

# “Merging Results Is No Easy Task”: An International Survey Study of Collaborative Data Analysis Practices Among UX Practitioners

Emily Kuang  
School of Information  
Rochester Institute of Technology  
Rochester, New York, USA  
ek8093@rit.edu

Xiaofu Jin  
Computational Media and Arts  
The Hong Kong University of Science  
and Technology (Guangzhou)  
Guangzhou, China  
xjinao@connect.ust.hk

Mingming Fan\*  
Computational Media and Arts  
The Hong Kong University of Science  
and Technology (Guangzhou)  
Guangzhou, China  
Integrative Systems and Design &  
Computer Science and Engineering  
The Hong Kong University of Science  
and Technology  
Hong Kong SAR, China  
mingmingfan@ust.hk

## ABSTRACT

Analysis is a key part of usability testing where UX practitioners seek to identify usability problems and generate redesign suggestions. Although previous research reported how analysis was conducted, the findings were typically focused on *individual analysis* or based on a small number of professionals in specific geographic regions. We conducted an online international survey of 279 UX practitioners on their practices and challenges while *collaborating during data analysis*. We found that UX practitioners were often under time pressure to conduct analysis and adopted three modes of collaboration: independently analyze different portions of the data and then collaborate, collaboratively analyze the session with little or no independent analysis, and independently analyze the same set of data and then collaborate. Moreover, most encountered challenges related to lack of resources, disagreements with colleagues regarding usability problems, and difficulty merging analysis from multiple practitioners. We discuss design implications to better support collaborative data analysis.

## CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing**.

## KEYWORDS

User experience; UX; Usability testing; Data analysis; Collaboration; Survey

\*Corresponding Author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](https://permissions.acm.org).

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-9157-3/22/04...\$15.00

<https://doi.org/10.1145/3491102.3517647>

## ACM Reference Format:

Emily Kuang, Xiaofu Jin, and Mingming Fan. 2022. “Merging Results Is No Easy Task”: An International Survey Study of Collaborative Data Analysis Practices Among UX Practitioners. In *CHI Conference on Human Factors in Computing Systems (CHI '22)*, April 29-May 5, 2022, New Orleans, LA, USA. ACM, New York, NY, USA, 16 pages. <https://doi.org/10.1145/3491102.3517647>

## 1 INTRODUCTION

Usability testing is a frequently employed user-centred design method for detecting usability problems [25, 110]. Despite being effective and useful, analyzing usability test sessions can be tedious, challenging, and time-consuming [12, 25, 29, 81]. When analyzing usability test sessions, user experience (UX) practitioners often need to attend to multiple behavioral signals in both the visual and audio channels of the recordings and conduct several tasks (e.g., look at user’s actions, listen to user’s verbalizations, and write their own annotations) simultaneously [12]. Furthermore, UX practitioners working in industry often face time pressure to deliver their analysis results [25, 29, 64]. Consequently, it is not uncommon that UX practitioners might miss important usability problems or misinterpret them [26, 38].

To address these challenges, UX practitioners are recommended to analyze data *collaboratively* to increase the completeness and reliability [25, 29, 30]. By bringing different perspectives from multiple practitioners together, collaboration could help overcome the “evaluator effect” [38]—a common phenomenon where different UX practitioners may uncover or interpret usability problems and their severity levels differently. Unfortunately, fewer than 30% of UX practitioners have a chance to collaborate with others to analyze the same usability test session due to practical constraints, such as limited resources [29]. Even if they do collaborate, they often have to put together an ad-hoc set of tools to support their analysis of test sessions and collaboration [25, 64].

To better understand how UX practitioners work in practice, researchers conducted various studies in the past decades [25, 30, 64, 81, 93]. While informative, these studies were typically based on small data samples (e.g., interview with eleven UX practitioners

[30]) or focused on the usage of specific methods, such as think-aloud protocols [25, 64]. For the few studies that did focus on general practices of UX practitioners (e.g., [29]), they offered limited insights into collaboration practices and purposes around data analysis among UX practitioners. Furthermore, almost all previous studies were conducted a decade ago except Fan et al.'s study [25], which was focused on think-aloud protocol usage instead of general practices. In the most recent decade, new technology breakthroughs, such as deep learning and virtual reality (VR), have spurred a variety of AI-enabled products (e.g., smart speakers) and VR applications. These emerging products enable new interaction methods and provide novel user experiences, which might require UX practitioners to adapt their analysis and collaboration practices. Indeed, new UX analytical platforms (e.g., UserTesting.com [105] and FullStory [32]) and data-informed design and analytical tools (e.g., MixPanel [70]), have also emerged in the meantime. As a result, there is a lack of understanding of how UX practitioners collaborate when analyzing usability test sessions in today's social and technological contexts. In addition, it is important to gain an understanding of the practices used during independent analysis as it lays the foundation for collaboration [30]. We take a first step to address this gap by exploring the following research questions (RQs):

- RQ1: What are the independent data analysis practices of UX practitioners?
- RQ2: What are the practices, challenges, and desired improvements for collaborative data analysis of UX practitioners?
- RQ3: How might UX practitioners' data analysis and collaboration practices be affected by their years of experience and team size?

To answer these RQs, we conducted an online survey study with a representative sample of 279 UX professionals who had different levels of UX experience from six continents in the world. The survey focused on the resources they used when analyzing usability test recordings and how they collaborated with fellow UX team members. The survey also collected new features that they hoped to have in new collaborative tools to better support their collaboration in the data analysis phase.

Our results show that about two thirds (66%) of UX practitioners were under time pressure to conduct analysis and most of them utilized structured formats and severity ratings when describing usability problems. The top three purposes of collaboration were to identify more usability problems, generate redesign suggestions, and improve reliability respectively. We also found that respondents with greater years of UX experience were more likely to consider the improvement of reliability when collaborating. Their collaboration happened in a variety of circumstances and most of them had encountered challenges related to lack of time, disagreements with colleagues regarding usability problems, and difficulty merging analysis from multiple practitioners. Based on the challenges that respondents encountered, we derive some design considerations that may benefit both UX practitioners and researchers by suggesting target areas for collaboration improvement and pointing out avenues for future research.

In sum, we make the following contributions in this work:

- We present a quantitative understanding of UX practitioners' collaboration practices and challenges based on a sample of 279 international UX professionals;
- We show potential correlations between their years of UX experience and team size with their analysis practices;
- We highlight design considerations for improving data analysis and collaboration among UX practitioners.

## 2 BACKGROUND AND RELATED WORK

Our work is motivated and informed by prior work on *UX analysis practices*, and *collaborative data analysis among UX practitioners*.

### 2.1 Analysis of Data Collected from Usability Test Sessions

Usability is defined as the extent to which a system enables users to achieve goals effectively, efficiently and with satisfaction, based on the intended context of use [43]. To improve system usability, UX practitioners must identify and address usability problems that users may encounter. Examples of usability evaluations include observations, cognitive walkthroughs, interviews, focus groups, heuristic evaluations, and usability testing [79]. The most frequently employed method for detecting such problems is through usability testing [25]. This is due to strong evidence that usability testing is effective in a wide variety of usage scenarios [81].

While usability testing has been shown to be effective and useful across various methods and domains, conducting sound usability testing is challenging. In addition to the challenges practitioners face during data collection [22, 81], they need to use a variety of tools to record the session for later analysis. All participants in an exploratory study on usability testing reported they video-record the session and take notes during the session [30]. Similarly, in other surveys of UX practitioners, 100% [29] and 89% [25] of the participants take observations notes, 73% [29] and 77% [25] of the participants video-record the session, and 70% [25] conduct post-task interviews.

During analysis, UX practitioners use these diverse sources of data to generate coherent descriptions of usability problems, which include possible causes, effects, and solutions [30], following both light-weight and formal analysis procedures. The Instant Data Analysis (IDA) technique is a light weight approach requiring practitioners to perform quick analysis immediately upon the conclusion of a usability testing session [50]. On the other hand, the User Action Framework (UAF) [2] and the Structured Usability Problems EXtraction (SUPEX) framework [15] describe formal standards for analysis. What's more, templates [14, 54] and guidelines [11] have been developed for organizing problem descriptions, prioritizing the problems (also known as the severity rating [58]), and deriving potential interventions. However, in practice, the analysis is often unstructured, incomplete with no identification of causes or solutions [81], and is informal and lack rigor [30]. Even if practitioners use structured formats, they tend to use homegrown formats developed by themselves or their companies [29].

In sum, there seems to be a gap between recommended guidelines and practices for UX data analysis. This might induce vague evaluation procedures that lead to bias or unclear problem criteria, which result in an overflow of reported issues [38]. As new methods

and tools have emerged since the majority of the studies regarding analysis practices were conducted a decade ago (e.g., [29, 30, 64, 81]), we seek to understand the current practices adopted by UX practitioners when performing independent data analysis, which is usually done prior to collaborating with other practitioners [30].

## 2.2 Collaborative Data Analysis among UX Practitioners

While collaboration during usability testing happens under various circumstances, such as between a group of participants [19], between UX practitioners and stakeholders [31, 96], and between UX practitioners and experts of complex domains [12], this paper addresses collaboration in the context of UX practitioners working with their colleagues during the data analysis stage.

**The Need for Collaboration.** Analyzing usability test recordings is challenging and time-consuming [12, 25, 29, 81]. UX practitioners need to pay attention to both the visual and audio channels of the recording while observing user’s actions and writing notes simultaneously [12]. However, in practice, they only have limited time and resources to complete analysis [29]. As a result, they have to juggle between achieving high reliability and validity in their analysis and completing their analysis efficiently [25]. One important way to cope with the tension is to collaborate with other UX practitioners to review test sessions together [25], which not only divides the workload but is also perceived to improve the thoroughness of problem detection [92].

Another benefit of collaboration is alleviating a persistent issue encountered by UX practitioners in UX data analysis—the “*evaluator effect*,” which refers to the fact that different practitioners tend to identify different sets of UX problems even when analyzing the same test session [38, 44]. This issue exists for novice and experienced practitioners, for problem detection and severity assessment, and for simple and complex systems [38]. The majority of practitioners indeed perceived the “*evaluator effect*” when merging their individual findings into group evaluations [39]. Collaboration among UX practitioners enables them to view the data from different perspectives, which could lead to an increase in both the reliability [38] and thoroughness of the problems identified [92]. A study has shown that practitioners working in isolation identified significantly fewer usability problems than a team of three to five evaluators [78]. In other words, collaboration allows a higher possibility that all the usability problems in a session can be found and that the analysis is more consistent across practitioners.

**Collaboration Practices.** Although many UX practitioners recognized the value of collaboration in improving reliability, few collaborated with others when analyzing the same test recording in practice [29]. In an exploratory study with eleven UX practitioners, only three reported having some form of collaboration [30]. More recently, a survey of 197 UX practitioners found that more than half (56%) of the participants analyzed data and wrote informal reports *alone* and 42% also analyzed data and wrote formal reports *alone* [25]. One important reason for limited collaboration is that UX practitioners often work on projects that have quick turnaround cycles or limited resources [12, 25, 64].

Another key part of collaboration between UX practitioners is the use of software tools to facilitate session review and analysis.

Collaborative software has been shown to support heuristic evaluation in groups, where it allowed groups of evaluators to reduce the number of duplicate usability problems, reach consensus earlier, and improve productivity [62]. Usability testing in general also benefits from collaborative software support. However, this software mainly served as a collaborative writing tool and did not provide any session review functionalities [61]. Moreover, it required the use of an external meeting software for distributed collaboration, which demonstrates the lack of *integrated* tools that support both analysis and communication among UX practitioners.

