Education for Information -1 (2024) 1–16 1
DOI 10.3233/EFI-240014
IOS Press
CORRECTED PROOF

User participation in digital accessibility evaluations: Reviewing methods and implications

María Inés Laitano[a,*], Roelvis Ortiz-Núñez[b] and Yudayly Stable-Rodríguez[c]
[a]*Laboratoire Paragraphe, Université Paris 8, France*
[b]*Doctoral Candidate, Escuela de Ciencias Sociales, Universidad Pontificia Bolivariana, Colombia*
[c]*Instituto de Información Científica y Tecnológica, Cuba*

Although laws and standardization bodies promote user participation in digital accessibility evaluations, people with disabilities still consider themselves excluded from this process. One reason could be the lack of systematized knowledge about evaluation methods involving users. This article seeks to understand how and for what purpose digital accessibility evaluations with user participation were conducted in the scientific literature from 2018 to 2021. Three types of user participation emerged: 1) user-based usability testing to evaluate task accomplishment, user reactions and interface qualities; 2) interviewing users to assess the local and social factors impacting digital service accessibility; 3) using questionnaires or crowdsourcing to check the compliance of certain interfaces with accessibility standards. Participants are primarily chosen based on their functional impairments and, to a lesser degree, their project-related skills, biographical information, technology habits, among other criteria. The comprehensive user insights gained with these methods are judged to be positive whereas the lack of representativeness of the selected user samples is found to be regrettable. The article finally discusses the definitions of accessibility and disability that underpin these methodologies.

Keywords: User participation, digital accessibility evaluation, disability, evaluation methods, usability test, interviews, questionnaires, user test, web accessibility

1. Introduction

A recent report by the European Union found that significantly more people with disabilities find the digital services of public bodies difficult to use than users in general. A significant number of disabled users and the organizations that represent them report little or no involvement by States in the implementation of digital accessibility (Bianchini et al., 2022, p. 7). The report concludes with several findings on this issue including the insufficient expertise of professionals, absent feedback mechanisms between users and public bodies, divergent evaluation methodologies and biased evaluations in terms of user profiles. However, user participation in accessibility evaluations has been a recommendation of the World Wide Web Consortium (W3C)'s Web Accessibility Initiative (WAI) since at least 2005 (WAI, 2005) although the

*Corresponding author: María Inés Laitano, Laboratoire Paragraphe, Université Paris 8, France. E-mail: mlaitano02@univ-paris8.fr.

evaluation methodology formalized in 2014 does not include it as a mandatory step (WAI, 2014). The WAI recommends conducting usability tests, selecting a diverse pool of users with varying disabilities and prior experiences with digital technologies.

The digital accessibility research community has been discussing the issue of evaluation methods for some time. In 2008, Brajnik insisted on the importance of implementing evaluation methods that characterize the context of use, such as heuristic walkthroughs or user tests and that differ from compliance audits in which the context of use is absent or very general in nature (Brajnik, 2008). According to the author, methods can be very different depending on whether they are analytical or empirical or the information used to deduce accessibility problems (observations of user behavior or opinions expressed by users or evaluators). Later Brajnik and his co-authors investigated the effect of evaluator expertise on compliance audits (Brajnik et al., 2010) and heuristic walkthroughs (Brajnik et al., 2011). However, to the best of our knowledge there are no articles that have expressly discussed user participation in digital accessibility evaluations.

To date, literature reviews examining digital accessibility evaluations have not placed a significant emphasis on user participation. Silva et al. (2019) compare accessibility problems detected by three types of methods – automated evaluations, manual expert inspections and user tests. For user tests, they only describe the assistive technologies used and the participants' disabilities. They conclude that automated evaluations are very limited since they detect less than 40% of the problems encountered. Nuñez et al. (2019) review web accessibility evaluations to determine the most commonly used evaluation methods. They find that automated evaluations are the most common method, although 55% of the evaluations reviewed implement user tests. These user tests evaluate accessibility standards or some customized indicator and some tests include participants who are experienced in the domain being evaluated. Campoverde-Molina et al. (2020) review empirical works that evaluate the web accessibility of educational environments. As in the previous case, they find that 80% of the papers perform automated evaluations, 12% manual evaluations carried out by experts or users and the remaining 8% a combination of both. For evaluations involving users, they only detail the assistive technologies used, the participants' profile (students in this case) or the functional disability they share. Ara et al. (2023) classify publications on web accessibility according to the type of engineering process implemented – requirements, problems, framework, testing, etc. They list some evaluation methods involving users (tests, questionnaires, etc.) without detailing how the evaluations are actually carried out in reality.

The aim of this article is to elucidate how user participation is carried out in digital accessibility evaluations and its underlying purposes. It adds to the body of knowledge concerning methods for digital accessibility evaluation by systematically reviewing empirical evidence related to user participation. The evaluations under consideration are sourced from scientific literature (as detailed in Section 2) because it pays particular attention to the methods used and their justification. Section 3 provides insights into the evaluation methods used, selected indicators, user profiles, evaluation environments, as well as the benefits and limitations of user participation. Sections 4 and 5 feature a discussion and conclusions regarding user participation in digital accessibility evaluations.

