What Makes A Good Reference Manager? A Quantitative Analysis of Bibliography Management Applications

Tongan Cai*
Chacha Chen*
Ting-Hao (Kenneth) Huang
Frank E. Ritter
College of Information Sciences and Technology, Pennsylvania State University
University Park, Pennsylvania, USA

ABSTRACT

Many researchers and students use reference managers to collect, manage, and format references and citations. While prior work has assessed these tools qualitatively, it is still unclear how to quantitatively evaluate reference managers. This paper starts to quantify the user effort required to use reference managers. We first collected surveys from 69 graduate students to understand their experience with reference managers, and then conducted user studies with 12 participants. In our study, each participant was asked to perform a standardized task using four popular reference managers: Mendeley, Zotero, EndNote, and RefWorks. We used RUI, a keystroke and mouse-move logger, to record the participants' activities and approximate their physical and mental effort. We also used pre- and post-study surveys to collect users' feedback and self-reported task load (as expressed by the NASA TLX Index.) The results showed that different reference managers require different levels of effort, and users generally prefer the tools that involve less effort. We also found that although reference managers share similar features, differences in presentation and organization matter. We conclude this work by providing a set of guidelines for both users and developers of reference managers.

CCS CONCEPTS

• **Human-centered computing** → **Usability testing**; *Activity centered design.*

KEYWORDS

reference managers, task analysis, keystroke, mouse click, mental effort, physical effort

ACM Reference Format:
Tongan Cai, Chacha Chen, Ting-Hao (Kenneth) Huang, and Frank E. Ritter. 2021. What Makes A Good Reference Manager? A Quantitative Analysis of Bibliography Management Applications. In *Asian CHI Symposium 2021 (Asian CHI Symposium 2021), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3429360.3468183

*Both authors contributed equally to this research. The order was chosen by coin flip.

Asian CHI Symposium 2021 , May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8203-8/21/05…$15.00
https://doi.org/10.1145/3429360.3468183

1 INTRODUCTION

Manually collecting and tracking citations and references in academic papers can be time-consuming and tedious. Researchers and students use reference managers to simplify these tasks. A reference manager is a computer application that helps users collect, manage, and format references and citations for academic purposes. Reference managers are often optimized for writing literature reviews [9]; for example, they can make cited papers' PDF files instantly viewable and searchable [10].

Which reference manager is the "best"? It is not always easy for users to select the best tool [6]. One popular source of references is the online ratings. The website G2[1] claims that the most popular reference manager is Mendeley, based on 170 reviews and a 4.3/5 rating. The editor of another website, Scribendi[2], asserts RefWorks as the top pick. Some work has also studied reference managers' impacts on their users' resulting texts, including the references and articles, both qualitatively [11, 12] and quantitatively [7]. However, the literature has little to say about how *users* interact with reference managers.

This paper demonstrates our preliminary study that compares four commonly used reference managers: Mendeley, Zotero, EndNote, and RefWorks. In particular, we measured the different levels of effort required to operate these tools and correlate the levels of effort to the user's preference. We first collected surveys from 69 graduate students to understand their experience with reference managers, and then conducted user studies with 12 participants. In our study, each participant was asked to perform a standardized task using four popular reference managers: Mendeley, Zotero, EndNote, and RefWorks. We used a keystroke and mouse-move logger to record the participants' activities and approximate their physical and mental effort. We also used pre- and post-study surveys to collect users' feedback and self-reported task load (as expressed by the NASA TLX Index [4].) The results showed that different software requires different levels of effort, and users generally prefer the tools that involve less effort. We also found that although reference managers share similar features, differences in presentation and organization matter. We conclude this work by providing a set of guidelines for both users and developers.

[1] G2: https://www.g2.com/categories/reference-management
[2] Scribendi:https://www.scribendi.com/advice/reference_management_software_solutions.en.html

The major contributions of our work are that:

- We explore an evaluation scheme for reference managers that combines qualitative evaluation of the software's features and functionalities with quantitative evaluation of a user's effort during a routine task, and associate the results with user preference. We believe this scheme can be applied to similar usability analysis or software evaluation tasks.
- Based on our results, we provide suggestions on how to choose a reference manager.
- We also create guidelines for developers to improve reference managers in the three most important areas to users: accuracy, effort, and functionality.

