Online Collaboration Applications Evaluated Based on Ease of Use

Applying HF/E methods to determine the usability of Web apps intended to improve and facilitate collaborative work reveals pros and cons.

By Lauren E. Margulieux, Dar-Wei Chen, Joseph D. McDonald, Keith R. Bujak, Thomas M. Gable, Cale M. Darling, Laura M. Schaeffer, & Laura H. Barg-Walkow

FEATURE AT A GLANCE:
Many online applications promise to unlock the power of collaborative work by connecting people from around the world. Finding an app that is appropriate for a task and easy to use, however, can be difficult. In this study, we evaluated 20 popular apps according to the basic work functions they accomplish (e.g., sharing files) and their adherence to the classic usability standards outlined by Nielsen and Molich. The results can help teams find apps that perform functions that are necessary for their tasks and ensure that the selected app is the best option for the team's productivity.

KEYWORDS:
computer systems, HCI, research methods, test and evaluation, usability, software application, collaborative work, team collaboration, Web app

As technology advances at ever-increasing speeds, so does the landscape of collaborative work. People are now expected to be able to communicate and work with others, even with substantial distances between them. The Internet has driven technology-mediated collaborative work by providing a platform for the many collaboration applications that provide functions to support people while completing tasks (for definitions, see Figure 1). As a result, the market for Internet-based collaboration applications, or Web apps, that support sharing work among people has grown rapidly. This growth has led to an abundance of choices of Web apps that facilitate a number of work functions.

Given the proliferation of Web apps, choosing the most effective one can be overwhelming. Some services provide ratings and evaluations of these apps to help people choose among them. These assessments often reflect ratings based on the app's primary function. For example, Skype has often been evaluated solely as a videoconferencing tool. However, because many apps are multifunctional, it is arguably more useful to rate apps by all the functions they provide. For example, Skype has an audiovisual function for videoconferencing and a synchronous communication function for instant messaging, so it should be evaluated for each function. Furthermore, ratings of apps rarely use empirically supported human factors/ergonomics (HF/E) principles to evaluate the apps.

To help users decide which Web apps they should use for collaborative work, we evaluated 20 apps based on commonly used usability heuristics outlined in Nielsen and Molich (1990) according to each of the functions that they serve. With the results

produced by these evaluations, readers should be able to find Web apps that carry out desired functions and adhere to usability standards.

USABILITY EVALUATIONS

Usability evaluations can be conducted with a variety of methods. For our purposes, we decided that heuristic evaluation would be the best method because of the large number of apps that we were evaluating.

Selecting functions and Web apps for evaluation. The specific tasks that people perform in collaborative work are as varied as the users. However, the general functions people need to complete are often similar across tasks. In our evaluation, we focused on apps' abilities to support these work functions. The functions analyzed in this evaluation were identified through task analyses of 35 collaboration apps identified through Internet searches for collaboration apps and reviews of such apps.

To ensure that the apps we recommend are accessible to readers and easy to adopt, the following criteria were used to select Web apps for evaluation from the 35 apps found:

- *Cost:* Only free Web apps were selected for evaluation. If a Web app used "free-mium" pricing (i.e., basic functions are free, but additional functions require payment), only the free functions were considered during evaluation.
- *Web presence:* Only Web apps with the highest Web presence (i.e., that could be found through various recent sources, such as user reviews, critic reviews, and forums) were selected for evaluation. This criterion ensures that none of the apps

Collaborative Work

Work shared among a group of people

↓ is supported by

Collaboration Applications

Applications that facilitate collaboration

↓ that provide

Functions

Purpose for which an application is used

↓ that facilitate completing
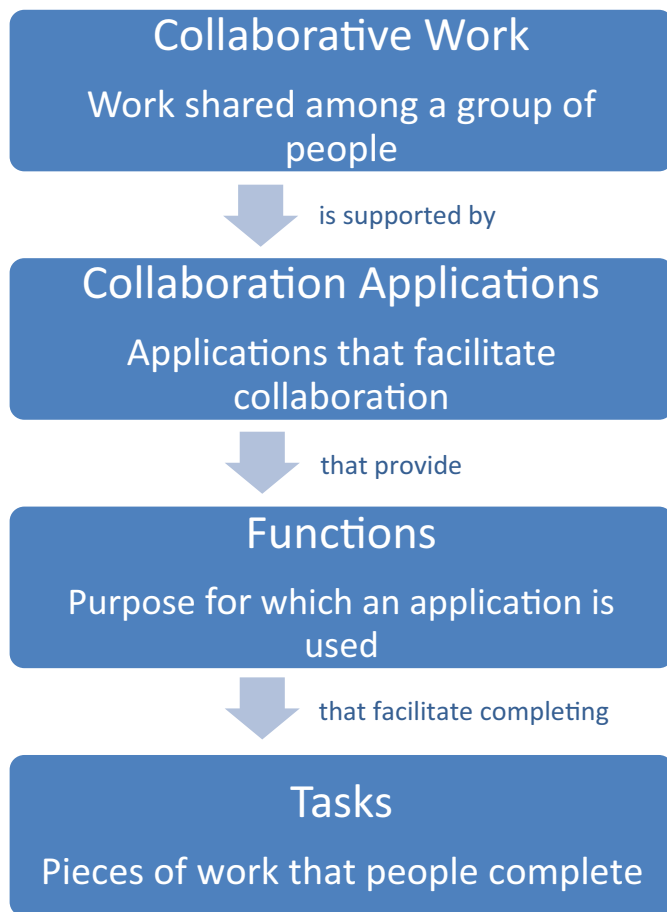
Tasks

Pieces of work that people complete

Figure 1. Relationships among and definitions of collaborative work, collaboration applications, functions, and tasks.