2. Methods

Searches were conducted on two international scientific information platforms, the Web of Science Core Collection and Scopus, in March 2022. The search process was guided by two main 'filters': the first one was thematic, focusing on literature related to digital accessibility, and the second one was methodological, targeting publications that evaluate accessibility. Initial searches conducted to refine the search string revealed that numerous publications did not incorporate methodological keywords in the Keywords fields, but rather in the Title field. Consequently, the decision was made to utilize the Title field instead of the Keywords field for filtering publications that conducted accessibility evaluations. The search string used on the Title was as follows: *(digital OR web) AND (accessibility OR "inclusive design" OR "design for all" OR "universal design") AND (evaluat* OR quality OR diagnostic OR usability OR assess* OR audit OR test* OR performance OR empiric* OR "case study" OR survey OR measure OR framework) NOT "universal design for learning"*. 2018 was identified as the most productive year in this area on the Web of Science followed by 2020 so the searches focused on papers published between 2018 and 2021. The search was restricted to papers in English, Spanish and Portuguese. After removing duplicates, a total of 128 references were identified.

Subsequently, a manual review of the papers was undertaken in which we excluded digital accessibility evaluations that did not involve users, such as automated evaluations or evaluations conducted by experts. Papers that were not evaluations, such as those related to the development of applications for evaluators and papers about accessibility in contexts other than digital accessibility (e.g. physical spaces or medical services) were also excluded. Papers that were not long works (i.e., short papers or communications, posters and abstracts) were also excluded. As a result, 16 papers were selected (Laitano et al., 2024).

The selected papers were examined to find responses to the following research questions: what evaluation methods were employed to involve users, what indicators were selected and what ways of evaluating them, what user profiles were involved, where the evaluations were conducted and what the authors of the papers identified as the advantages and limitations of user participation? The findings of this analysis are presented in the following section.

3. Results

3.1. Selected papers overview

The analysis by period, considering the total number of selected papers ($N = 16$),

Table 1
Publication types and languages of selected papers ($N = 16$). Author's own work

Publication type	Publication language	Number of papers	Proportion of selected papers
Conference paper	English	7	43.75%
	Spanish	1	6.25%
Article	English	7	43.75%
	Portuguese	1	6.25%
Total		**16**	**100%**

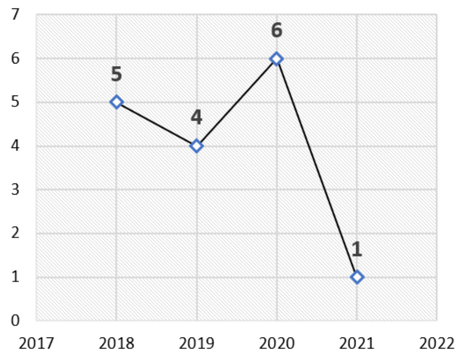

Fig. 1. Number of selected papers ($N = 16$) by year. Author's own work.

shows a limited growth in the publications number related to user participation in digital accessibility evaluations, which indicates the marginal place that this topic has in the literature. The years with the highest publications number were 2020 and 2018 with 37.5% and 31.25% of publications respectively, followed by 2019 with 25%; In contrast, the year with the lowest scientific production was 2021 with only 6.25% (Fig. 1).

Regarding the distribution of scientific production according to publication type and language, Table 1 shows that only conference papers and articles were recorded, with the same number of publications ($n = 8$). Regarding language, there is a predominance of publications in English (87.5%), followed by a minimal proportion in Spanish (6.25%) and Portuguese (6.25%). When comparing the language distribution of publications with the countries where the evaluations were conducted (Fig. 2), it can be observed that there is no correspondence between the diversity of languages in the covered countries and the publications language, which is consistent with the predominance of English as the lingua franca of scientific communication.

3.2. Evaluation methods employed to involve users

The method most commonly found in the selected literature is the user-based usability test (56.25%, see Table 2). In this method, users guided by a moderator individually engage in a series of tasks across one or more digital interfaces with

Table 2
Evaluation methods employed to involve users in selected papers ($N = 16$). Author's own work

Evaluation method	Number of papers using the method	Proportion of selected papers
User-based usability test	9	56.25%
Controlled experiment	3	18.75%
Interview	2	12.5%
Questionnaire	1	6.25%
Crowdsourcing	1	6.25%
Total	**16**	**100%**

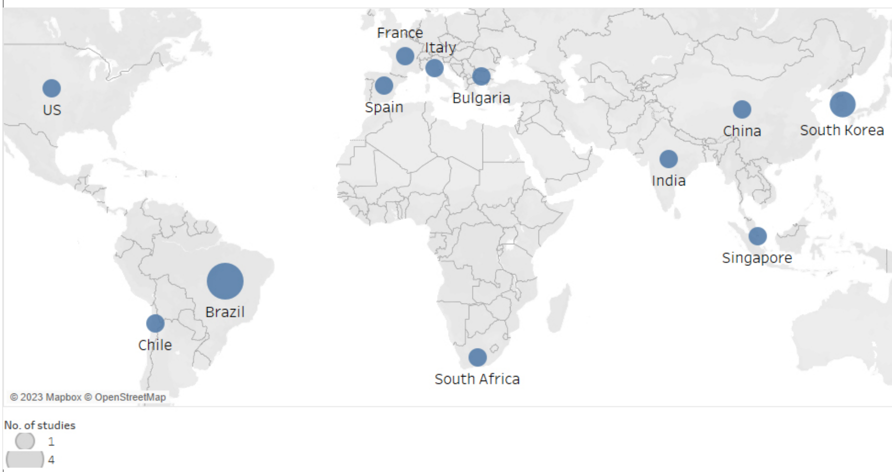

Fig. 2. Number of selected papers ($N = 16$) by country where the evaluations were conducted. Author's own work.

