

Blind and Visually Impaired Users: A methodological reflection on a large-scale project

Iris Xie¹, Shengang Wang², Hyun Seung Lee³, Tae Hee Lee⁴

^{1,2,3,4}University of Wisconsin-Milwaukee

Abstract

The purpose of this study is to review and reflect upon the qualitative and quantitative methods applied and lessons learned in the process of completing the project “Creating accessibility and usability of digital library (DL) guidelines to support blind and visually impaired (BVI) users.” This study focuses on the following issues: 1) The recruitment of BVI users, scholars/experts, and DL developers and associated challenges and solutions; 2) The use of data collection methods for diverse DL stakeholders and associated challenges and solutions; 3) The application of data analysis techniques and associated challenges and solutions. It is a challenge to recruit large numbers of BVI participants, and it is an effective approach to establish connections with trusted organizations and institutions - particularly BVI-related organizations - to promote the studies. Multiple data collection methods (questionnaires, think-aloud protocols, transaction logs, diaries, interviews, and focus groups) were used throughout the research project. As BVI users utilize screen readers to understand DL interfaces and simultaneously express their actions and thoughts, the challenge associated with think-aloud protocols can be resolved by offering orientation training and clear instructions. Both quantitative and qualitative data analysis methods were used throughout the project. Descriptive analysis is the most commonly employed quantitative method since statistical analysis cannot be performed due to small sample size. For qualitative analysis, it is critical for sighted researchers to understand BVI users by collaborating with blind researchers and carefully reviewing the video and transcript data to enhance the open coding process. Future research on the methodology applied in BVI-related studies is also discussed.

Keywords:

Qualitative methods, Quantitative methods, Blind and visually impaired users, reflection, Digital library, Design guidelines



1. Introduction

As digital technologies diversify the means of encountering information in different formats, researchers have continuously worked to improve accessibility and usability for people with disabilities, such as the blind and visually impaired (BVI) users. BVI users refer to individuals who experience partial or complete blindness or visual impairment and utilize assistive technologies such as screen readers to interact with desktop and mobile devices. The sight-centered information retrieval (IR) design has created unique challenges for BVI users. To address these challenges, the research team conducted multiple studies to develop accessibility and usability guidelines for digital libraries (DLs).

Previous studies applied various methods of recruiting BVI participants and other stakeholders (e.g., experts, scholars, and developers) to meet the needs of the studies, such as working with associated organizations and distributing recruitment flyers and advertisements within BVI communities (Lazar et al., 2007; Yoon et al., 2016). However, not all studies provided detailed information on the recruitment process for BVI participants and the challenges of recruiting a large sample of participants (Armano et al., 2018; Bhardwaj & Kumar, 2017; Buimer et al., 2019; Huang, 2018). Moreover, relatively few scholars and developers were involved in BVI studies compared to BVI participants (Byerley & Beth Cambers, 2002; Gonçalves et al., 2018).

Researchers have employed various methods such as surveys, interviews, focus groups, and think-aloud methods to collect data to investigate BVI users' behavior when using IR systems. These methods have helped identify various challenges BVI users face when interacting with information retrieval systems. Studies have used both quantitative and qualitative methods for data analysis, such as descriptive statistical analysis, t-tests, correlation analysis, and open coding to provide insights into BVI users' experiences with IR systems and to identify areas for design improvement in these systems (Bhardwaj & Kumar, 2017; Conway et al., 2012; Gooda Sahib et al., 2012).

Although there is ample information on recruitment, data collection, and data analysis, there is a lack of details on how these methods were applied, the challenges during the research process, and how these challenges were resolved. This paper provides our review of and reflection on the methods applied and lessons learned during the research project that focused on creating guidelines supporting BVI users in DLs. It aims to contribute to BVI research by revisiting the recruitment strategies, data collection methods, and analysis techniques employed during a large-scale project on developing the Digital Library Accessibility and Usability Guidelines (DLAUG). It is critical for advancing our understanding of potential challenges and solutions for sampling, data collection, and analysis on BVI-related studies. This paper addresses the following issues:

- 1) The recruitment of BVI users, scholars/experts, and DL developers and associated challenges and solutions;
- 2) The employment of data collection methods for diverse DL stakeholders and associated challenges and solutions;
- 3) The application of data analysis techniques and associated challenges and solutions.

2. Literature Review

2.1 Participant recruitment in BVI research

Different types of sampling methods have been applied to recruit BVI users and other stakeholders (e.g., experts, scholars, and developers) in BVI research. Some researchers have used purposeful sampling methods. For example, Bodaghi et al. (2014, 2016) purposefully selected a research university that focused on students with special needs, including visually impaired (VI) students. Some researchers relied on the help of relevant organizations to assist with BVI participant recruitment. Lazar et al. (2007) recruited participants on-site at the National Federation of the Blind (NFB) Convention; in addition, recruitment materials were sent out to relevant email listservers with the help of the NFB. Yoon et al. (2016) recruited VI users via an email advertisement through Minnesota State Services for the Blind and Associated Colleges of the Twin Cities Disability Services. Bodaghi et al. (2017) approached VI student participants using a user list provided by a library. Some BVI studies did not provide details about how they recruited BVI participants (Bhardwaj & Kumar, 2017; Buimer et al., 2019; Gonçalves et al., 2018; Huang, 2018).

The number of BVI participants varies among BVI studies. In general, the sample size is relatively small for BVI research. Byerley and Beth Chambers (2002) recruited two blind users to assess the accessibility of two databases, and Armano et al. (2018) had four BVI participants to test the accessibility of a learning management system. Buimer et al. (2019) had eight BVI participants to investigate whether an emotion recognition system worked equally well in realistic and laboratory conditions. Some studies had moderate numbers of BVI participants. For example, 20 blind people participated in Gonçalves et al.'s (2018) evaluation study of e-commerce websites' accessibility and usability. Only sparse research has involved a large number of BVI participants, mostly for survey research. For example, Bhardwaj and Kumar (2017) reached out to 475 visually impaired students and received responses from 95 of them.

