A Validation Test of an Accessibility Evaluation Method

Yu Ting Sun$^{(\boxtimes)}$, Kim-Phuong L. Vu, and Thomas Z. Strybel

California State University, Long Beach, USA
Yuting.sun@csulb.edu

Abstract. Digital materials have become a popular medium for information access, and attracted a diverse group of users, such as college students, who benefit from the low cost of, and portable access to, the materials. However, college students with disabilities may have difficulty accessing electronic materials if the materials were not developed appropriately. Laws and standards provide guidance on making digital documents accessible, but these regulations are implemented slowly. As a result, published materials on the market may have accessibility issues. Efforts have been made to produce evaluation methods for eBooks. For example, automated tools have been used to check for accessibility aspects in multiple studies, but using automated tools to evaluate accessibility of electronic materials is not enough due to the complexity of the checkpoints. Thus, human evaluators are needed. This study assessed a newly developed accessibility evaluation methodology that was designed for e-textbooks, and examined whether books that were rated as higher in accessibility versus books that were rated lower in accessibility resulted in differences in user experience and performance. This study consisted of 6 students with visual impairments and 6 students with normal or corrected-to-normal vision, who read and interacted with eBooks. User experience and performance were measured using subjective questionnaires, reading time, and accuracy to content-related questions. We found differences in user experience ratings for eBooks that were rated as high or low in accessibility; however, we found no differences in users' task performance as a function of the accessibility level of the eBook.

Keywords: Accessibility evaluation · Accessibility · Evaluation criteria · Usability evaluation · e-textbook · e-text book usability · e-textbook accessibility · eBook accessibility · eBooks

1 Introduction

Accessibility in general means the extent to which the complete use of any resource is available to the greatest number of people [1]. Accessibility involves two issues: whether users with disabilities can access electronic information and how document content functions with assistive technologies [2]. When websites meet the goal of serving all user groups, they are accessible for people with a diverse range of hearing, movement, visual, and cognitive abilities. "The UN Convention on the Rights of Persons with Disabilities recognizes access to information and communications technologies, including the Web, as a basic human right" [3]. It is important that websites

and documents provide equal access and opportunities to everyone, including people with diverse disabilities.

Over the past decade, digital materials such as eBooks have become increasingly popular for many user groups such as college students because they are low in cost and can be easily accessed. To promote the use of eBooks, legislation was enacted in California that directed the California higher education system (California State University, University of California, and California Community Colleges) to collaborate to achieve "the goal of making higher education in California more affordable by providing faculty and students access to free and lower-cost instructional materials" [4]. This legislation supports students' access to e-textbooks for free and should, as a result, increase the use of eBooks in the classroom.

With the increasing prevalence of digital learning materials and the integration of computers in all facets of education, it is important that these materials be made accessible to populations with disabilities. Accessibility can be achieved with electronics resources because the material can be coded in a manner that makes the content easier to format and manipulate (e.g., change font size or color) than their physical, printed counterparts.

Accessibility standards are used to help designers and developers identify and address accessibility issues. For instance, the Web Content Accessibility Guidelines (WCAG) 2.0 from the World Wide Web Consortium (W3C) developed standards and guidelines for web accessibility. These standards are also used in Australia, Canada, and many other countries [2]. The WCAG guidelines are helpful for developers as they set standards for web accessibility. However, these guidelines lack an evaluation component and there is no validated evaluation tool to help users and designers understand why a particular website is or is not accessible. In fact, the W3C recognized the lack of validation on these accessibility guidelines and acknowledge that further work on the topic is needed [5]. Thus, although laws and standards for making digital documents accessible exist, they are being implemented slowly and published materials on the market can have accessibility issues.

Several accessibility evaluation tools are available on the market for people to use for conducting web evaluations, but there has been no real test of the effectiveness of each of these tools as a means for evaluating eBooks [6]. This is especially the case for eBooks produced in EPUB and PDF formats. Accessibility evaluations need human expertise to determine whether a book is actually accessible, even with the use of accessibility tools. Sun et al. [7] and Chan et al. [8] developed a methodology for evaluators to use to determine accessibility of various formats of electronic textbooks.