the primary aim of assessing the ease of use of these interfaces. It has been used in various contexts such as a high school examination system (Leria et al., 2018), a video playback application (Funes et al., 2018), a university's institutional and e-learning websites (Maboe et al., 2018), three massive open online course providers (Park et al., 2019), two university-based platforms (Shachmut & Deschenes, 2019), ten public healthcare websites (Yi, 2020), three audio production workstations (Pedrini et al., 2020), a university library website (Galkute et al., 2020) and an educational game for undergraduate students (de Oliveira et al., 2021). The number of participants involved in these tests varies, ranging from one (Pedrini et al., 2020) to 25 (Yi, 2020).

Other studies (18.75%) conduct usability testing but their primary objective is to extrapolate the findings to all interfaces and not just those specifically involved in the test, classifying them within the realm of what Human-Computer Interaction terms "controlled experiments" (Lazar, 2017, Chapter 2). This method necessitates a larger pool of participants compared to conventional usability tests. For instance, Giraud et al. (2018) demonstrated that filtering redundant and irrelevant information improves website usability for blind users based on a sample of 76 participants.

Alonso-Virgos et al. (2020) validate a set of web usability guidelines for users with Down syndrome on a sample of 25 participants. Nogueira et al. (2019) studied the emotional impact of usability barriers in responsive web design on blind users on a sample of 18 participants.

Certain studies incorporate a usability test or experiment in conjunction with another method either to triangulate the results following the test or to shape the experiment's design prior to its execution. For instance, De Oliveira et al. (2021) use a focus group subsequent to individual tests to validate the observations made during the tests. Alonso-Virgos et al. (2020) opt for the utilization of a questionnaire to select the design guidelines that will subsequently be integrated into the experiment.

Interviews (12.5%), questionnaires (6.25%) and crowdsourcing (6.25%) are less commonly used methods for engaging users in accessibility evaluations. Kameswaran and Muralidhar (2019) carry out semi-structured interviews with the aim of gaining insights into the social factors influencing the accessibility of digital payment systems for individuals with visual impairments in metropolitan India. Lim et al. (2020) conduct contextual interviews (Lazar, 2017, Chapter 8.5.2) to understand more about how individuals with disabilities use government e-services in Singapore. Sabev et al. (2020) distribute an electronic questionnaire featuring questions aligned with WCAG compliance criteria. Meanwhile, Song et al. (2018) harness crowdsourcing to evaluate manual WCAG criteria with the input of 50 non-expert workers.

3.3. Indicators and techniques used to assess digital accessibility

An examination of the indicators assessed within usability tests and controlled experiments finds a range of such indicators encompassing task accomplishment (e.g., execution speed), aspects related to the user (cognitive load, attention, emotions) and indicators associated with the interface itself (flexibility, learnability, robustness, etc.). To assess these indicators, the moderator takes specific measurements during the test, observes relevant behavior, or prompts the user to evaluate specific indicators. Table 3 provides a comprehensive breakdown of the indicators and techniques used in usability tests and controlled experiments. The indicator of efficiency appeared in 7 (22%) instances; the indicators of ease/difficulty in performing tasks in 3 (9%) instances; the indicators of WCAG compliance, satisfaction, and flexibility in two instances (6%) each; and the other indicators in one instance (3%) each, with the latter category including the following: strategies used to complete tasks, cognitive load, attention, emotional reactions, familiarity, support functionality, learnability, recoverability, robustness, visibility, sufficiency of instructions, ease of navigation, preferred input and feedback, achievability, effective communication, and minimum necessary physical effort.

In studies without usability testing, the collected indicators or data differ based on the method and study objectives involved. For instance, interviews aim to comprehend the social and local factors influencing the accessibility of specific digital services. Kameswaran and Muralidhar (2019) conducted interviews exploring digital

Table 3
Indicators and techniques used to assess digital accessibility in usability tests or controlled experiments. Author's own work