Compared with BVI participants, the number of scholars or experts involved in each BVI study is low. For example, there were fewer than five experts involved in the studies conducted by Byerley and Beth Chambers (2002), Conway et al. (2012), and Gonçalves et al. (2018). Gonçalves et al. (2018) invited three usability experts who were professors in relevant fields to perform the heuristic evaluation. In some studies, the authors themselves were also experts. The authors in Conway et al.'s (2012) study served as accessibility experts. Byerley

and Beth Chambers (2002) were two sighted librarians who tested the databases for their study based on their experiences of supporting BVI users.

2.2 Data collection methods in BVI research

Focusing on BVI-related studies in LIS, Xie et al. (2021c) found that a variety of data collection methods have been used, and the most common ones included questionnaires, interviews, assessment reports, transaction logs, observation, think-aloud, document analysis, and focus groups. In terms of LIS research concerning disability and accessibility, Hill (2013) pointed out that accessibility testing, questionnaires, and interviews were most often used for collecting data.

When conducting empirical BVI studies, researchers have used questionnaires, interviews, focus groups, think-aloud, log data, and assessment reports. Questionnaires have been widely used in BVI research. Bhardwaj and Kumar (2017) created a structured questionnaire to collect data concerning VI students' challenges when using assistive tools. Gonçalves et al. (2018) used two questionnaires for 20 blind participants: the first collected information about participants' computer literacy and use of communication and information technology, and the second focused on blind participants' satisfaction with assessed websites. Menzi-Çetin et al. (2017) asked VI students to complete satisfaction questionnaires to evaluate the accessibility of a university website.

Yoon et al. (2016) used both pre-interviews and exits interview with their VI participants: the pre-interviews collected general information about Web use and experienced challenges, while exit interviews asked participants to offer retrospective thoughts on their assessment experiences. Similarly, Gonçalves et al. (2018) conducted post-test interviews to assess the accessibility and usability of e-commerce websites. Buimer et al. (2019) employed exit interviews to collect information concerning VI participants' experiences using an emotion recognition system. Some researchers used both interviews and focus groups in their BVI studies. For example, Bodaghi et al. (2013) examined VI students' experiences and perceptions of library study carrels. Later, Bodaghi et al. (2016) used both interviews and focus groups to investigate VI library users' perceptions of librarian empathy.

Researchers employed think-aloud to gather feedback from BVI participants. For example, to better understand blind people's challenges with travel sites, Babu (2013) asked five blind participants to think aloud while they performed an airline flight search task. Yoon et al. (2016) encouraged six VI participants to narrate their thoughts when accessing library websites via screen readers.

The use of log data have been adopted in some BVI studies. For example, Van der Geest et al. (2014) used a video camera to record participants' actions to assess their skills. Similarly, log data, specifically "video recording and emotion logs," were used by Buimer et al. (2019, p. 5). Log data were also collected by Gonçalves et al. (2018) to assess websites' accessibility and usability. As an

alternative to log data, diaries have also been employed in BVI research, though not as often as other methods. Lazar et al. (2007) asked blind participants to keep diaries of their frustrations when using the Web. Notably, they used diaries in rich text format and also tested the diary forms via pilot testing to improve technical functionality (e.g., compatibility in different technical environments) and wording.

Researchers have also involved BVI users or experts to assess website compliance. Rømen and Svanæs (2012) invited seven disabled users, including three BVI users, to assess the usefulness of the Web Content Accessibility Guidelines (WCAG). Conway et al. (2012) used both expert manual evaluation and user tests to assess library website accessibility. In some studies, researchers only relied on automatic tools to produce assessment reports. For example, Yang et al. (2020) employed online website evaluation tools (e.g., WAVE, AChecker) to assess Ivy League library websites.

2.3 Data analysis methods in BVI research

In BVI research in LIS, Xie et al. (2021c) found that descriptive statistics were the most frequently used, followed by unspecified qualitative analysis methods, t-test, thematic analysis, ANOVA, correlation analysis, open coding, and Chi-square, as well as other less frequent methods. Different types of data analysis methods have been employed by researchers when investigating issues concerning BVI users.

Descriptive statistics have been the most common quantitative analysis method used in BVI studies. Conway et al. (2012) presented the frequency numbers of websites and errors based on the results from the automatic evaluation and manual evaluation by experts and people with disabilities, including BVI users. Descriptive statistics were used to show the distribution of responses to different questionnaire questions, e.g., frequency of accessing online resources and frequency of different types of difficulties faced by BVI students (Bhardwaj & Kumar, 2017).

In addition to descriptive statistics, other quantitative methods, such as t-test, ANOVA, and Chi-square, have also been applied. T-test was mainly used for comparison between BVI and sighted users. Gooda Sahib et al. (2012) included both VI and sighted searchers. They used t-test to compare the differences between the groups in their search behaviors (e.g., length of queries, number of search results viewed) and used the Chi-square test to check the differences between the two groups in their awareness and use of support features. On some occasions, t-test was also used to compare performance results from before or after tests.

In their exploratory research, Xie et al. (2015) employed open coding to identify BVI users' help-seeking situations in a digital library. Open and axial coding methods were used by Gooda Sahib et al. (2012, 2015) to explore BVI participants' information seeking and searching. In addition to open coding,

other qualitative analysis methods were also used in BVI studies, such as thematic analysis and content analysis. For example, Bodaghi et al. (2014) identified two main themes regarding VI students' perceptions of library student volunteers, including volunteers as academic saviors and volunteers as social connectors. One of the critical problems in qualitative studies on BVI research is that some of the papers did not specify their specific qualitative data analysis methods (Xie et al., 2021c).

