresses: Sajjadur Rahman, sajjadur@megagon.ai, Megagon Labs, USA; Pao Siangliulue, pao@b12.io, B12, USA; Adam Marcus, marcua@marcua.net, B12, USA.

2573-0142/2020/10-ART169 $15.00
https://doi.org/10.1145/3415240

and team collaboration [4]. Addressing the challenges in creating action plans is thus crucial to having high-quality outcomes.

Checklists are effective mechanisms for communicating plans and guidelines. From the cockpit [16] to the operating room [29], they are a well-researched way to ensure that a basic set of steps is taken. More recently, they have even been shown to provide guard rails on projects in creative fields such as design [15]. While communicating general-purpose best practices is powerful, it's equally important to capture situation-specific details. What outcome does this patient want to optimize for? How does that client want their design to stand out? Action plans are advanced checklists that capture such details and provide a step-by-step breakdown of how a team should complete a given project [66, 67]. Each step of the action plan is an action item (task) and may contain situation-specific details such as requirements derived from a conversation. While checklists and action plans have proven effective at enforcing adherence to a general-purpose set of best-practices, what's less clear is how to construct a set of project-specific steps.

In this paper, we study how to create project-specific action plans for projects that involve multiple stakeholders. Specifically, we study how teams of project managers and designers create and operationalize action plans for client-specific web design projects. In a needfinding study, we study teams of project managers and web designers who build websites for clients and largely coordinate tasks through semi-structured notes and comments in shared Google documents and Slack groups. We identify challenges in the following areas: information capture through note-taking, lack of structure in the action plans, conflicting and redundant information in multiple sources of information, and error-prone work in assigning ownership and priority. These patterns of plan synthesis, assignment, and prioritization for multiple stakeholders can be seen in a wide variety of fields. In a medical setting, teams of experts combine intake forms, interviews, evaluations, and tests in supporting a patient. In an educational setting, several teachers and staff collaborate on creating education plans for students with diverse needs, combining multiple forms of evaluation, feedback, and documentation.

Having uncovered challenges in our needfinding study, we present *MixTAPE*[1], a **mix**ed-initiative **t**eam-based **a**ction **p**lanning **e**nvironment that facilitates the creation of action plans for a team of collaborators. *MixTAPE* features a semi-structured note-taking environment for capturing notes during synchronous conversations that can later be converted to tasks for various team members. In generating tasks from these notes, *MixTAPE* also automatically generates and integrates tasks it derives from structured sources of information such as preference forms a client submits. In combining the multiple sources of information, *MixTAPE* utilizes two classifiers for automatically identifying an owner for and assigning a priority to each generated task. Finally, *MixTAPE* presents an interface for correcting and iterating on the automatically generated and classified tasks before introducing those tasks to the collaborators.

We evaluate *MixTAPE* in an observational study of 32 real-world new website design projects at B12. The study follows two project managers who combine notes taken in *MixTAPE* during client conversations, creating followup tasks for project collaborators including themselves, clients, and eight designers who design and build the websites for the clients. Through quantitative measures, we determine that *MixTAPE* improves several measures of task quality including comprehensiveness and consistency, while reducing the overall time to create the project plan for the team. Through interviews and surveys, we identify opportunities for further improvement, and discuss broader implications for the design of systems that facilitate the creation of action plans from multiple structured and semi-structured sources of information. In summary, our contributions include:

[1]Similar to a mixtape of music tracks, *MixTAPE* combines tasks (music) from multiple sources (artists) into an authoritative record for collaboration (enjoyment).

- A needfinding study that identifies several pain points in creating action plans for a team of collaborators, including: a) difficulty in capturing critical details during live conversations, b) challenges in synthesizing a detailed plan across multiple sources of both structured and unstructured information, c) inconsistent presentations of tasks for each team member, and d) labor-intensive and error-prone steps like assigning owners and prioritizing tasks for each owner.
- A mixed-initiative system called *MixTAPE* that addresses the pain points identified in the study through several mechanisms: a) a semi-structured and autocompleting note-taking environment that helps project managers capture details from client conversations, b) an automatic task synthesis mechanism which combines semi-structured notes and other sources of structured information to create an action plan, c) a structured representation in which action items are grouped into several predefined categories, and d) an automatic process for assigning owners and priorities to each task.
- An empirical, mixed-methods evaluation of *MixTAPE* as project managers and designers use the system to design and build websites for 32 B12 clients. We quantitatively determine that, compared to a previously unstructured process, *MixTAPE* generates 1.45X as many tasks that are more consistent and descriptive, while providing better task context to the stakeholders and reducing the time to create a project plan by 33.70%. Through interviews and surveys, we find that participants rate *MixTAPE* highly overall. Moreover, we identify the broader implications and future opportunities for project plan generation in a mixed-initiative setting. Finally, we explain which of *MixTAPE*'s concepts can be generalized to a broad set of tasks and team types.

2 MOTIVATING EXAMPLE: WEBSITE DESIGN

To better situate our work, we provide a motivating example based on the coauthors' experience at B12[2]. B12's project managers coordinate with teams composed of various experts (e.g., designers, copywriters) that collaborate on the creation of client websites. When a client first starts working with B12, there are several sources of information about that clients' preferences. First, a salesperson maintains a set of notes on their early conversations to scope out a project. Second, the client may interact with B12's software product, filling out several aesthetic preferences and content audit forms, which result in largely structured sources of information. Finally, a project manager conducts a kickoff call with the client over the phone, taking semi-structured notes on the client's preferences, existing content, and desired functionality.

Prior to the introduction of *MixTAPE*, designers would be presented with various sources of structured and semi-structured information upon joining a new project. The project manager's notes would be presented in a Google Document called *Team Messages*, and were considered the ultimate source of truth. The other sources of information would be programmatically compiled into a *client brief* that presented all sales notes and client preferences. Designers would ask questions in the Team Messages document and on Slack, and project managers would clarify in the document, appending new details through follow-on conversations with the client. As the design team sent iterations to the client, notes and subsequent feedback would result in more document-based clarifications and comment threads. The Team Messages document was ostensibly an action plan, with project managers adding work to be done, and designers seeking clarification and identifying completed work in the document. Google Documents was selected over traditional tools like Asana [7] and Trello [61] for recording action plans because it afforded quick note-taking capabilities for

[2]https://www.b12.io/

information capture while providing various formatting options (e.g., bolding, coloring, nesting of bullets, strikethroughs) to subsequently add structure.

Naturally, with multiple sources documenting client preferences at different times, the sources would contain conflicting information. Team Messages not only identified what to do, but also contained specifications on which other sources to ignore. With an implicit action plan without the affordances of an explicit checklist, different project managers would provide different structures and techniques to prioritize key features of their Team Messages document. Similarly, designers would utilize different approaches to clarifying uncertainty or denoting the completion of a task. In this multi-source and not fully standardized environment, team members would at times miss key details, apply different priorities to their work, or follow the wrong direction when encountering conflicting information. In Section 4, we summarize the results of a needfinding study that identifies the key challenges project managers faced while working in this setup. To reduce mistakes and frustration, B12 researchers searched for a system to more explicitly generate a unified action plan from the varied sources of project information.

3 RELATED WORK

In this section, we review related work on note-taking and task management, discuss how such work inspired the design of *MixTAPE*, and identify the differences in modes of operation between related work and *MixTAPE*.

3.1 Live note-taking assistance

A common technique for documenting details is the practice of note-taking during conversations. While many tools have been developed to aid in live note-taking [20, 24, 26, 35], we are particularly interested in tools that situate conversations in the proper context, discussed next.

Note-taking and conversation-aware intelligence. Action plans often contain situation-specific details derived from synchronous conversations, *e.g.*, team brainstorming sessions and client meetings. Capturing situation-specific details synchronously while maintaining a conversation's context is challenging due to the cognitive effort and split attention required to take notes [53, 54, 65]. To make note-taking easier in such a setting, Tilda [68] enables people to collaboratively enrich their chat conversation by annotating, grouping, and linking messages in a variety of ways, such as tagging via emojis or inline notes. Tilda uses the markup left by participants to structure the chat stream into a skimmable summary. MeetBot [47] uses even simpler annotations—chat participants can add action items to a running list of notes using hashtag commands. Inspired by these tools, *MixTAPE* introduces a markdown-like language to capture different components of a conversation, *i.e.*, tasks, comments, within a semi-structured note-taking pane. However, the goal of the *MixTAPE* note-taking pane is to create a structured to-do list for multiple stakeholders of a project whereas these tools create comprehensible summaries of group chats that contain a variety of items, like questions, decisions, and action items. Moreover, unlike these tools, the sole contributor to notes in *MixTAPE* is a project manager capturing notes during real-time voice conversations.

Context-aware suggestions during note-taking. Automatic suggestion of text can greatly improve user efficiency in text entry and is extensively used in typing-based interfaces like search bars in search engines [9, 57], messaging apps [18, 33, 34], and email [17, 19, 36]. Smart email and messaging applications are conversationally aware, providing suggestions tailored to conversation dynamics. Similar to these applications, *MixTAPE*'s note-taking pane automatically suggests context-aware text as users take notes. However, unlike the aforementioned smart applications, live note-taking is often constrained by time and requires users to quickly take notes as well as

actively participate in conversations. In this paper, we identify challenges in utilizing auto-complete features in such a multi-modal setting.

3.2 Task management

The research community has explored the problem of task creation and management from several points of views—from personal and collaborative task management systems to intelligent agents for assisting in task creation, tracking, and completion. We now discuss different systems developed from these perspectives while contextualizing the *MixTAPE* setting and goals.