Sun et al. [7] reviewed the tools that were available for evaluating various aspects of eBooks by using a SkillsCommons checklist as a guide. The 15 SkillsCommons accessibility checkpoints were developed by the California State University and Multimedia Educational Resource for Learning and Online Teaching (MERLOT) program, and have been used to evaluate accessibility of electronic text content and media on the web. These checkpoints are based on guidelines for web accessibility and experience from subject matter experts and developers [9]. Sun et al. [7] used the checkpoints to develop evaluation methodologies and scoring system. The evaluation method was developed in two phases: tools review phase and manual creation. In the tools evaluation phase, the researchers reviewed existing tools and methods that could

be used to assess the 15 SkillsCommons checkpoints. Appropriate tools were selected and used in the manual creating phase. Step-by-step manuals were created based on comparisons of functions and compatibility with different operating systems. These manuals were used in all eBook evaluations.

Three different computations, composite score, average score, and weighted average score, were generated to represent the overall accessibility level of e-textbooks [8]. Weighted average scores were selected to be the best scoring metrics after working with accessibility SMEs to determine severity weightings for each checkpoint in terms of its overall accessibility impact on users.

The purpose of this study was to validate the newly developed accessibility evaluation method and metrics developed by Sun et al. [7] and Chan et al. [8]. This study was designed to address the question of whether a book rated high in accessibility provides a better user experience and performance for users with disabilities than a book rated low in accessibility. Different types of content [e.g., science, technology, engineering, and mathematics (STEM) vs. non-STEM] were also examined to determine whether it influences user experience and performance.

2   Method

Participants. Participants in this study were six students with visual impairments recruited from Disabled Student Services (DSS) at California State University Long Beach, through fliers posted at CSULB and the surrounding community, or through the snowball technique. Six students without visual impairments were recruited through fliers and snowball technique as well. All participants were compensated for their time at $15 per hour. A screening protocol was administered to identify participants with visual impairments.

Materials. Four e-textbooks from the Cool4ed website (www.cool4ed.org) were selected. All four books were in the same format (EPUB, HTML, or PDF) and two were in the same general subject area. Based on data collected in pilot testing, chapters from the four books were selected based on the difficulty of the materials to ensure that the material were comparable in terms of difficulty. Comprehension questions were presented at the end of each chapter to assess users' performance. The difficulty of the comprehension questions was also determined through pilot testing. The topics of two of the chapters were STEM related, and the topics of the two other chapters were non-STEM related. As shown in Table 1, one STEM and one Non-STEM chapter were determined to be high in accessibility based on the method and scoring metric from Chan et al. [8] and Sun et al. [7]; the remaining two chapters were rated low in accessibility.

Morae was used to record participants' interactions with the eBooks. Reading times were recorded by screen recorders and times for responding to questions were recorded by the researcher. A pretest questionnaire was administered prior to the experiment. Subjective questionnaires and the System Usability Scale (SUS) [11] were used to assess user experience after each condition. All questionnaires were presented orally or visually, based on participant preference, at the end of each session.

Table 1. Chapters used in this study categorized by accessibility level and content.

Accessibility level	Content	Chapter topic	Accessibility score*
High	STEM	Water	9.26
	Non-STEM	Charles II and restoration colonies	9.12
Low	STEM	Energy and metabolism & ATP	4.99
	Non-STEM	Britain in the new world	4.49

* Scores based on method from Sun et al., 2016 [7].

Procedure. The testing took place over two days. Prior to the experiment, participants were asked to provide informed consent and to complete a pretest questionnaire. During the experiment, participants were asked to read a chapter from each book and then answer questions on the content of the chapter. Participants were allowed to refer to the book when answering the reading comprehension questions. Assistive technologies such as JAWS, ZoomText, and NVDA were available for participants if requested.

A counterbalancing scheme was used to assign the order of the four books to participants. For each book, the corresponding questions were administered orally to students with visual impairments or on paper for students without visual impairments. Immediately after responding to the comprehension questions, participants were asked to complete the SUS and eBook experience questionnaire. Participants were then given a short break, after which they completed the same process for the second book and repeated the procedures for the third and fourth books on the second day.