2.4 Methodological reflection papers

Beyond conducting research projects and applying research methods in specific studies, researchers have further provided their reflections on their research projects and adopted research methods. Notably, some reflection papers focus on specific research methods. For example, Day and Thatcher (2009) reflected on the methodological issues of using diaries in a longitudinal study, especially the challenges they encountered. In particular, some challenges of diaries were noted, such as diary duration issues and difficulty motivating participants to keep diaries. Al-Yateem (2012) reflected on qualitative interviews in his research projects and suggested strategies for enhancing interview quality for other researchers (e.g., selecting appropriate interview settings and explaining research goals and procedures to participants). Based on the design and implementation of their narrative research project, George and Selimos (2018) ascertained that their narrative approach considered both public and personal narratives based on media and interview data.

In addition to methodological reflections, researchers have also reflected on other general aspects of their research projects. For example, Johnson (2009) broadly discussed the involvement of people with intellectual disabilities in inclusive research and the impact of inclusive research on people with intellectual disabilities based on a national research project. Liu and Burnett (2022) reflected on issues in collaborative intercultural research, such as language strategy, ethical payments, and anonymity. Furthermore, previous BVI studies also call for the need to reflect on research methods applied in specific projects or studies.

3. Data Sources

The data for this study is mainly based on the research project "Creating digital library DL design guidelines on accessibility, usability, and utility for blind and visually impaired (BVI) users" funded by the Institute of Museum and Library Services (IMLS). Document analysis and group discussions were used for data collection. Documents for this study include published and unpublished manuscripts, research data (e.g., think-aloud protocols, transaction logs, and coding schemes), and associated Institutional Review Board (IRB) materials (e.g., recruitment flyers, consent forms, and research instruments). Figure 1 illustrates the IMLS project studies and associated sampling, data collection, and data analysis methods.

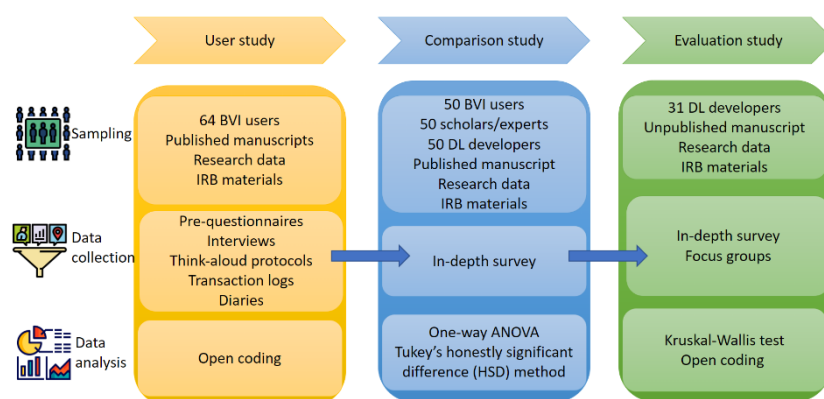


Figure 1. IMLS project studies and associated research methods

The main user study aimed to identify different types of help-seeking situations that BVI users encounter when interacting with DLs (Xie et al., 2021b). Based on the findings of the user study, a draft of the design guidelines was developed. In addition to the main user study, three associated user studies were conducted before the IMLS project. The first study identified the three most frequent help-seeking situations for BVI users and associated user, system, task, and interaction factors (Xie et al., 2018). The second examined whether implementing help features in DLs reduced BVI users' help-seeking situations (Xie et al., 2020). The last compared blind and sighted users' orientation tactics in DLs (Xie et al., 2021a). The purpose of the comparison study was to solicit feedback on the drafted design guidelines. Specifically, three types of stakeholders (i.e., BVI users, scholars/experts, and DL developers) reviewed the guidelines in terms of different criteria (i.e., importance, clarity, relevance, and usefulness) and provided suggestions via an in-depth survey. In the evaluation study, DL developers assessed five major DLs according to the guidelines to check how well those DLs complied with the proposed guidelines. In addition, DL developers offered suggestions on how to improve the guidelines.

Document analysis is a data collection method to gather existent documents, which can be seen as a stand-alone method or a method complementary to other methods (e.g., interviews and focus groups) (Flick, 2019). Document analysis was based on the sampled research materials. Two group discussions were held among the research team. Four members of the research project team participated in two group discussions. Each group discussion lasted about one hour. Before each group discussion, a list of discussion questions was prepared. During the first group discussion, the researchers reflected on what sampling methods (participant recruitment), data collection methods, and data analysis methods were used for the project and discussed the strengths, limitations, and lessons learned of the methods used. During the second group discussion, the researchers further expanded their initial discussions about the research methods

used and clarified any unclear points from the first focus group to enhance the understanding of the strengths, limitations, and lessons learned of the methods used.

4. Methodological Reflection

4.1 Recruitments of BVI users and scholars/experts and DL developers

In the initial stage of the study, BVI users were recruited from local organizations such as local industries or nonprofit organizations, as this specific study consisted of face-to-face interviews and search task sessions using the JAWS screen reader. The recruitment flyers were distributed to different regional blind associations. However, focusing only on the regional area caused difficulty recruiting a desired number of participants. To overcome this limitation, the researchers tried to recruit participants at a national level and went to the NFB Convention. The potential participants were sought before the convention. One problem was the non-attendance of some participants who had expressed their interest to participate in the study. To resolve the problem, one tactic was to have a backup list. Using the list, the researchers could quickly replace no-show participants with others from the waiting list.