Personal task management. Bellotti et al. [11] conducted one of the most relevant ethnographic studies of task management. The study presented a number of implications for the design of task management tools based on the task management practice of their participants. To address these implications, a lightweight task management tool, TaskVista, was developed. The tool contained intuitive visualizations for collecting and managing tasks. Similar to TaskVista, tasks in *MixTAPE*—for a given project—appear as a to-do list and project managers can change task priority by moving tasks up and down the list. Remember The Milk (RTM) [10] is a web-based personal task manager that couples basic task list functionality available as a web application with additional services including mobile applications, location services and various integrations with other applications. RTM allows users to set task priorities, move tasks into different lists, and delete tasks. Moreover, users can update the task status as completed or postponed, a feature also available in *MixTAPE*. While the design of our task list interface draws inspiration from the aforementioned personal task management tools, the application domain of *MixTAPE* is different from these tools. In *MixTAPE*, tasks are tied to a project rather than personal to-dos, multiple stakeholders have ownership of specific tasks in a project, and stakeholder actions are at times dependent on the completion of tasks by collaborators. Moreover, tasks in *MixTAPE* are synthesized by combining multiple sources of information in a mixed-initiative manner, whereas task creation in the aforementioned systems is manual and doesn't necessarily deal with multiple sources.

Collaborative task management. Managing tasks involving multiple stakeholders requires dealing with interdependent tasks [13]. In addition, people often need to operate over multiple projects and responsibilities in parallel [46] each with its own collaborative demands. Email was once extensively used to manage collaborative tasks [12, 63]. However, delays between messages, the need for back and forth exchanges, and lack of native support to track pending replies makes managing tasks in this manner both ineffective and inefficient. The communication challenges are addressed by many modern collaboration platforms such as Slack [58] and Teams [49]. However, the goal of these systems is to enable communication among project collaborators, not facilitate creation, management, and tracking of project tasks or action plans. The built-in communication mechanism in *MixTAPE* shares similarities with comment threads where stakeholders leave comments under their assigned tasks for other collaborators. This interaction can be made more instantaneous by integrating real time communication capabilities similar to Slack or Teams.

The problem of interdependent tasks is also addressed by many commercial tools. For example, Tracks [60] allows an administrator user to manage multiple users similar to project managers in *MixTAPE*. Other tools such as Asana [7], Trello [61], Forecast [25], and ClickUp [22] focus on the problem of task tracking for project management purposes. Users of these systems are introduced to a personalized view of their to-dos, similar to the role-specific views in *MixTAPE*. Administrators (project managers) can assign tasks and add priorities manually, whereas in *MixTAPE*, ownership and prioritization of a task are assigned automatically and then curated by project managers. Commercial systems don't have *MixTAPE*'s notion of multiple sources of information for action

plan generation, and thus require manual conversion from notes or structured data sources into task lists.

Task Manager [40] takes a broader approach toward collaborative task management and supports task coordination, tracking, resource sharing, and communication via messages. However, the system did not undergo any formal user evaluation, making it difficult to determine the extent to which a tool like this can support task coordination in practice. Feedback from participants in our user study indicates that embedded communication channels in task management systems are desirable and can act as a one-stop-shop for collaborative task management by removing the added overhead of using an external communication platform.

Intelligent task management. Intelligent tools and capabilities for supporting task management have garnered significant interest in recent years. Tools like Forecast [25] and ClickUp [22] can automatically estimate task deadlines and assign tasks to team members. *MixTAPE*, on the other hand, adopts a mixed-initiative approach where automatically assigned tasks are further reviewed by project managers to correct potential assignment errors. Towel [23] is a task management application that couples a user's to-do list with a *project execution assistant* (PExA) [50] that can complete a learned set of task capabilities on behalf of the user. Users can perform modifications on to-dos such as grouping, tagging, checking (completing) in Towel while delegating tasks to PExA. However, task creation in Towel is still manual, unlike in *MixTAPE*, which automatically creates tasks from captured notes and briefs. Moreover, only simple tasks can be delegated to PExA like setting reminders or monitoring deadlines. One interesting area of future work is to combine the assistance *MixTAPE* offers in creating multi-participant action plans with the lightweight assistants offered by PExA or other commercial offerings including Siri from Apple [6], Cortana from Microsoft [48], Alexa from Amazon [5], or Google Assistant [28].

3.3 Task creation for plan synthesis

We now discuss different approaches for synthesizing action plans.

Synthesizing action plans for projects. Prior work on action plan synthesis mostly focused on outsourcing plan creation to crowd workers [37, 39] or friends [1] for a variety of tasks like creative writing and health behavior planning.

The TaskGenies [39] system provides crowd-powered action plans for common tasks. To provide an action plan, the system first searches for a similar task in it's task database using NLP techniques. If a similar task with an action plan is found, this plan is returned to the user. If there is no similar task in the database, the system crowdsources the creation of a new plan. As crowdworkers enter each step of an action plan in a web-based interface, TaskGenies uses NLP to suggest steps that other workers created for similar tasks. Similarly, as project managers take notes in the note-taking pane, *MixTAPE* suggests similar notes that were entered for other projects. The *MixTAPE* autocomplete suggestion model is built from an expert-sourced task description corpus. Kaur et al. [37] also leverages expert-sourcing to create action plans for creative writing tasks. First a list of 18 writing tasks, called a vocabulary, were compiled by expert-sourcing. These tasks range from being about the mechanics and organization to the semantics of a document. Given a document and the vocabulary, crowdworkers are asked to add comments to the document related to the mechanics, organization, and semantics of the document. These comments are then compiled into action plans. Similar to the vocabulary for writing tasks, *MixTAPE* divides tasks in an action plan into predefined high level themes, *e.g.*, structure, content, design. While action plans for both TaskGenies and creative writing tasks are synthesized by crowdworkers in an offline fashion, in *MixTAPE*, project managers engage in live conversation with clients and the choice of action items are dictated by client requirements. The action plans are then synthesized automatically from the collected notes. Unlike existing

systems, tasks within a *MixTAPE* action plan are automatically assigned to specific stakeholders and have varying degrees of priority.

We discussed earlier how action plan synthesis in existing project management software [7, 22, 25, 61] is manual and cumbersome. These tools do not facilitate the initial rapid capture of unstructured notes in low-latency situations such as phone calls. Their rigid task structure makes the initial note capture challenging. Moreover, to synthesize tasks, project managers have to enter an explicit level of detail on areas such as task ownership or priority. *MixTAPE*, on the other hand, features an initial note-taking mode followed by automatic task creation from notes by intelligent agents and task curation by project managers.

User control vs. automation. The tension between giving users direct control and automating actions on their behalf is long-discussed in HCI research [56]. Direct manipulation systems [7, 61] grant users more control in creating action plans, but require additional time for tasks like assigning roles and priorities. Given the action plans, Forecast [25] and ClickUp [22] can automatically estimate task deadlines and assign tasks based on the workload and role of the team members. Automatic suggestion of text is extensively used in typing-based interfaces like search bars in search engines [9, 57], messaging apps [18, 33, 34], and email [17, 36]. Mixed-initiative systems use such techniques to offer the best of both worlds with principles on when an automated system should proactively take action, and when a user should [31]. *MixTAPE* lives in this middle ground between direct manipulation and automation. This approach streamlines the process of action plan synthesis without sacrificing control.

4 NEEDFINDING STUDY

To understand how project managers formulated plans for projects using the existing structured client brief and call notes, we conducted a needfinding study in which we shadowed two project managers at B12. We observed how each project manager created plans for two new client websites. Specifically, the study consisted of three phases: 1) shadowing project managers during kickoff calls as they took notes and after the calls as they synthesized action plans, 2) conducting a reflective session for each call, and 3) a semi-structured interview exploring various pain-points. A detailed description of the process and questions is available as supplementary documentation.

During the shadowing sessions, we documented how project managers conducted the kickoff calls and synthesized the information they collected from multiple sources to create action plans. Each shadowing session lasted about an hour, and we recorded the screen interactions of the project managers as well as conversations with clients. We explicitly asked for permission from clients to record the sessions.

Based on the information collected from the shadowing session, we conducted a reflective interview with each project manager. We asked questions on the software used, the arrangement of the software on the screen, the note-taking strategy, the approach to synthesizing an action plan, and the challenges involved in resolving issues like conflicting, redundant, or missing information across various sources.

After the shadowing and reflective sessions, we conducted a semi-structured interview. Questions included "What are the challenges in creating action plans with the current setup," "How do you deal with potential conflicts among various sources of information," "How do you synthesize multi-source information into action plans," and "How can the current setup be enhanced to improve your note-taking experience?"

4.1 Collaboration and roles

Since our ultimate goal was to support existing structure and process, we observed the roles that each stakeholder played, and when collaboration opportunities arose in the existing process. We observed that the resulting action plans were collaborative: project managers, designers, and clients were assigned action items and desired awareness of other collaborators' progress. The creation of the action plans, however, was not collaborative: project managers would coordinate with designers and clients, identify next steps, and manage the creation and communication of action items, deadlines, and ownership.

Beyond the single creator/multiple collaborators style of the action plans, there is further nuance with respect to expertise. Project managers are *creator experts*: they not only formulate an action plan for the team, but are also experts in the execution and vocabulary of the project. Designers are *participant experts*: while they don't create the action plan, they have as much or more experience in the execution and vocabulary of each project. Clients are *participant non-experts*: like designers, they don't create the action plan, but simultaneously are pivotal in its success (e.g., if a project relies on a client sharing their biographical information, no one else can do this for the client) while not being experts in the terminology or process of creating a website. As we considered the design of *MixTAPE*, understanding the organization and limitations of the collaborative process it needed to facilitate was critical. In Section 9, we discuss future work on how stakeholders of varying roles and expertise can collaborate and iterate over action plans created by *MixTAPE*.

4.2 Challenges observed in the study

Our study revealed some drawbacks with the existing experience. Project managers follow a script during kickoff calls while capturing free-form notes. Client requirements can be diverse, however, and the conversation can quickly change course despite the script. Such dynamicity makes it challenging for project managers to coordinate between the script and notes during calls. For example, while discussing the aesthetics of a website, a client may ask for specific content in a section. As the notes lack structure, project managers lose track of where to put the new information. As a result, details get lost in the unstructured text noted during the call.

While synthesizing multiple sources of information (*e.g.*, the client brief and call notes) into action plans, project managers face several challenges. Project managers spent time on areas such as resolving conflicts among the sources, removing redundant information, and writing concise notes. For example, a client might ask for a different color palette during a call than was discussed with a salesperson. To provide a designer with clear directions, the project manager might add a note to *ignore the color request in the client brief*. Given the quantity of information the designer has to read, however, there is no guarantee they will capture every detail.