presented here are used obscurely, maintained sporadically, or likely to be discontinued in the future. This criterion also relies on the current market; therefore, the apps with the highest Web presence are likely those that provide the most common functions with the most ease of use. Though this type of selection has benefits, it is likely that specialty apps or recently developed apps were not included in this evaluation.

Table 1 lists the 20 Web apps selected for evaluation and the functions that they support. For apps that support multiple functions, each function was rated independently. Table 1 also includes the usability scores that each app received for each function as a percentage of points earned in the evaluation.

The 10 heuristics (Nielsen & Molich, 1990) listed in Table 2 were used to evaluate the usability of each app's functions. These heuristics allow for a qualitative analysis. A qualitative analysis was necessary for this evaluation because it provided details that users need to choose an app that is most effective for their specific tasks.

However, we also wanted to provide quantitative usability scores for each app to make comparisons between them easier. Instead of using a separate quantitative measurement,

such as the System Usability Scale (SUS) by Brooke (1996), we developed a scale for Nielsen and Molich's (1990) heuristics to maintain consistency across measures. Every function was rated on each of the 10 usability heuristics using a 4-point scale. After pilot testing with 3-, 4-, 5-, and 7-point scales, the raters found the 4-point scale to provide enough variability to distinguish between user experiences while maintaining the highest reliability across raters. The criteria were as follows:

- 4 points = no issues
- 3 points = minor issues, overall usage of app not hindered
- 2 points = issues exist that create some difficulty in using the app
- 1 point = issues that cause abandonment of the app

Conducting the evaluation. Seven raters with master's degrees in engineering psychology conducted the heuristic evaluations. The raters practiced the evaluation by collectively discussing factors that affected how Nielsen and Molich's (1990) heuristics applied to each function of the Web apps.

Next, raters individually practiced completing a few practice evaluations and compared their resulting scores in pairs. After these practice evaluations, raters conducted the usability evaluations independently. For each functional group (e.g., sharing files), one rater conducted the evaluations for all the apps within that group to ensure consistency across ratings most likely to be compared with one other. For example, all the apps that are used for sharing files were evaluated by the same rater.

Table 3 provides an example of the evaluation conducted for each function of the Web apps.

RESULTS

Table 1 summarizes the overall usability scores for functions of the apps. Additional details about the scores are available in Figures 2 through 11. The scores were calculated for each function by totaling the points on each of the 10 heuristics (see Table 2). Because 10 heuristics were scored on a 4-point scale, the maximum raw score for a function was 40 points. The usability scores are represented as percentages.

The lowest usability score across all apps was 27 out of 40 (68%), which is just below an average score of 3 out of 4, meaning that apps, at worst, had minor issues that did not hinder overall usability. This score means that the least usable app had a relatively high score. In other words, none of the apps recommended in this article is in general, difficult to use.

To help potential users select apps that best suit their particular needs, the most prominent features and limitations that are unique to each app are discussed in the "Pro" and "Con" columns. (Note: Specific points for each heuristic, like those shown in Table 3, are available by request from the authors.)

CONCLUSION

Technology-facilitated team collaboration is becoming increasingly prevalent in today's work environments.

Table 1. Usability Scores for Each Free Web App in Its Respective Functions, Based on the Usability Heuristics by Nielsen and Molich (1990)

App	Sharing Files	Draw and Whiteboard	Audiovisual	Synchronous Text	Asynchronous Text	Scheduling Calendar	Edit and Write	Polling Members	Schedule an Event	Assign Duties
30 Boxes	—	—	—	—	—	75	—	—	—	—
Agilewords	73	—	—	—	—	—	85	—	—	98
Cisco WebEx	—	78	90	95	—	—	—	—	—	—
CoSketch	—	88	—	—	—	—	—	—	—	—
Doodle Poll	—	—	—	—	—	—	—	—	90	—
DropBox	88	—	—	—	—	—	88	—	—	—
Gmail	—	—	—	—	95	—	—	—	—	—
Google Calendar	—	—	—	—	—	90	—	—	—	—
Google Hangouts	—	—	90	90	—	—	—	—	—	—
Google Drive	85	—	—	—	—	—	98	—	—	—
Pidgin	—	—	—	83	—	—	—	—	—	—
Skype	—	—	88	83	—	—	—	—	—	—
Trello	—	—	—	—	—	—	—	90	—	90
When2Meet	—	—	—	—	—	—	—	—	68	—
WhenIsGood	—	—	—	—	—	—	—	—	75	—
Windows Hotmail	—	—	—	80	88	—	—	—	—	—
Yahoo! Calendar	—	—	—	—	—	85	—	—	—	—
Yahoo! Groups	—	—	—	—	—	—	—	85	—	—
Yahoo! Mail	—	—	—	—	80	—	—	—	—	—
Zoho Docs	85	—	—	—	—	—	100	—	—	—