For the IMLS project, BVI users were recruited from BVI-related organizations in the United States. The recruitment flyers were distributed to local BVI organizations, the national BVI organization, and libraries serving BVI users. In particular, the flyer was targeted to the NFB to reach the general BVI population. By recruiting at the national level, the researchers were able to get a representative sample of the BVI population for creating BVI accessibility and usability guidelines and gather more generalizable data. Moreover, the researchers were able to collect data on BVI users using diverse screen readers other than JAWS.

Some participants were also recruited using snowball sampling, distributing flyers through BVI researchers who are actively engaged with the organization and partners and advisory board members of the project. This recruitment method was used due to the unique BVI participants' characteristics. One researcher noted the importance of distributing flyers through trusted organizations and personnel, saying that "It's more about through a formal channel or through a personal referral because, as BVI users, they are normally vulnerable, so they less trust those messages that are not from the formal channel or from somebody they trust."

One of the challenges in BVI recruitment is reminding participants that the pre-requirements must be met to participate in the study. For example, the participants must have some experience using screen readers and the Internet to participate. However, participants mentioned some cases where participants had difficulty completing the tasks because they were incapable of using the technology. One researcher stated that "I remember there were some issues, like

the person actually didn't have much skill at all...because we asked, we have a pre-requirement about they had to have used the Internet for three years." To resolve this problem, the research team stressed the importance of double-checking that participants satisfied the pre-requirements for the study.

Scholars, experts, and DL developers were recruited to participate in the study by Xie et al. (2022). The research team conducted literature searches in academic databases to identify scholars relevant to the study. The contact information of scholars with publications related to DLs and the accessibility and usability of systems was collected to send invitation flyers. The accessibility and usability expert participants were initially recruited by distributing invitation flyers to related domain listservs for DL developers. Additional invitation flyers were sent to DL developers at various academic libraries across the United States based on US News' top 200 universities. DL Developers for each academic library were identified through the Google search engine. Moreover, the project team reached out to the advisory board members and used their connections to provide the team with recommendations for people who would be eligible for this study.

Even with the above steps, it proved challenging to recruit a sufficient number of scholars/experts and DL developers to participate. To resolve this, the research team applied snowball sampling technique by sending personal invitation emails to individuals known to research this area and to individuals recommended by the advisory board members. Additionally, the team received recommendations from participants of other individuals who could be suitable to participate in this study. Moreover, the team searched on LinkedIn to identify individuals who specify themselves as accessibility and usability experts or DL developers.

Another challenge was related to checking and verifying qualifications, such as whether prospective participants were truly DL developers. To verify DL developers' qualifications, the team searched through the web to locate information about the participants. For example, one researcher noted that "I needed to carefully go through the websites to see whether I can find them in departments related to digital library creation..." For some, the participants were asked to provide the team with their official employment email for verification. The team noted that, for future research, a potential solution would be to find participants during relevant conferences and seek researchers who are interested in this topic and are willing to participate.

4.2 Data collection methods

For the user study, multiple data collection methods were used throughout the research project, including questionnaires, interviews, think-aloud protocols, transaction logs, and diaries. For the user studies, questionnaires (both pre- and post-questionnaires), interviews (pre- and post-search interviews), think-aloud protocols, transaction logs, and diaries were employed (Xie et al., 2018, 2020, 2021a, 2021b).

Pre-questionnaires allowed participants to provide information concerning their demographic characteristics, including Internet experience, search skills, and system knowledge (Xie et al., 2018, 2020, 2021a, 2021b). Pre-search interviews were used to solicit information about participants' behaviors when using the Internet and their perceptions of help features. Post-search interviews were also used to allow participants to offer additional information about their perceptions of DL usability and help features (Xie et al., 2020). Participants were asked to perform various types of search tasks (i.e., orientation, know-item search, specific information search, and exploratory search) in different DLs (Xie et al., 2018, 2021a, 2021b). They were encouraged to think aloud when interacting with DLs, and their think-aloud and transaction logs were recorded by Morae (Xie et al., 2018, 2020, 2021a, 2021b). The advantage of using Morae software is that it is specifically designed for human-computer interaction (HCI) research. It can capture all log and think-aloud data and has data analysis functions. However, Morae is only installed on specific desktop devices, requiring researchers to invite BVI participants to the research lab to participate in the study.

One challenge of using think-aloud was related to participants' varying degrees of expressiveness, with some participants talking very little. To deal with this issue, one researcher pointed out that "one strategy that works is to remind participants to keep talking." In addition, some participants did not want to continue working on the search tasks even though they did not find relevant information. In this case, the researchers would encourage them to try different ways to complete the search tasks. Another challenge was related to the impact of participants' prior knowledge on completing search tasks. Some participants answered some search questions based on their prior knowledge without actually finding relevant items. One researcher noted that "some participants already knew the date of Gettysburg Address" when working on a search task. It is important to make sure that participants must read or listen to the content of items to answer relevant search questions.

In order to collect data nationally, diaries were also used as a data collection method. In Xie et al. (2021b), participants in the offsite group, who were not able to use Morae, were instructed to keep diaries to record their search processes when working on search tasks. Diaries helped collect data remotely from participants. If diaries were not used, the researchers had to travel across the country to collect data. However, using diaries also has challenges. Some participants did not follow the diary instructions but only provided brief feedback about their encountered problems during search tasks in their diaries. Some of them did not recognize specific problems that they encountered in the search process. According to one researcher, "We actually called them, you know, explained to them and helped them to identify specific problems." It was critical to call participants to further explain the instructions and solicit more detailed feedback. Another limitation was that some participants did not reply to

researchers' reminder emails in time. Setting an absolute deadline for participants will be a solution. Accuracy of the diary data may also be an issue. Some BVI participants mentioned that they asked others to help with the tasks and write the diaries when working on the search tasks. One researcher noted that "when participants did it immediately, it seemed that they used help from somebody else." The research team realized that diaries might not be an effective approach for BVI studies.