Moreover, the presentation of synthesized action plans varies across project managers, which can be confusing for designers. For example, one plan presents a hierarchical action plan with bullets and sub-bullets, while another includes a high-level summary at the top of the document, keeping a flat list of bullets toward the bottom. Project managers vary their action plan structure across projects, adding further uncertainty.

Despite the variance across action plans, we noted some underlying structure that project managers were trying to communicate. Project managers varyingly used color, bolding, capitalization, nesting, and text order to communicate *priority*. For example, project managers used various combinations of bold text, red text, and entirely capitalized text to prefix high-priority work that had to be completed even if other work could not be (e.g., "**HIGH PRIORITY:** Replace hero image of doctor with one of a dentist"). When work appeared in bulleted lists, higher-priority work tended to appear higher in the list than lower-priority work. Furthermore, nesting was used to group similar

tasks (e.g., a parent bullet called *Typos* with a list of bullets indented underneath it containing typos in the copy).

Project managers would sometimes further annotate the collaborator (e.g., designer, client, or themselves) that had to complete a particular task, indicating *ownership*. This tended to happen either through task prefixing (e.g., "Designer: Replace logo once client sends it" and "Client: Send updated logo") or by nesting multiple bullets owned by a collaborator under a parent bullet with their role (e.g., a bullet called "Designer" followed by indented bullets depicting all of the work the designer had to do). Finally, collaborators would indicate *completion* of work using strikethroughs, highlights, comments, or changing bullets into checkmarks. Project managers identified challenges in consistently providing these annotations within time constraints.

When asked about their thoughts on a system that would draft an owner- and priority-annotated action plan from multiple sources of project information, project managers voiced a preference for a mixed-initiative approach in which they could review the automatically generated action plans. In considering the functionality to provide in such a system, the core challenges identified in the needfinding study included:

C1. Capturing all the crucial details of a project during live note-taking can be difficult.

C2. The lack of structure and consistency in presenting action plans across project managers and projects can result in confusion amongst collaborators.

C3. Synthesizing multi-source information may result in conflicting and redundant action plans.

C4. While project managers see benefit in annotating priority and ownership of individual tasks, this practice is time-consuming and error-prone.

4.3 Design implications

From the observed challenges, we identified the following design needs.

D1. In-context note-taking interface: Contextual information is important in a chat conversation [68]. During a live call with a client, a project manager had to switch back and forth between the client's website, the client brief, and Team Messages where they take notes (**C1**). Putting the note-taking interface in the same context as the other work products allows project managers to access relevant information quickly.

D2. Quick text input: During a live conversation, typing every detail out is cumbersome (**C1**). In our pilot study, project managers had to repeatedly type out similar action items (e.g., "change the color palette of the website," "wait for content from the clients") across different projects. Reducing the keystrokes required to type out these repeated tasks could save project managers' precious time during the call.

D3. Flexible yet structured plan creation: Generated plans should be standardized to prevent confusion among collaborators (**C2**). Tasks should be organized in a consistent way across projects. Similar and relevant tasks could be grouped together to reduce context switching. Further, task descriptions and presentation should be consistent. However, plan creation also needs to accommodate the specific needs of individual projects. The system should support described structures with a certain degree of flexibility.

D4. Many sources, one task list: With multiple sources of information, project managers and collaborators had to communicate through other mechanisms like Slack to resolve conflicts or remove redundant tasks (**C3**). This problem could be solved by providing a single source of truth and a protocol to communicate task status.

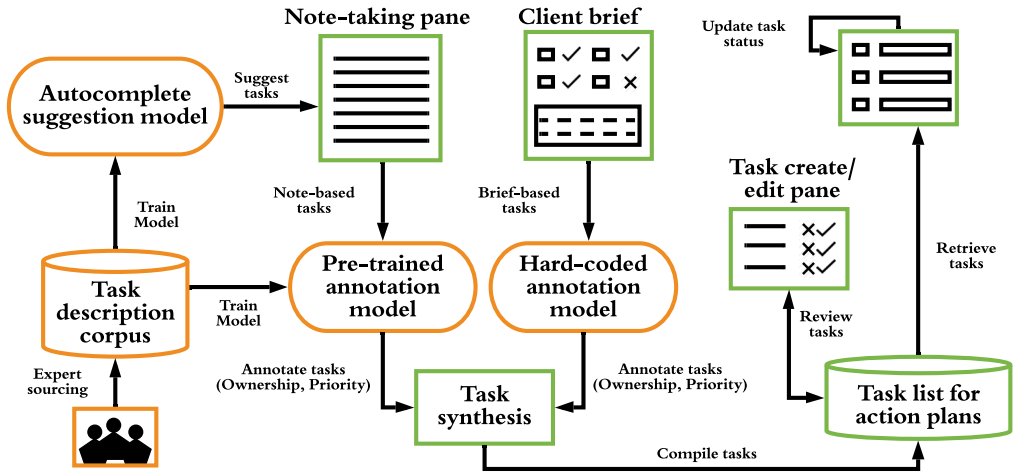


Fig. 1. The *MixTAPE* workflow. Components that required labeling and offline model generation (*e.g.*, the task description corpus and dependent models) are highlighted in orange.

D5. Automatic task annotation as a first pass: Project managers spent a lot of time assigning tasks to collaborators and prioritizing the tasks (**C4**). Some tasks can be automatically annotated with collaborator assignment and priority. However, not all task annotation can be resolved in such manner due to specific requirements of different projects. In this case, a mixed-initiative system where the annotations are first automatically generated but can still be corrected by a human offers the best of both worlds [31].

5 SYSTEM DESCRIPTION

To address the design needs identified in the needfinding study, we implemented *MixTAPE* as a web application using the *Django* web framework on the backend and the *React* frontend library, relying on *scikit-learn* for model training. We now provide a brief overview of how *MixTAPE* operates and then describe its components in further detail.

5.1 *MixTAPE* overview

Figure 1 provides an overview of the *MixTAPE* workflow for a given project. Before launching *MixTAPE* at B12, we first collected approximately one thousand tasks—created by Project Managers—by analyzing the Team Messages documents of prior projects. The tasks were then labeled with ownership (project manager, designer, client) and priority (high, medium, low) to create a *task description corpus*. One of the paper co-authors labeled the tasks, *i.e.*, assigned task ownership and priority, after discussion with the project managers to understand their approach. The labeled tasks were then used to train classification models to automate various components of *MixTAPE* (discussed in Section 5.2 and 5.3). These corpus- and model-building efforts occurred offline and are highlighted in orange in Figure 1.

As discussed in Section 2, given a project, two sources of information are compiled: a) a *client brief* from structured survey responses submitted by the client, and b) *kickoff call notes* compiled by the project manager. Two types of tasks, *brief-based* and *note-based* are automatically created from

these sources respectively. Each task is then assigned an owner and a priority. The task synthesis process is discussed in Section 5.3. The project manager then reviews the annotated tasks using a *task creation/editing pane* and corrects any inconsistencies in assigned ownership or priority. Following task review, each collaborator can view their assigned tasks in their respective *task view*s and update the task status as they complete their work. We now describe these steps and interface components in detail.

5.2 Semi-structured note-taking

MixTAPE provides a large text area for project managers to take notes related to a project (Figure 2a). A project manager can toggle the note-taking pane with a shortcut. The note-taking pane is embedded in the website editing application (**D1**). The text area is prepopulated with section headers of different types of tasks (e.g., website *Structure* and *Design*). These section headers impose a structure on project managers' note-taking (**D3**).

The note-taking experience offers a syntax for project managers to describe actionable tasks. Project managers take note of a task by prepending the task description with a special symbol ("*" or "-"). They can also add details to a task by prepending the note with "**" or "- -" below the task to which the note belongs. This structured format, inspired by markdown, enables project managers to quickly capture various aspects of conversations with clients while maintaining the structure required for task creation (**D2**, **D3**).

To further keep the task contents consistent (**D3**), especially across different project managers, *MixTAPE*'s note-taking interface offers autocomplete suggestions that project managers can choose to accept or ignore (Figure 2a). Autocomplete suggestions also allows users to capture notes with fewer keystrokes (**D2**). *MixTAPE*'s autocomplete suggestion model is built from the task description corpus discussed in Section 5.1. We first generate n-grams from the task description corpus, where n = 1,2,3, and remove infrequent n-grams. We then compute the point-wise mutual information (PMI) [21] of the n-gram collection. As the project manager types their notes in the pane, a ranked list of top-k n-grams is displayed based on the most recently typed w = 3 words. The list is ranked by the PMI of the candidates retrieved by the model.

5.3 Task creation

When the project managers finish taking notes, they can click on the "Convert to tasks" button (Figure 2a) at the bottom of the pane to create an action plan. *MixTAPE* synthesizes information from the just-captured call notes and the previously provided structured client brief information (Figure 2b) to create an action plan for the project (**D4**).

The action plan synthesis algorithm works as follows. First, the call notes are parsed by their section, with a new task created for each "*" or "-" and subsequent lines of notes containing "**" or "- -" added to the details of that task. These converted call notes are called *note-based tasks* for the remainder of the paper. Another list of tasks, called *brief-based tasks* for the rest of the paper, is automatically created from the client brief. To generate brief-based tasks, a collection of Python functions read a structured representation of the client brief (Figure 2b) and create task text using a template that matches the representation. For example, in Figure 2b, the client selects that they hate the fonts, so *MixTAPE* maps the selection to a "Change [*something*]" template and creates a task for designers to "Change the fonts" in Figure 2c.