Note. Each function was rated across apps using a 4-point scale for each of the 10 heuristics, such that 40 points was the maximum raw score for any individual function of an app. Usability scores are shown as percentages of the maximum raw score. If the function was not evaluated for an app, the cell has a dash.

Table 2. Usability Heuristics and Their Definitions (Nielsen & Molich, 1990)

Heuristic	Definition
Visibility of system status	The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.
Match between system and the real world	The system should speak the users' language, with words, phrases, and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.
User control and freedom	Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support *undo* and *redo*.
Consistency and standards	Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.
Error prevention	Even better than good error messages is a careful design that prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.
Recognition rather than recall	Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.
Flexibility and efficiency of use	Accelerators – unseen by the novice user – may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.
Aesthetic and minimalist design	Dialogues should not contain information that is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.
Help users recognize, diagnose, and recover from errors	Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.
Help and documentation	Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, be focused on the user's task, list concrete steps to be carried out, and not be too large.

Table 3. Example Usability Evaluation for Synchronous Text Communication in App X

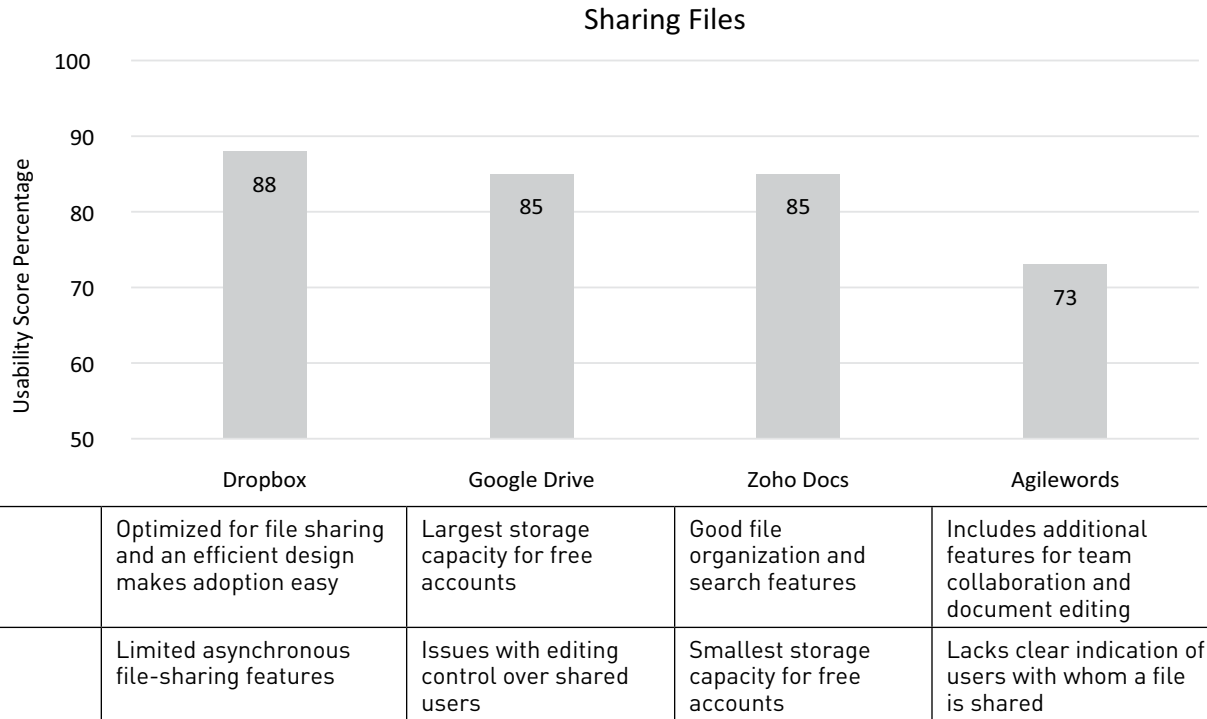| Heuristic	Score	Issues (if any)
Visibility of system status	3	No "idle/away" designation
Match between system and the real world	4	
User control and freedom	3	Deleted messages not retrievable
Consistency and standards	4	
Error prevention	3	No spell-check function
Recognition rather than recall	2	Cannot view multiple chat windows simultaneously
Flexibility and efficiency of use	3	Contacts can be sorted only in two ways (alphabetical and online/offline)
Aesthetic and minimalist design	3	Some ad clutter
Help users recognize, diagnose, and recover from errors	4	
Help and documentation	4	

Sharing Files

	Dropbox	Google Drive	Zoho Docs	Agilewords
Pro	Optimized for file sharing and an efficient design makes adoption easy	Largest storage capacity for free accounts	Good file organization and search features	Includes additional features for team collaboration and document editing
Con	Limited asynchronous file-sharing features	Issues with editing control over shared users	Smallest storage capacity for free accounts	Lacks clear indication of users with whom a file is shared

Figure 2. Usability scores for Webs apps that perform the "sharing files" function.