For the comparison study among three types of stakeholders, Xie et al. (2022), administered an in-depth survey to 150 participants from three types of stakeholders (i.e., BVI users, scholars/experts, and DL developers). Participants were instructed to rate the associated guidelines based on their perceived importance, clarity, relevance, and usefulness concerning different help-seeking situations as well as offering qualitative justifications. The in-depth survey provided not only quantitative rating data but also qualitative feedback that helped revise the guidelines. The challenge of using an in-depth survey was the time commitment. There were participants complaining about how time-consuming it was to finish the survey. In particular, some scholar participants mentioned their challenges of finishing the in-depth survey while they were doing their research. Our solution to this challenge was to ask each participant to rate half of the guidelines instead of the entire set of comprehensive guidelines. Also, a PDF file of the guidelines was created for the convenience of participants to check the guidelines and make associated notes.

For the DL evaluation study, an in-depth survey and online focus groups were employed to collect data from DL developer participants (Xie et al., 2023). An in-depth assessment survey was created, and the questions were organized by different types of help-seeking situations. Each DL developer participant assessed to what degree one of the five specified DLs (i.e., Library of Congress Digital Collections, Digital Public Library of America, HathiTrust, Artstor, and National Science Digital Library) complied with the guidelines for each help-seeking situation based on a 7-point Likert scale. Moreover, they provided specific examples of violations against the guidelines and good techniques and features used by the DL they assessed. Using an in-depth survey helped reveal the compliance status of assessed DLs with the proposed guidelines. The qualitative feedback about DL features helped add more relevant examples to the guidelines. The Canvas learning management system was used for the online focus groups of DL developers. The focus groups were asynchronous, allowing DL participants to post discussion messages when they were available. In the focus groups, participants discussed the guidelines they followed in developing DLs and the challenges they encountered when following the guidelines. Participants were not worried about their identities being revealed. In the asynchronous focus groups, participants used their subject number and did not show their faces nor use their voices.

The main challenge of using an in-depth survey was the time commitment as mentioned in the comparison study. To help relieve participants' burdens, one researcher noted that "we divided the lengthy guideline files into two halves, and each participant only needed to work on one half instead of the whole guideline file."

For the focus groups, the researchers served as moderators to check the discussions. One researcher noted that "I was regularly checking the discussions to make sure that all participants were actively involved in the discussion." Reminder email messages were sent to participants if they did not post any messages under some discussion questions. Participants were also reminded to post their discussion messages as early as they could so that other participants would have time to review and respond. Another challenge was that some participants did not discuss all aspects of the questions. The moderator carefully checked all of the messages and reminded participants to answer the questions to which they did not respond.

4.3 Data analysis methods

Quantitative analysis was used in all four user studies, including three user studies that preceded the IMLS project. The latter expanded the research of the previous studies. The analysis aimed to identify the frequency of help-seeking situations faced by BVI users and compare their behavior across different user groups (Xie et al., 2018, 2020, 2021a, 2021b). Descriptive analysis was utilized to determine the frequency of help-seeking situations in these studies (Xie et al., 2018, 2020, 2021a, 2021b). Furthermore, t-tests and Mann-Whitney U tests were used in the studies to compare two user groups (Xie et al., 2020, 2021a). Specifically, these tests were applied to compare the orientation tactics between BVI users and sighted users (Xie et al., 2021a), as well as the help-seeking situations encountered between a control group of BVI users and an experimental group of BVI users (Xie et al., 2020). In addition, Pearson's correlation coefficient analysis was employed to assess the presence of relationships between a specific type of situation and different main-level factors, including user, system, task, or interaction (Xie et al., 2018). For the comparison study, one-way ANOVA was used to test the significant mean difference among the three groups (Xie et al., 2022). Specifically, this test was applied to compare the effect of participant types on the perceived relevance, clarity, and usefulness of guidelines concerning different types of help-seeking situations. The Kruskal-Wallis test was employed in the DL evaluation study to compare the mean differences among the compliance rates of the DLAUG in each situation, as collected by the assessment questionnaires. The mean rank obtained from the Kruskal-Wallis test was used to identify the guidelines with the lowest compliance rating among existing DLs, which were subsequently selected for further qualitative analysis.

One challenge faced when using quantitative studies is the issue of sample size. The minimum sample size for quantitative research is typically 30, as fewer than 30 may not be suitable for quantitative analysis (Gravetter & Wallnau, 2013). However, it is generally recommended to have as large a sample size as

possible, as larger sample sizes can be more easily generalized or transferred to other contexts. As a result, previous user studies conducted with BVI users may have had limitations due to the inability to acquire a sufficient sample size to perform complex quantitative analysis, as noted by one researcher, “Everything is related to the normalization and sample size.” Therefore, acquiring an adequate sample size in studies conducted with BVI participants may be challenging due to the nature of the data and guidelines.

To address the limitations posed by sample size and types of data, non-parametric tests such as the Mann-Whitney U test and the Kruskal-Wallis test were applied when the data was not normally distributed. A researcher explained the reason for this, stating, “we use the non-parametric tests because of the sample size.” There were two main reasons why the studies utilized the Mann-Whitney U test and the Kruskal-Wallis test. First, non-parametric tests can handle data measured on an ordinal or nominal scale, which is inappropriate for use with parametric tests that require interval or ratio scale data (Field, 2013). Second, non-parametric tests, including the Mann-Whitney U test and Kruskal-Wallis test, do not rely on assumptions of normality or homogeneity of variance, which can be particularly useful when dealing with small sample sizes or non-normal distributions (Harris, 2001). This means that non-parametric tests can be more robust than parametric tests in situations where the data does not meet the assumptions required for the use of parametric tests.