Both sets of tasks are grouped into relevant sections (e.g., Structure, Design) and are assigned a priority. Each task consists of a title, an optional note, and an owner (Figure 2c). The owner of a task can be any of the project collaborators, (i.e., the project manager, client, or designer). *MixTAPE* uses a mixed-initiative approach to assign an owner and a priority level to a task (**D5**). It first automatically assigns the task owner and the priority of the task (high, medium, and

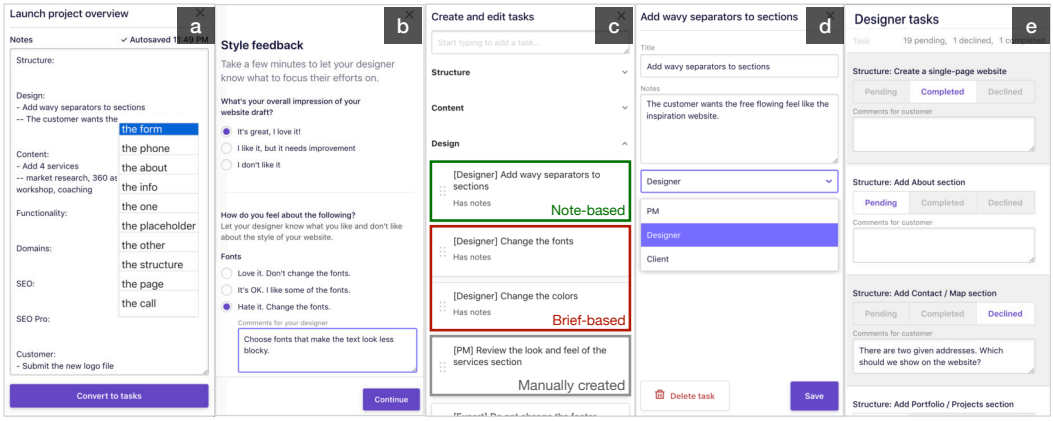


Fig. 2. *MixTAPE* components (HTML/CSS edited slightly for presentation): a) semi-structured note-taking pane with autocomplete support, b) survey responses that are compiled into client brief, c) task editing pane for review and management of action plans, d) task detail pane for modifying a task, and e) task view for designers.

low). Afterward, the project manager can correct the generated task details. The automated task assignment mechanism combines hard-coded rules (*e.g., any tasks under the "Other" section will be assigned to the project manager*) and predictions from our pre-trained models as described next.

Owner assignment models. To train a model for assigning an owner to a task, we utilized the task description corpus mentioned earlier. Besides task description, each entry within the corpus contained additional information such as the section of the task, and we manually provided class labels for owner for each task. We experimented with a number of feature combinations for our candidate classifiers, empirically using task section labels and a TF-IDF vector representation [2] of task descriptions as features. We selected a classifier by performing a cross-validated grid search on an 80% training set over a variety of model families (including Decision Trees [55], Linear SVC [14], and non-linear SVM [32]), hyper-parameters (*e.g.*, C, gamma, degree), optimizing for accuracy. The top performing model, an SVM with a polynomial kernel, had an accuracy of 0.8740 on a separate 20% test set (precision = 0.8760, recall = 0.8749, and F1 score = 0.8740).

Task priority assignment. Using the same task description corpus and features, we manually labeled the priority (high, medium, or low) for each task description. High priority tasks are must-do (e.g., "Add the MailChimp form embed code"). Medium priority tasks should be done (e.g., "Show blogs by categories"). Low priority tasks are nice-to-have (e.g., "[Customer is] seeking more minimalist design"). We conducted a similar analysis to that of the owner assignment models. We found the SVM with a polynomial kernel to again be the best candidate with a test set accuracy of 0.807 (precision = 0.816, recall = 0.807, and F1 score = 0.809). The selected feature inputs to the models were the task owner label and TF-IDF vector representation of the task description. The priority labels for this model were on a scale of 1 (low priority) to 3 (high priority).

On larger (and potentially unlabeled) datasets, features like semantic vectors [44] (e.g., Word2Vec), when used in combination with more complex models, could have further improved the classification performance. Nonetheless, we found the model we created to be sufficient for *MixTAPE*, and the experimental results in Table 2 of Section 7 show that *MixTAPE*'s models performed well in live projects.

5.4 Task editing pane

After *MixTAPE* automatically synthesizes tasks, the note-taking pane is replaced by a task editing pane showing a list of tasks grouped by section (Figure 2c). On the task editing pane, project managers can create new tasks, edit existing tasks, remove irrelevant tasks, and reorganize and reorder tasks to reflect their priorities. They can reassign a task to a different team member if it was incorrectly assigned.

5.5 Task view

Project managers and designers can view the action plan on a task view page (Figure 2e). The task view page shows lists of tasks categorized by their task owner. Each row shows the task title, notes, and status of the task (completed, pending, or declined). When a task is completed, a user can mark the task as completed, or leave an explanation in the text area for the client on why they elected to decline the task.

6 EVALUATION STUDY

In order to evaluate whether *MixTAPE*'s features described in Section 5 address the challenges identified in the needfinding study of Section 4, we conducted a study spanning six weeks with the following research questions:

RQ1. How did *MixTAPE* impact the resulting action plans and the process to generate them?

RQ2. How did *MixTAPE* and its components affect participants' experience in capturing, synthesizing, and executing on action plans?

Participants. Our evaluation involved two types of participants: designers and project managers. We recruited eight designers who regularly design websites for B12's clients. All designers are independent contractors who work remotely and have between several months and several years of experience with B12. Designers were all initially trained with B12 at least several months before the introduction of *MixTAPE*. The designers were compensated with their normal hourly rate for all work and research activities.

We recruited the same two project managers who participated in the needfinding study to participate in the evaluation study. Both project managers had several months of experience with B12's project delivery model prior to the introduction of *MixTAPE*. Note that the project managers are not co-authors.

Because our study involved client-specific requests, we also requested informed consent from each client to allow researchers to track the progress of their projects.

Projects. Each project in the study involved the creation of a website for a new B12 client. The study involved 32 website design projects launched using *MixTAPE*. We also collected information on 69 projects that were launched prior to the introduction of *MixTAPE*. For these 69 projects, we analyzed project logs and Team Messages Google documents with call notes described in Section 2. We refer to the setup before *MixTAPE* as Baseline throughout the evaluation.

Project manager study procedure. We first introduced *MixTAPE* to each project manager in separate warm-up sessions where we walked them through the entire process from note-taking to action plan generation. Moreover, we shared a video tutorial that explained the complete project manager experience in *MixTAPE*. We then shadowed each project manager on a number of kickoff calls to observe their interaction with the system. Based on these shadowing sessions, we made light experience improvements, fixed a few bugs, and formulated questions for reflective sessions in the next phase of the study. Project managers recorded screen and audio captures of each kickoff

call and their post-kickoff project plan synthesis. We used these recordings to guide qualitative observations as well as reflective sessions.

At the end of the first week of the study, we conducted a reflective interview with the project managers to obtain early feedback on *MixTAPE* and also asked questions about their workflows, interactions with the system, and challenges they encountered in their use. Note that no new changes were made to *MixTAPE* based on the reflective interview. Finally, we conducted a semi-structured interview to understand their overall experience in using *MixTAPE*, the impact of the new features introduced, the differences from the Baseline experience, and the implications of introducing the new experience. Note that instead of waiting for the study to conclude after six weeks, the interview was conducted as soon as a project manager finished creating action plans for all of their allocated projects.

Designer study procedure. We introduced designers to features of *MixTAPE* using a video tutorial. We conducted two types of surveys over the course of the study. The first survey was conducted after designers created a first version of the website on each project. Designers rated their experience with *MixTAPE* on eight metrics: *accuracy* of following client requirements, *comprehensiveness* of the action plans, *quality* of the generated action plans, degree of *contradiction* among action items, level of *frustration*, *mental demand*, ease of *planning* for a project, and project *completion success*. The ratings were on Likert scales from 1 (most negative experience) to 7 (most positive experience). We also asked them an open-ended question on their overall experience in using *MixTAPE*.

At the end of the study, the designers participated in a final survey where we asked them to rate their overall experience in the Baseline system experience as well as their experience with *MixTAPE* on the aforementioned eight metrics as well as a ninth: project *completion speed*. The survey presented each question in a neutral/factual way in a randomized order across surveys. Designers also answered open-ended questions on the positive and negative aspects of each experience.

Presentation of quotes. In presenting participant responses in Sections 7 and 8, we replaced B12-specific terminology with terms introduced in this paper, *e.g.*, client brief, call notes. These replacements are denoted by square brackets (*[]*) in the quotes.

Quantitative evaluation. In addition to the qualitative interview and survey data, we quantitatively analyzed and compared the action plans created in both experiences. We also measured the statistical significance of the comparisons where relevant. As the number of projects for each experience, *i.e.*, *MixTAPE* (32) and Baseline (69), were different, we utilized Welch's t-test [62]. For each t-test, we report the mean (μ_X, μ_B) and standard deviation (std_X, std_B) of both experiences, the t-statistic (positive when $\mu_B < \mu_X$), and p-value (statistically significant when $p < 0.01$). Furthermore, we provide summary statistics of usage logs to understand how *MixTAPE* was utilized.

Limitations. While rigorous, our study has several limitations that can be strengthened by future larger-scale and broader-domain studies.

The first limitation is in sample size: we conducted the study with two project managers and eight professional web designers on 32 website design projects for B12's clients, comparing those to 69 prior projects. While a larger sample would have allowed us to perform more definitive quantitative analysis, we combined this analysis with qualitative survey, interview, instrumentation, and screen/audio recording data to result in justifiable inferences that can be explained with multiple data sources.

Another limitation of the study is in its focus on web design projects. While testing in other domains is prudent, the observations and reflections that result from the study entirely focus on aspects of the note-taking and project plan creation experience in *MixTAPE*, and we focus on observations that are not specific to web design.

The final limitation is that, due to the realities of deploying new monolithic experiences in real-world work settings, our study does not randomize the usage of the pre-*MixTAPE* and post-*MixTAPE* experience. To control for this, we utilize a needfinding study to identify challenges in the pre-*MixTAPE* experience and present the systems in a neutral and randomized fashion whenever discussing them with participants. We acknowledge the fact that even after randomizing the order of the question set by system, the designer survey may still introduce recency bias toward *MixTAPE*. One alternative could have been to provide the survey on Baseline to the designers before they started using *MixTAPE*. However, since the designers were independent contractors, we were able to select participants only after they started picking up projects and managed corresponding tasks using the newly launched *MixTAPE*. Therefore, such bias was unavoidable in our study environment.