Drawing and Whiteboarding

	CoSketch	Cisco WebEx
Pro	Minimalist design makes adoption easy	Integrates well with audiovisual conferencing function
Con	No help feature	No warnings for deleting work and limited drawing options
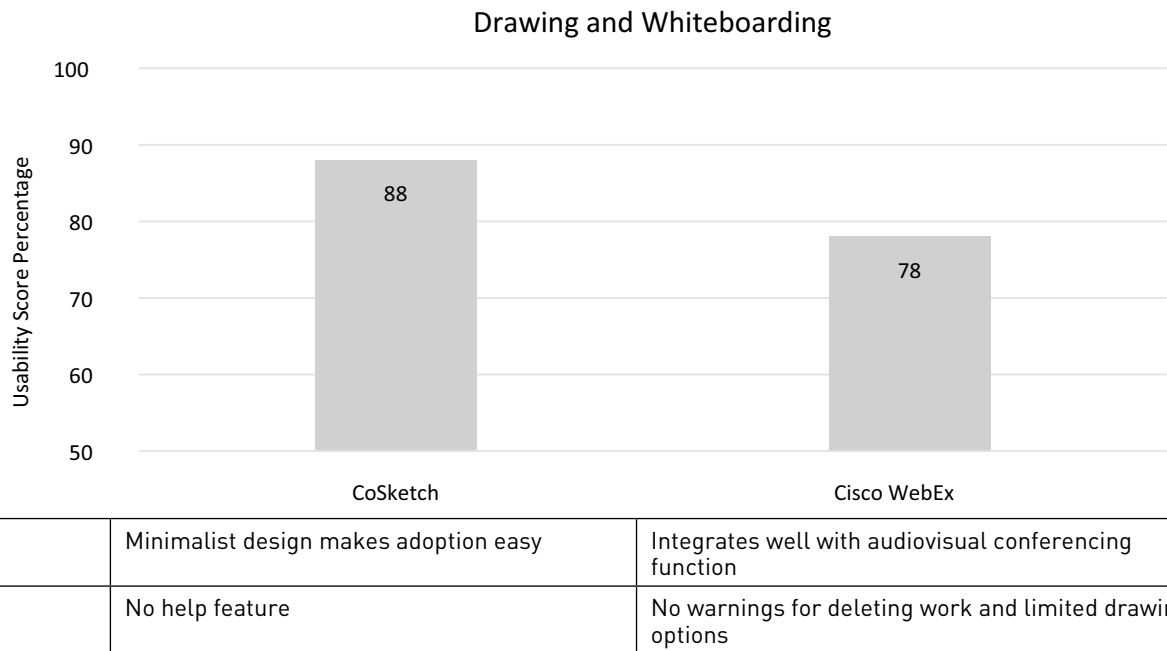
Figure 3. Usability scores for Webs apps that perform the "drawing and whiteboarding" functions.