Numerous commercial tools are also used by UX practitioners in industry. Although *offline tools*, such as Morae [100], Noldus Viso [80], and Silverback [95], can be installed on a local machine and allow for reviewing sessions with functionalities like note-taking and marking events on video progress bar, they do not support collaboration well. In contrast, *online tools* (e.g. UserTesting [105] and FullStory [32]) allow for more flexible collaboration, where multiple UX practitioners can gain access to the same session recordings. However, such tools provide limited data analysis capabilities, which are comprised of session playback, note-taking, and tagging, and practitioners still need to communicate through an external platform to discuss their analysis results. Consequently, UX practitioners often have to put together a generic set of tools on an ad hoc basis that are not designed specifically for their needs [25, 64], such as screen recorders, eye-trackers, prototyping tools, web analytics tools, spreadsheets, text editors, and presentation tools [29], to collaboratively capture and analyze various sources of data.

In sum, despite the necessity of collaboration, UX practitioners face many challenges for collaboration in practice. Although previous research reported how UX practitioners conducted analysis on user evaluations in general [7, 25, 29, 30], the findings did not provide specific insights into practices and factors affecting collaboration and its associated challenges. Therefore, to help identify opportunities to improve collaboration among UX practitioners, we conducted a survey study to first understand UX practitioners’ practices during independent analysis (RQ1), then their practices, challenges, and desired improvements for collaborative analysis in detail (RQ2), and finally how their years of experience and team size might affect analysis and collaboration practices (RQ3).

## 3 METHOD

We employed a web-based questionnaire survey to reach a broader set of UX practitioners globally since analysis practices are expected to vary among practitioners. This method allowed us to get a diverse sample of respondents at different stages of their career across six continents. Our choice of method is in line with previous survey studies of usability practices [25, 30, 64]. Below we present details about our survey design, recruitment methods, and respondents.

### 3.1 Survey Design

To address the research questions, the survey was separated into three sections: (1) current independent data analysis practices, (2) practices and challenges of collaboration, and (3) new features for collaboration tools to assist with data analysis. The questions sought to understand background data (practitioner’s work experience and evaluation context), practices and tools used for usability

**Table 1: Summary of respondent demographics (N = 279)**

| Gender                  |             | Size of UX Team           |             |
|-------------------------|-------------|---------------------------|-------------|
| Female                  | 141 (50.5%) | 1 person                  | 84 (30.1%)  |
| Male                    | 115 (41.2%) | 2 - 5 people              | 148 (53.1%) |
| Prefer not to say       | 22 (7.9%)   | 6 - 10 people             | 26 (9.3%)   |
| Prefer to self-describe | 1 (0.4%)    | 11 and above              | 21 (7.5%)   |
| Work Location           |             | Years of Experience in UX |             |
| North America           | 119 (42.7%) | Less than 1 year          | 77 (27.6%)  |
| Asia                    | 70 (25.1%)  | 1-2 years                 | 74 (26.5%)  |
| Europe                  | 64 (22.9%)  | 3-5 years                 | 69 (24.7%)  |
| South America           | 21 (7.5%)   | 6-9 years                 | 30 (10.8%)  |
| Africa                  | 3 (1.1%)    | 10 or more years          | 29 (10.4%)  |
| Oceania                 | 2 (0.7%)    |                           |             |

problem identification, and perceived challenges and desired improvements to support collaboration practices. The majority of survey questions were fixed response items, such as single choice, multiple-choice (i.e. select-all-that-apply), and Likert-scale ratings, which were required. To supplement the fixed responses and gain a deeper understanding, there were eight optional free text questions that sought explanations for why practitioners chose certain responses. These questions also asked about what are the greatest challenges practitioners encounter, how they handle disagreements with colleagues, and how their collaboration could be better supported. For the free response questions, the average length of the answers was three sentences. They were first coded independently by two authors before meeting to discuss and group them into themes. The questions in sections (2) and (3) regarding collaboration practices and challenges were not shown to respondents who answered "1 person (just myself)" for the size of UX team since those questions were not applicable to their work experiences.

### 3.2 Testing the Survey Tool

To pretest the survey, We followed Dillman's suggested three stage process [20]. First, the survey was reviewed by colleagues to uncover potential misinterpretations and additional questions that may have been overlooked. Next, we discussed the survey plan with UX practitioners to ensure the sufficient motivation for the survey as well as communication clarity. Finally, we performed pilot-testing with four UX practitioners to identify any flaws in the survey questions and distribution platform to ensure that the length is appropriate.

### 3.3 Respondent Recruitment

The survey was approved by the authors' institutional review board (IRB) and was deployed in late 2020. It was conducted through Qualtrics and was accessed by all respondents through an anonymous link. The respondents were required to be UX practitioners with prior experience conducting usability evaluations. The respondents were recruited through several channels; invitations were distributed via mailing lists of local UXPA (i.e., User Experience Professional Association) chapters around the world (which requires qualification verifications for their members), UX Professional groups on LinkedIn, and industry connections. As incentives to participate in the survey, all respondents who recorded their

e-mail address could get early access to the project results and were included in a lottery for \$30.

### 3.4 Data Quality

We took multiple precautions to ensure data quality. We provided plain descriptions of terms to explain the options, such as "collaborative writing: team members can make annotations on same recording". Respondents were informed that data collection was fully anonymous to avoid bias. Respondents who chose to leave their email address for the lottery were assured that the e-mail addresses would be separated from the questionnaire data. We also discarded certain responses from the total 309 attempts of the questionnaire to ensure data quality. We excluded 30 responses for leaving after the first few questions (26), declining to provide consent (3), and selecting "Strongly Agree" across all Likert scales, which suggests that each statement was not read carefully (1).

### 3.5 Respondent Demographics

We received a total of 279 valid responses. Average survey completion time was 6.8 minutes ( $Md^1 = 6.2$ ). Table 1 shows a summary of respondents' demographics and work experience. They worked across six continents with the largest number of respondents in North America, which was likely because it shared the largest number of companies and was where the survey was distributed from. Nonetheless, there were reasonable numbers of respondents in Asia and Europe. One limitation was the small number of responses from other continents, which will be discussed in Sec. 6.

We received a near equal distribution for respondents with less than 1 year, 1-2 years, and 3-5 years of prior work experience, making up a total of 79% with 5 years or less of experience working in UX. Over half (53%) of the respondents worked on teams consisting of 2-5 people, with only 17% reporting they worked with teams greater than 5 people. Interestingly, 30% of the respondents were the only UX practitioner on their team.

## 4 FINDINGS

In the process of analyzing usability testing sessions, it is typical that individual analysis is done prior to collaboration [30]. Thus, it is important to gain an understanding of the practices used in

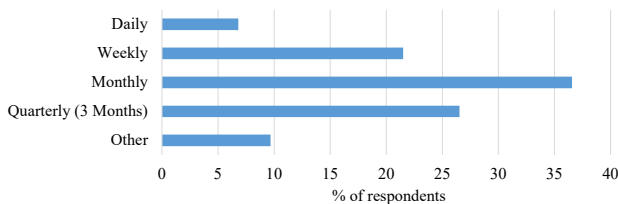
<sup>1</sup>Md = median

the individual portion. In this section, we will first discuss their independent analysis practices (RQ1), then we will discuss in greater detail their collaboration (RQ2), and correlations with prior UX experience and team size (RQ3).

#### 4.1 RQ1: Independent Data Analysis Practices

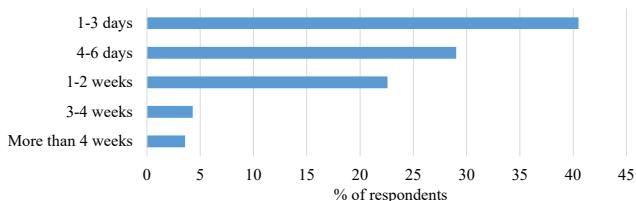
Our analysis revealed two themes about individual data analysis practices: *frequency and time spent on usability testing* and *resources used for usability problem identification*.

**4.1.1 Frequency and Time Spent on Usability Testing.** Figure 1 shows that the highest portion of the respondents conducted usability testing monthly (37%), followed by quarterly (27%), and weekly (22%). Among the 24 respondents who chose “Other”, over half (59%) reported that the frequency varies depending on the schedule of the projects and that some projects may need multiple rounds of testing.



**Figure 1: Frequency of conducting usability testing** ( $N = 279$ , *single – select*)

Figure 2 shows that 41% of the respondents only had *1-3 days* to analyze data collected from usability testing sessions and 29% of them had *4-6 days*. Since the majority (70%) of the respondents needed to complete their analysis within one week of the usability testing session, there is significant time pressure on most UX practitioners. Indeed, a similar proportion (66%) of respondents “somewhat agreed” or “strongly agreed” that they found it *time-consuming to conduct data analysis* ( $Md=4$ ,  $IQR^2=1$ ).

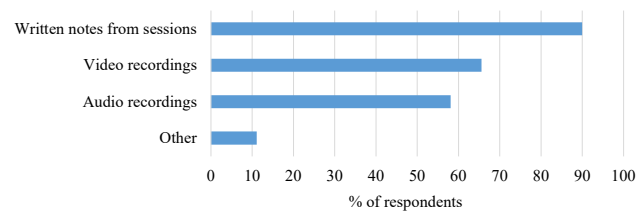


**Figure 2: Time spent on analyzing data collected from usability testing sessions** ( $N = 279$ , *single – select*)

**4.1.2 Resources Used for Usability Problem Identification.** Figure 3 shows that the majority (90%) of the respondents reviewed written notes from sessions (either their own or their team members’ observational notes), 66% of them reviewed video recordings, and

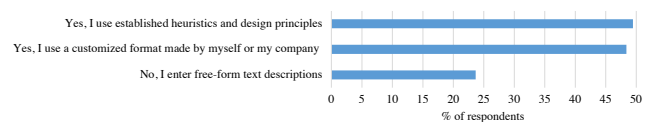
58% of them reviewed audio recordings. 11% of the respondents reported other resources including automated transcripts, qualitative data from interviews, testing software (e.g., UserTesting [104], UserZoom [106]), heatmaps and polls, and periodic industry reports.

This is in line with the findings of a 2020 international survey showing that 89% of respondents reviewed observation notes and 77% reviewed session recordings (both audio and video) [25]. However, an earlier interview study from 2010 found that the majority of respondents relied mainly on their notes and memory for analysis instead of doing a full analysis of recordings [30]. This suggests that more UX practitioners use recordings more often nowadays than before, which could be due to recent advancements in technology and remote usability testing. However, written notes remain the most used method, highlighting the importance in the quality of the original notes from sessions.



**Figure 3: Resources reviewed when conducting data analysis** ( $N = 279$ , *multi – select*)

Figure 4 shows that almost half (49%) of the respondents used established heuristics and design principles. Similarly, almost half (48%) followed a customized format. Since respondents were allowed to select multiple options, they could have used a combination of established and customized formats as well.

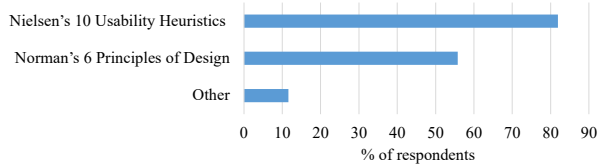


**Figure 4: Structured formats used when describing identified problems** ( $N = 279$ , *multi – select*)

Specifically, as shown in Figure 5, Nielsen’s 10 usability heuristics [77] were the most popular among the respondents with the majority (82%) of them using it when detecting and describing problems. Over half (56%) of them used Norman’s 6 design principles [82]. Furthermore, respondents also reported using other UX evaluation guidelines and standards, including the International Usability and UX Qualification Board [107], the McCUE questionnaire for UX assessment [68], the Gestalt Principles of Design [51], heuristics for mobile computing [6], and the ISO 9241-110 Standard for Ergonomics of Human-System Interaction [43].