In many ways, our specific study design falls directly out of the study limitations: given a small number of participants and the introduction of a new system, we motivate the system design with a needfinding evaluation, and evaluate the design by comparing the resulting system to the original Baseline quantitatively and qualitatively. This form of formative evaluation is well-supported by the literature in studies of novel systems like CrowdCrit [45], Critter [15], and Tilda [68]. Additionally, the conditions for Baseline are well-founded: in comparing a new system for adding structure to unstructured Slack conversations, the authors of Tilda compared its outcomes to those of a Google document for accomplishing the same goal, much like the Team Messages Google document used by B12.

7 RQ1. IMPACT OF *MIXTAPE* ON ACTION PLANS

Overall, *MixTAPE* enabled faster action plan synthesis while improving the consistency and comprehensiveness of tasks compared to Baseline. Moreover, our comparative survey revealed that participants found action plans generated by *MixTAPE* are of higher quality than those generated by Baseline.

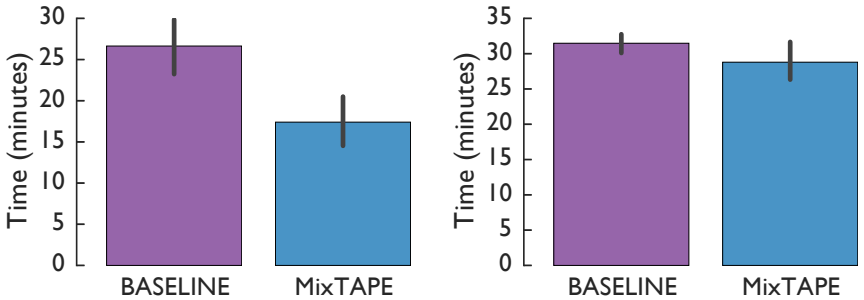
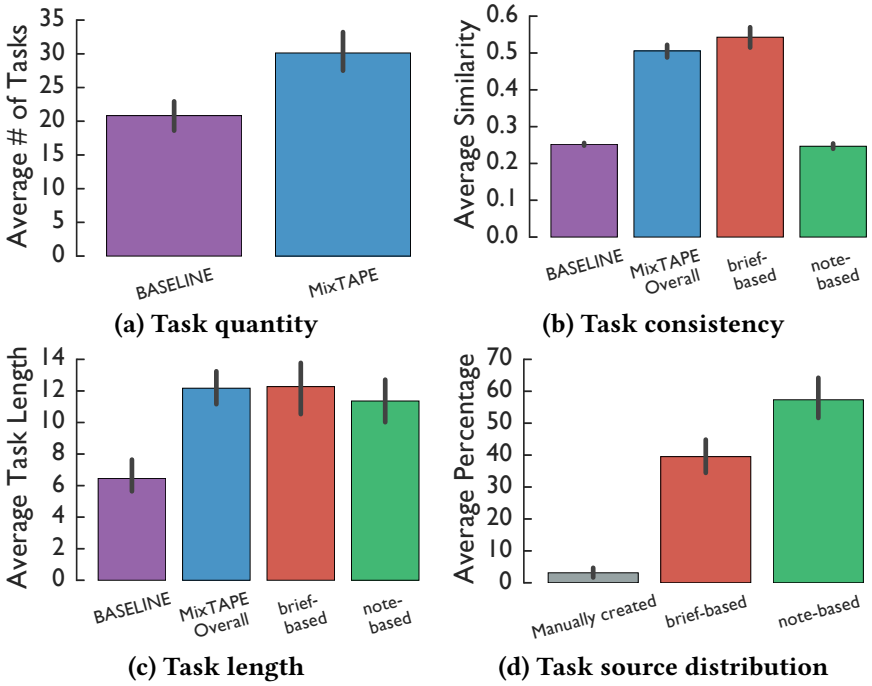


Fig. 3. *MixTAPE* enabled (a) faster action plan synthesis. However, (b) note capture time stayed almost the same. Error bars throughout paper represent 95% confidence intervals.

7.1 Faster action plan synthesis

Figure 3a shows the time spent on average by project managers for plan synthesis using Baseline and *MixTAPE*. The error bars for the bar plots, presented throughout the paper, were computed based on a 95% confidence interval. Both the note-capture and plan synthesis times were measured by B12 's activity logging application, enabling fine-grained time-tracking for all project manager activities.

In Figure 3a, we see that project managers spent significantly less time generating the plans across projects after their kickoff calls with *MixTAPE*—a nine-minute (33.70%) reduction from Baseline (μ_B = 27.09, std_B = 14.21, μ_X = 17.96, std_X = 7.58, t = -2.7130, $p < 0.01$). As fully loaded project managers can kick off up to five projects a day, this amounts to over 45 minutes per day that project managers can utilize for other activities. Figure 3b shows average note capture time for both experiences. *MixTAPE* did not significantly improve the time to capture notes during the kickoff call with clients (μ_B = 31.71, std_B = 6.81, μ_X = 29.21, std_X = 8.04, t = -1.4103, $p > 0.01$). We discuss qualitative reasons for this, such as a lack of use of the autocomplete feature, in Section 8.1.



(a) Task quantity **(b) Task consistency**

(c) Task length **(d) Task source distribution**

Fig. 4. *MixTAPE* generated (a) more tasks that were, (b) more consistent and (c) more descriptive, compared to Baseline. Instead of following multiple documents, designers viewed the tasks within a single source of truth (d).

7.2 More comprehensive action plans

Figure 4a shows the average number of tasks created per project, with Baseline and *MixTAPE*. *MixTAPE* generated more tasks (1.45*X*) than the Baseline (μ_B = 20.83, std_B = 11.93, μ_X = 30.12, std_X = 8.29, t = 4.1321, $p < 0.01$). Designers also found *MixTAPE*'s action plans to be more comprehensive than those generated through Baseline (see Table 1). One designer (D2) commented: "*Ideally [with MixTAPE], I know I've pretty much completed the project once I've completed each task.*" Moreover, project managers noticed a reduction in the number of changes clients requested after the initial website design based on the action plans created by *MixTAPE*: "*I feel like I've had more positive launches [with MixTAPE] lately. I haven't gotten any client feedback like the website is missing something. That is a comment I used to get with Team Messages*" (PM1).

7.3 Consistent yet descriptive tasks

Standardized language helps teams operate more effectively [30, 51, 52], so consistent language across projects is important. To measure the consistency of tasks across projects for a given tool, we first concatenated the tasks created for a given project into a document. Given a document, we performed tokenization, stopword removal, and lemmatization to create high-quality TF-IDF vector representation [2]. The TD-IDF vector representation offsets the impact of high frequency words. To measure pairwise similarity between two documents, we used the cosine similarity metric [27] as it is less susceptible to document size, and therefore more resilient to variance in number of tasks and task detail. Given a collection of term vectors, the average of the pairwise cosine similarity provides an estimate of the consistency of the language within that collection. Figure 4b shows the average similarity of the task collections across projects launched by both Baseline and *MixTAPE*—the overall consistency *doubled* ($2.04X$) with the introduction of *MixTAPE* ($\mu_B = 0.25$, $std_B = 0.18$, $\mu_X = 0.51$, $std_X = 0.19$, $t = 30.4320$, $p < 0.01$). Figure 4b further shows the diversity measure for tasks created from individual sources of information. We see that the brief-based tasks—created from the client brief—largely contributed to the standardization of language, compared to the less structured note-based tasks created from the call notes. We exclude the manually created tasks from this chart, as the quantity of such tasks (3.13%) was too low to draw conclusions compared to note-based (57.34%) and brief-based (39.53%) tasks (see Figure 4d).

Figure 4c shows the average length of tasks created by both systems across projects. For *MixTAPE*, Figure 4c shows the average length of the brief-based and note-based tasks. Overall, the tasks created by *MixTAPE* were more descriptive ($1.88X$ longer) than Baseline ($\mu_B = 6.46$, $std_B = 5.97$, $\mu_X = 12.17$, $std_X = 2.97$, $t = 5.2294$, $p < 0.01$). Combined with the task consistency results (Figure 4), even though *MixTAPE* generated tasks are now more verbose/descriptive, they are more consistent across projects, making it easier for the designers to interpret tasks. One of the project managers (PM2) commented: "*I think it's probably much easier for the designer to understand something consistent and structured across projects. Versus in the Team Messages, where they have to deal with different types of notes from different managers.*" Each task in *MixTAPE* may contain additional notes (added by '**' or '- -'). While this increased the level of task detail compared to the Baseline, the notes provided contextual information regarding tasks.

7.4 Quality of the action plans

Table 1 shows the results of the final survey in which designers rated their experience with Baseline and *MixTAPE* (question order was randomized across participants). Notably, *MixTAPE* had a higher average rating than Baseline for all the metrics.

As confirmed in Table 1, the designers found the tasks generated using Baseline to be highly contradictory, whereas *MixTAPE* generated less contradictory tasks. Similar sentiment was echoed by one project manager (PM1): "*With the new tool (MixTAPE), I have noticed I have been getting less clarifying questions from the designers.*" According to Table 1, the designers rated the quality of tasks generated using *MixTAPE* much higher compared to Baseline. One designer (D5) commented: "*It (MixTAPE) is seamless and gets to the point. Less time consuming and easy to understand.*" According to Table 1, even though designers did not have any negative experience in either tool for a number of metrics, *e.g.*, ease of planning, frustration, mental demand, completion success, a higher percentage of designers had positive experience with *MixTAPE*, compared to Baseline. One designer (D1) commented: "*While I was able to complete quality projects in a timely manner using Team Messages, it was less straightforward and dependable. I'd keep looking at each document multiple times to ensure I completed everything. [MixTAPE] resolved a lot of that back and forth, which saved some time and allowed me to start designing the websites immediately.*"

Table 1. Final survey results for *MixTAPE*. Ratings are on a scale of 1 (*i.e.*, most negative) to 7 (*i.e.*, most positive). Participants rated their experience with *MixTAPE* as very positive for all the metrics.