Object	Indicator	Assessment technique
Task	Efficiency = speed at which tasks are completed	Task execution time (Alonso-Virgos et al., 2020; Galkute et al., 2020; Giraud et al., 2018; Leria et al., 2018; Yi, 2020). Observed by the moderator (Maboe et al., 2018).
	Efficiency = accuracy and completeness of task completion	Percentage/number of tasks completed (Galkute et al., 2020; Park et al., 2019). Dropout rate and error rate (Giraud et al., 2018; Leria et al., 2018).
	Satisfaction = comfort, confidence, usefulness and pleasure with which tasks are completed	Rated by the user at the end of the test (Giraud et al., 2018; Leria et al., 2018).
	Ease/difficulty in performing tasks	Comments the user makes aloud while performing the tasks (Shachmut & Deschenes, 2019) or at the end of the test (Funes et al., 2018). Observed by the moderator (Maboe et al., 2018).
	Strategies used to complete tasks	Observed by the moderator (Galkute et al., 2020).
User	Cognitive load	French version of the NASA Raw Task Load indeX questionnaire and dual-task paradigm (Giraud et al., 2018).
	Attention	Eye-tracking (Alonso-Virgos et al., 2020).
	Emotional reactions	Affect Grid questionnaire (Nogueira et al., 2019).
Interface	Familiarity = capacity to connect prior knowledge in comprehending the interface	Observed by the moderator (Maboe et al., 2018).
	Support functionality = capacity to prevent errors or facilitate easy error recovery	
	Learnability = ease or simplicity in learning how to use the interface	
	Recoverability = ability of the interface to detect user errors and allow for their reversal or correction	
	Robustness = ensures that no harm is caused to the user during the process of rectifying errors	
	Visibility = presence of clear instructions for users regarding what actions they should take	
	Sufficiency of instructions	Comments the user makes aloud while performing the tasks (de Oliveira et al., 2021).
	Ease of navigation	
	Preferred input and feedback	Rated by the user at the end of the test (Funes et al., 2018).
	Achievability = possibility to execute the key functionalities of the product	Quantitatively rated by the user at the end of the test (Pedrini et al., 2020).
	WCAG compliance	Comments the user makes aloud while performing the tasks or at the end of the test (Yi, 2020). Observed by the moderator (Nogueira et al., 2019).
	Flexibility = different possibilities to complete tasks	Observed by the moderator (Maboe et al., 2018). Content analysis on user comments and responses (Park et al., 2019).
	Effective communication	Content analysis on user comments and responses (Park et al., 2019).
	Minimum necessary physical effort	

payment practices in public transport apps in metropolitan India. They observed, for instance, that payment preferences correlated with payment behaviors in the country, individuals' financial status, trust in digital payments, immediacy and the limitations of cash payments. Their discussion encompasses the role of digital payments in the relationship between drivers and passengers, as well as the interactions of passengers with assistants or strangers nearby. Lim et al.'s (2020) interviews were focused on understanding how individuals with disabilities in Singapore utilize government e-services. The authors studied the use of assistive technology, common issues encountered when using government e-services, internet usage practices and emotional aspects.

Publications that use questionnaires (outside the realm of usability tests) do so to facilitate access to a larger pool of participants. For instance, Alonso-Virgos et al. (2020) gathered 120 responses from individuals with Down syndrome regarding their leisure activities, activities affected by their condition, daily challenges related to listening, communication, etc., their practices, assistive technologies and obstacles they face while using the Web. Sabev et al. (2020) developed a questionnaire concerning WCAG conformance at levels A and AA. Respondents were asked to answer questions like "Does the site include a site map in an accessible HTML format?" (Ibid, p. 137). Lastly, the sole study utilizing crowdsourcing techniques (Song et al., 2018) formulated tasks for outsourcing based on WCAG criteria requiring human judgment.

3.4. User profiles involved in evaluations

Almost all selected papers (93.75%) provide a characterization of the users based on their functional disabilities. Visual impairments are notably prevalent which is likely to be linked to the predominantly graphical nature of digital information (Lim et al., 2020). The level of detail provided varies: from visual impairment in general (Pedrini et al., 2020) to the specification of five levels of visual acuity (Funes et al., 2018) and the common distinction between blindness and low vision (de Oliveira et al., 2021; Galkute et al., 2020; Giraud et al., 2018; Kameswaran & Muralidhar, 2019; Leria et al., 2018; Lim et al., 2020; Maboe et al., 2018; Nogueira et al., 2019; Park et al., 2019; Sabev et al., 2020; Song et al., 2018; Yi, 2020). Disabilities involving color vision (Sabev et al., 2020), auditory functions (Lim et al., 2020; Maboe et al., 2018; Song et al., 2018), intellectual functions (Alonso-Virgos et al., 2020) and motor functions (Song et al., 2018), particularly hand movement (Maboe et al., 2018), are also present in the selected papers, although to a lesser extent. Only one study detailed users concerning their physical conditions, specifically mild left hemiparesis (Lim et al., 2020).

The only study that did not characterize users based on their disabilities (6.25%) selected individuals who self-identified as proficient users of assistive technologies or regular users of captions and transcriptions (Shachmut & Deschenes, 2019).

The markers complementing this initial description consist of typical sociodemo-

Table 4
Complementary characteristics used to describe users involved in evaluations (in addition to disabilities and sociodemographic factors). Author's own work

Type of characteristic	User characteristic
Biographical information	Congenital or acquired blindness (Nogueira et al., 2019).
	Length of disabled experience (Park et al., 2019).
	Length of experience of smartphone use (Park et al., 2019).
	Length of experience in the use of computers and assistive technologies (Nogueira et al., 2019).
Skills relevant to the evaluation project	Previous experience with the interface to be evaluated (Galkute et al., 2020; Park et al., 2019).
	Previous involvement in usability or accessibility projects (Nogueira et al., 2019; Song et al., 2018).
	Professional experience in the domain of the interface to be evaluated (Pedrini et al., 2020).
	Web accessibility professional experience (Yi, 2020).
	Knowledge of the foreign language in which the interface to be evaluated is written (Park et al., 2019).
Characteristics of personal equipment	Smartphone ownership (Park et al., 2019).
	Smartphone operating system (Kameswaran & Muralidhar, 2019).
	Screen reader used (Galkute et al., 2020; Nogueira et al., 2019).
Technology-related habits	Frequency of web use (Yi, 2020).
	Experience with digital games (de Oliveira et al., 2021).

graphic indicators (gender, age, education level, employment status, city of residence), as well as biographical information, skills relevant to the evaluation project, characteristics of personal equipment and technology-related habits (see Table 4).