3   Results and Discussion

Reading time and accuracy on the reading tasks, and ratings to user experience surveys with the chapters, were submitted to separate mixed-ANOVAs with the factors of accessibility level (High vs. Low) and chapter content (STEM vs. Non-STEM) being within-subjects and user group (students with visual impairments vs. Students without visual impairments) being between-subjects.

Performance. There was a significant difference between user groups in terms of reading performance. The students without visual impairments read the materials more quickly than the students with visual impairments. Students with visual impairments took longer because they had to employ assistive technologies to access the books. A significant interaction between user group and content level was found on accuracy scores, as shown in Fig. 1. Accuracy for students without visual impairments was unaffected by content, but accuracy for students with visual impairments was higher for non-STEM books than for STEM books. The book content (STEM vs. non-STEM) made a difference only in accuracy (proportion of correct answers to the comprehension questions).

User Experience. Subjective experiences of participants were measured with a standard scale of usability, the System Usability Scale (SUS), and with custom rating
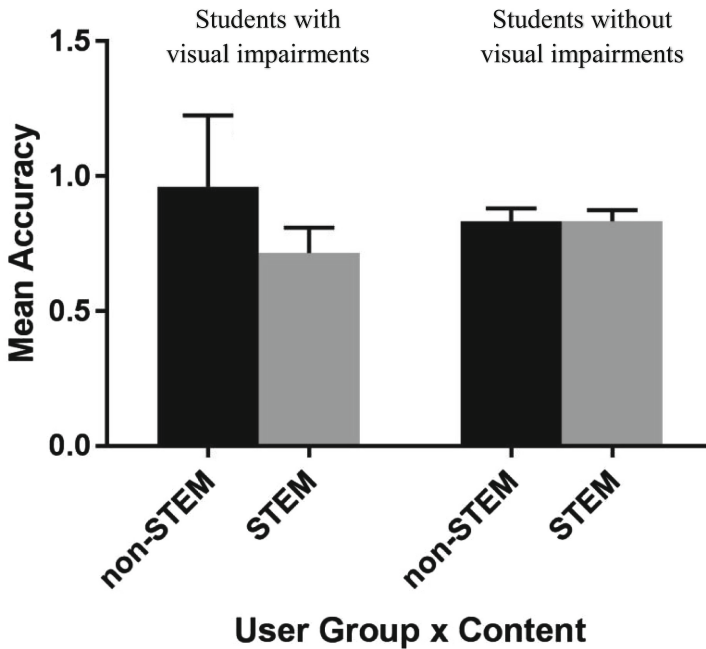
Fig. 1. Mean accuracy as a function of user group and content. There were no differences in accuracy between science, technology, engineering, and mathematics (STEM) and non-STEM books for students without visual impairments, but students with visual impairments showed higher accuracy for Non-STEM books than STEM books. Error bars represent one standard error of the mean (SEM).

questions. SUS scores have been shown to be a reliable and valid tool for measuring usability [10] and [11]. The custom questions asked participants about their satisfaction with the eBook, views on the accessibility of the chapter, their likelihood of purchasing an eBook from the same publisher, their willingness to purchase the eBook just read, and acceptability of the eBook being a required text in a course. These were administered after each chapter was completed.

As shown in Fig. 2, students without visual impairments produced higher SUS scores than the students with visual impairments. Moreover, chapters rated high in accessibility produced higher SUS scores compared to chapters rated low in accessibility. According to Tullis and Albert [12], SUS scores above 70 are considered to have acceptable usability, scores between 50 and 70 are considered to be of marginal usability, and scores below 50 are unacceptable in terms of usability. Therefore, the books rated high in accessibility used here were shown to have acceptable usability, with a mean SUS score of 77.81; the books rated low in accessibility produced marginal usability scores, with a mean SUS score of 57.29.