Mertic	*MixTAPE*		Baseline	
	Mean	Variance	Mean	Variance
Accuracy	**6.00**	0.33	4.71	0.24
Quality	**6.00**	0.33	4.43	0.62
Degree of Contradiction	**6.00**	0.67	3.71	2.24
Ease of Planning	**6.14**	1.14	4.57	0.62
Frustration	**5.86**	1.14	4.57	0.29
Mental Demand	**6.14**	0.48	5.00	1.00
Comprehensiveness	**5.71**	1.24	4.29	0.57
Speed of Completiom	**5.86**	1.14	4.43	0.62
Successful Completion	**6.14**	1.14	5.43	0.62

7.5 Accurate task assignment and prioritization

Table 2 summarizes the performance of both types of owner classification and task priority assignment models: pre-trained and hard-coded. The pre-trained owner classifier achieved an overall accuracy of 96.38%, suggesting that project managers largely agreed with *MixTAPE*'s classification of task owner. To evaluate a priority classifier, we measured the rank correlation between the order of the tasks per section as generated by *MixTAPE* with the eventual order (rank) of the tasks after project managers reviewed the generated action plans. We used the *Kendall's Tau* [41] metric for this purpose, where the closer the value to 1, the more highly correlated to ranked groups are. Again, *MixTAPE*'s models demonstrated high degree of correlation with project managers' arrangement of tasks. Surprisingly, the accuracy (and rank correlation) statistics of the hard-coded tasks were not 100% (and 1). In some instances, project managers added tasks meant for the client in the "Other" section during note-taking and then moved the tasks during plan synthesis using the task editing pane. However, *MixTAPE* was programmed to assign any task added to the "Other" section to project managers (see Section 5.3), which contributed to a small number of mistakes to the overall accurate assignments.

8 RQ2. PARTICIPANT EXPERIENCE WITH *MIXTAPE*

We now identify how *MixTAPE* affected participants' experiences, and how they perceived its various components.

8.1 Semi-structured note-taking

The semi-structured note-taking pane helped project managers maintain context while typing notes. One project manager (PM1) mentioned—"*I like how it's (MixTAPE) more structured than Team Messages in terms of what goes where.*" However, the project managers revealed that they didn't use the auto-complete feature at all and found the feature to be distracting, given such a high-intensity typing environment. In Figure 3b of Section 7, we see that there was no significant improvement in note-taking time using *MixTAPE*, which is consistent with the lack of adherence to autocomplete. Ultimately, the time to capture notes was dominated by the length of the kickoff call, which did not change with the introduction of *MixTAPE*. In Section 9, we discuss how the autocomplete experience can be improved.

Table 2. Owner assignment and task prioritization.

Model Type	Owner (Accuracy)	Priority (Kendall's Tau)
Pre-trained	0.9638	0.9145
Hard-coded	0.9508	0.9598

8.2 Task creation: assignment and prioritization.

Overall, project managers preferred the synthesis experience with *MixTAPE*. One project manager (PM2) commented "*Converting to tasks using the Team Messages was harder. [MixTAPE] was a lot easier as it automatically created a structured set of tasks, which I prefer.*" We presented the accuracy of the owner and priority classification model in Table 2. Project managers echoed observations that classifier accuracy was high and found the automatic assignment of team members to tasks very helpful compared to Baseline: " *I think it (MixTAPE) does owner assignment correctly more often than not. I had to manually separate the tasks out in the Team Messages*" (PM1). Even though their confidence in the classification model increased with time, project managers still reviewed the assignments during the synthesis process: "*I am feeling more confident as I am using the tool (MixTAPE) more and more. I do double check though*" (PM2).

8.3 Task editing pane

Organization. Project managers liked how the tasks were grouped into sections as it made plan synthesis easier. One project manager (PM2) said, "*[MixTAPE] structures everything very clearly for the designer and they can easily understand what's needed. It is very difficult to miss things.*"

Conflict resolution. As *MixTAPE* aggregates two sources of information, some of the synthesized tasks may be conflicting or redundant. Nearly 20% of tasks were deleted on average, the majority of which were brief-based. The task editing pane enables project managers to resolve conflicts through edits and deletion without having to write additional comments for the designers as they had to in Baseline (see Section 4). One of the project managers (PM1) said, "*Often times the [brief-based tasks] conflicted with the [note-based tasks]. Once I resolve those, I feel like I don't need to let the designer know anything else, which is nice.*"

Overall, project managers had to update the description of a very small percentage of tasks per project (less than 4% on average). The same observation applies to both brief-based and note-based tasks. The semi-structured note-taking pane partly played role in this outcome, as one project manager (PM2) commented: "*With the new [note-taking pane] you need to follow a structure, which I think is good. Because it means that you have to be very precise in what you type and make sure it makes sense.*" Project managers noted that they would have preferred if structured data was pre-populated in the notes, thereby enabling the resolution of conflicts during note-taking rather than during synthesis: "*Having the [client brief] imported in the note would help me to update things during the [kickoff call] and not post-kickoff where I have to delete tasks*" (PM2).

8.4 Task view

Project managers used the task view component to track the progress of a project, and designers used it for following the action plan. We now discuss the experience of both groups of participants. Table 3 summarizes the results of the per-project designer survey completed after the delivery of each project with *MixTAPE*. For all the metrics, the designers rated their experience as very positive (around 6/7) on average.

Executing on the action plans. The project managers preferred the *MixTAPE*-generated structured action plans as designers were presented with a strict outline. One project manager (PM1)

Table 3. Per-project survey results for *MixTAPE*. Ratings are on a scale of 1 (*i.e.*, most negative) to 7 (*i.e.*, most positive). Participants rated their experience with *MixTAPE* as very positive for all the metrics.

Metric	Mean	Variance
Accuracy	6.07	0.88
Completion Success	6.03	1.28
Comprehensiveness	6.13	1.36
Contradiction	6.07	0.71
Frustration	6.07	0.90
Mental Demand	6.06	1.56
Planning	6.05	1.34
Quality	6.06	0.94

commented, "*With [MixTAPE], it is more straightforward to enforce a process. It's miles ahead in terms of what we can do for quality control and consistency than with just Team Messages. I appreciate that as a project manager.*" Similarly, all of the designers appreciated the fact that the task view presented the tasks as a checklist which brought more organization and coherence to how they approached a project compared to Baseline. One designer (D2) commented, "*It's great to have a project broken down into its component steps, which forces specificity more so than Team Messages which could possibly be vague or missing information.*"

A single source of truth. As shown in Figure 4d, both brief-based and note-based tasks contributed significantly to the final action plan. Note that the chart shows the average percentage of tasks generated from each source after the deletion of tasks. With Baseline, designers had to move back and forth between these sources to formulate the final action plan themselves, which was cumbersome. One designer (D5) commented: "*I like that what should be included or not included works seamlessly into one embedded checklist. I think [task view] is a more efficient and scalable way to input all requirements into one source without going back and forth between the [client brief] and Team Messages.*"

Tracking progress. With the ability to track the status of each task, project managers found it easier to manage the progress of a project: "*With [MixTAPE], it's clearer to me what the designer has completed and what's pending. Whereas in the Team Messages, it was really ambiguous—some designers didn't interact with the document at all. Some would strike through completed tasks, some would check off—everybody did different things*" (PM2). Another project manager (PM1) thought *MixTAPE* ensured more accountability: "*I think in this new experience (MixTAPE), it is easier to hold the designers accountable. If they haven't completed a task [before submitting a project], I can point that [out] to them.*"

Coordinating across iterations. With *MixTAPE*-generated action plans as the single source of truth, both project managers and designers wanted to use the action plan as the basis for coordination in subsequent iterations of a project. However, they found *MixTAPE* difficult to use for iteration. For each iteration, project managers needed to add tasks manually, losing the speed of the note-taking experience and the automated task prioritization. Moreover, unless mentioned explicitly, it was difficult to identify which iteration a given task belonged to. One project manager commented: "*When we need further iterations from the client, I feel I have less control and I have a harder time communicating with the designer*" (PM2). Our per-project surveys revealed that designers shared the same sentiment: "*Very hard to find tasks between iterations (with MixTAPE). If there's notes added after a [new request] in the second version, it's hard to tell which comment belongs to what version*" (D7). Some designers ($N = 4$) mentioned that they preferred back and forth

communication via the comment option in the Team Messages Google document. One designer (D8) commented—"*I liked the ability to discuss specific parts via Google doc native comments. It was pretty handy sometimes to see a full log of comments plus initial [first version] description in a single thread.*"

9 DISCUSSION AND CONCLUSION

In this paper, we have introduced *MixTAPE* and shown how its semi-structured note-taking environment, automated task synthesis algorithm, and task owner and priority classifiers positively impact the creation of collaborative action plans. Quantitatively, we find *MixTAPE* to increase the quantity and consistency of tasks while reducing the time to synthesize a project plan. Qualitatively, study participants identify it as positively impacting their experience across several measures compared to Baseline, a largely unstructured experience. In this section, we discuss the implications of our findings to systems involved in the broader space of project plan generation. We also discuss how the concepts discussed in *MixTAPE* can be generalized to other domains.

9.1 Implications for mixed-initiative planning tools

9.1.1 Adapting by observing. The principal of mixed initiative interfaces [31] argues that such systems should "*continue to learn by observing a user's goals and needs.*" We now discuss how this philosophy applies to planning tools like *MixTAPE* based on our observations from the study.

Structure imposing process. Our needfinding study was informed by an existing process. Our resulting system design viewed *MixTAPE* as a tool that supported this existing process. In introducing *MixTAPE*, we unexpectedly found during the project manager shadowing sessions that the introduction of the system *resulted in new process standards*. For example, as the system exposed designers and clients to structured notes, questions arose around the best way to effectively take notes. For example, whereas an outline with a bullet like "logo" made sense in a previously unstructured set of notes, taking more action-oriented notes such as "Upload your logo" became more effective. As the system more explicitly presented tasks in sections like *Structure* and *Content*, practical questions around where to place various tasks resulted. These questions of process weren't new, but in presenting tasks in structured form, the questions were newly important. As researchers, we initially opted to observe these questions arise without interference, but realized that participants sought guidance and eventually assisted in the iterative generation of a process document on note-taking and task-creation best practices. As *MixTAPE* is deployed in other domains, stakeholders' requirements may evolve similarly as they use the system and adjust their mental models to the newly introduced process.