The involvement of users without disabilities is noted in two different scenarios – a control group consisting of sighted individuals whose outcomes are compared with those generated by non-sighted individuals (Nogueira et al., 2019) and among 21 out of the 50 participants who are outsourced for the manual evaluation of WCAG, aiming to diversify perspectives (Song et al., 2018).

Certain papers explicitly outline their recruitment strategy: through organizations targeting people with disabilities (Alonso-Virgos et al., 2020; Song et al., 2018), via online disability-related listservs and forums (Giraud et al., 2018; Kameswaran & Muralidhar, 2019), through personal connections and snowball sampling (Kameswaran & Muralidhar, 2019), among users of a specialized Braille library and students from a special education institute (Song et al., 2018).

The issue of participant compensation is also addressed in some papers (Funes et al., 2018; Kameswaran & Muralidhar, 2019; Maboe et al., 2018; Shachmut & Deschenes, 2019; Song et al., 2018; Yi, 2020). When provided, it typically involves a one-time payment based on the time spent on participation.

3.5. Space settings used in evaluations

The physical setting for the evaluations is correlated to the evaluation method employed. Usability tests, controlled experiments and crowdsourcing were conducted in simulated environments that participants were required to visit (81.25% of selected papers, see Table 2) while interviews and questionnaire occurred in more real or personal spaces.

Simulated environments are often referred to as usability labs (Funes et al., 2018; Nogueira et al., 2019; Park et al., 2019; Shachmut & Deschenes, 2019) and can be situated within a university (Maboe et al., 2018) or located in a place familiar to participants, such as the Federation of the Blind in Korea (Yi, 2020). Some studies utilize the participants' personal devices (Galkute et al., 2020; Park et al., 2019), while others make use of laboratory equipment (Alonso-Virgos et al., 2020; Funes et al., 2018; Pedrini et al., 2020; Yi, 2020). Shachmut and Deschenes (2019) propose an intermediate solution namely lab equipment configured based on user preferences, with user-provided input peripherals.

In scenarios involving the utilization of laboratory equipment, the choice of screen reader, web browser and operating system for the test becomes a crucial consideration. Funes et al. (2018) provided users the option between two screen readers (NVDA or JAWS) and two web browsers (Firefox or Internet Explorer). Pedrini et al. (2020) adhered to the American Foundation of the Blind's recommendations, selecting the NVDA screen reader for Windows OS and the native VoiceOver reader for Mac OS. Similarly, Yi (2020) mandated the use of Internet Explorer and the Sense Reader screen reader, as did Nogueira et al. (2019), who enforced the JAWS screen reader.

Specific space settings based on cultural features were not observed although the evaluations did cover diverse countries (Fig. 2).

3.6. Advantages of user participation

The studies analyzed enumerate various benefits associated with involving users in accessibility evaluations. They emphasize the insight these methods offer into people's genuine needs (Alonso-Virgos et al., 2020), the challenges they encounter (Yi, 2020), their mental models (de Oliveira et al., 2021), their digital competencies (Leria et al., 2018), their ways of interacting with digital interfaces (Sabev et al., 2020) and the crucial role that social interactions and collaboration play in accessibility conditions (Kameswaran & Muralidhar, 2019).

User participation also facilitates understanding of the relationship between accessibility and the technical environment of use such as hardware, browser, screen reader configuration, or voice quality (Leria et al., 2018). Similarly, they can assess the compatibility of assistive technologies with the latest digital advances (Nogueira et al., 2019).

According to the literature, evaluating accessibility with users yields positive outcomes: it identifies opportunities for improving the assessed product (de Oliveira et al., 2021; Galkute et al., 2020); the results provide a foundation for prioritizing accessibility issues or criteria (Alonso-Virgos et al., 2020; Lim et al., 2020); and interacting with users enables test moderators to grasp tangible accessibility implications (Shachmut & Deschenes, 2019).

Finally, the authors identify the financial and reliability advantages of methods that involve users. The results generated by these methods contain a high level of detail (Yi, 2020), are objective (Alonso-Virgos et al., 2020) and serve as empirical evidence of user requirements (Park et al., 2019). In the case of crowdsourcing, Song et al. (2018) argue that hiring users for evaluating conformance criteria which require human judgment is more cost-effective than employing experts.

3.7. *Limitations of user participation*

Among the limitations of user participation, concerns about the representativeness of participants frequently emerge. In some studies, the number of participants and the types of disabilities recruited are considered unrepresentative of the entire population with disabilities (Galkute et al., 2020; Lim et al., 2020; Park et al., 2019). For instance, the study focusing on the Indian population with visual impairment suggests that a sample dominated by males of high socioeconomic status and formal education does not represent that population adequately (Kameswaran & Muralidhar, 2019). Additionally, the representativeness of the evaluated interfaces is mentioned as a limitation, for example, studying chauffeur-driven transport vehicle services might not be adequate to draw conclusions about the accessibility of digital payments in general (Kameswaran & Muralidhar, 2019).