Audiovisual Communication
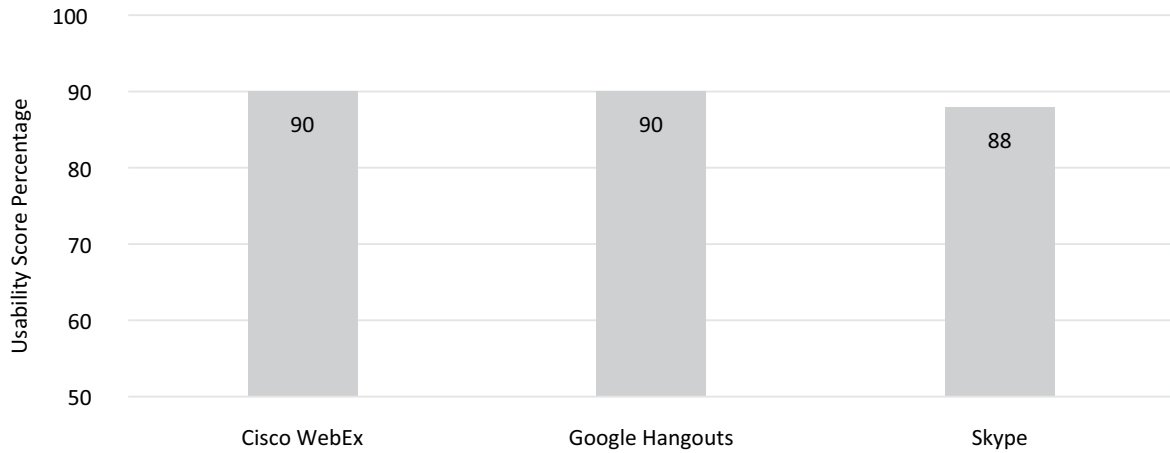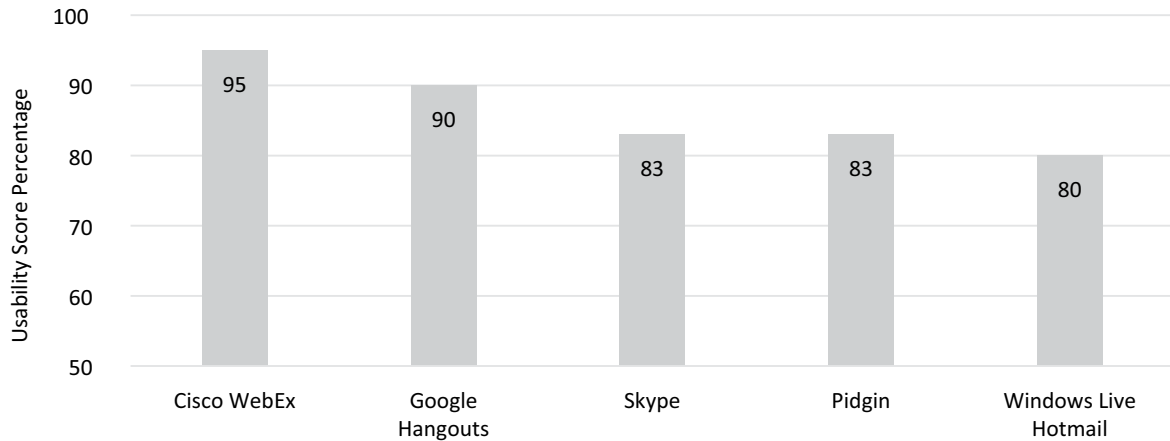

	Cisco WebEx	Google Hangouts	Skype
Pro	Intuitive interface design	Interface is aligned with what users expect from Google	Decent amount of help and troubleshooting options
Con	Host of call has large amount of control relative to other members of call	Call history not readily available	"Current speaker" designations not always clear; process of adding people to existing call is not user-friendly

Figure 4. Usability scores for Webs apps that perform the "audiovisual communication" function.

Synchronous Text Communication

	Cisco WebEx	Google Hangouts	Skype	Pidgin	Windows Live Hotmail
Pro	"Meet" and "schedule" functions are readily available from chat windows	Minimalist design; all messages logged in e-mail account	Buttons are in predictable and consistent locations	All old messages logged in some form	Aesthetic design
Con	Alert sounds are the same for incoming message and successfully sent outgoing message	Limited file-sharing capability	No "idle" designation; cannot see multiple chat windows at same time	File transfer not supported for many major platforms (e.g., Facebook)	Alerts for new messages not salient; available contacts not shown by default

Figure 5. Usability scores for Webs apps that perform the "synchronous text communication" function.

Asynchronous Text Communication
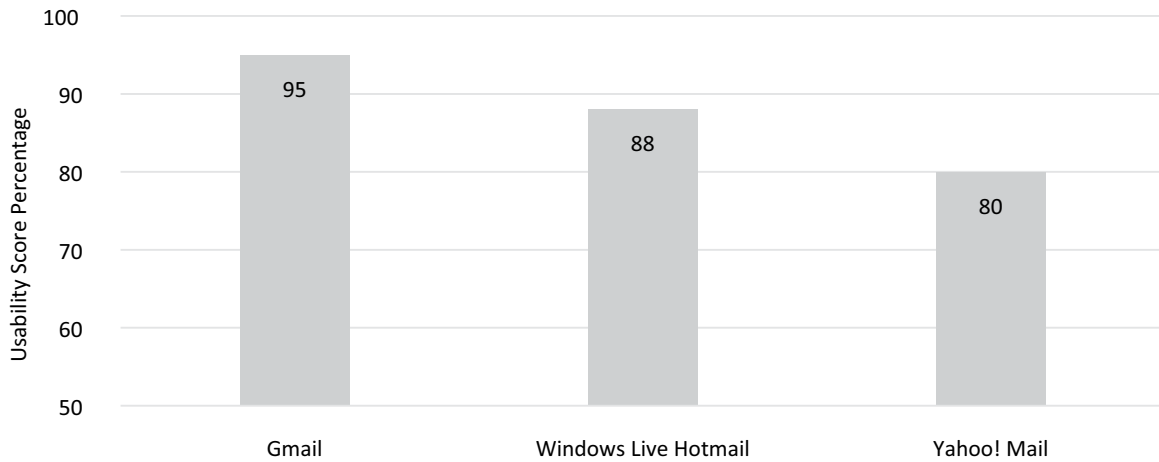
	Gmail	Windows Live Hotmail	Yahoo! Mail
Pro	High usability and functionality	High usability and functionality	Simple app that is easy to learn
Con	Service is occasionally interrupted for a few minutes	Does not provide some advanced features	Error prevention and recovery is limited