When using these standard heuristics, principles, and guidelines, respondents also modified them to cater to their companies’ needs. The free responses revealed their reasons, which included adding more context to the heuristics to make it clearer to non-experts. Many of them did not change the core meaning of the heuristics but

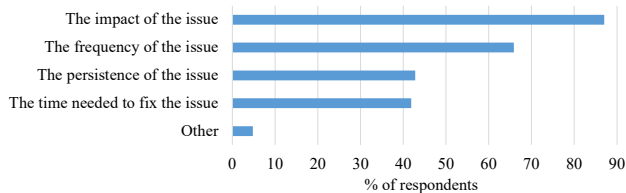
<sup>2</sup>IQR = inter-quartile range, which is the difference between the 75th and 25th percentiles



**Figure 5: Usage of established heuristics or design principles** ( $N = 138$ , *multi – select*)

rather the language, which allowed them to better communicate the findings to stakeholders and other audiences.

In addition to identifying UX problems, most (75%) of the respondents rated the severity of the problems. Since assigning severity classifications was suggested by prior work [30], our findings showed the different factors that respondents considered while doing so (see Figure 6). Specifically, the majority (87%) of them considered “the impact of the issue”, followed by “the frequency of the issue” (66%), “the persistence of the issue” (43%), and “the time needed to fix the issue” (42%). Moreover, a small percentage of respondents (5%) reported also considering the potential cost to fix the issue, the scope of the issue, and whether it had any relation to previously uncovered problems.

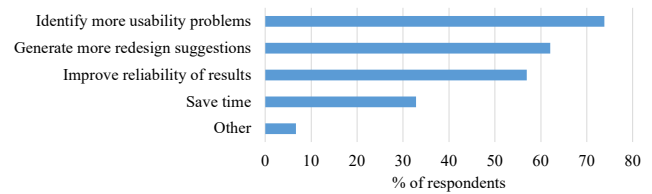


**Figure 6: Factors considered when assigning severity levels** ( $N = 208$ , *multi – select*)

## 4.2 RQ2: Collaborative Data Analysis Practices, Challenges, and Desired Improvements

**4.2.1 Collaboration Purposes and Circumstances.** The top three purposes for collaboration, as shown in Figure 7, were to *identify more usability problems* (74%), *generate more redesign suggestions* (62%), and *improve reliability of results* (47%). This finding suggests that respondents were aware that collaboration could increase the completeness and reliability of their analysis, which would be beneficial to address the “evaluator effect” [38].

In this survey, *improving reliability of results* was only the third-frequently chosen option which is in contrast to prior findings of this being the top driver for collaboration [29]. One explanation for this could stem from the format of the question. In the prior survey, it was a single-choice question where 47% of respondents selected *improving reliability of results*. In this survey, it was selected by over half (57%) of the respondents. This result suggests that most of respondents did pay attention to the reliability of their analysis. However, it was less frequently considered as compared to more result-oriented purposes such as increasing the quantity of

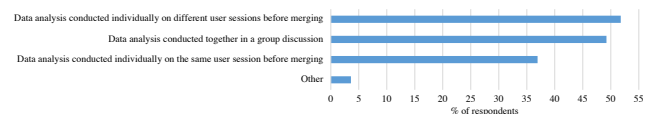


**Figure 7: Purposes of collaboration between UX practitioners** ( $N = 195$ , *multi – select*)

usability problems and redesign suggestions. Additionally, respondents provided free-form responses in terms of their purposes for collaboration, and the top responses were to “save money”, “consolidate paper work”, and “ensure everyone’s voices are heard.” The awareness of the importance of collaboration was demonstrated by respondents who mentioned that *different perspectives are critical to identifying as many issues and potential solutions as possible*.

As shown in Figure 8, most of the collaboration between colleagues happened when they conducted data analysis *individually on different user sessions before merging* (52%). This resembles a *divide and conquer* approach and allows the UX practitioners to save time as it reduces the workload for a single practitioner. The second most common approach was to conduct analysis *together in a group discussion* (49%). Furthermore, over half (52%) of these discussions took place *directly after* the usability testing session, 46% took place *within a week*, and only 2% took place *more than a week after*. This is in line with our earlier finding that the majority (70%) of respondents only have less than a week to analyze data from usability testing (see Figure 2).

In contrast, only 37% of the respondents conducted data analysis *individually on the same user session before merging*. This phenomenon was echoed in the free-form responses as many respondents mentioned that even though they worked on a team with other UX practitioners, only one person was responsible for analyzing all usability testing sessions for a single product in the company.

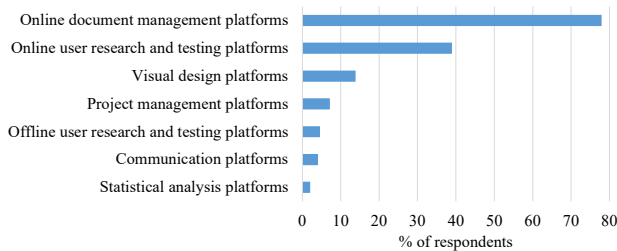


**Figure 8: Circumstances of collaboration between UX practitioners** ( $N = 195$ , *multi – select*)

Another vital factor to successful collaboration is having effective communication between colleagues [48, 98]. Thus, we asked respondents what type of communication they use most often to discuss work-related content with their colleagues. Almost three-quarters (73%) of the respondents chose synchronous methods such as meetings or phone calls, whereas 27% use asynchronous methods such as emails. Indeed, their choice of communication method is in line with their preference as the majority (68%) selected “some-what agree” or “strongly agree” to the prompt “I prefer to talk to my colleague directly rather than type through chat” ( $Md=4$ ,  $IQR=2$ ). In their explanations, respondents mentioned that “there is always a

need for synchronous communication in between asynchronous discussions” and that “it’s hard to put together a presentation or align on findings without real-time discussions.”

**4.2.2 Platforms that Support Collaboration.** The respondents used a total of 31 different platforms for collaboration while conducting data analysis of user sessions, which were grouped into various categories (see Figure 9).



**Figure 9: Different categories of platforms used for collaboration while conducting data analysis ( $N = 195$ , multi-select)**

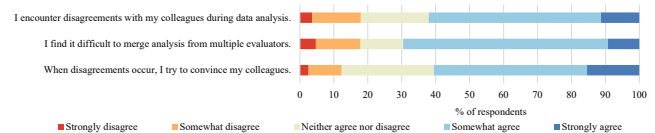
The majority (78%) of respondents used *online document management platforms* which included Google Workspace (e.g., Docs, Sheets, Slides, and Forms) [33] and Microsoft SharePoint [66]. The next popular category was *online user research and testing tools*, which was used by 39% of the respondents. Examples include UserTesting [105], UserZoom [106], Useberry [103]. These platforms provide functionalities like note-taking and marking events on the video progress bar on top of basic usability test support such as screen recording, survey administration, and results exporting. Respondents often used combinations of these top two categories. For example, one respondent described that they used Google Slides and Sheets to collaborate during data analysis since it was fast and real-time, but still found it easier to share a link to the UserTesting study when reviewing videos and participant data.

Although *offline user research and testing tools* offer similar functionalities, they do not support flexible collaboration. For example, it is difficult for multiple UX practitioners to gain access to the same session recordings when stored locally. In fact, only 5% of the respondents currently used offline platforms, such as Morae [100], which will likely continue to decrease as online remote user testing platforms gain more popularity [74]. Other categories include *visual design platforms* such as Miro [69], Mural [42], and Figma [28] (14%), *project management platforms* such as Jira [4], Confluence [3], and Trello [5] (7%), and *communication platforms* such as Microsoft Teams [67] and Zoom [116] (4%).

Respondents had mixed feelings about the use of many tools to support their collaborative data analysis. On the one hand, respondents mentioned that they like having the option of many tools as it gives them the choice to use something that they are comfortable with. On the other hand, respondents felt that “if there are too many tools, it can get confusing especially if there isn’t a main one the team has agreed to rely on most.”

**4.2.3 Challenges of Collaboration.** As shown in Figure 10<sup>3</sup>, respondents encountered various challenges during collaboration. Almost two-thirds (62%) of respondents encountered situations where they disagreed with their colleagues during data analysis (Md=4, IQR=1). Moreover, finding it difficult to merge analysis from multiple evaluators was also considered as a challenge by over two-thirds (70%) of respondents (Md=4, IQR=1). When disagreements occurred, 61% of them would make a case for their decisions and try to convince their colleagues (Md=4, IQR=1).

Regarding their approaches to managing disagreements, most of the respondents (82%) would *have a discussion in-person or through video/voice call* and 33% would *have a discussion through chat*. Other strategies include *involving another person as the mediator* (24%) and *asking another person to be the tie-breaker* (23%). Some respondents provided additional strategies in the free responses, which includes *approaching management or a supervisor with both sets of findings and asking them to make the decision*. One respondent stated that they rarely disagreed on a usability problem, but rather the priority of a problem. When this was the case, they “refer[red] to the main goal of the study and then reflect[ed] on what the client would find most helpful and actionable.” Another response was creating an anonymous board to list down the issues and then voting on it at the next group meeting. In the case where a designer was part of the discussion, one respondent mentioned they would let the designer of the test product make the final call. Interestingly, one respondent wrote “have manager make decision to go with my advice over the others”, which suggests that they would try to convince their manager to listen to their advice first. If they are under time pressure and still can’t reach an agreement on a certain issue, respondents mentioned two strategies in particular: (1) asking the leader to make the final decision, and (2) presenting their results with an explanation that covers both interpretations since “most likely one person isn’t wrong, it is simply that the interpretations are based on different viewpoints.”



**Figure 10: Likert scale rating of challenges encountered when conducting analysis ( $N = 195$ )**

The responses for the greatest challenges that UX practitioners have encountered when collaborating on data analysis were analyzed and grouped into the following themes:

#### Data-related challenges:

- (1) *Data storage and access issues:* Respondents found it difficult to store all the data and analysis associated with usability testing in “a visually organized, quick and easily accessible way for follow-up discussions” with colleagues.
- (2) *Data analysis challenges:* Respondents mentioned some aspects of data analysis were especially tedious or difficult,

<sup>3</sup>This divergent color scheme was based on an established color-blind friendly palette [101].

**Table 2: Analysis and collaboration practices that have significant differences for varying years of UX experience**

| Collaboration practices                                    |   |
|--|---|
| Found it time consuming to conduct data analysis           | $\chi^2(16) = 32.1, N = 279, p < .01, V = .17 **$ |
| Used a customized problem description format               | $\chi^2(4) = 15.2, N = 279, p < .01, V = .23 **$  |
| Collaboration purpose is to improve reliability of results | $\chi^2(4) = 11.6, N = 279, p < .05, V = .20*$    |

such as “clean[ing] up software produced transcripts”, “conduct[ing] qualitative analysis of text”, “setting the frame for insight extraction”, “synthesizing findings into actionable recommendations”, and “remember[ing] a specific point of time after the session has ended”.