Open coding was the only qualitative analysis method applied for the IMLS project. It was used to identify help-seeking situations (Xie et al., 2018, 2021a, 2021b), factors (Xie et al., 2021a), and coping tactics (Xie et al., 2021b). In addition, it helped identify types of violations against the DLAUG, guidelines used by DL developers in practice, and types of challenges DL developers encounter when following guidelines based on the assessment questionnaires and focus group data from DL developers. Open coding was selected because the BVI studies conducted by the research team were exploratory research. No research had been done in a DL context for supporting BVI users, and thus there was no pre-defined coding scheme available for use. For that reason, the research team had to use open coding to identify the situations from scratch.

The challenge of using open coding was to make sure that codes accurately reflected associated content and had appropriate definitions and labels. At the initial stage of coding, researchers encountered difficulty coming up with the right labels and definitions. As one researcher mentioned, “the most challenging part for me was at the beginning stage of coding; I was not sure how BVI users navigate or use DLs.” Another researcher echoed that “it is important to start slowly and get familiar with the data.” In addition to analyzing the transcript text, researchers also checked the context by viewing the recorded videos of BVI users’ search processes to decide on the best way to classify and label situations and other aspects.

To ensure the accuracy of open coding results, inter-coder reliability and group discussions were applied. A researcher who is blind also participated in many rounds of group discussions to help enhance the research team's understanding of BVI users' experiences when using DLs. Another challenge was that it took researchers much time to carefully review the analyzed documents when any modification was made to the coding scheme. Each iteration was time-consuming, although it helped researchers get a clearer picture of the results as the analysis proceeded.

There are a few lessons learned about using open coding. The first one is to "get familiar with everything before delving into it." The second one is to pay attention to specific contexts of actions and "make sure those coding categories are mutually exclusive." The third lesson is to ensure the reliability of coding results using appropriate methods. As one researcher put it, "We applied multiple methods to make sure the accuracy of the coding results."

5. Discussion

This reflection paper presents an in-depth analysis of sampling, data collection, and data analysis processes in studies related to BVI users. Compared to previous research, this paper is one of the few that illustrates the research process in detail in a large-scale research project. Moreover, it reveals the challenges that researchers encountered in the research process and the associated solutions employed.

5.1 Recruiting participants through trusted formal and informal channels

Recruiting BVI participants, particularly in large numbers, has always been a challenge (Armano et al., 2018; Buimer et al., 2019). Previous research mainly used a purposeful sampling method, and researchers recruited participants from the national blind association and universities (Bodaghi et al., 2017; Lazar et al., 2007; Yoon et al., 2016). This research team employed multiple sampling strategies, including the snowball strategy in addition to the purposeful sampling method. For recruitment, the research team recruited not only participants from national blind associations but also local blind associations. In addition, libraries and museums that support BVI users served as contacts for recruitment. The partners and advisory board members of the project played critical roles in helping find participants for the studies as well. Sixty-four BVI users successfully participated in the user study.

Interestingly, the sample size in prior studies was even smaller for scholars, developers, and experts involved in BVI research; many only had five or fewer participants (Byerley & Beth Chambers, 2002; Gonçalves et al., 2018). Recruiting scholars/experts and developers proved challenging mainly because of their limited numbers and busy schedules. In this project, we successfully recruited 50 scholars/experts and 50 DL developers for the comparison study and 31 DL developers for the evaluation study.

This study demonstrated that challenges exist in recruiting BVI users and scholars, experts, and developers for BVI studies. For BVI user recruitment, the key is trust. Researchers must establish connections with trusted organizations and institutions to promote their studies. For scholar, expert, and developer recruitment, the key is finding people interested in the topic and acknowledging its importance. Another challenge is related to pre-qualifications and validation of the requirements. In the digital world, researchers need to pay special attention to making sure their requirements are clearly stated (e.g., using screening-questionnaires) and verified (e.g., checking identities of developer participants) to ensure the recruitment of qualified participants.

5.2 Collecting data via multiple methods considering participants' needs

For data collection, researchers have applied various methods, such as questionnaires (Bhardwaj & Kumar, 2017; Gonçalves et al., 2018; Menzi-Çetin et al., 2017) and interviews, along with less popular ones, such as think-aloud protocols (Babu, 2013; Yoon et al., 2016), diaries (Lazar et al., 2007), log data (Buimer et al., 2019; Van der Geest, 2014), focus groups (Bodaghi et al., 2016), etc. This project employed almost all of the data collection methods: questionnaires, interviews, think-aloud protocols, transaction logs, diaries, in-depth surveys, and focus groups. For the user study, one unique aspect is that both pre- and post-data collection methods involved questionnaires and interviews. Therefore, both data related to general perceptions about IR system use and specific DL systems were taken into consideration. Another unique aspect is that these data collection methods are interrelated. For example, the researchers used diaries as an alternative method for think-aloud protocols and transaction logs for BVI users. The third unique aspect is that we used asynchronous focus groups via learning management software for the evaluation study. This was quite an effective data collection approach.