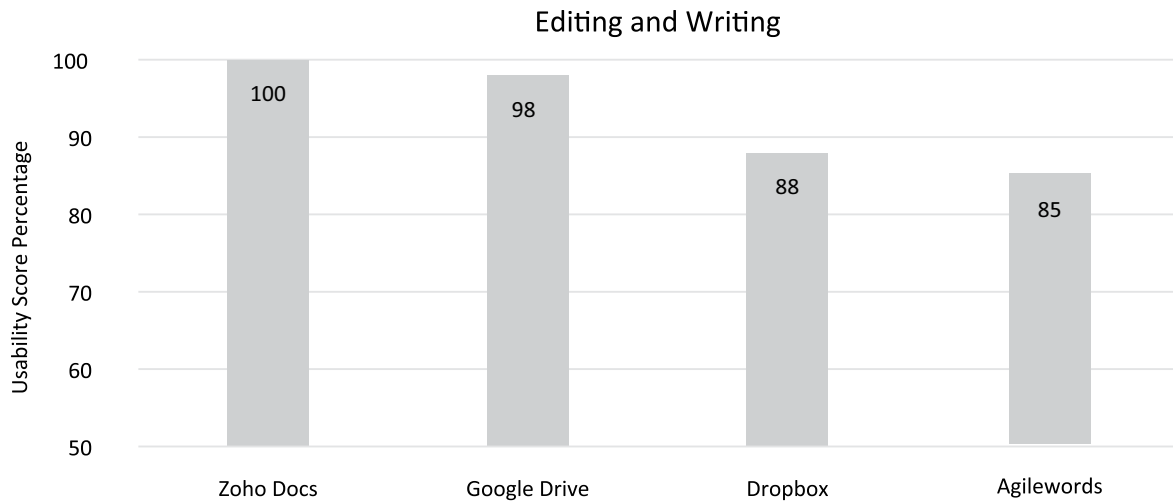Figure 6. Usability scores for Webs apps that perform the "asynchronous text communication" function.

Scheduling Calendar
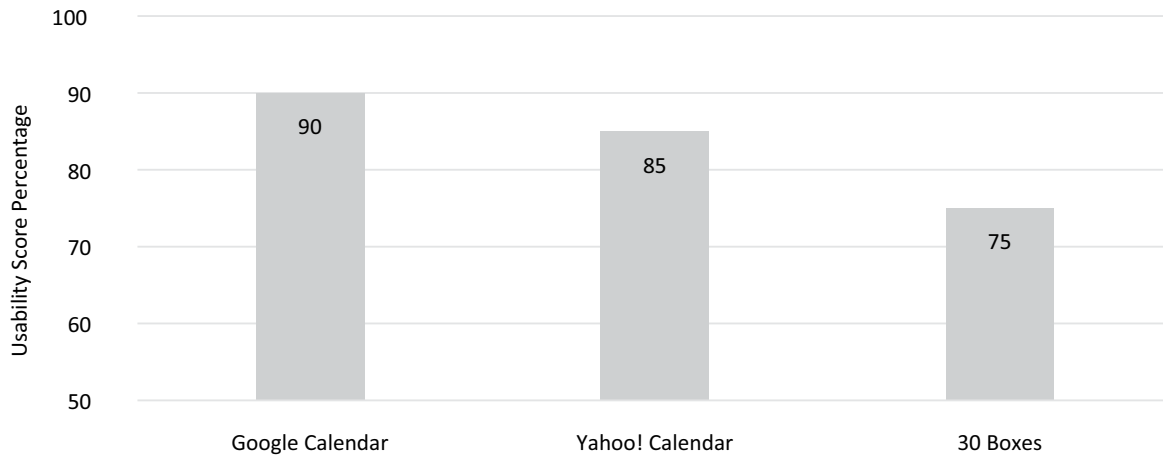
|	Google Calendar	Yahoo! Calendar	30 Boxes
Pro	Easy-to-navigate interface; shortcuts are easily memorized and applied	Easily navigated and visually pleasing design	Unique approach with dialogue-based interface
Con	Cannot cut and paste events	Standard process of confirming edits of event can get annoying	Sequence for adding events can be difficult to remember

Figure 7. Usability scores for Webs apps that perform the "scheduling calendar" function.

Editing and Writing

| Pro	Syncing abilities with other file-editing services; synchronous file editing	Synchronous file editing in addition to asynchronous file editing	Hierarchical system of files and folders allows for organization	Ability to assign different editing capabilities to different users
Con	Limited user base; some issues with importing different file types	Limited feature and file type functionality	Can be effortful to reconcile different versions of a document	Compatible only with word documents; challenges with navigation/saving documents

Figure 8. Usability scores for Webs apps that perform the "editing and writing" function.