#### Coworking and communication challenges:

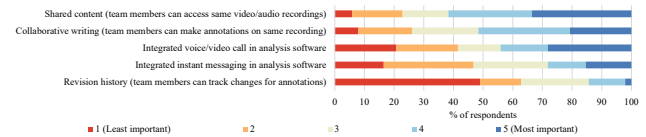
- (1) *Inconsistent coding strategies*: Respondents had difficulty reaching agreement on what codes or labels the transcript should be used to analyze texts. Respondents felt that “it’s not always clear how to analyze and it’s hard to change in the middle.” It was also difficult to “merge data analysis from many researchers and notify others about the updates”.
- (2) *Contrasting priorities*: Respondents struggle to convince colleagues to have a “problem-first” mentality instead of rushing to a solution without careful research and analysis. Another respondent even had difficulty “getting everyone in the room to listen to the analysis due to different priorities and values between colleagues and clients”.
- (3) *Mismatch of experience*: Respondents were not confident in their colleagues’ capabilities to analyze data. The team consisted of varying levels of “maturity when it comes to UX understanding”, leading to inefficiencies in collaboration such as *extra time [needed] to set the frame for analysis*.
- (4) *Biased feedback*: Respondents found that getting honest opinions from their team was sometimes a challenge as their colleagues were “biased to give positive feedback”.

#### Environmental challenges:

- (1) *Lack of resources*: Many respondents did not have enough time or budget for multiple people analyze each recording. One of them even needed to “convince stakeholders that it requires more time than they think” and another mentioned that there was “no budget for real analysis”. We also found that 30% of the respondents were the only UX practitioner on their team and one mentioned “lack of a second opinion” as their biggest challenge. This illustrates the prevalence of resource constraints, where they have to work alone without others to double check their analysis results.
- (2) *Remote work issues*: Respondents mentioned that “communicating with the team is challenging when we can’t speak” due to working from home. The process of getting quick feedback on a usability problem from the team requires waiting for a response or setting up a meeting, which causes delays and inefficiencies. The COVID-19 pandemic—the time at which this survey was conducted—might affect this finding as many workplaces shifted from in-office to working from home [9]. We will discuss it in Section 6.
- (3) *Limitations of software tools*: Some respondents were frustrated with managing “so many columns in the Excel sheet and it’s too tiring to go back and forth [between columns]”.

Others were used to paper analysis where they can “put multiple different color post-its or highlight the sheets with notes”. However, these features are missing in some analysis software and they feel the need for a “quick way to relate different types of data”.

**4.2.4 Desired Features in New Collaboration Tools.** We also wanted to understand what features respondents would prefer in an all-inclusive user testing analysis tool that supports collaboration between UX practitioners. They assigned each of the five collaboration features a value from 1 (least important) to 5 (most important), shown in Figure 11.



**Figure 11: Rating of desired collaboration features in the data analysis tool (N = 195)**

From the most important to the least, the features were 1) *shared content (team members can access same video/audio recordings)* (Md=4, IQR=2), 2) *collaborative writing (team members can make annotations on the same recording)* (Md=4, IQR=2), 3) *integrated voice/video call in analysis software* (Md=3, IQR=3), 4) *integrated instant messaging in analysis software* (Md=3, IQR=2), and 5) *revision history (team members can track changes for annotations)* (Md=2, IQR=2).

Overall, more than half (58%) of the respondents “somewhat agreed” or “strongly agreed” that they needed new tools to support their data analysis and collaboration process (Md=4, IQR=1). The main suggestions for improvements gathered from the free-form responses include:

- Being able to schedule meetings and attach materials in the invitation to team members from within the data analysis tool so that the materials are readily available and organized.
- Having an easier way to export findings and notes from the data collection tool to a usable format like Word.
- Having one spot to consolidate the findings with coworkers in addition to the notes and session recordings.

### 4.3 RQ3: Correlations

We conducted Pearson’s chi-squared test for sets of categorical data to evaluate how likely it is that any observed differences arose by chance [16]. We also utilized Cramer’s V as a measure of the strength of the association between variables (effect size) [16].

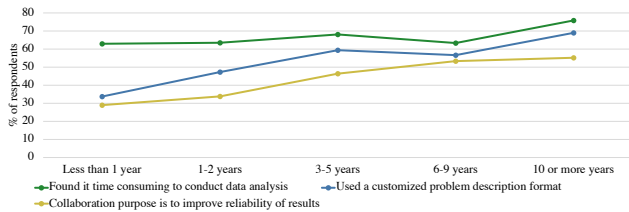
**4.3.1 Years of UX Experience.** Table 2 shows the collaboration practices that have significant differences between respondents with



**Table 3: Analysis and collaboration practices that have significant differences between people on varying sizes of UX teams**

| Collaboration practices                           |   |
|---|---|
| Used a customized problem description format      | $\chi^2(2) = 8.3, N = 195, p < .05, V = .15*$   |
| Included severity ratings of usability problems   | $\chi^2(2) = 7.6, N = 195, p < .05, V = .14*$   |
| Managed disagreements by having a mediator        | $\chi^2(2) = 7.3, N = 195, p < .05, V = .14*$   |
| Tried to convince colleagues when in disagreement | $\chi^2(8) = 22.2, N = 195, p < .01, V = .12**$ |

different years of UX experience<sup>4</sup>. Fig 12 provides a visualization of these observed trends.



**Figure 12: Line chart showing three significant correlations according to years of UX experience.**

First, as years of UX experience grew, respondents were more likely to agree and strongly agree that *it is time-consuming to conduct data analysis*. Although overall 66% of respondents felt it was time-consuming, the highest proportion (76%) occurred in respondents with over 10 years. One reason may be that experienced respondents have a better understanding of the workload and difficulty in conducting a thorough data analysis based on their years of experience; it could also be that they tend to have other responsibilities that take up their time such as managing projects and meeting with stakeholders.

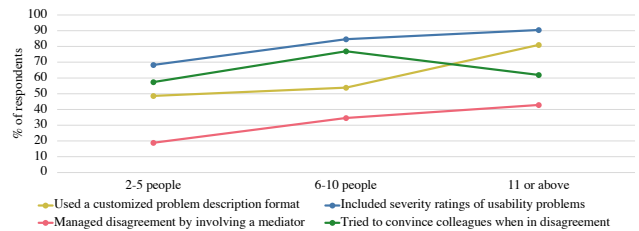
Second, the more years of UX experience that respondents had, the more likely they might *use a customized format* made by themselves or their companies when composing analysis reports. Almost 70% of respondents with 10 or more years of UX experience used a custom format, compared to only 34% with less than a year. Practitioners with more experience tended to modify heuristics to meet specific needs while novice practitioners followed pre-established formats and guidelines such as Nielsen’s Heuristics [77].

Lastly, respondents with greater years of UX experience were more likely to treat *“improve reliability of results” as a purpose for collaboration* since over half (55%) of respondents with 10 or more years of experience and 6-9 years (53%) selected it compared to just 29% of respondents with less than 1 year and 34% of 1-2 years. This may be because more experienced practitioners were more aware of the importance of reliability in the analysis results.

**4.3.2 Size of UX Team.** Table 3 shows the collaboration practices that have significant differences between respondents on different size UX teams, which is visualized in Fig 13.

First, respondents on a larger team were more likely to *use a customized format made by themselves or their companies*. The majority of respondents on a team of 11 and above used a self-defined customized format (81%), about half of respondents on a team of

<sup>4</sup>significant levels: \*: p<.05, \*\*: p<.01, \*\*\*: p<.001



**Figure 13: Line chart showing four significant correlations according to team size.**

6-10 used it (54%), and less than half of respondents on a team of 2-5 used it (49%).

Similarly, respondents on a larger team were more likely to *include ratings of severity for the problems identified*. 91% of respondents on a team of 11 and above included it, whereas only 68% of them on a team of 2-5 people did. These two trends may be because larger teams tend to have more resources to create custom heuristics or modify existing ones to fit their needs and prioritize the severity of the identified problems.

Another trend seen in Fig 13 is that respondents on a larger team were significantly more likely to *involve another person as the mediator when managing disagreements between colleagues*. 43% of respondents on a team of 11 or more people would do so, compared to just 35% of respondents on a team of 6-10 people and 19% of respondents on a team of 2-5 people. This may be because larger teams tend to have more people available to play the role of a mediator. For teams with only two people, involving another person takes more effort since they might need to find someone from another team and explain the context on the disagreement.

Lastly, respondents on a team of 6-10 people (77%) were more likely to *make a case for their decisions and try to convince their colleague when disagreements occur* compared to 57% of them on a team of 2-5 people and 62% of them with 11 and above.

## 5 DISCUSSION

In this section, we discuss the key findings and present related design considerations for creating faster-paced and more reliable analysis and collaboration among UX practitioners.

### 5.1 Modes of Collaboration

For the respondents whose companies have the resources for multiple UX practitioners, collaboration generally falls under three modes: the most frequent mode used by over half (52%) of the respondents is to *independently analyze different portions of the data first with little or no overlap before collaboration (i.e., divide and*

*conquer*). This mode highlights how UX practitioners cope with the tension between reliability and efficiency of data analysis [25]. However, while evaluators may collectively find more problems, each usability test session is essentially only reviewed by one evaluator, who might miss problems due to the “evaluator effect” [38].

The second-most frequently used (49%) mode is to *collaboratively analyze the sessions with little or no independent analysis (i.e., group analysis)*. The prevalence of this type of group analysis echoes the findings of Følstad et al [29]. However, its advantages and disadvantages are not well-understood as there is a gap in the literature on conducting analysis in groups [29]. Law and Hvannberg studied the process of consolidating usability problems individually vs. collaboratively in a within-subjects design and revealed a tendency towards collaborative settings [56]. However, a follow-up between-subjects study found no significant differences between practitioners in the two settings, suggesting that the consolidation process does not benefit from positive group decision effects [40]. It is worth noting that participants from both studies were novice practitioners (i.e. students with limited or no background knowledge in UX), so their results might not be generalizable to practitioners with more experience. Indeed, our results (Table 2) also suggest that years of UX experience may affect their collaboration practices. As a result, the benefits of group analysis remain unclear for UX practitioners who have a variety of experience. In order to preserve the reliability of the results, it is critical to investigate best-practice guidelines and procedures for group analysis.

The third mode is to *independently analyze the same set of data and then collaborate*. Although prior research suggested that this type of collaboration could improve reliability [38] and is recommended [29], our study found that only 37% of respondents practiced this collaboration mode. One potential reason might be that it requires more human effort for the same amount of data as it requires more than one practitioner to look at the same session. Furthermore, these practitioners would have to spend additional time consolidating their analyses. Indeed, 70% of our respondents felt it was difficult to merge analysis from individual practitioners, which may lead to more time costs.

**5.1.1 Design Consideration Based on Collaboration Modes. Establish processes for analysis in groups and balance the trade-offs between efficiency and validity.** We uncovered the three modes of collaboration among UX practitioners. However, it remains an open question of what are the advantages and disadvantages of the three collaboration modes in terms of improving completeness and reliability of usability testing results. Although the third mode was recommended by prior work [29], such mode has higher time costs compared to the other two modes. The other modes are at odds with recommended practices where evaluation performance should be based on thoroughness, validity, and reliability [30, 37] with the aim of minimizing the “evaluator effect” [38]. However, in reality, the time and resources of a particular company should also be taken into account. As previously mentioned, the lack of time and resources is a key challenge that respondents faced, which was echoed by previous surveys of UX practitioners in industry [12, 25, 29, 81]. Thus, one possible advantage for group analysis could be to save time as more people may lead to faster identification of usability problems and having in-situ discussions

could avoid the need of a second meeting to merge the results from different practitioners. Due to the trade-offs between time and reliability, one important question is to help UX practitioners determine when they have conducted sufficient analysis (e.g. at what point can they stop). For example, Nielsen and Landauer presented a model showing the proportion of usability problems in an interface found using various numbers of practitioners [78]. The proportion increased as the number of practitioners increased, thus the use of at least three practitioners was recommended [78]. However, this model was based on independent analysis of the same data from multiple practitioners. Thus, it would be interesting to investigate methods to extend this model for analysis conducted in groups. Future research should explore ways of collaboration that achieve a balance between validity and efficiency in UX data analysis, which is a key challenge that UX practitioners face [25].