The study reveals one challenge associated with data collection on BVI-related studies regarding think-aloud protocols. BVI users had to use screen reader software to understand DL interfaces and simultaneously express their actions and thoughts. In order to facilitate that, BVI users were trained during the orientation tasks and got used to thinking aloud during their search process. In addition, researchers reminded them to keep talking when they forgot to do so. Another challenge was related to the diary. To collect data naturally across the United States, the diary method was employed in the user study. Nevertheless, diary entries and their accuracy were the two critical issues. BVI users had difficulty identifying the problems they encountered in their search process, and sometimes they asked someone else to fill in the diary for them. One solution is for researchers to provide more guidance and clarification via phone. Another solution is to collect data via online meetings (e.g., Microsoft Teams or Zoom) so think-aloud data and transaction logs could be collected remotely. The challenge related to the comparison study was how to define importance, relevance, clarity, and usefulness clearly when three types of DL stakeholders

had different perceptions based on their own experiences. For future studies, we need to provide clearer definitions for the key variables.

5.3 Analyzing quantitative and qualitative data by enlarging sample size and systematically reviewing data

Even though quantitative data analysis methods have been applied to BVI research, descriptive analysis is the most commonly employed quantitative method (Bhardwaj & Kumar, 2017; Conway et al., 2012). Some of the BVI studies focus on the comparison between BVI and sighted users (Gooda Sahib et al., 2012). The same scenario applies to this project. Additionally, the comparison is also conducted between the control group and the experimental group of BVI users. Interestingly, the issues with quantitative analysis are not related to data analysis; instead, they are more associated with sampling and data collection. One challenge of quantitative analysis is the small sample size. Because of that, fewer statistical analysis methods could be employed. Although this project recruited quite a large sample of BVI users (64), BVI participants were divided into multiple groups to explore multiple DLs, which limited the employment of statistical analysis. In addition, non-parametric tests, such Mann Whitney U test and Kruskal-Wallis test were used rather than t-tests and ANOVA tests. Increasing sample size and using ordinal or nominal scales to collect data should enable researchers to perform more parametric tests.

For qualitative analysis, the main issue of previous BVI research is that many published papers did not specify their qualitative data analysis methods, making it difficult for other researchers to understand the research process and validate research findings (Xie et al., 2021c). Open coding, thematic analysis, and qualitative content analysis were also employed as typical qualitative analysis methods (Bodaghi et al., 2014; Gooda Sahib et al., 2012, 2015; Xie et al., 2015). Like this project, the exploratory nature of many BVI studies requires researchers to review all the data to develop categories, which can be particularly challenging for sighted researchers without BVI experience. Collaborating with blind researchers and repeatedly reviewing the video and transcript data can help improve the open coding process.

6. Conclusions

This reflective paper not only highlights sampling strategies, data collection, and data analysis techniques but also identifies the challenges associated with the research process derived from a large-scale BVI project. Most importantly, the researchers proposed solutions to deal with problems that occurred related to BVI research. While BVI papers present their methodologies for their research studies in a concise form, this in-depth analysis of the research process offers a great opportunity for the research team to thoroughly review the entire research process and reflect on the lessons learned. The findings of this paper will enhance the methodology part of BVI research.

The challenges of recruitment of BVI users, scholars/experts, and system developers call for an increase in the sample size through national/local BVI and professional organizations and institutions. Furthermore, checking the qualifications of participants is essential in the recruitment process. Data collection challenges inform researchers that think-aloud protocols can be used for BVI studies as long as researchers provide a training section and remind participants to keep talking throughout the data collection process. Compared to diaries, online meetings can be a better alternative for collecting think-aloud data and transaction logs remotely. In addition, it is critical to define variables clearly when using surveys across multiple types of stakeholders. The challenges of quantitative data analysis can be resolved by increasing the sample size and using ordinal or nominal scales to collect data, and the challenges of qualitative data analysis can be dealt with by sighted researchers working with BVI researchers and analyzing the data more systematically.

Acknowledgement

The authors thank IMLS Leadership Grants for Libraries for funding this project and Dr. Rakesh Babu for his contribution to the project.

References

- Al-Yateem, N. (2012). The effect of interview recording on quality of data obtained: A methodological reflection. *Nurse Researcher*, 19(4), 31–35.
- Armano, T., Borsero, M., Capietto, A., Murru, N., Panzarea, A., & Ruighi, A. (2018). On the accessibility of Moodle 2 by visually impaired users, with a focus on mathematical content. *Universal Access in the Information Society*, 17, 865–874.
- Babu, R. (2013). Understanding challenges in non-visual interaction with travel sites: An exploratory field study with blind users. *First Monday*, 18(12). <https://doi.org/10.5210/fm.v18i12.4808>
- Bhardwaj, R. K., & Kumar, S. (2017). A comprehensive digital environment for visually impaired students: User's perspectives. *Library Hi Tech*, 35(4), 542–557.
- Bodaghi, N. B., Cheong, L. S., & Zainab, A. N. (2016). Librarians empathy: Visually impaired students experiences towards inclusion and sense of belonging in an academic library. *The Journal of Academic Librarianship*, 42(1), 87–96.
- Bodaghi, N. B., Cheong, L. S., Zainab, A. N., & Riahikia, M. (2017). Friendly librarians: The culture of caring and inclusion experiences of visually impaired students in an academic library. *Information Development*, 33(3), 229–242.
- Bodaghi, N. B., Zainab, A. N., & Zainab, A. N. (2013). My carrel, my second home: Inclusion and the sense of belonging among visually impaired students in an academic library. *Malaysian Journal of Library & Information Science*, 18(1), 39–54.
- Bodaghi, N. B., Binti Awang-Ngah, Z., & Abdullah, N. (2014). Student volunteers as academic saviours and social connectors among the visually impaired in an academic library. *Libri*, 64(1), 40–48.
- Buimer, H., Schellens, R., Kosteljik, T., Nemri, A., Zhao, Y., Van der Geest, T., & Van Wezel, R. (2019). Opportunities and pitfalls in applying emotion recognition software for persons with a visual impairment: Simulated real life conversations. *JMIR mHealth and uHealth*, 7(11), e13722.
- Byerley, S. L., & Beth Chambers, M. (2002). Accessibility and usability of Web-based library databases for non-visual users. *Library Hi Tech*, 20(2), 169–178.