2 RELATED WORK IN COMPARATIVE STUDIES OF REFERENCE MANAGERS

Most previous work evaluating reference managers conducted surveys to compare user opinion on different managers. Lorenzetti and Ghali's [9] survey on reference manager usability emphasized ease-of-use issues when integrating the software with other programs and the sharing of reference databases among researchers. Some reports compared functional features for different reference managers, including device compatibility, system requirements, and price [14]. Zhang adopted a similar idea [15], in which each tool was analyzed in terms of its features in collaboration as well as accessing, collecting, organizing, citing, and formatting citations. Hensley's work [5] compared the benefits and drawbacks of each program from a librarian's perspective. Other works created evaluation metrics, such as error rate in importing/exporting [2] and average importing time for bibliography entries [13]. Basak [1] quantified and compared how well different programs imported fields from Google Scholar, and visualized the data with radar plots. However, as discussed in the Introduction, few reports have considered the amount of effort required for users to complete a routine task in reference managers software.

3 PRE-STUDY SURVEY

To begin our analysis of reference managers, we conducted a pre-study survey[3] acted as a qualitative pre-study to the user study that follows and to identify use cases and potential issues for current reference managers. Respondents were recruited using snowball sampling[4]. We asked a set of demographic questions (level of study, experiences with reference managers, etc.) and only included those who had previous experience with reference managers.

3.1 Survey Results and Discussion

3.1.1 Demographics. A total of 69 participants completed our pre-study survey. The majority of the participants were graduate students (15% bachelor / 29% master / 56% doctoral). 61% of the participants had previous experience with reference managers. We include all the results from the participants.

3.1.2 Use Cases. In our pre-study survey, we started by asking a free-form question on the most common use cases of reference managers if they have used them before. By examining the reported use cases, we found that common uses of reference managers included research purposes, class projects, paper organization, citation organization, reading papers, and sharing papers. The majority (72%) of reported use was for research.

3.1.3 Previous Experience of Reference Managers. Common reference managers include Mendeley, EndNote, Zotero, RefWorks, and EasyBib. The top four reference managers in our study, in order of user preference, were Mendeley (38%), Zotero (27%), EndNote (20%), and RefWorks (5%).

3.1.4 Usage frequency. 40% of our survey respondents used reference managers several times a week, and 17% used them every day. Our survey results also shows that people were satisfied with their current reference managers. They believed their preferred reference manager to be fast, efficient, and easy to use, as shown in Figure 1(b).

3.1.5 Functionality. We show the functionalities that survey respondents cared about most in Figure 1(c).

4 USER STUDY

To quantitatively investigate the efficiency of the reference managers, we conducted a user study to analyze the physical and mental effort users expended when using those applications to perform a routine reference-management task. This portion of the study involved 12 participants who volunteered. The user study went through an ethical review and was IRB-approved.

4.1 Investigated Reference Managers

The user study investigated reference management applications Mendeley, Zotero, EndNote and RefWorks[5].

- **Mendeley** is a free reference manager by Elsevier. In 2018, it was estimated that there are already 5,000,000 users of Mendeley[6]. During our user study, we observed that the auto-completion of information is not always correct and can be misleading sometimes. For newer conference papers or preprint papers, Mendeley consistently failed to find the correct information from its database, which results in errors.
- **Zotero** is also a free, open-source reference management application by the Corporation for Digital Scholarship. It has similar workflow logic as Mendeley, but appears to be better at importing and indexing PDF inputs. Zotero creates new entries with very good accuracy and efficiency. Users can also have personal accounts for syncing and managing references.
- **EndNote** is a commercial, closed-source reference management application by Clarivate Analytics. EndNote has its own file format (*.Data and *.enl) users can use to import/-export. It also has a "database search" integration feature in its full version. It integrates with Google Scholar such that references can be imported in one click.

[3]The full questionnaire can be found in the Supplementary Material.
[4]We emailed the survey to direct contacts and invited them to pass the survey on to friends. Responses are from STEM-major students at US universities.