## 5.2 Tools Used for Analysis

Our findings suggest that there is a lack of resources in the types of tools available to support both data analysis and collaboration at the same time. The 31 different tools used by the respondents fell into seven categories (e.g. *online document management platforms, online user research and testing tools, and visual design platforms*). The variety in the types of tools provides evidence that respondents have to switch between different platforms. This may cause overhead in trying to review data, conduct analysis, and then communicate and merge these findings with colleagues. During this process, there could be many difficulties such as storing and keeping track of separate recordings and analysis documents. These *data storage and access issues* were mentioned as a key challenge by respondents. Respondents also had mixed feelings about the use of many tools. Despite being flexible, having multiple tools was also confusing.

**5.2.1 Design Consideration Based on Analysis Tools. Develop one integrated environment to support UX practitioners in reviewing usability test sessions, performing data analysis, and collaborating.** One of the main improvements that respondents suggested was “*having one spot to consolidate the findings with coworkers.*” Building a platform that allows UX practitioners to review usability test sessions, conduct data analysis, and communicate with their team would also mitigate the overhead of needing to use multiple tools. Numerous commercial tools have been developed to support UX practitioners with conducting usability test and reviewing test session data, including UserZoom [106] and FullStory [32]. While these tools allow multiple team members to access the same recordings, their data analysis capabilities are mostly limited to session playback, note-taking, and mouse point clouds. Thus, it is important to design tools to meet the unique analysis needs that generic tools (e.g., Google Sheets [33]) do not support well, such as adding heuristics, which was a common practice among 76% of the respondents and have been shown to be effective in enabling practitioners to uncover more usability problems [55]. In addition, such tools should consider to visualize subtle behavioral patterns that might indicate UX problems [24, 27], such as the sentiment of users’ words and their abnormal tones. What’s more, such tools should also integrate collaboration features (e.g. dimension coverage of data [88, 115], radar views [35], synchronized annotation timeline [90]) to support UX practitioners to discuss and consolidate their

individual analyses while still having access to both the raw session recordings and their analysis notes in one integrated environment. For example, collaborative data analysis can benefit from *displaying data dimension coverage of history* [88, 89] or highlights of previously investigated data in a graph visualization [115]. In addition, radar views provide awareness of where one is working relative to their remote collaborators in a virtually shared workspace [35]. A synchronized annotation timeline supports collaborative video annotations by distributing work and sharing results simultaneously [90]. Collaborative writing tools such as Google Docs [33] have implemented variations of these features, such as showing edit history and collaborators' cursor locations to increase workspace awareness. However, more research is warranted to explore how to integrate these features into future tools that support analysis of usability testing data like video recordings and meet the unique needs of UX practitioners.

In terms of communication needs, more respondents preferred *integrated voice/video call in analysis software* rather than *integrated instant messaging*. However, current collaborative writing tools mainly employ the use of comments, while relying on external meeting software for synchronous collaboration. An explanation for the observed preference could be because voice/video calls are considered as *richer* according to Media Richness Theory since it contains auditory (and possibly visual) signals that facilitate faster connections [17]. In contrast, text-based media are considered less personal but more efficient when used during task-oriented conversations [111]. Another point to consider is that voice/video calls can only support synchronous collaboration whereas instant messaging supports both synchronous and asynchronous (where one can reply at a later time) [63]. Future designers should consider these trade-offs when developing an integrated system for collaboration and data analysis as different forms of communication have been shown to affect patterns of interaction between collaborators and effectiveness of work outcome [63, 83].

### 5.3 Implications of Prior UX Experience

The correlations that we found in Sec. 4.3 suggest that prior UX experience is related to certain data analysis and collaboration practices. As years of experience grow, respondents were more likely to agree that *it is time-consuming to conduct analysis on recordings* and also more likely to treat *“improve reliability of results” as a purpose for collaboration*. Higher awareness of the time-consuming nature of analysis and the importance of reliability could potentially lead to more thorough and robust analysis results. This is supported by prior studies showing that novice practitioners, who had not received training on user interface design principles, performed considerably worse than UX specialists and much worse than double specialists, i.e. those having expertise in user interface design and the domain of the software system [76]. In relation to collaboration, survey respondents mentioned the lack of confidence in their colleagues' capabilities to analyze data as a key challenge. Since the team consisted of varying levels of *“maturity when it comes to UX understanding”*, it led to extra time needed to set the frame of analysis while collaborating.

#### 5.3.1 Design Considerations Based on UX Experience. **Effectively integrate team members with different qualifications and**

**UX background.** The behavioral model of group performance proposed by Sauer et al. states that the advantage of groups compared to individuals stems from increased task expertise [91]. Thus, the team performance may improve only if a team succeeds with integrating members with different qualifications, perspectives, and knowledge [18]. For example, teams could participate in certain interventions like *team training*, which have been shown to improve work team effectiveness through goal-setting and team building [10]. Future work, such as in-depth qualitative studies, can be conducted to further explore the reasons behind other collaboration trends observed with years of UX experience.

### 5.4 Impact of Emerging Technologies on UX Practices

**5.4.1 Recent Technological Changes.** Over the past ten years, technology has advanced at a rapid pace and UX practices must adapt to these fast-growing innovative environments. There has been an emergence of start-up companies, which frequently guide their software development using agile practices [86]. These companies are typically limited in human resources, leading to difficulties in hiring experienced UX practitioners and inadequate usability testing practices [52]. One survey found that the top hindrance to UX practices in start-up companies was the short turnaround time for analysis [94], which is a challenge echoed by our survey results. Furthermore, 32% of UX practitioners in start-up companies also experienced “communication and collaboration gaps between UX and other professionals” [94]. Thus, how to make UX practices more lightweight to match agile practices and timeboxed sprints without compromising validity remains a key challenge [25, 41, 94].

In addition to changes in company structure, the popularization of everyday technology, such as wireless headphones, smartwatches, smart speakers, and VR headsets, has sparked new research on evaluating the UX of these products. For example, researchers conducted focus groups and surveys to determine attributes affecting the UX of headphones [45], longitudinal studies to improve smartwatch wearability [46], and in-depth interviews to investigate the UX of smart speakers [112]. Furthermore, traditional questionnaires like the System Usability Scale (SUS) [59] were insufficient for VR headsets, which require additional considerations like motion sickness [1]. Thus, researchers have developed custom scales for measuring the UX of VR headsets [114]. In term of analysis resources, our survey showed that the majority of the respondents still used Nielsen's usability heuristics [77] and Norman's design principles [82]. However, they also modified them to cater to their specific products. The need to update heuristics for new technology is evident in recent studies that utilized custom usability principles to evaluate smartwatches [13]. Recently, researchers also proposed a set of adapted heuristics for conversational agents [53] and for VR products [75], based on testing and feedback by experts. Future work should develop sets of reusable heuristics that apply to new categories of products.

Aside from evaluating the UX of new products, recent advances in VR/AR technology have also been leveraged to develop new UX design and analysis methods [49]. For example, respondents mentioned that they used physical post-its during analysis and found this feature missing in software tools. Researchers recently

designed a VR system for writing and organizing virtual post-it notes that overcomes physical limitations like lack of space [57]. Another AR application was developed to help users find specific post-its, cluster information, and visualize cluster metrics using AR overlays [99]. Although research demonstrated the effectiveness of such methods, they lack collaboration considerations and have limited adoption in industry. Thus, future work could explore how to enhance the collaboration capabilities of VR/AR systems for UX analysis and smoothly integrate them into existing practices.

For the detection of UX problems, recent advances in AI show that it is possible to automate parts of this process. Researchers leveraged AI to assess the usability of digital interfaces [34, 36, 47, 84, 85]. For example, user interaction events were utilized to create machine learning (ML) classifiers to detect usability issues of websites [34, 85], mobile applications [47], and VR applications [36]. However, these methods were primarily based on textual interaction logs, which are only available for specific types of interfaces. Depending on click and scroll maps limits their applicability to products without network connectivity such as coffee makers. In contrast, we found that respondents mainly used written notes, audio, and video recordings when conducting data analysis. Similar to how UX practitioners analyzed users' behaviors in recordings, researchers developed an AI that could infer when the participant encountered a problem during a think-aloud test session by analyzing the subtle speech patterns in their verbalizations [23]. While promising, such method needs improvements in accuracy and explainability to enhance users' trust [60]. Thus, to better help UX practitioners without disrupting their current workflow, AI systems might be designed to play an assistive role in analyzing the data and resources that UX practitioners are currently using including session recordings, written notes, and transcripts.

**5.4.2 Towards Human-AI Collaboration for UX Analysis.** Recent research has suggested that users exhibit subtle speech patterns such as abnormal pitch and speech rate when encountering usability problems [24]. Two-thirds of our respondents reviewed video recordings, which echoes the findings of prior work showing that UX practitioners use multi-modal information from both the acoustic and visual channels [12]. When they attend to many signals simultaneously, they may not notice these subtle speech patterns. Thus, there is an opportunity for AI to assist practitioners in detecting and highlighting subtle speech patterns. In fact, recently, researchers began to design AI-assisted UX analysis tools, such as VisTA [26] and CoUX [97], to help UX practitioners with analyzing usability test sessions by presenting automatically-extracted problem-indicators to them. Although VisTA demonstrated the promise of human-AI collaboration for UX evaluation, it was designed to assist individual practitioners. As a collaborative tool, CoUX was evaluated with pairs of UX practitioners, but the AI assistance was limited to video analysis [97]. Thus, future work should investigate ways to integrate AI to support the human-human collaboration process for UX evaluation. For example, one particular challenge experienced by over two-thirds of the respondents was merging analysis from multiple evaluators. The AI system could possibly act as a tie-breaker which is a strategy employed by almost one-quarter of the respondents. Another possible role for the AI

system is to learn from conversations between practitioners and recognize disagreements to make appropriate suggestions. Thus, more research is warranted to determine whether and to what extent should AI be involved in the analysis process, since many human collaborators face challenges in understanding AI's capabilities or envisioning what it might be [113].

In addition to the state-of-the-art research prototypes, some commercially available tools that our respondents used already contain features derived from AI and ML. For example, analytical platforms like UserTesting [105] and PlaybookUX [87] offer sentiment analysis which detects whether the participant is expressing positive, negative, or neutral sentiment. Furthermore, UserTesting also provides "smart tags" where predefined situations such as confusion, dislike, or suggestion are automatically labelled [105]. There are also data-informed analytical tools like MixPanel [70] which provides correlation analysis on retention data, and UXTesting [108] which automatically detects participants' emotions. Since these features are relatively new, their effectiveness is yet to be validated. Moreover, the UX community may benefit from integrating more subtle behavioral patterns (e.g., verbalization and speech indicators of UX problems [24, 27]) into existing tools.