Conway, V., Brown, J., Hollier, S., & Nicholl, C. (2012). Website accessibility: A comparative analysis of Australian national and state/territory library websites. *The Australian Library Journal*, 61(3), 170–188.

Day, M., & Thatcher, J. (2009). “I’m really embarrassed that you’re going to read this...”: Reflections on using diaries in qualitative research. *Qualitative Research in Psychology*, 6(4), 249–259.

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. Sage.

Flick, U. (2019). *An Introduction to Qualitative Research (6th Edition)*. Sage.

George, G., & Selimos, E. D. (2018). Using narrative research to explore the welcoming of newcomer immigrants: A methodological reflection on a community-based research project. *Forum: Qualitative Social Research*, 19(2), 119–138.

Gonçalves, R., Rocha, T., Martins, J., Branco, F., & Au-Yong-Oliveira, M. (2018). Evaluation of e-commerce websites accessibility and usability: An e-commerce platform analysis with the inclusion of blind users. *Universal Access in the Information Society*, 17, 567–583.

Gooda Sahib, N., Tombros, A., & Stockman, T. (2012). A comparative analysis of the information-seeking behavior of visually impaired and sighted searchers. *Journal of the American Society for Information Science and Technology*, 63(2), 377–391.

Gooda Sahib, N., Tombros, A., & Stockman, T. (2015). Evaluating a search interface for visually impaired searchers. *Journal of the Association for Information Science and Technology*, 66(11), 2235–2248.

Gravetter, F. J., & Wallnau, L. B. (2013). *Statistics for the Behavioral Sciences (9th ed.)*. Wadsworth.

Harris, R. J. (2001). *A Primer of Multivariate Statistics*. Psychology Press.

Hill, H. (2013). Disability and accessibility in the library and information science literature: A content analysis. *Library & Information Science Research*, 35(2), 137–142.

Huang, H. (2018). Blind users' expectations of touch interfaces: Factors affecting interface accessibility of touchscreen-based smartphones for people with moderate visual impairment. *Universal Access in the Information Society*, 17, 291–304.

Johnson, K. (2009). No longer researching about us without us: A researcher's reflection on rights and inclusive research in Ireland. *British Journal of Learning Disabilities*, 37(4), 250–256.

Lazar, J., Allen, A., Kleinman, J., & Malarkey, C. (2007). What frustrates screen reader users on the web: A study of 100 blind users. *International Journal of Human-Computer Interaction*, 22(3), 247–269.

Liu, X., & Burnett, D. (2022). Insider-outsider: Methodological reflections on collaborative intercultural research. *Humanities and Social Sciences Communications*, 9(1), 1–8.

Menzi-Çetin, N., Alemdağ, E., Tüzün, H., & Yıldız, M. (2017). Evaluation of a university website's usability for visually impaired students. *Universal Access in the Information Society*, 16, 151–160.

Rømen, D., & Svanæs, D. (2012). Validating WCAG versions 1.0 and 2.0 through usability testing with disabled users. *Universal Access in the Information Society*, 11, 375–385.

Van der Geest, T., van der Meij, H., & Van Puffelen, C. (2014). Self-assessed and actual Internet skills of people with visual impairments. *Universal Access in the Information Society*, 13, 161–174.

Xie, I., Babu, R., Castillo, M., & Han, H. (2018). Identification of factors associated with blind users' help-seeking situations in interacting with digital libraries. *Journal of the American Association for Information Science and Technology*, 69(4), 514–527.

Xie, I., Babu, R., Joo, S., & Fuller, P. (2015). Using digital libraries non-visually: Understanding the help-seeking situations of blind users. *Information Research: An International Electronic Journal*, 20(2), n2.

Xie, I., Babu, R., Lee, H. S., Wang, S., & Lee, T. H. (2021a). Orientation tactics and associated factors in the digital library environment: Comparison between blind and sighted users. *Journal of the Association for Information Science and Technology*, 72(8), 995–1010.

Xie, I., Babu, R., Lee, T. H., Wang, S., & Lee, H. S. (2021b). Coping tactics of blind and visually impaired users: Responding to help-seeking situations in the digital library environment. *Information Processing and Management*, 58(5), 102612.

Xie, I., Babu, R., Wang, S., Lee, H. S., & Lee, T. H. (2022). Assessment of digital library design guidelines to support blind and visually impaired users: A study of key stakeholders' perspectives. *The Electronic Library*, 40(6), 646–661.

Xie, I., Babu, R., Lee, T., Castillo, M., You, S. & Hanlon, A. (2020). Enhancing usability of digital libraries: Designing help features to support blind and visually impaired users. *Information Processing and Management*, 57(3), 102110.

Xie, I., Wang, S., & Saba, M. (2021c). Studies on blind and visually impaired users in LIS literature: A review of research methods. *Library & Information Science Research*, 43(3), 101109.

Xie, I., Wang, S., Lee, T. & Lee, H. (Unpublished). Accessibility of Digital Libraries. School of Information Studies, University of Wisconsin-Milwaukee.

Yang, W., Zhao, B., Liu, Y. Q., & Bielefield, A. (2020). Are Ivy League library website homepages accessible?. *Information Technology and Libraries (Online)*, 39(2), 1–18.

Yoon, K., Dols, R., Hulscher, L., & Newberry, T. (2016). An exploratory study of library website accessibility for visually impaired users. *Library & Information Science Research*, 38(3), 250–258.