[5]Software versions employed in this study include Mendeley 1.19.4, Zotero 5.0.94, EndNote X9.2, and the web-version of RefWorks.
[6]https://www.elsevier.com/connect/ten-years-of-mendeley-and-whats-next

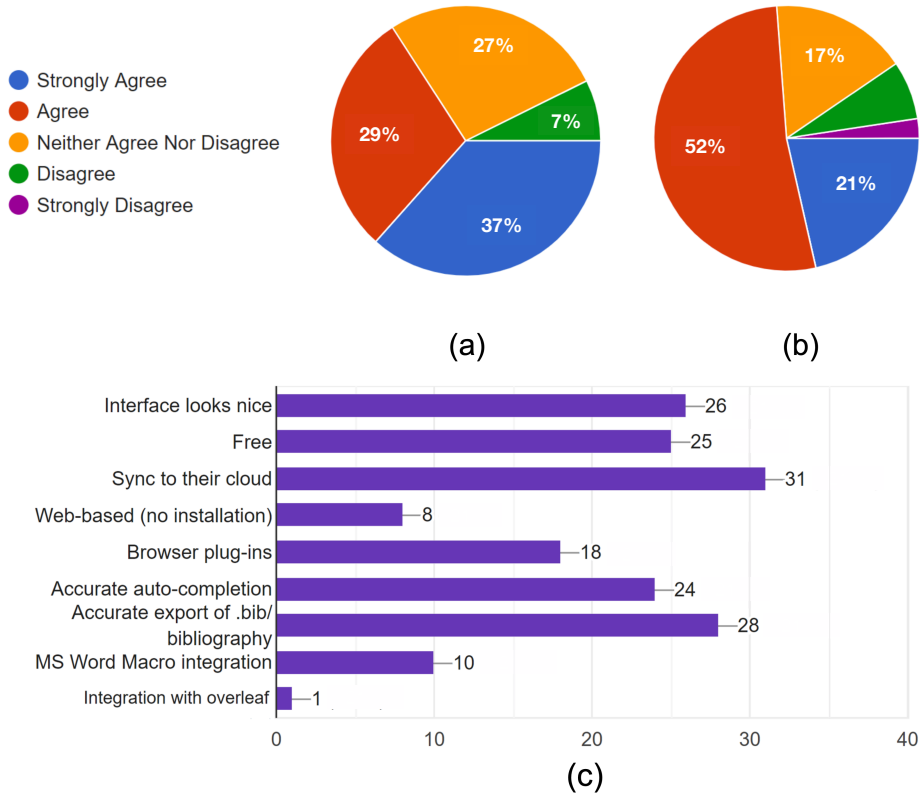

Figure 1: Pre-study survey: (a) How much do you agree with this statement: "My preferred reference manager is fast and efficient." (b) How much do you agree with this statement: "My preferred reference manager is easy to use without hassles." (c) "How do you tell if a reference manager is good? Select all that apply."

- **RefWorks** is a web-based reference management application produced by ProQuest that requires an institutional subscription. RefWorks provides strong auto-completion, requiring only the title information of the paper, and does not entail any installation. Also, the BibTeX export of RefWorks can be directly copied to the clipboard, which is very efficient for LaTeX users.

4.2 User Study Design

We designed the following tasks, which emulate the steps to create a bibliography from a set of titles, to analyze the physical effort users expend to perform the task in each of the four applications[7].

Goal. The participant was given a list of titles and the corresponding PDF files without author or publisher information elaborated. The goal was to create a full bibliography in a correct bibliography format with accurate information. The steps of the process are:

(1) The participant imported the PDF files to the library and used the reference manager's auto-complete function to fill in only the title of each source. (This method worked for Mendeley and RefWorks.)
(2) After obtaining the references, the participant checked that the information provided by the reference manager was accurate. If not, they made corresponding changes for the entries (date format/journal name/author name).
(3) The participant output a bibliography list of papers in a .bib file that included every reference.

Participants performed the task in each of the four programs, where the order of programs was randomized for each participant. Each participant was asked to follow the required task and operate the application on a laptop running Windows 10 while their keystrokes, button presses, and mouse moves were recorded by the RUI (Recording User Input) keystroke logger [8]. To calculate mouse movements, we adopted a sum of Euclidean distances between each recorded move.