Adapting to usage norms. Tasks are at the very core of the *MixTAPE* design, where each task corresponds to a single stakeholder action. During the course of our study, we noticed a recurring theme of project managers creating a high level task (e.g., "- Add a team page") with subsequent notes (e.g., "- - Attach the CEO's photo") that were actually tasks. Allowing tasks to nest would better facilitate a scenario where tasks are small and actionable while still logically grouped, which prior work shows can increase adherence [15]. While our research was on the synthesis of action plans, we also got to observe how participants completed the tasks in those plans. Two features of commercial task-tracking systems that participants requested were dependencies (e.g., "Replace the client's logo" depends on a client completing "Upload your logo") and future-facing tasks (e.g., "Discuss chat integration in two months"). Synthesizing and classifying both task types, *i.e.*, nested and dependent tasks, is an interesting problem for future work. While this may require redefining what constitutes a task in *MixTAPE* and redesigning various aspects (*e.g.*, the markdown language used for note-taking), such iterative refinements are expected a mixed-initiative system.

9.1.2 Toward collaborative action planning. We now discuss how to effectively facilitate collaboration in such a setting based on our observations from the study.

Communication among stakeholders. One task-related observation from the study is that as tasks became the basis for collaboration, participants wanted tasks to become the basis for conversation. Participants asked for two forms of conversation threads on each task: one for internal conversation (*e.g.*, for project managers and designers to clarify details instead of using tools like Slack) and the other for external conversation (*e.g.*, for designers to explain to clients why they made certain decisions). Future studies on *MixTAPE* may explore how stakeholders collaboratively augment action plans in the absence of standardization (motivated by concepts of boundary objects [42, 59], which are artifacts that are shared by different stakeholders but viewed or used differently by each of them) and how people with different levels of expertise and roles (as discussed in Section 4.1) collaborate on action plan creation (motivated by Enact [43] that studies collaboration between expert stakeholders).

Note-taking in multiple phases. As described and implemented, *MixTAPE* imposes a rigid notes-then-tasks usage. In practice, the process our participants followed required more than one phase of note-taking. After the team completed their tasks, they would send the client a website, receive feedback, and complete followup tasks. Action plan synthesis systems should support iteration, where a project manager can take multiple rounds of notes and synthesize them into tasks to be completed in several phases.

9.1.3 Automation's role in task synthesis. Finally, we discuss how automation may impact task synthesis in a mixed-initiative setting and how to further improve the experience.

Autocomplete in a multi-modal setting: Despite the prevalence of autocomplete in modern text-based interfaces in products like GMail or Amazon, autocomplete was the least used feature of *MixTAPE*. On reflection, n-gram ($n = 1, 2, 3$) autocompletion is incompatible with a multi-modal note-taking environment, where actively listening to clients while capturing their key points leaves little cognitive effort or attention to also read autocomplete suggestions [53, 54, 65]. Therefore, mixed-initiative planning tools need to accommodate alternative mechanisms to improve the note-taking experience. One alternate experience design can be to suggest common complete tasks rather than common phrases to reduce cognitive burden. Furthermore, with the advent of speech recognition systems, the role of transcribing tasks can be delegated to an automated agent to further reduce the load of the project managers and improve the quality of the call notes.

Enhancements to automated task synthesis: Given the negative result on autocomplete usage, it is reasonable to question the effect of automation on project outcomes. Still, note conversion provided structure for the resulting action plans. Automatic task synthesis replaced the previously tedious process of manual owner and priority assignment, reducing plan synthesis time by 33% in the process. In a mixed-initiative manner, project managers further curated action plans to resolve conflicts and correct automation errors, resulting in high-quality action plans as identified by several participants.

While project managers appreciated automatic task synthesis, they were frustrated when tasks generated from their notes, *i.e.*, note-based tasks, were redundant with those that the machine generated, *i.e.*, brief-based tasks. In the future, planning tools like *MixTAPE* can incorporate automated quality assurance modules to automatically identify similar or duplicate tasks after task synthesis, and flag conflicts or redundancies for the project managers to handle. A similar idea has been adopted in Critter [15] to maintain the quality of checklists for creative design tasks.

Towards automated task execution. As discussed in Section 3, current intelligent task management systems [23] delegate very simple tasks to project execution assistants. It is worth exploring the

design of a task manager that can delegate more advanced tasks that do not require creative thinking (*e.g.*, "change font color to blue") to automated execution assistants so that human collaborators can allocate more time for creative tasks. Designing such a system that facilitates human-AI collaboration offers interesting research challenges [64].

9.2 Generalizability of *MixTAPE*

We now explain how the concepts and techniques proposed in this paper generalize to other domains. More specifically, we discuss which *MixTAPE* components can be used as-is and which components would require domain-specific modifications.

The process of action plan creation involving multiple stakeholders can be seen in a wide variety of fields. Asana, a popular task management tool, now has *fifty thousand* organizations as paying customers and is used by teams of collaborators in diverse domains such as finance, telecommunication, digital media, and software industry [8]. We also discussed plan synthesis for other settings like medical and education setting in Section 1. In all of these settings, various sources of information are combined manually by project managers to create an action plan with tasks having varying degrees of priority and assigned to various stakeholders. To this end, *MixTAPE* provides a *framework* for combining various information sources and automating the task synthesis process.

After successful deployment on website creation projects, *MixTAPE* has now been deployed to manage other types of projects at B12: copywriting projects and website enhancement projects. These projects have different processes and team structures from the website creation projects. For example, a copywriting project (e.g., writing a blog post, editing product descriptions) is completed by a team of a project manager, a copywriter, and sometimes a designer. So far, the teams have successfully adapted *MixTAPE* to a different workflow and a different set of collaborators. However, these deployments were made several months after the user study was concluded and the implications of launching *MixTAPE* in these teams have not been formally evaluated.

While we are confident that the lessons from *MixTAPE* apply to the creation of action plans for a broad variety of teams, it's worth considering how to reduce the domain-specific work required to apply the system to other domains. The majority of the online components of *MixTAPE* (*i.e.*, the task synthesis algorithm, task create/edit pane, and task view) are relevant for any domain that requires team-based task creation and management. While the markdown language-based note-taking pane is also generalizable, the structure of the pane will vary based on the domain. On the other hand, training models for auto-completion and owner/priority classification would require creation of a domain-specific corpus. However, the process of corpus creation (via expert-sourcing) and model training may remain the same.

One challenge with task management in such a setting is that the types of tasks may evolve as new projects with different clients are launched. As a result, all three trained models (*i.e.*, the autocomplete suggestions, owner assignment, and priority assignment) may become obsolete and need to be retrained. These would require expert-sourcing the task description corpus creation again, which can be tedious. One alternative can be to leverage reinforcement learning [38] to augment the models—as project managers review and relabel tasks in the task create/edit pane, these changes can be propagated upstream to reward/penalize the classification models. This concept is also generalizable to other domains.

MixTAPE is a first step toward mixed-initiative project management. We look forward to a future in which deep system support for semi-structured note-taking, automated task synthesis, and task classification form the basis of any collaborative toolkit.

ACKNOWLEDGMENTS

Deploying these research concepts in industry required a significantly larger cast of characters than the author list on this paper does justice. The project managers and designers who served as our study participants, in addition to tirelessly working on amazing client outcomes, set aside time for experimentation and reflection. Our anonymous reviewers provided valuable feedback: their suggestions immensely improved the quality of the paper. Murat Jumashev and Aijan Junusheva were instrumental in the tricky implementation of *MixTAPE*. Ryan Langlois and Meredith Blumenstock synthesized and iterated on spectacular wireframes based on many rounds of input and feedback, and Meredith provided helpful feedback on our interview/survey designs. Shreya Bamne helped prototype the initial versions of *MixTAPE*'s classification models, and Daniel Haas helped frame our explanation of the models and model training. Vicki Borja, Megan Farmer, and Chris Basil served as critical subject matter experts. Daniela Retelny and Nitesh Banta provided high-level feedback and guidance throughout the project. We're grateful for all of their contributions to this work.

REFERENCES

[1] Elena Agapie, Lucas Colusso, Sean A Munson, and Gary Hsieh. 2016. Plansourcing: Generating behavior change plans with friends and crowds. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 119–133.

[2] Akiko Aizawa. 2003. An information-theoretic perspective of tf–idf measures. *Information Processing & Management* 39, 1 (2003), 45–65.

[3] David Allen. 2015. *Getting things done: The art of stress-free productivity*. Penguin.

[4] Teresa Amabile and Steven Kramer. 2011. *The progress principle: Using small wins to ignite joy, engagement, and creativity at work*. Harvard Business Press.

[5] Amazon. 2009. Alexa. https://developer.amazon.com/alexa/. [Online; accessed 20-September-2019].

[6] Apple. 2009. Siri. https://www.apple.com/ios/siri/. [Online; accessed 20-September-2019].

[7] Asana. 2008. Asana. https://asana.com/. [Online; accessed 20-September-2019].

[8] Asana. 2008. Asana Users. https://blog.asana.com/2018/09/asana-company-updates-2018/. [Online; accessed 20-September-2019].

[9] Ziv Bar-Yossef and Naama Kraus. 2011. Context-sensitive query auto-completion. In *Proceedings of the 20th international conference on World wide web*. ACM, 107–116.

[10] Web based task and time management. 2004. Remember the Milk. https://www.rememberthemilk.com/. [Online; accessed 20-September-2019].

[11] Victoria Bellotti, Brinda Dalal, Nathaniel Good, Peter Flynn, Daniel G Bobrow, and Nicolas Ducheneaut. 2004. What a to-do: studies of task management towards the design of a personal task list manager. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 735–742.

[12] Victoria Bellotti, Nicolas Ducheneaut, Mark Howard, and Ian Smith. 2003. Taking email to task: the design and evaluation of a task management centered email tool. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 345–352.

[13] Victoria Bellotti, Nicolas Ducheneaut, Mark Howard, Ian Smith, and Rebecca E Grinter. 2005. Quality versus quantity: E-mail-centric task management and its relation with overload. *Human–Computer Interaction* 20, 1-2 (2005), 89–138.

[14] Asa Ben-Hur, David Horn, Hava T Siegelmann, and Vladimir Vapnik. 2001. Support vector clustering. *Journal of machine learning research* 2, Dec (2001), 125–137.