There are also limitations related to the scope of the results. Usability tests may not identify issues related to content code that automated evaluations can, such as HTML H1 tags that lack text (Galkute et al., 2020). Additionally, usability tests are typically conducted in a single session which means they may not uncover problems related to prolonged usage patterns or analyze user learning (Park et al., 2019).

Other limitations pertain to the execution of usability testing. Yi (2020) acknowledges that this method demands more effort and time compared to automated tests or expert evaluations. Bringing a non-sighted participant to the test session can pose a genuine challenge particularly if the testing site has multiple entrances and is located on a university campus without a specific street address (Shachmut & Deschenes, 2019). Configuring Wi-Fi or screen recording can be time-consuming when using participants' personal computers (Shachmut & Deschenes, 2019). Conversely, if a screen reader and operating system that the user is not familiar with are imposed for the test, the evaluation results will be significantly biased (Pedrini et al., 2020).

4. Discussion

The evaluation method and the chosen indicators reveal much about the underlying definition of accessibility, as emphasized by Brajnik (2008), as well as the user's role in the assessment. In the reviewed usability tests, accessibility aligns with usability, which is defined by task accomplishment indicators, user abilities and emotions and digital interface qualities. This perspective is akin to that of the Person-Environment

Model (Iwarsson & Stahl, 2003), where usability should assess four dimensions: (1) the individual's functional capacity, (2) the barriers in the target environment, in relation to the standards available but also based on user subjectivity, (3) the tasks the person must perform within that environment and (4) the extent to which the individual's needs can be fulfilled in that environment in terms of task performance. It is worth noting that the individual is considered here in their individuality, without accounting for their interpersonal relationships and the observed environment is solely physical.

Moreover, works employing the interview method emphasize the socially constructed aspect of accessibility. Accessibility is elucidated here by the conventions and habits of the local context, by interpersonal relationships that mediate or accompany technology (such as the driver or a stranger in the case of digital payments), by the person's history, by the characteristics of the activity without technology mediation, among other factors. Accessibility aligns in this perspective with social participation, as defined by the Disability Creation Process Model (Fougeyrollas et al., 2019). Social participation occurs in daily life situations where personal factors interact with physical and social environmental factors. The field of web accessibility also refers to a process, shaped by political, sociocultural and technical factors (Cooper et al., 2012). This perspective thus underscores the situated and evolving nature of accessibility.

Studies that solicit users to evaluate compliance with WCAG via questionnaires or crowdsourcing seem questionable to us. Although conformity assessments demand human judgment in verifications that cannot be automated (WAI, 2014), the evaluators' level of expertise plays a critical role in assessment quality (Brajnik et al., 2011). Considering that is relatively unlikely that all recruited users are experts in WCAG and conformance assessments (unless this is an explicit selection criterion), user participation in these cases does not seem appropriate.

The reviewed usability tests focusing on indicators related to the digital interface align with the recommendations made by the WAI for such tests. The WAI suggests collecting errors related to accessibility barriers instead of indicators like task execution time or user satisfaction (WAI, 2005). This recommendation likely stems from the distinction made by the WAI between accessibility, pertaining to individuals with disabilities and usability, concerning all users (WAI, 2010). However, other studies (Aizpurua et al., 2014) demonstrate that users do not perceive accessibility in terms of interface barriers and it is the test moderator who "translates" the actions or statements of the user into interface issues. Consequently, the researcher has a lower likelihood of influencing the results when the indicators assess the user or task execution.

The description of users involved in the examined evaluations reveals underlying conceptions about what constitutes disability. The prevailing form of description alludes to the participants' functional impairments and places disability within the individual's body, evoking what is commonly referred to as the medical model of disability (Marks, 1997). This same model is evident in studies that establish a control group by selecting individuals without functional impairments. Conversely, the conception of disability leans toward a social perspective when participants are characterized by aspects of the person's living environment, such as personal technological equipment, or by non-health-associated personal factors like skills, life history, or habits. As mentioned previously, the social model asserts that situations of disability stem from the interplay between the individual's factors and those present within their social living environment (Fougeyrollas et al., 2019).

The spatial aspect of evaluations is predominantly discussed in usability tests. Given that this method focuses on the person-physical environment relationship, it is common to consider the physical location of the evaluation, the configuration of computer equipment (both hardware and software) and the user interfaces selected for testing. The user's familiarity with the testing environment is emphasized in some studies as a means to replicate real-world conditions and ensure user comfort, thereby minimizing potential testing biases (Aizpurua et al., 2014). Conversely, in other studies, this factor appears to be of lesser importance, with some even proposing specially designed interfaces for testing that do not resemble any real-life counterparts.

5. Conclusions

This article reviewed scientific literature where users evaluated digital accessibility, aiming to understand their methods and objectives. Three evaluation types were identified. The first involves users in usability testing to evaluate tasks accomplishment, user reactions and interface qualities. The second type is based on interviews of users to assess the local and social factors impacting digital service accessibility. The third type involves users checking the compliance of certain interfaces with accessibility standards using questionnaires or crowdsourcing. Participants are primarily chosen based on their functional impairments and, to a lesser degree, their project-related skills, biographical information or technology habits, among other criteria. The authors appreciate the comprehensive user insights gained while also regretting the lack of representativeness of the selected users sample.