4.3 User Study Analysis

As the 12 participants performed the task, RUI recorded the mouse coordinates of the mouse on the user's screen as well as each keystroke and mouse click. To represent the number of operations users took to perform the task, we report the total mouse distance (MDistance, in pixels), number of mouse clicks (MClicks), and the total time elapsed (Time). We present the results of RUI for different reference managers in Table 1. Each cell contains the averaged

[7]We specified different detailed instructions for each reference management application. Full instructions for each application can be found in the Supplementary Material.

Table 1: The results of RUI for different reference managers. We measured the physical efforts by Mouse distance (MDistance, $x10^3$ Pixels), number of mouse clicks (MClicks, clicks), and elapsed time (Time, seconds), respectively. Corresponding standard deviations are also reported. Lower values indicate better efficiency. The lowest values in each column are in bold.

	MDistance	MClicks	Time
Mendeley	85.5 ± 30.00	109.33 ± 36.98	325.96 ± 77.03
Zotero	**57.6 ± 26.30**	67.25 ± 26.62	**251.21 ± 92.17**
EndNote	112.5 ± 27.40	138.50 ± 37.89	408.12 ± 44.67
RefWorks	65.8 ± 22.40	**62.45 ± 20.02**	294.81 ± 37.92

statistics (mean) we obtained through the RUI recordings, and we also calculate the corresponding variance for each statistics.

Overall, Zotero required substantially less mouse distance and elapsed time, and RefWorks was the best with respect to number of mouse clicks. Both Zotero and RefWorks were efficient in terms of auto-indexing the PDF files with accurate information extraction.

Mendeley had an average performance on all three measures. Interestingly, EndNote required the most actions and time for the given task because it does not automatically extract information from uploaded PDFs. Instead, it requires the user to manually modify the information needed for our task.

4.4 Mental Effort Analysis with NASA TLX

In addition to the performance measures, we also asked the 12 user-study participants to self-evaluate the task load of each reference manager using the NASA TLX seven-point Likert scale across six categories [3, 4]. We asked the participants to rate workload based on their experience after using each reference manager. Figure 2 summarizes the results.

For all six aspects evaluated, the order is consistent: Zotero has the lowest reported workload, RefWorks the second, Mendeley the third, and EndNote the highest. The reason for the ordering might be the automatic indexing PDF feature for Zotero and RefWorks. Mendeley constantly failed to extract accurate information from PDFs, which necessitated manual correction for bibliography entries. EndNote does not support information extraction from PDFs, so it is the most demanding for our given task.

4.5 Retrospective Survey

We conducted an exit survey of the user-study participants to better understand their experiences with all four reference managers. Interestingly, more than a half of the participants (58%) changed their mind from the pre-study survey report of their favorite reference manager. In the pre-study survey, those participants initially preferred EndNote and Mendeley (see Table 2). In the retrospective survey, they preferred Zotero and RefWorks.

In terms of speed, 33% of the participants thought current reference managers are slow and should be improved. Before the their exposure to all four reference managers, 55% of the pre-study survey participants reported that adding a new citation takes less than a minute (which is generally not the case), and 14% of the participants had no idea about how long this would take. After participating the user study, 58% of the participants reported that it took approximately one to two minutes to add a new citation for users using their preferred reference managers. 33% of the participants agreed that the current efficiency of reference managers is low and shall be improved, yet 42% of the participants thought the program were slow but okay, while 17% disagree. Those who disagree think efficiency matters and should current reference managers should be improved. Others (8%) indicated that there are more important issues (correctness) than efficiency.

Table 2: Users' preferred reference manager before (left) and after the user study

	Before	After
Mendeley	38%	8%
EndNote	50%	8%
Zotero	12%	34%
RefWorks	0	50%

We calculated the correlation coefficients between six NASA indices and three RUI measurements and found very high correlations between the variables (Figure 3), which supports the validity of our user-effort measurement approach.