[15] Aditya Bharadwaj, Pao Siangliulue, Adam Marcus, and Kurt Luther. 2019. Critter: Augmenting Creative Work with Dynamic Checklists, Automated Quality Assurance, and Contextual Reviewer Feedback. (2019).

[16] Daniel Boorman. 2001. Today's electronic checklists reduce likelihood of crew errors and help prevent mishaps. *ICAO Journal* (2001).

[17] Sasha P Caskey, Robert G Farrell, Dimitri Kanevsky, and Tara N Sainath. 2016. Automatically generating email subject lines. US Patent 9,356,889.

[18] Fanglin Chen, Kewei Xia, Karan Dhabalia, and Jason I Hong. 2019. MessageOnTap: A Suggestive Interface to Facilitate Messaging-related Tasks. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 575.

[19] Mia Xu Chen, Benjamin N Lee, Gagan Bansal, Yuan Cao, Shuyuan Zhang, Justin Lu, Jackie Tsay, Yinan Wang, Andrew M Dai, Zhifeng Chen, et al. 2019. Gmail smart compose: Real-time assisted writing. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2287–2295.

[20] Patrick Chiu, John S Boreczky, Andreas Girgensohn, and Don Kimber. 2001. LiteMinutes: an Internet-based system for multimedia meeting minutes. *WWW* 1 (2001), 140–149.
[21] Kenneth Ward Church and Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational linguistics* 16, 1 (1990), 22–29.
[22] ClickUp. 2016. ClickUp. https://clickup.com/. [Online; accessed 20-September-2019].
[23] Kenneth Conley and James Carpenter. 2007. Towel: Towards an Intelligent To-Do List.. In *AAAI Spring Symposium: Interaction Challenges for Intelligent Assistants*. 26–32.
[24] Richard C Davis, James A Landay, Victor Chen, Jonathan Huang, Rebecca B Lee, Francis C Li, James Lin, B CBM III, and MN Schleimer. 1999. Price, and B. *N. Schilit. NotePals: Lightweight Note Sharing by the Group, for the Group. CHI* (1999), 338–45.
[25] Forecast. 2016. Forecast. https://www.forecast.app/. [Online; accessed 20-September-2019].
[26] Werner Geyer, Heather Richter, and Gregory D Abowd. 2005. Towards a smarter meeting record-capture and access of meetings revisited. *Multimedia Tools and Applications* 27, 3 (2005), 393–410.
[27] Wael H Gomaa and Aly A Fahmy. 2013. A survey of text similarity approaches. *International Journal of Computer Applications* 68, 13 (2013), 13–18.
[28] Google. 2009. Google Assistant. https://assistant.google.com/. [Online; accessed 20-September-2019].
[29] Brigette Hales, Marius Terblanche, Robert Fowler, and William Sibbald. 2007. Development of medical checklists for improved quality of patient care. *International Journal for Quality in Health Care* 20, 1 (2007), 22–30.
[30] Patricia Hickey, Jean Anne Connor, Bethany Trainor, Marcie Brostoff, Richard Blum, Kathy Jenkins, and Eileen Stuart-Shor. 2012. Implementation of an organization-wide standardized communication initiative. *Journal of Communication in Healthcare* 5, 1 (2012), 32–39.
[31] Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 159–166.
[32] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. 2003. A practical guide to support vector classification. (2003).
[33] Facebook Inc. 2017. M Now Offers Suggestions to Make Your Messenger Experience More Useful, Seamless and Delightful. https://bit.ly/2o1M2PQ. [Online; accessed 20-September-2019].
[34] Google Inc. 2016. Google Allo—A smart messaging app. https://allo.google.com/. [Online; accessed 20-September-2019].
[35] Matthew Kam, Jingtao Wang, Alastair Iles, Eric Tse, Jane Chiu, Daniel Glaser, Orna Tarshish, John Canny, and John Canny. 2005. Livenotes: a system for cooperative and augmented note-taking in lectures. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 531–540.
[36] Anjuli Kannan, Karol Kurach, Sujith Ravi, Tobias Kaufmann, Andrew Tomkins, Balint Miklos, Greg Corrado, Laszlo Lukacs, Marina Ganea, Peter Young, et al. 2016. Smart reply: Automated response suggestion for email. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 955–964.
[37] Harmanpreet Kaur, Alex C Williams, Anne Loomis Thompson, Walter S Lasecki, Shamsi T Iqbal, and Jaime Teevan. 2018. Creating Better Action Plans for Writing Tasks via Vocabulary-Based Planning. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 86.
[38] S Sathiya Keerthi and B Ravindran. 1994. A tutorial survey of reinforcement learning. *Sadhana* 19, 6 (1994), 851–889.
[39] Nicolas Kokkalis, Thomas Köhn, Johannes Huebner, Moontae Lee, Florian Schulze, and Scott R Klemmer. 2013. Taskgenies: Automatically providing action plans helps people complete tasks. *ACM Transactions on Computer-Human Interaction (TOCHI)* 20, 5 (2013), 27.
[40] Thomas Kreifelts, Elke Hinrichs, and Gerd Woetzel. 1993. Sharing to-do lists with a distributed task manager. In *Proceedings of the Third European Conference on Computer-Supported Cooperative Work 13–17 September 1993, Milan, Italy ECSCW'93*. Springer, 31–46.
[41] Mirella Lapata. 2006. Automatic evaluation of information ordering: Kendall's tau. *Computational Linguistics* 32, 4 (2006), 471–484.
[42] Susan Leigh Star. 2010. This is not a boundary object: Reflections on the origin of a concept. *Science, Technology, & Human Values* 35, 5 (2010), 601–617.
[43] Germán Leiva, Nolwenn Maudet, Wendy Mackay, and Michel Beaudouin-Lafon. 2019. Enact: Reducing Designer–Developer Breakdowns When Prototyping Custom Interactions. *ACM Transactions on Computer-Human Interaction (TOCHI)* 26, 3 (2019), 19.
[44] Kevin Lund and Curt Burgess. 1996. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior research methods, instruments, & computers* 28, 2 (1996), 203–208.
[45] Kurt Luther, Jari-Lee Tolentino, Wei Wu, Amy Pavel, Brian P. Bailey, Maneesh Agrawala, Björn Hartmann, and Steven P. Dow. 2015. Structuring, Aggregating, and Evaluating Crowdsourced Design Critique. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 473–485. https://doi.org/10.1145/2675133.2675283

[46] Gloria Mark, Victor M Gonzalez, and Justin Harris. 2005. No task left behind? Examining the nature of fragmented work. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 321–330.
[47] MeetBot. 2017. Debian. https://wiki.debian.org/MeetBot/. [Online; accessed 20-September-2019].
[48] Microsoft. 2009. Cortana. https://developer.amazon.com/alexa/. [Online; accessed 20-September-2019].
[49] Microsoft. 2017. Teams. https://www.microsoft.com/en-us/microsoft-365/microsoft-teams/group-chat-software. [Online; accessed 20-September-2019].
[50] Karen Myers, Pauline Berry, Jim Blythe, Ken Conley, Melinda Gervasio, Deborah L McGuinness, David Morley, Avi Pfeffer, Martha Pollack, and Milind Tambe. 2007. An intelligent personal assistant for task and time management. *AI Magazine* 28, 2 (2007), 47–47.
[51] Björn Nagel, Daniel Böhnke, Volker Gollnick, Peter Schmollgruber, Arthur Rizzi, Gianfranco La Rocca, and Juan J Alonso. 2012. Communication in aircraft design: Can we establish a common language. In *28th International Congress of the Aeronautical Sciences*. 1–13.
[52] Michelle O'Daniel and Alan H Rosenstein. 2008. Professional communication and team collaboration. In *Patient safety and quality: An evidence-based handbook for nurses*. Agency for Healthcare Research and Quality (US).
[53] Stephen T Peverly, Vivek Ramaswamy, Cindy Brown, James Sumowski, Moona Alidoost, and Joanna Garner. 2007. What predicts skill in lecture note taking? *Journal of Educational Psychology* 99, 1 (2007), 167.
[54] Annie Piolat, Thierry Olive, and Ronald T Kellogg. 2005. Cognitive effort during note taking. *Applied cognitive psychology* 19, 3 (2005), 291–312.
[55] S Rasoul Safavian and David Landgrebe. 1991. A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics* 21, 3 (1991), 660–674.
[56] Ben Shneiderman and Pattie Maes. 1997. Direct manipulation vs. interface agents. *interactions* 4, 6 (1997), 42–61.
[57] Milad Shokouhi. 2013. Learning to personalize query auto-completion. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 103–112.
[58] Slack. 2013. Slack. https://slack.com/. [Online; accessed 20-September-2019].
[59] Susan Leigh Star. 1989. The structure of ill-structured solutions: Boundary objects and heterogeneous distributed problem solving. In *Distributed artificial intelligence*. Elsevier, 37–54.
[60] Tracks. 2009. Remember the Milk. http://www.getontracks.org/. [Online; accessed 20-September-2019].
[61] Trello. 2011. Trello. https://trello.com/. [Online; accessed 20-September-2019].
[62] Bernard L Welch. 1947. The generalization ofstudent's' problem when several different population variances are involved. *Biometrika* 34, 1/2 (1947), 28–35.
[63] Steve Whittaker. 2005. Supporting collaborative task management in e-mail. *Human–Computer Interaction* 20, 1-2 (2005), 49–88.
[64] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In *Proceedings of the 2020 chi conference on human factors in computing systems*. 1–13.
[65] Alexander Seeshing Yeung, Putai Jin, and John Sweller. 1998. Cognitive load and learner expertise: Split-attention and redundancy effects in reading with explanatory notes. *Contemporary educational psychology* 23, 1 (1998), 1–21.
[66] Jeffrey M Zacks and Barbara Tversky. 2001. Event structure in perception and conception. *Psychological bulletin* 127, 1 (2001), 3.
[67] Jeffrey M Zacks, Barbara Tversky, and Gowri Iyer. 2001. Perceiving, remembering, and communicating structure in events. *Journal of experimental psychology: General* 130, 1 (2001), 29.
[68] 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), 196.

Received January 2020; revised June 2020; accepted July 2020