Polling Group Members
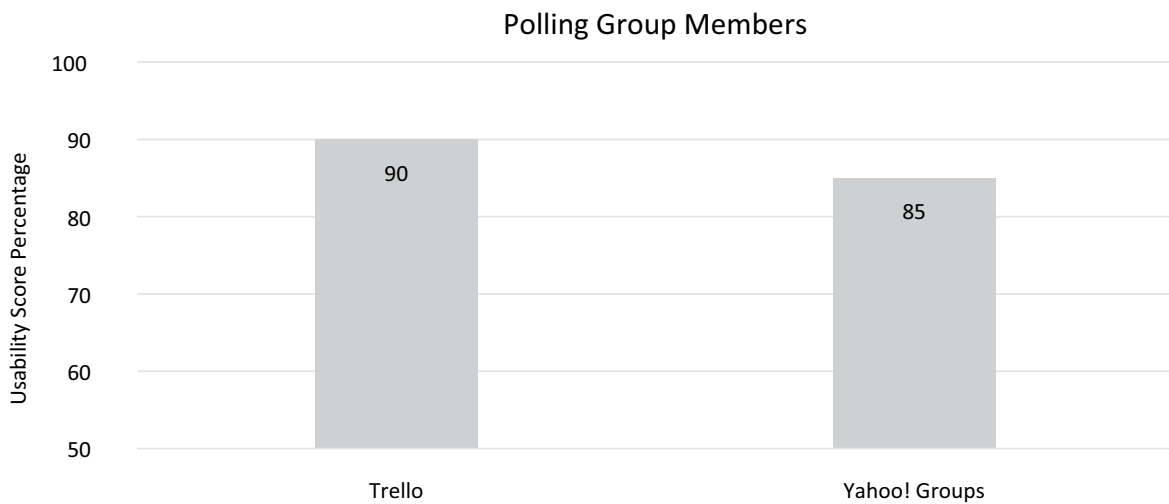
| Pro	Highly customizable	Familiar feel for Yahoo users
Con	Functionality revolves around a "cards" layout that takes time to learn	Ads are inserted in between posts, making the interface cluttered

Figure 9. Usability scores for Webs apps that perform the "polling group members" function.

Scheduling an Event
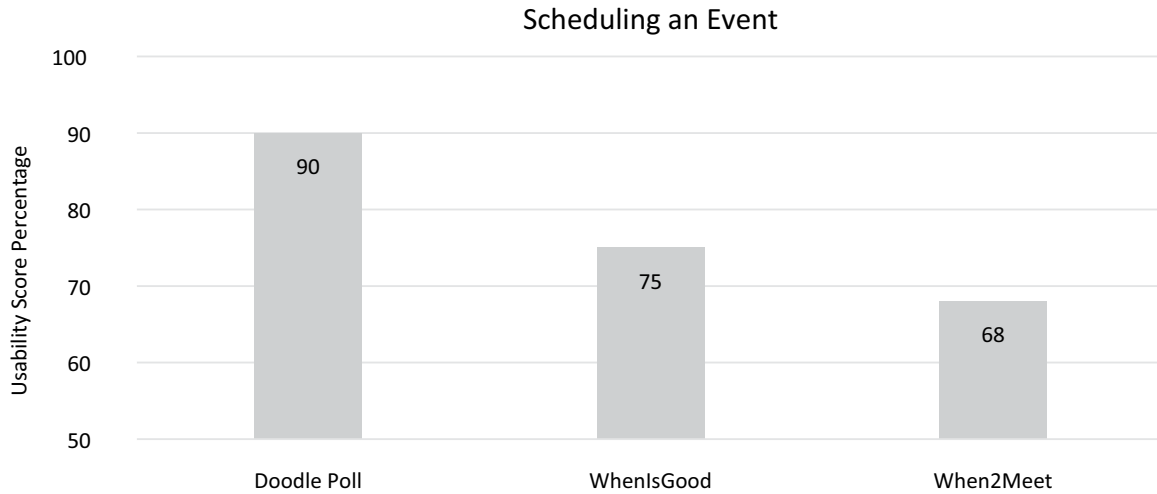
	Doodle Poll	WhenIsGood	When2Meet
Pro	Very flexible and easy-to-understand options	Ideal for large number of time slots due to ease of setup	Other people can see current results and reply based on that information
Con	Not ideal for a large number of time slots as creation process can be lengthy	Limited in customization and other options	Limited options; cannot undo or edit an event after creating it

Figure 10. Usability scores for Webs apps that perform the "scheduling an event" function.