5 IMPLICATIONS AND GUIDELINES

Based on our study, we would like to provide suggestions to users on choosing a suitable reference manager. For developers, we propose guidelines on how to improve the efficiency of the current reference managers.

5.1 Choosing a Reference Manager

In Table 3, we summarize the important features for commonly used reference managers in terms of five main categories: import file, database connectivity, complete and correct basic information, export bibliography, and operating system support.

All of the reference managers show complete support for import file formats. Zotero's diverse database connection and strong PDF indexing ability result in fast and accurate performance in our user study, which will suit users who have a large number of PDFs to import. If the user only has titles available (as opposed to files), RefWorks' auto-completion works elegantly while maintaining fewer errors and mismatches than Mendeley. Both Mendeley and Zotero support group sharing of libraries via email invitation, but RefWorks needs organization authorization, which is not the most ideal tool when collaborating.

In the following, we further list several factors that the participants mentioned in their responses regarding selecting reference managers.

Cost. If "Free" is a must, EndNote and RefWorks might not be the choice.

OS and Devices Compatibility. All four applications support Windows and OS X operating systems, but only Mendeley and Zotero can be used in Linux. As for mobile applications, Mendeley and EndNote provide an iOS apps. Only Mendeley provides an app for

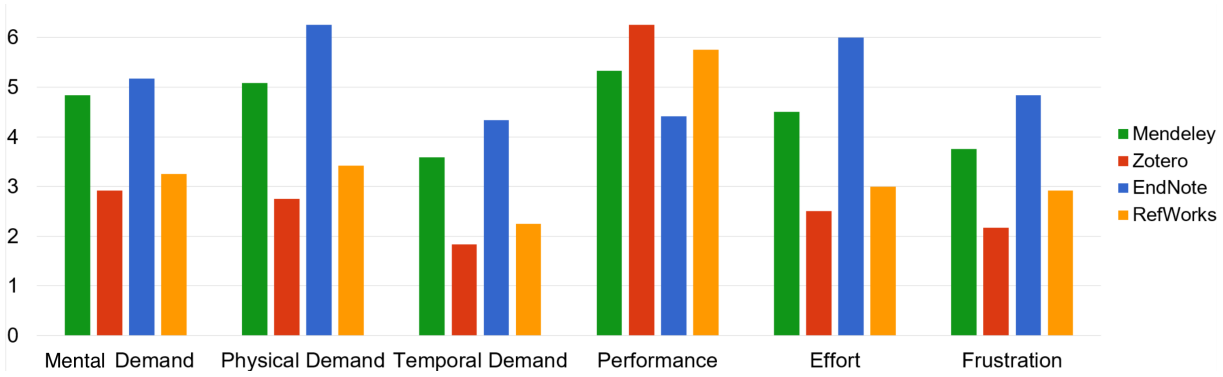

Figure 2: Participants' ratings on the NASA TLX workload index on a 1-7 scale. Lower values correspond to less demand, effort, and frustration. On the Performance scale, the higher the better. In general, Zotero and RefWorks have lower workload demand.

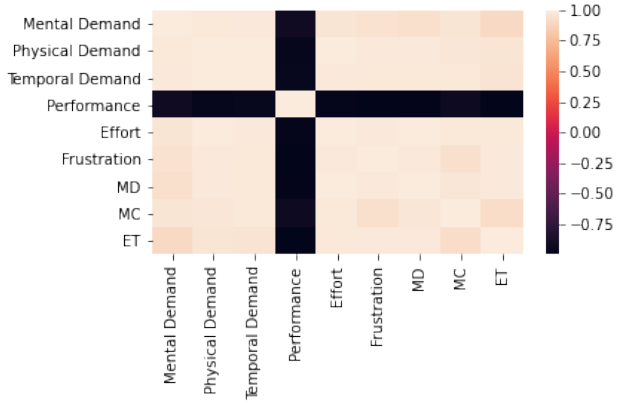

Figure 3: The covariance matrix heatmap between measures (NASA TLX indices and RUI results). Note that the performance has high negative correlation with others because higher performance score indicates better performance, while lower scores for other metrics indicates lower efforts and demands .

Table 3: Comparing functionalities and features among applications.