The results reveal diverse perspectives on accessibility and disability which are occasionally in conflict with the social model of disability. In this context, we believe there is an unexplored research avenue in digital accessibility evaluations that delve into the personal and contextual factors influencing accessibility, with particular emphasis on the evolving nature of this phenomenon. Users, as experts in their own experiences, will play a pivotal role in these assessments.

Acknowledgments

This Cuban-French collaboration has been funded by the French Ministry of Europe and Foreign Affairs and the French Ministry of Higher Education, Research and Innovation within the framework of the Hubert Curien Carlos J. Finlay 2021 cooperation programme (project number 47098PF).

References

Aizpurua, A., Arrue, M., Harper, S., & Vigo, M. (2014). Are users the gold standard for accessibility evaluation? *Proceedings of the 11th Web for All Conference*, pp. 1-4. doi: 10.1145/2596695.2596705.

Alonso-Virgos, L., Baena, L.R., & Crespo, R.G. (2020). Web accessibility and usability evaluation methodology for people with Down syndrome. En A. Rocha, B.E. Perez, F.G. Penalvo, M. del Mar Miras, & R. Goncalves (Eds.), *Iberian Conference on Information Systems and Technologies, CISTI* (Vols. 2020-June). IEEE Computer Society. doi: 10.23919/CISTI49556.2020.9140858.

Ara, J., Sik-Lanyi, C., & Kelemen, A. (2023). Accessibility engineering in web evaluation process: A systematic literature review. *Universal Access in the Information Society*. doi: 10.1007/s10209-023-00967-2.

Bianchini, D., Galasso, G., Gori, M., Codagnone, C., Liva, G., Stazi, G., Laurin, S., & Switters, J. (2022). *Study supporting the review of the application of the Web Accessibility Directive (WAD)*. Publications Office of the European Union. https://data.europa.eu/doi/10.2759/25194.

Brajnik, G. (2008). Beyond Conformance: The Role of Accessibility Evaluation Methods. En S. Hartmann, X. Zhou, & M. Kirchberg (Eds.), *Web Information Systems Engineering – WISE 2008 Workshops*, Springer, pp. 63-80. doi: 10.1007/978-3-540-85200-1_9.

Brajnik, G., Yesilada, Y., & Harper, S. (2010). Testability and validity of WCAG 2.0: The expertise effect. *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 43-50. doi: 10.1145/1878803.1878813.

Brajnik, G., Yesilada, Y., & Harper, S. (2011). The expertise effect on web accessibility evaluation methods. *Human-Computer Interaction*, *26*(3), 246-283. doi: 10.1080/07370024.2011.601670.

Campoverde-Molina, M., Lujan-Mora, S., & Garcia, L.V. (2020). Empirical studies on web accessibility of educational websites: A systematic literature review. *IEEE Access*, *8*, 91676-91700. doi: 10.1109/ACCESS.2020.2994288.

Cooper, M., Sloan, D., Kelly, B., & Lewthwaite, S. (2012). A Challenge to Web Accessibility Metrics and Guidelines: Putting People and Processes First. *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility*, pp. 201-204. doi: 10.1145/2207016.2207028.

de Oliveira, R.N.R., Belarmino, G.D., Rodriguez, C., Goya, D., da ROCHA, R.V., Venero, M.L.F., Benitez, P., & Kumada, K.M.O. (2021). Development and evaluation of usability and accessibility of an educational digital game prototype aimed for people with visual impairment [Desenvolvimento e avaliação da usabilidade e acessibilidade de um protótipo de jogo educacional digital para pessoas com deficiência visual]. *Revista Brasileira de Educacao Especial*, *27*, 847-864. doi: 10.1590/1980-54702021v27e0190.

Fougeyrollas, P. (2021). Classification internationale 'Modèle de développement humain-Processus de production du handicap' (MDH-PPH, 2018). *Kinésithérapie, la Revue*, *21*(235), 15-19. doi: 10.1016/j.kine.2021.04.003.

Fougeyrollas, P., Boucher, N., Edwards, G., Grenier, Y., & Noreau, L. (2019). The disability creation process model: A comprehensive explanation of disabling situations as a guide to developing policy and service programs. *Scandinavian Journal of Disability Research*, *21*(1), 25-37. doi: 10.16993/sjdr.62.

Funes, M.M., Trojahn, T.H., Fortes, R.P.M., & Goularte, R. (2018). Gesture4all: A framework for 3D gestural interaction to improve accessibility of web videos. *Proceedings of the ACM Symposium on Applied Computing*, 2151-2158. doi: 10.1145/3167132.3167363.

Galkute, M., Rojas P, L.A., & Sagal M, V.A. (2020). Improving the web accessibility of a university library for people with visual disabilities through a mixed evaluation approach. En G. Meiselwitz, *12th International Conference on Social Computing and Social Media, SCSM 2020, held as part of the 22nd International Conference on Human-Computer Interaction, HCII 2020: Vol. 12194 LNCS*, Springer.

Giraud, S., Therouanne, P., & Steiner, D.D. (2018). Web accessibility: Filtering redundant and irrelevant information improves website usability for blind users. *International Journal of Human-Computer Studies*, *111*, 23-35. doi: 10.1016/j.ijhcs.2017.10.011.