## 6 LIMITATIONS AND FUTURE WORK

Our survey respondents were from six continents with varying levels of UX experience. Thus, the findings are expected to provide a reasonably informative perspective of practices and challenges in collaboration during the data analysis phase among UX practitioners around the world. However, the number of responses from different continents was mismatched. While the majority of the data came from North America, Asia, and Europe, relatively fewer responses were from Latin America, Africa, and Oceania. Thus, one should be cautious about generalizing the findings to UX practitioners. Indeed, our research calls for more work to investigate UX practices and challenges in under-represented regions of our survey, such as countries in Africa, Latin America, and Oceania. Furthermore, studying and comparing the practices and challenges of UX practitioners in different countries could inform the UX community about the impact of culture on UX data analysis and collaboration, which will further help the community design customized collaborative UX data analysis strategies and tools.

This survey was conducted during the COVID-19 pandemic, so there is a possibility that the findings would not be representative of all time. During the pandemic, certain issues may be at the forefront of the respondents' minds, such as remote collaboration being brought up as their greatest challenge. As articles have shown how the COVID-19 pandemic has changed product users [73], future work can also explore its impacts on UX practitioners. Furthermore, we focused on collaboration between UX practitioners within the same company and found correlations of analysis behaviors based on the in-house UX team size. However, cross-company collaboration is also a common practice for companies who cannot afford to hire expertise and maintain testing facilities [21] or have products that are sold in different countries or languages [72]. Prior research suggested that cross-company collaboration happened in the form of outsourcing participant recruitment [72], small subtasks such as transcription [8], or complete evaluations to usability testing

services [21, 71]. These collaboration modes included tasks that were not directly related to data analysis, which was the focus of our research, but may lead to other challenges such as language barriers, sub-quality reports, and deviations from expected practices [21, 71, 72]. Thus, future work should further explore the state of cross-company collaborations in recent years with a focus on how the COVID-19 pandemic may have affected these practices.

In this survey, we focused on the analysis of data collected from usability testing, which only covers the pragmatic aspect of UX research. However, the UX community believes it is also valuable to consider the hedonic aspects, which refers to the momentary pleasures we experience when interacting with a product [65, 109]. Furthermore, there is research on how UX is related to eudaimonia, in which users gain meaning from need fulfillment [65] and how perceived beauty and aesthetics impact UX [102, 109]. Based on these aspects, future work should investigate how UX practitioners collect and analyze data for evaluating the overall product quality. From these findings, we can propose tools to better assist with discovering hedonic and eudaimonic issues in addition to task-oriented usability issues.

Our survey study primarily provides a quantitative understanding of UX practitioners' current practices, challenges, and opportunities in data analysis and collaboration even though it also offers qualitative insights via open-ended questions. To better understand the rich reasons behind the phenomena that our study revealed, more qualitative research, such as in-depth interviews with UX practitioners from different industrial sections in different continents, should be conducted. In particular, it is worth investigating different standards and principles that UX practitioners use for problem identification and merging and their considerations for choosing one over the other. Furthermore, as disagreements tend to occur during collaboration, it is valuable to gain a deeper understanding of their strategies for managing disagreements, in particular regarding whether there is a problem and what the severity level should be. Last but not least, as our study revealed that they would have a discussion with other practitioners to resolve disagreements, it would be interesting to explore the content of their conversations when discussing disagreements to better understand the nature of the disagreements and how they manage to gain a consensus.

## 7 CONCLUSION

We have conducted an international survey to understand the practices and challenges of collaboration in the context of conducting data analysis on usability test sessions. Based on the responses of 279 participants who had varying UX experience and worked in different geographic locations, we found that UX practitioners collaborate to primarily *identify more usability problems and generate redesign suggestions*, and to *improve reliability of results* to a lesser extent. We identified three modes of collaboration: *independently analyze different portions of the data with little or no overlap and then collaborate (i.e., divide and conquer)*, *collaboratively analyze the session with little or no independent analysis (i.e., group analysis)*, and *independently analyze the same set of data and then collaborate*. Although the third mode was recommended by the literature, it was least adopted among our respondents because it was perceived to be more time-consuming than the other two. Moreover, most of the

respondents encountered challenges related to lack of time as 70% needed to complete analysis within 1 week, impacting their choice of collaboration modes. These findings highlight an opportunity to address the trade-offs between efficiency and validity of analysis results. What's more, respondents also experienced disagreements with colleagues regarding usability problems and difficulty in merging analysis from multiple practitioners. Our survey findings could potentially inform UX practitioners about how their colleagues perceive collaboration during data analysis. In addition, our findings reveal opportunities for developing better methods and tools to facilitate collaboration during analysis, for example, developing an integrated platform to support both analysis and collaboration in one place, establishing group analysis procedures, supporting both synchronous and asynchronous collaboration, and integrating team members with various skill levels.

## ACKNOWLEDGMENTS

We thank all our respondents for taking the time to share their experiences.

## REFERENCES

- [1] Shelly L. Ames, James S. Wolffsohn, and Neville A. McBrien. 2005. The Development of a Symptom Questionnaire for Assessing Virtual Reality Viewing Using a Head-Mounted Display. *82, 3 (2005)*, 168–176. <https://doi.org/10.1097/01.OPX.0000156307.95086.6>
- [2] Terence Andre, H Hartson, Steven Belz, and Faith McCreary. 2001. User Action Framework: A Reliable Foundation for Usability Engineering Support Tools. *International Journal of Human-Computer Studies* 54 (Jan. 2001), 107–136. <https://doi.org/10.1006/ijhc.2000.0441>
- [3] Atlassian. 2021. Confluence: Your Remote-Friendly Team Workspace. <https://www.atlassian.com/software/confluence>.
- [4] Atlassian. 2021. Jira: Issue & Project Tracking Software. <https://www.atlassian.com/software/jira>.
- [5] Atlassian. 2021. Trello. <https://trello.com>.
- [6] Enrico Bertini, Silvia Gabrielli, Stephen Kimani, Tiziana Catarci, and Giuseppe Santucci. 2006. Appropriating and Assessing Heuristics for Mobile Computing. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '06)*. Association for Computing Machinery, New York, NY, USA, 119–126. <https://doi.org/10.1145/1133265.1133291>
- [7] Ted Boren and Judith Ramey. 2000. Thinking Aloud: Reconciling Theory and Practice. *IEEE Transactions on Professional Communication* 43, 3 (Sept. 2000), 261–278. <https://doi.org/10.1109/47.867942>
- [8] Nis Bornoe, Louise Barkhuus, Barry Brown, and Malcolm Hall. 2011. TagPad for iPad – Designing a Support Tool for Interview Studies. In *11th Danish Human-Computer Interaction Research Symposium* (Frederiksberg, Denmark), 49–52.
- [9] Erik Brynjolfsson, John J. Horton, Adam Ozimek, Daniel Rock, Garima Sharma, and Hong-Yi TuYe. 2020. COVID-19 and Remote Work: An Early Look at US Data. <https://doi.org/10.3386/w27344> Series: Working Paper Series.
- [10] Martina Buljac-Samardzic, Connie M. Dekker-van Doorn, Jeroen D. H. van Wijngaarden, and Kees P. van Wijk. 2010. Interventions to Improve Team Effectiveness: A Systematic Review. *Health Policy* 94, 3 (March 2010), 183–195. <https://doi.org/10.1016/j.healthpol.2009.09.015>
- [11] Miranda G. Capra. 2006. *Usability Problem Description and the Evaluator Effect in Usability Testing*. Ph.D. Dissertation. Virginia Tech, Blacksburg, VA.
- [12] Parmit K. Chilana, Jacob O. Wobbrock, and Andrew J. Ko. 2010. Understanding Usability Practices in Complex Domains. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10*. ACM Press, Atlanta, Georgia, USA, 2337–2346. <https://doi.org/10.1145/1753326.1753678>
- [13] Jaemin Chun, Anind Dey, Kyungtaek Lee, and SeungJun Kim. 2018. A qualitative study of smartwatch usage and its usability. *28, 4 (2018)*, 186–199. <https://doi.org/10.1002/hfm.20733> [\\_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hfm.20733](https://onlinelibrary.wiley.com/doi/pdf/10.1002/hfm.20733).
- [14] Gilbert Cockton, Gilbert Woolrych, Alan, Hindmarch, and Mark. 2004. Reconditioned Merchandise: Extended Structured Report Formats in Usability Inspection. <https://doi.org/10.1145/985921.986083>
- [15] Gilbert Cockton and Darryn Lavery. 1999. A Framework for Usability Problem Extraction. In *INTERACT*, Vol. 1999. 347–355.
- [16] Jacob Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed ed.). L. Erlbaum Associates, Hillsdale, NJ.