Task	Mendeley	Zotero	EndNote	RefWorks
Import File				
Manually Add	•	•	•	•
Add Entry from PDF	•	•	•	•
Import from .bib/.RIS etc	•	•	•	•
*Browser Plug-in	•	•	•	•
Database Connectivity				
ArXiv	•	•	•	
CiteSeer	•	•	•	
IEEE Xplore		•	•	
PubMed	•	•	•	•
Complete & Correct Basic Information				
Manually Change	•	•	•	•
Online Detail Filling	•			•
PDF information extraction	•	•		•
Export Bibliography				
Different formats	•	•	•	•
Export .bib	•	•	•	•
*MS Word Integration	•	•	•	•
Operating System Support				
Windows	•	•	•	•
MacOS	•	•	•	•
Linux	•	•		
iOS App	•		•	
Android App	•			

Android. For mobile users or cross-platform users, RefWorks is highly recommended.

Local v.s. Web-Based Software. From the responses collected in the pre-study survey, 17% of our participants said they mainly use reference managers to store their papers and work as PDF readers. These users tend to prefer local software over web-based ones.

Integration with Other Applications. Other strategies users may adopt to improve efficiency include (but are not limited to) the Chrome plug-ins for Zotero and Mendeley, Google Scholar one-click import for all four applications, and citation integration provided by major publishers on the article's webpage. These strategies help import new reference entries more accurately, efficiently, and easily.

5.2 Guidelines for Developers

Based on our results, we have developed guidelines for the development of reference managers concerning four interconnected aspects: *(i)* accuracy , *(ii)* user effort, *(iii)* functionality, and *(iv)* promotion.

Accuracy. In our user study and retrospective post-study survey, participants identified errors and mismatches during the import process as their main issue. These errors greatly slow the process and appears to impact the user's trust in that tool. Also, the auto-completion and online search functions sometimes redirect to a prior version or to other papers with very similar names, which requires users to manually correct the entries. Developers should prioritize accuracy when developing PDF indexing and matching protocols. Challenges to tackle include the citation of pre-print

papers, papers with the same titles by different authors, and differentiating conference proceedings from journal papers. Besides keeping back-end query sources up to date, if the algorithm reports low confidence in the result, it should ask users to step in and make sure the details are correct.

User effort. Both the results of RUI tracking and the retrospective survey show that users tend to prefer the application that takes the least physical effort to operate (Zotero). Based on our observations, the major discrepancy comes from the importing step and the correction of errors. While step-by-step instructions and options can benefit new users, experienced users prefer to achieve the goal within minimal steps and effort. Instead of providing all the options in the right-click drop list, simplified lists will help users navigate more quickly. Features like "one-click" exports also increase efficiency for experienced users who know what they are doing. It is also a good idea to keep all the processes in one interface, as switching between different interfaces increases both physical and mental effort, and slows the process.

Functionality. These four reference management applications have very similar design in features and functionalities, but vary in user interfaces. Users can import a reference entry in Mendeley, Zotero, and RefWorks by dragging a file to the interface. However, all participants experienced a longer wait time when using Mendeley. Under non-ideal connection bandwidth, RefWorks also takes a lot longer to respond. Zotero finishes in seconds.

To suit the needs of users who only use reference managers as a PDF manager/reader, developers should improve the built-in PDF readers to support better labeling, commenting, and highlighting tools. Other important features include Google Scholar integration, team collaboration, and multi-device/multi-OS support. It is never a bad idea to support these features, but accuracy and execution time should be prioritized so new functionalities do not harm overall efficiency.

Promotion. Given the chance to experiment with four applications, 58% of the user study participants changed their program preference. We conclude that the variation in user preferences can result from unfamiliarity with other reference management applications. One of the participants asserted before the study, "I have been using EndNote for many years. It's not free, but convenient to use." After trying the four applications, he reported, "Zotero and RefWorks are so good! I'd probably stop my EndNote subscription." Promotion should allow users to try the application, either by providing tutorials for institutional subscribers or online sessions for novices, instead of mere advertisements.