Iwarsson, S., & Stahl, A. (2003). Accessibility, usability and universal design – Positioning and definition of concepts describing person-environment relationships. *Disability and Rehabilitation*, *25*(2), 57-66. doi: 10.1080/dre.25.2.57.66.

Kameswaran, V., & Muralidhar, S.H. (2019). Cash, digital payments and accessibility – A case study from India. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW). doi: 10.1145/3359199.

Laitano, M.I., Ortiz-Núñez, R., & Stable-Rodríguez, Y. (2024). User participation in digital accessibility evaluations: reviewing methods and objectives [Data set]. *Zenodo*. doi: 10.5281/zenodo.10908703.

Lazar, J. (2017). *Research methods in human computer interaction* (2nd edition). Elsevier.

Leria, L.A., Filgueiras, L.V.L., Ferreira, L.A., & Fraga, F.J. (2018). Autonomy as a quality factor in the development of an application with digital accessibility for blind students in large-scale examinations. En L.G. Chova, A.L. Martinez, & I.C. Torres (Eds.), *11TH International Conference of Education, Research and Innovation (ICERI2018)*, pp. 5724-5731.

Lim, Z.Y., Chua, J.M., Yang, K., Tan, W.S., & Chai, Y. (2020). Web accessibility testing for Singapore government e-services. *Proceedings of the 17th International Web for All Conference, W4A 2020*. 17th International Web for All Conference, W4A 2020. doi: 10.1145/3371300.3383353.

Maboe, M.J., Eloff, M., & Schoeman, M. (2018). The role of accessibility and usability in bridging the digital divide for students with disabilities in an e-learning environment. En J. van Niekerk & B. Haskins (Eds.), *ACM International Conference Proceeding Series*, Association for Computing Machinery, pp. 221-228. doi: 10.1145/3278681.3278708.

Marks, D. (1997). Models of disability. *Disability and Rehabilitation*, *19*(3), 85-91. doi: 10.3109/09638289709166831.

Nogueira, T.D.C., Ferreira, D.J., & Ullmann, M.R.D. (2019). Impact of accessibility and usability barriers on the emotions of blind users in responsive web design. En V.F. de Santana, A. Kronbauer, J.J. Ferreira, V. Vieira, R.L. Guimaraes, & C.A.S. Santos (Eds.), *IHC 2019 – Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems*. Association for Computing Machinery, Inc. doi: 10.1145/3357155.3358433.

Nuñez, A., Moquillaza, A., & Paz, F. (2019). Web Accessibility Evaluation Methods: A Systematic Review. En A. Marcus & W. Wang, *8th International Conference on Design, User Experience and Usability, DUXU 2019, held as part of the 21st International Conference on Human-Computer Interaction, HCI International 2019: Vol. 11586 LNCS*. Springer Verlag.

Park, K., So, H.J., & Cha, H. (2019). Digital equity and accessible MOOCs: Accessibility evaluations of mobile MOOCs for learners with visual impairments. *Australasian Journal of Educational Technology*, *35*(6), 48-63. doi: 10.14742/ajet.5521.

Pedrini, G., Ludovico, L.A., & Presti, G. (2020). Evaluating the Accessibility of Digital Audio Workstations for Blind or Visually Impaired People: *Proceedings of the 4th International Conference on Computer-Human Interaction Research and Applications*, pp. 225-232. doi: 10.5220/0010167002250232.

Sabev, N., Georgieva-Tsaneva, G., & Bogdanova, G. (2020). Research, analysis and evaluation of web accessibility for a selected group of public websites in Bulgaria. *Journal of Accessibility and Design for All*, *10*(1), 124-160. doi: 10.17411/jacces.v10i1.215.

Shachmut, K., & Deschenes, A. (2019). Building a Fluent Assistive Technology Testing Pool to Improve Campus Digital Accessibility (Practice Brief). *Journal of Postsecondary Education and Disability*, *32*(4), 445-452.

Silva, C.A., de Oliveira, A.F.B.A., Mateus, D.A., Costa, H.A.X., & Freire, A.P. (2019). Types of problems encountered by automated tool accessibility assessments, expert inspections and user testing: A systematic literature mapping. *Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems*, 1-11. doi: 10.1145/3357155.3358479.

Song, S., Bu, J., Artmeier, A., Shi, K., Wang, Y., Yu, Z., & Wang, C. (2018). Crowdsourcing-based web accessibility evaluation with golden maximum likelihood inference. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW). doi: 10.1145/3274432.

W3C Web Accessibility Initiative (WAI). (2005). *Involving Users in Evaluating Web Accessibility*. Web Accessibility Initiative (WAI). https://www.w3.org/WAI/test-evaluate/involving-users/.

W3C Web Accessibility Initiative (WAI). (2010). *Accessibility, Usability, and Inclusion*. Web Accessibility Initiative (WAI). https://www.w3.org/WAI/fundamentals/accessibility-usability-inclusion/.

W3C Web Accessibility Initiative (WAI). (2014). *Website Accessibility Conformance Evaluation Methodology (WCAG-EM) 1.0*. http://www.w3.org/TR/WCAG-EM/.

Yi, Y.J. (2020). Web accessibility of healthcare Web sites of Korean government and public agencies: A user test for persons with visual impairment. *Universal Access in the Information Society*, *19*(1), 41-56. doi: 10.1007/s10209-018-0625-5.