- [17] Richard Daft and Robert Lengel. 1986. Organizational Information Requirements, Media Richness and Structural Design. *Management Science* 32 (May 1986), 554–571. <https://doi.org/10.1287/mnsc.32.5.554>
- [18] Carsten De Dreu and Laurie Weingart. 2003. Task Versus Relationship Conflict, Team Performance, and Team Member Satisfaction: A Meta-Analysis. *The Journal of applied psychology* 88 (Sept. 2003), 741–9. <https://doi.org/10.1037/0021-9010.88.4.741>
- [19] Gert-Jan de Vreede, Ann L. Fruhling, and Anita Chakrapani. 2005. A Repeatable Collaboration Process for Usability Testing. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*. 46–46. <https://doi.org/10.1109/HICSS.2005.46> ISSN: 1530-1605.
- [20] Don Dillman. 2000. *Mail and Internet Surveys: The Tailored Design Method*. Vol. 2.
- [21] Ian Douglas. 2008. Global spread of usability expertise. In *Proceedings of the 2008 Euro American Conference on Telematics and Information Systems* (New York, NY, USA) (*EATIS '08*). Association for Computing Machinery, 1–4. <https://doi.org/10.1145/1621087.1621115>
- [22] Joseph S. Dumas and Beth A. Loring. 2008. *Moderating Usability Tests: Principles and Practices for Interacting*. Elsevier.
- [23] Mingming Fan, Yue Li, and Khai N. Truong. 2020. Automatic Detection of Usability Problem Encounters in Think-Aloud Sessions. *ACM Transactions on Interactive Intelligent Systems* 10, 2 (June 2020), 1–24. <https://doi.org/10.1145/3385732>
- [24] Mingming Fan, Jinglan Lin, Christina Chung, and Khai N. Truong. 2019. Concurrent Think-Aloud Verbalizations and Usability Problems. *ACM Transactions on Computer-Human Interaction* 26, 5 (Sept. 2019), 1–35. <https://doi.org/10.1145/3325281>
- [25] Mingming Fan, Serina Shi, and Khai N Truong. 2020. Practices and Challenges of Using Think-Aloud Protocols in Industry: An International Survey. *Journal of Usability Studies* 15, 2 (2020), 85–102.
- [26] Mingming Fan, Ke Wu, Jian Zhao, Yue Li, Winter Wei, and Khai N. Truong. 2020. VisTA: Integrating Machine Intelligence with Visualization to Support the Investigation of Think-Aloud Sessions. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (Jan. 2020), 343–352. <https://doi.org/10.1109/TVCG.2019.2934797>
- [27] Mingming Fan, Qiwen Zhao, and Vinita Tibdewal. 2021. Older Adults' Think-Aloud Verbalizations and Speech Features for Identifying User Experience Problems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (*CHI '21*). Association for Computing Machinery, New York, NY, USA, Article 358, 13 pages. <https://doi.org/10.1145/3411764.3445680>
- [28] Figma. 2021. Figma: The Collaborative Interface Design Tool. <https://www.figma.com/>.
- [29] Asbjørn Følstad, Effie Lai-Chong Law, and Kasper Hornb. 2012. Analysis in Practical Usability Evaluation: A Survey Study. In *Proceedings of the 30th SIGCHI Conference on Human Factors in Computing Systems - CHI '12*. ACM Press, Austin, Texas, 2127–2136. <https://doi.org/10.1145/2207676.2208365>
- [30] Asbjørn Følstad, Effie Lai-Chong Law, and Kasper Hornbæk. 2010. Analysis in Usability Evaluations: An Exploratory Study. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (NordCHI '10)*. Association for Computing Machinery, New York, NY, USA, 647–650. <https://doi.org/10.1145/1868914.1868995>
- [31] Ann L. Fruhling and Gert-Jan de Vreede. 2006. Collaborative Usability Testing to Facilitate Stakeholder Involvement. In *Value-Based Software Engineering*, Stefan Biffl, Aybüke Aurum, Barry Boehm, Hakan Erdogmus, and Paul Grünbacher (Eds.). Springer. [https://doi.org/10.1007/3-540-29263-2\\_10](https://doi.org/10.1007/3-540-29263-2_10)
- [32] FullStory. 2021. FullStory | Robust Analytics, Session Replay, Heatmaps, Dev Tools. <https://www.fullstory.com>.
- [33] Google. 2021. Google Workspace: Business Collaboration Tools. <https://workspace.google.com>.
- [34] J. Grigera, Alejandra Garrido, J. Rivero, and G. Rossi. 2017. Automatic Detection of Usability Smells in Web Applications. *Int. J. Hum. Comput. Stud.* (2017). <https://doi.org/10.1016/j.ijhcs.2016.09.009>
- [35] Carl Gutwin and Saul Greenberg. 2002. A Descriptive Framework of Workspace Awareness for Real-Time Groupware. *Computer Supported Cooperative Work* 11, 3 (Nov. 2002), 411–446. <https://doi.org/10.1023/A:1021271517844>
- [36] Patrick Harms. 2019. Automated Usability Evaluation of Virtual Reality Applications. *ACM Transactions on Computer-Human Interaction* 26, 3 (April 2019), 14:1–14:36. <https://doi.org/10.1145/3301423>
- [37] H. Rex Hartson, Terence Andre, and Robert Williges. 2003. Criteria For Evaluating Usability Evaluation Methods. *Int. J. Hum. Comput. Interaction* 15 (Feb. 2003), 145–181. [https://doi.org/10.1207/S15327590IJHC1501\\_13](https://doi.org/10.1207/S15327590IJHC1501_13)
- [38] Morten Hertzum and Niels Ebbe Jacobsen. 2001. The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. *International Journal of Human-Computer Interaction* 15, 1 (2001), 183–204. [https://doi.org/10.1207/S15327590IJHC1501\\_14](https://doi.org/10.1207/S15327590IJHC1501_14)
- [39] Morten Hertzum, Rolf Molich, and Niels Ebbe Jacobsen. 2013. What You Get Is What You See: Revisiting the Evaluator Effect in Usability Tests. *Behaviour & Information Technology* 33, 2 (April 2013), 144–162. <https://doi.org/10.1080/0144929X.2013.783114>
- [40] Rebekka Hoffmann, Anna Helga Jónsdóttir, and Ebba Thora Hvannberg. 2019. Consolidation of Usability Problems With Novice Evaluators Re-Examined in Individual vs. Collaborative Settings. *Interacting with Computers* 31, 6 (April 2019), 525–538. <https://doi.org/10.1093/iwc/iwz034>
- [41] Laura Hokkanen, Kati Kuusinen, and Kaisa Vááñänen. 2016. Minimum Viable User EXperience: A Framework for Supporting Product Design in Startups. In *Agile Processes, in Software Engineering, and Extreme Programming (Cham) (Lecture Notes in Business Information Processing)*, Helen Sharp and Tracy Hall (Eds.). Springer International Publishing, 66–78. [https://doi.org/10.1007/978-3-319-33515-5\\_6](https://doi.org/10.1007/978-3-319-33515-5_6)
- [42] Tactivos Inc. 2020. MURAL: A Digital Workspace for Visual Collaboration. <https://www.mural.co/>.
- [43] ISO. 2020. ISO 9241-110:2020 Ergonomics of Human-System Interaction. <https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/07/52/75258.html>.
- [44] Niels Jacobsen, Niels E, Morten Hertzum, Morten, Bonnie John, and Bonnie E. 1998. The Evaluator Effect in Usability Tests. <https://doi.org/10.1145/286498.286737>
- [45] Rasmus Jensen, Nikolaj Lauridsen, Andreas Poulsen, Casper Tofte, and Flemming Christensen. 2016. Analysis of Subjective Evaluation of User Experience with Headphones. Audio Engineering Society. <https://www.aes.org/e-lib/browse.cfm?elib=18345>
- [46] Hayeon Jeong, Hee-yung Kim, Rihun Kim, Uichin Lee, and Yong Jeong. 2017. Smartwatch Wearing Behavior Analysis: A Longitudinal Study. 1, 3 (2017), 60:1–60:31. <https://doi.org/10.1145/3131892>
- [47] JongWook Jeong, NeungHoe Kim, and Hoh Peter In. 2020. Detecting Usability Problems in Mobile Applications on the Basis of Dissimilarity in User Behavior. *International Journal of Human-Computer Studies* 139 (July 2020), 102364. <https://doi.org/10.1016/j.ijhcs.2019.10.001>
- [48] Alexander Jones and Volker Thoma. 2019. Determinants for Successful Agile Collaboration between UX Designers and Software Developers in a Complex Organisation. *International Journal of Human-Computer Interaction* 35, 20 (Dec. 2019), 1914–1935. <https://doi.org/10.1080/10447318.2019.1587856>
- [49] Lee Kent, Chris Snider, James Gopsill, and Ben Hicks. 2021. Mixed reality in design prototyping: A systematic review. 77 (2021). <https://doi.org/10.1016/j.destud.2021.101046>
- [50] Jesper Kjeldskov, Jesper, Mikael Skov, Mikael B, Stage, and Jan. 2004. Instant Data Analysis: Conducting Usability Evaluations in a Day. <https://doi.org/10.1145/1028014.1028050>
- [51] Kurt Koffka. 2013. *Principles Of Gestalt Psychology*. Routledge.
- [52] Kati Kuusinen, Martin Kjølbøye Sørensen, Nicklas Mandrup Frederiksen, Niclas Kildahl Laugesen, and Soren Holm Jul. 2019. From Startup to Scaleup: An Interview Study of the Development of User Experience Work in a Data-Intensive Company. In *Human-Centered Software Engineering (Cham) (Lecture Notes in Computer Science)*, Cristian Bogdan, Kati Kuusinen, Marta Kristín Lárusdóttir, Philippe Palanque, and Marco Winckler (Eds.). Springer International Publishing, 3–14. [https://doi.org/10.1007/978-3-030-05909-5\\_1](https://doi.org/10.1007/978-3-030-05909-5_1)
- [53] Raina Langevin, Ross J Lordon, Thi Avrahami, Benjamin R. Cowan, Tad Hirsch, and Gary Hsieh. 2021. Heuristic Evaluation of Conversational Agents. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Number 632. Association for Computing Machinery, 1–15. <https://doi.org/10.1145/3411764.3445312>
- [54] Darryn Lavery, Gilbert Cockton, and Malcolm P. Atkinson. 1997. Comparison of Evaluation Methods Using Structured Usability Problem Reports. *Behaviour & Information Technology* 16, 4-5 (Jan. 1997), 246–266. <https://doi.org/10.1080/014492997119824>
- [55] Effie Lai-Chong Law and Ebba Thora Hvannberg. 2004. Analysis of Strategies for Improving and Estimating the Effectiveness of Heuristic Evaluation. In *Proceedings of the Third Nordic Conference on Human-Computer Interaction (NordCHI '04)*. Association for Computing Machinery, New York, NY, USA, 241–250. <https://doi.org/10.1145/1028014.1028051>
- [56] Effie Lai-Chong Law and Ebba Thora Hvannberg. 2008. Consolidating Usability Problems with Novice Evaluators. In *Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges (NordCHI '08)*. Association for Computing Machinery, New York, NY, USA, 495–498. <https://doi.org/10.1145/1463160.1463228>
- [57] Joon Hyub Lee, Donghyeok Ma, Haena Cho, and Seok-Hyung Bae. 2021. Post-Post-it: A Spatial Ideation System in VR for Overcoming Limitations of Physical Post-it Notes. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. Number 300. Association for Computing Machinery, 1–7. <https://doi.org/10.1145/3411763.3451786>
- [58] James R. Lewis. 2006. Usability Testing. In *Handbook of Human Factors and Ergonomics*. John Wiley & Sons, Ltd, Chapter 46, 1267–1312. <https://doi.org/10.1002/9781118131350.ch46>
- [59] James R. Lewis. 2018. The System Usability Scale: Past, Present, and Future. 34, 7 (2018), 577–590. <https://doi.org/10.1080/10447318.2018.1455307> Publisher:

- Taylor & Francis eprint: <https://doi.org/10.1080/10447318.2018.1455307>.
- [60] Bingjie Liu. 2021. In AI We Trust? Effects of Agency Locus and Transparency on Uncertainty Reduction in Human-AI Interaction. (2021). Issue zmb013. <https://doi.org/10.1093/jcmc/zmb013>
- [61] Paul B. Lowry, Conan Albrecht, James D. Lee, and Jay F. Nunamaker. 2002. Users Experiences in Collaborative Writing Using Collaboratus: An Internet-Based Collaborative Work. *IEEE Computer Society*, 21–21. <https://doi.org/10.1109/HICSS.2002.993879>
- [62] Paul B. Lowry and Tom L. Roberts. 2003. Improving the Usability Evaluation Technique, Heuristic Evaluation, Through the Use of Collaborative Software. 284. <https://doi.org/10.2139/ssrn.666224>
- [63] Mark Mabrito. 2006. A Study of Synchronous Versus Asynchronous Collaboration in an Online Business Writing Class. *American Journal of Distance Education* 20, 2 (June 2006), 93–107. [https://doi.org/10.1207/s15389286ajde2002\\_4](https://doi.org/10.1207/s15389286ajde2002_4)
- [64] Sharon McDonald, Helen M. Edwards, and Tingting Zhao. 2012. Exploring Think-Alouds in Usability Testing: An International Survey. *IEEE Transactions on Professional Communication* 55, 1 (March 2012), 2–19. <https://doi.org/10.1109/TPC.2011.2182569>
- [65] Elisa D. Mekler and Kasper Hornbæk. 2016. Momentary Pleasure or Lasting Meaning? Distinguishing Eudaimonic and Hedonic User Experiences. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA). Association for Computing Machinery, 4509–4520. <https://doi.org/10.1145/2858036.2858225>
- [66] Microsoft. 2021. Microsoft SharePoint: Build Team Intranets & Share Files. <https://www.microsoft.com/en-ca/microsoft-365/sharepoint/collaboration>.
- [67] Microsoft. 2021. Microsoft Teams: Group Chat Software. <https://www.microsoft.com/en-ca/microsoft-teams/group-chat-software>.
- [68] Michael Minge and Manfred Thüring. 2018. The MeCUE Questionnaire (2.0): Meeting Five Basic Requirements for Lean and Standardized UX Assessment. In *Design, User Experience, and Usability: Theory and Practice (Lecture Notes in Computer Science)*, Aaron Marcus and Wentao Wang (Eds.). Springer International Publishing, Cham, 451–469. [https://doi.org/10.1007/978-3-319-91797-9\\_33](https://doi.org/10.1007/978-3-319-91797-9_33)
- [69] Miro. 2021. Miro: An Online Visual Collaboration Platform for Teamwork. <https://miro.com/>.
- [70] MixPanel. 2021. MixPanel: Product Analytics for Mobile, Web, & More. <https://mixpanel.com/>.
- [71] Rolf Molich. 2011. The Quest for Quality: Usability Testing Assessment. In *11th Danish Human-Computer Interaction Research Symposium* (Frederiksberg, Denmark), 56–59.
- [72] Rolf Molich, Susan Dray, and David Siegel. 2004. Tips and tricks for a better international usability test. In *Extended abstracts of the 2004 conference on Human factors and computing systems - CHI '04* (Vienna, Austria). ACM Press, 1606. <https://doi.org/10.1145/985921.986168>
- [73] Kate Moran. 2020. *COVID-19 Has Changed Your Users*. <https://www.nngroup.com/articles/covid-changed-users/>
- [74] Kate Moran and Kara Pernice. 2020. Remote Moderated Usability Tests: Why to Do Them. <https://www.nngroup.com/articles/moderated-remote-usability-test-why/>.
- [75] Rabia Murtza, Stephen Monroe, and Robert J. Youmans. 2017. Heuristic Evaluation for Virtual Reality Systems. 61, 1 (2017), 2067–2071. <https://doi.org/10.1177/1541931213602000> Publisher: SAGE Publications Inc.
- [76] Jakob Nielsen. 1992. Finding Usability Problems through Heuristic Evaluation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '92)*. Association for Computing Machinery, New York, NY, USA, 373–380. <https://doi.org/10.1145/142750.142834>
- [77] Jakob Nielsen. 1994. 10 Usability Heuristics for User Interface Design. <https://www.nngroup.com/articles/ten-usability-heuristics/>.
- [78] Jakob Nielsen and Thomas K. Landauer. 1993. A Mathematical Model of the Finding of Usability Problems. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (CHI '93)*. Association for Computing Machinery, New York, NY, USA, 206–213. <https://doi.org/10.1145/169059.169166>
- [79] Jakob Nielsen and Robert L. Mack (Eds.). 1994. *Usability inspection methods*. John Wiley & Sons, Inc.
- [80] Noldus. 2020. Record & Annotate - Recording Options and Easy Annotation. <https://www.noldus.com/visio/record-annotate>.
- [81] Mie Nørgaard and Kasper Hornbæk. 2006. What Do Usability Evaluators Do in Practice? An Explorative Study of Think-Aloud Testing. In *Proceedings of the 6th Conference on Designing Interactive Systems (DIS '06)*. Association for Computing Machinery, New York, NY, USA, 209–218. <https://doi.org/10.1145/1142405.1142439>
- [82] Donald A. Norman. 2002. *The Design of Everyday Things*. Basic Books, Inc., USA.
- [83] Vanessa Y. Oviedo and Jean E. Fox Tree. 2021. Meeting by Text or Video-Chat: Effects on Confidence and Performance. *Computers in Human Behavior Reports* 3 (Jan. 2021), 100054. <https://doi.org/10.1016/j.chbr.2021.100054>
- [84] Asil Oztekin, Dursun Delen, Ali Turkyilmaz, and Selim Zaim. 2013. A Machine Learning-Based Usability Evaluation Method for eLearning Systems. *Decision Support Systems* 56, C (Dec. 2013), 63–73.
- [85] Fabio Paternò, Antonio Giovanni Schiavone, and Antonio Conti. 2017. Customizable Automatic Detection of Bad Usability Smells in Mobile Accessed Web Applications. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '17)*. Association for Computing Machinery, New York, NY, USA, 1–11. <https://doi.org/10.1145/3098279.3098558>
- [86] Nicolò Paternoster, Carmine Giardina, Michael Unterkalmsteiner, Tony Gorschek, and Pekka Abrahamsson. 2014. Software development in startup companies: A systematic mapping study. 56, 10 (2014), 1200–1218. <https://doi.org/10.1016/j.infsof.2014.04.014>
- [87] PlaybookUX. 2021. *PlaybookUX: Scalable User Testing & Interview Software*. <https://www.playbookux.com/>
- [88] Ali Sarvghad and Melanie Tory. 2015. Exploiting Analysis History to Support Collaborative Data Analysis. In *Proceedings of the 41st Graphics Interface Conference (GI '15)*. Canadian Information Processing Society, CAN, 123–130.
- [89] Ali Sarvghad, Melanie Tory, and Narges Mahyar. 2017. Visualizing Dimension Coverage to Support Exploratory Analysis. *IEEE transactions on visualization and computer graphics* 23, 1 (Jan. 2017), 21–30. <https://doi.org/10.1109/TVCG.2016.2598466>
- [90] Aziret Satybaldiev, Peter Hevesi, Marco Hirsch, Vitor Fortes Rey, and Paul Lukowicz. 2019. CoAT: A Web-Based, Collaborative Annotation Tool. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers (UbiComp/ISWC '19 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 814–818. <https://doi.org/10.1145/3341162.3345592>
- [91] Chris Sauer, D. Ross Jeffery, Lesley Land, and Philip Yetton. 2000. The Effectiveness of Software Development Technical Reviews: A Behaviorally Motivated Program of Research. *IEEE Transactions on Software Engineering* 26, 1 (Jan. 2000), 1–14. <https://doi.org/10.1109/32.825763>
- [92] Andrew Sears. 1997. Heuristic Walkthroughs: Finding the Problems Without the Noise. *International Journal of Human-Computer Interaction* 9, 3 (Sept. 1997), 213–234. [https://doi.org/10.1207/s15327590ijhc0903\\_2](https://doi.org/10.1207/s15327590ijhc0903_2)
- [93] Qingxin Shi. 2008. A Field Study of the Relationship and Communication between Chinese Evaluators and Users in Thinking Aloud Usability Tests. In *Proceedings of the 5th Nordic Conference on Human-Computer Interaction Building Bridges - NordiCHI '08*. ACM Press, Lund, Sweden, 344. <https://doi.org/10.1145/1463160.1463198>
- [94] Sofia A. M. Silveira, Joelma Choma, Roberto Pereira, Eduardo M. Guerra, and Luciana A. M. Zaina. 2021. UX Work in Software Start-Ups: Challenges from the Current State of Practice. In *Agile Processes in Software Engineering and Extreme Programming (Cham) (Lecture Notes in Business Information Processing)*, Peggy Gregory, Casper Lassenius, Xiaofeng Wang, and Philippe Kruchten (Eds.). Springer International Publishing, 19–35. [https://doi.org/10.1007/978-3-030-78098-2\\_2](https://doi.org/10.1007/978-3-030-78098-2_2)
- [95] Silverback. 2019. Silverback 3. <https://silverbackapp.com/>.
- [96] Mark Simpson. 1991. The Practice of Collaboration in Usability Test Design. 38, 4 (1991), 527–531. <http://www.jstor.org/stable/43095829> Publisher: Society for Technical Communication.
- [97] Ehsan Jahangirzadeh Soure, Emily Kuang, Mingming Fan, and Jian Zhao. 2021. CoUX: Collaborative Visual Analysis of Think-Aloud Usability Test Videos for Digital Interfaces. (2021), 1–11. <https://doi.org/10.1109/TVCG.2021.3114822> Conference Name: IEEE Transactions on Visualization and Computer Graphics.
- [98] Rebecca Stern, Aisling Kelliher, Winslow Burleson, and Lisa Tolentino. 2008. Sharing the Squid: Tangible Workplace Collaboration. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems (CHI EA '08)*. Association for Computing Machinery, New York, NY, USA, 3369–3374. <https://doi.org/10.1145/1358628.1358859>
- [99] Hariharan Subramonyam, Steven Drucker, and Eytan Adar. 2019. Affinity Lens: Data-Assisted Affinity Diagramming with Augmented Reality. <https://doi.org/10.1145/3290605.3300628>
- [100] TechSmith. 2020. Morae 3 Tutorials. <https://www.techsmith.com/tutorial-morae-current.html>.
- [101] Paul Toh. 2021. Colour schemes and templates. <https://personal.sron.nl/~pault/#sec:diverging>.
- [102] Alexandre N. Tuch, Sandra P. Roth, Kasper Hornbæk, Klaus Opwis, and Javier A. Bargas-Avila. 2012. Is beautiful really usable? Toward understanding the relation between usability, aesthetics, and affect in HCI. 28, 5 (2012), 1596–1607. <https://doi.org/10.1016/j.chb.2012.03.024>
- [103] Useberry. 2021. Useberry: Codeless Prototype Analytics. <https://www.useberry.com/>.
- [104] UserTesting. 2020. UserTesting: How to Analyze and Share Results. <https://help.usertesting.com/hc/en-us/articles/115003371891-How-to-analyze-and-share-results>.
- [105] UserTesting. 2021. UserTesting: The Human Insight Platform. <https://www.usertesting.com/>.

- [106] UserZoom. 2021. UserZoom: Actionable UX Insights for Better Digital Experiences. <https://www.userzoom.com/>.
- [107] UXQB. 2021. International Usability and UX Qualification Board. <https://uxqb.org/en/>.
- [108] UXTesting. 2022. *UXTesting: The Best and Most Intelligent User Experience Solution for Enterprises*. <https://www.uxtesting.io/uxtesting>
- [109] Paul van Schaik and Jonathan Ling. 2008. Modelling user experience with web sites: Usability, hedonic value, beauty and goodness. 20, 3 (2008), 419–432. <https://doi.org/10.1016/j.intcom.2008.03.001>
- [110] Karel Vredenburg, Ji-Ye Mao, Paul W. Smith, and Tom Carey. 2002. A Survey of User-Centered Design Practice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '02)*. Association for Computing Machinery, New York, NY, USA, 471–478. <https://doi.org/10.1145/503376.503460>
- [111] Joseph B. Walther. 1996. Computer-Mediated Communication: Impersonal, Interpersonal, and Hyperpersonal Interaction. *Communication Research* 23, 1 (Feb. 1996), 3–43. <https://doi.org/10.1177/009365096023001001>
- [112] Xin-Ting Xiao and Seung-In Kim. 2018. A Study on the User Experience of Smart Speaker in China - Focused on Tmall Genie and Mi AI Speaker -. 16, 10 (2018), 409–414. <https://doi.org/10.14400/JDC.2018.16.10.409> Publisher: The Society of Digital Policy and Management.
- [113] Qian Yang, Alex Scuito, John Zimmerman, Jodi Forlizzi, and Aaron Steinfeld. 2018. Investigating How Experienced UX Designers Effectively Work with Machine Learning. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. Association for Computing Machinery, New York, NY, USA, 585–596. <https://doi.org/10.1145/3196709.3196730>
- [114] Mengli Yu, Ronggang Zhou, Huiwen Wang, and Weihua Zhao. 2019. An evaluation for VR glasses system user experience: The influence factors of interactive operation and motion sickness. 74 (2019), 206–213. <https://doi.org/10.1016/j.apergo.2018.08.012>
- [115] Jian Zhao, Michael Glueck, Petra Isenberg, Fanny Chevalier, and Azam Khan. 2018. Supporting Handoff in Asynchronous Collaborative Sensemaking Using Knowledge-Transfer Graphs. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (Jan. 2018), 340–350. <https://doi.org/10.1109/TVCG.2017.2745279>
- [116] Zoom. 2021. Video Conferencing, Web Conferencing, Webinars, Screen Sharing - Zoom. <https://zoom.us/>.