Assigning Duties
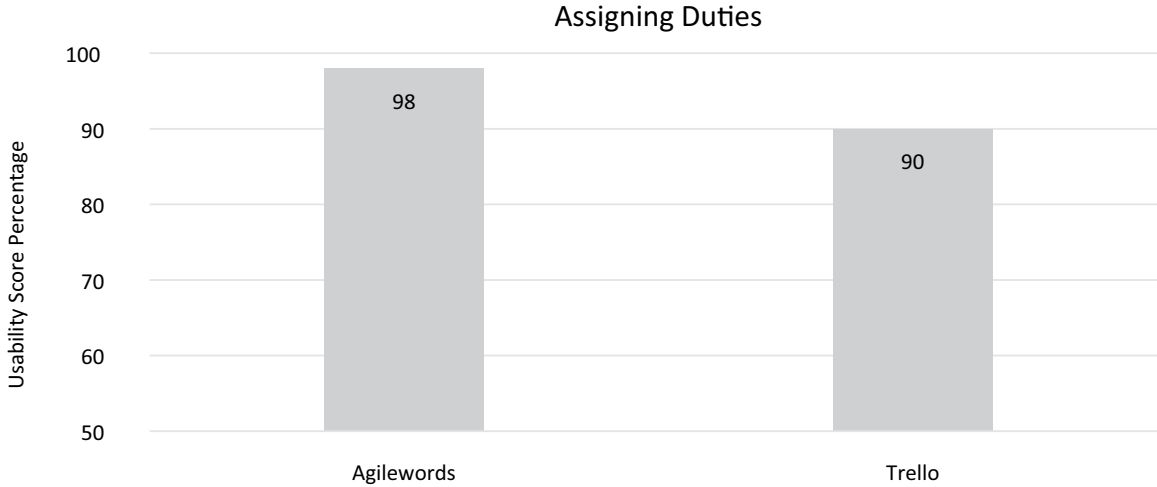
	Agilewords	Trello
Pro	Ideal for group documents because of clear visibility of edits	Drag-and-drop interface makes for easy manipulation
Con	Cannot easily move documents from one folder to another; must download and then upload to new folder	Members are not automatically notified when they have been assigned a task or the task has been modified

Figure 11. Usability scores for Webs apps that perform the "assigning duties" function.

To quickly facilitate such collaboration, teams often turn to free, Web-based apps. Because of the recent proliferation of such tools, determining which app will most effectively meet the team's needs has become a challenging task. We addressed this challenge by systematically evaluating 20 popular Web apps by first enumerating their functionality and then providing a usability score for each function. An app with strong usability features will likely be the best app for facilitating work.

The results of our analyses are presented in easily indexed charts, categorized by typical functions performed by teams. These graphics are designed to help teams choose among the range of available apps for specific collaborative functions by providing a comparison of the usability of functions across the evaluated Web apps. It is important to note that these ratings were assigned using empirically motivated HF/E heuristics, which provides scientifically backed insight into the usability of the Web apps.

Given the ongoing proliferation of Web apps, our method can be used to evaluate new tools against existing tools. This method could be easily applied to other types of Web apps, regardless of the specific function of the apps. Authors of future research could apply our methods to recommend apps beyond only those that are collaboration focused.

When a team is deciding on whether or not to adopt an online app to facilitate team collaboration, the usability of the app may not be the only consideration. For example, the compatibility of the app with other technologies may be important. However, we propose that using our ratings to select apps from a plethora of options will help ensure that the selected app will contribute to, rather than detract from, effective communication across team members.

REFERENCES

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). London, UK: Taylor & Francis.

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. In *Proceedings of the CHI'90 Conference on Human Factors in Computing Systems* (pp. 249–256). New York, NY: ACM Press.

Lauren E. Margulieux is a doctoral student in the engineering psychology program at the Georgia Institute of Technology. She received her master of science degree in engineering psychology from Georgia Tech. Her research interests include educational technology, blended and flipped instruction, and online learning, particularly for higher education in the areas of science, technology, engineering, and mathematics (STEM).

Dar-Wei Chen is a doctoral student in the engineering psychology program at the Georgia Institute of Technology. He received his master of science degree in engineering psychology from Georgia Tech. His research interests include interface design, metacognition, multimedia learning, and interactive materials.

Joseph D. McDonald is a doctoral student in the engineering psychology program at the Georgia Institute of Technology. He received his master of science degree in engineering psychology from Georgia Tech. His research interests include understanding how organizational, team-level, and individual-level factors interact with technology to bring about system-level changes.

Keith R. Bujak is a senior user experience engineer for the John Deere Technology Innovation Center. He received his PhD in engineering psychology from the Georgia Institute of Technology. His research interests include mobile app development for logistics systems, the applications of wearable technology, and training simulations for workforce improvement.

Thomas M. Gable is a doctoral student in the engineering psychology program at the Georgia Institute of Technology. He received his master of science degree in engineering psychology at Georgia Tech. His research interests include multimodal displays, multitasking, driving, human–computer interaction, and usability.

Cale M. Darling is a doctoral student in the engineering psychology program at the Georgia Institute of Technology. He received his master of science degree in engineering psychology from Georgia Tech. His research interests include situation awareness, visual attention, and display design.

Laura M. Schaeffer is a doctoral student in the engineering psychology program at the Georgia Institute of Technology. She received her master of science degree in engineering psychology from Georgia Tech. Her research interests include instructional design, the role of technology in education, and aiding students during problem solving.

Laura H. Barg-Walkow is a doctoral student in the engineering psychology program at the Georgia Institute of Technology. She received her master of science degree in engineering psychology from Georgia Tech. Her research interests include medical human factors, interruptions, errors, and training, particularly in the domain of health care.