6 LIMITATIONS

We recognize the limitations of this work. Our participants do not represent the range of users of the applications. Senior graduate students, research professors, and librarians can be expected to use such applications much more heavily than others, and their feedback could lead to different study results. The measurement of the physical effort of users is based on a small, simple task which may not comprehensively reveal the applications' full functionalities or all users' needs. More advanced measurements should be developed to provide more insights into the applications' usability.

7 CONCLUSION

Reference managers are widely used in academia and the research community to support the query and management of references, and is of great convenience in the development of manuscripts. This paper started to develop a way of quantitatively evaluating the quality of reference management applications to correlate the users' physical and mental efforts in using these applications to their preference for such applications. From that data, we provided suggestions to users on how to select a suitable reference manager and guidelines to developers on how to improve reference managers.

The proposed method improved the traditional three-stage method (pre-study survey/participant study/post-study survey) by asking participants to estimate their mental workload in the retrospective survey. The idea of evaluating the users' mental effort for a specific task is, to the best of our knowledge, not widely adopted in the comparative study of applications, and we believe this method can reveal more accurate feedback from participants. This approach can be applied to a wide range of applications, and this report can serve as a template.

REFERENCES

[1] Sujith Kumar Basak. 2014. A comparison of researcher's reference management software: Refworks, Mendeley, and EndNote. *Journal of Economics and Behavioral Studies* 6, 7 (2014), 561–568.

[2] Ron Gilmour and Laura Cobus-Kuo. 2011. Reference management software: A comparative analysis of four products. *Issues in Science and Technology Librarianship* 66 (2011), 63–75.

[3] Sandra G Hart. 2006. NASA-Task Load Index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[4] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Psychology*. Vol. 52. Elsevier, 139–183.

[5] Merinda K. Hensley. 2011. Citation management software: Features and futures. *Reference & User Services Quarterly* 50, 3 (2011), 204–208.

[6] Camille Ivey and Janet Crum. 2018. Choosing the right citation management tool: Endnote, Mendeley, RefWorks, or Zotero. *Journal of the Medical Library Association: JMLA* 106, 3 (2018), 399.

[7] Jiří Kratochvíl. 2017. Comparison of the Accuracy of Bibliographical References Generated for Medical Citation Styles by EndNote, Mendeley, RefWorks and Zotero. *The Journal of Academic Librarianship* 43, 1 (2017), 57–66. https://doi.org/10.1016/j.acalib.2016.09.001

[8] Urmila Kukreja, William E. Stevenson, and Frank E. Ritter. 2006. RUI: Recording user input from interfaces under Windows and Mac OS X. *Behavior Research Methods* 38, 4 (2006), 656–659. https://doi.org/10.3758/BF03193898

[9] Diane L Lorenzetti and William A Ghali. 2013. Reference management software for systematic reviews and meta-analyses: an exploration of usage and usability. *BMC Medical Research Methodology* 13, 1 (2013), 1–5.

[10] Thomas L Mead and Donna R Berryman. 2010. Reference and PDF-manager software: complexities, support and workflow. *Medical Reference Services Quarterly* 29, 4 (2010), 388–393.

[11] Lambodara Parabhoi, Arabinda Kumar Seth, and Sushanta Kumar Pathy. 2017. Citation management software tools: A comparison with special reference to Zotero and Mendeley. *Journal of Advances in Library and Information Science* 6, 3 (2017), 288–293.

[12] Kisman Salija, Rahmat Hidayat, and Andi Anto Patak. 2016. Mendeley impact on scientific writing: Thematic analysis. *International Journal on Advanced Science, Engineering and Information Technology* 6, 5 (2016), 657–662.

[13] S. Santharooban and Lavanya Lavakumaran. 2018. A comparative analysis of reference managers: EndNote and Mendeley. In *Proceedings of the 2018 International Conference of University Librarians Association of Sri Lanka*. 96–107.

[14] Jesús Tramullas, Ana Isabel Sánchez-Casabón, and Piedad Garrido-Picazo. 2015. Studies and analysis of reference management software: A literature review. *El profesional de la información* 24, 5 (2015), 680–688.

[15] Yingting Zhang. 2012. Comparison of select reference management tools. *Medical Reference Services Quarterly* 31, 1 (2012), 45–60.