For the custom usability questions, all users reported better experiences when using the books rated high in accessibility compared to books rated low in accessibility. As shown in Table 2, all users reported that books rated higher in accessibility were
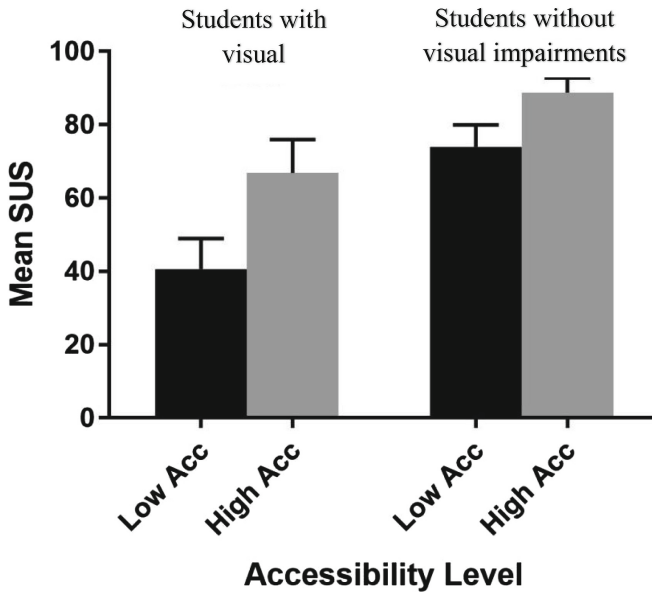
Fig. 2. Mean difference values representing System Usability Scale (SUS) scores. There was a significant difference in SUS scores between user group. SUS scores for the students with visual impairments were significantly lower than those with non-impaired vision. Error bars represent one standard error of the mean (SEM).

significantly more usable ($p = .007$), provided greater levels of satisfaction ($p = .006$), and were more accessible ($p < .001$), compared to books that were rated lower in terms of accessibility. Moreover, for books rated higher in accessibility, all users stated that they would be more likely to use other books from the same publisher ($p = .020$), would purchase the book that they had just read ($p < .001$), and felt more positive about the book being required in a class ($p < .001$). These results support the conclusion that the accessibility level of books made a difference in user experience and suggest that the accessibility evaluation method is capturing components influencing the users' experiences with eBooks.

The fact that both groups of users were affected by the accessibility levels could be explained by their comments when asked how they felt about the books and when asked to list what they liked or disliked about the books. For the books rated low in accessibility, both groups of users reported that there was too much advertising around the content, which was distracting. Both groups of users also reported that the contrast for some of the content made it difficult to read.

However, students with visual impairments reported significantly lower ratings on all scales, as shown in Table 3. Students without visual impairments reported better experiences with these eBooks probably because the accessibility features did not impact their experience as much. Students with visual impairments commented on the difficulty of navigating through the document, and difficulty in seeing some portions of

Table 2. Users subjective experience ratings for high vs. low accessibility level eBooks.

Question re: User Experience	High accessibility level	Low accessibility level
"Rate your overall experience" (1 = unsatisfied; 5 = satisfied)	M = 3.83, SD = 1.05	M = 2.96, SD = 1.27
"Do you think this e-book is accessible?" (0 = no; 1 = somewhat; 2 = yes).	M = 1.75, SD = 0.68	M = 1.04, SD = 0.62
"How likely will you use an e-book that is published from the same publisher?" (1 = never; 5 = very often)	M = 3.54, SD = 1.25	M = 2.75, SD = 1.23
"Would you be willing to pay to purchase this e-book?" (0 = no; 1 = yes)	M = 0.79, SD = 0.42	M = 0.21, SD = 0.42
"How would you feel if this e-book was a required text?" (0 = negative; 1 = positive)	M = 0.79, SD = 0.42	M = 0.21, SD = 0.42

Table 3. Users subjective experience ratings for students with visual impairments vs. students without visual impairments.

Question re: User Experience	Students with visual impairments	Students without visual impairments
"Rate your overall experience" (1 = unsatisfied; 5 = satisfied)	M = 2.92, SD = 1.44	M = 3.88, SD = 0.74
"Do you think this e-book is accessible?" (0 = no; 1 = somewhat; 2 = yes).	M = 1.17, SD = 0.82	M = 1.62, SD = 0.58
"How likely will you use an e-book that is published from the same publisher?" (1 = never; 5 = very often)	M = 2.54, SD = 1.32	M = 3.75, SD = 0.94
"Would you be willing to pay to purchase this e-book?" (0 = no; 1 = yes)	M = 0.38, SD = 0.49	M = 0.63, SD = 0.49
"How would you feel if this e-book was a required text?" (0 = negative; 1 = positive)	M = 0.38, SD = 0.50	M = 0.63, SD = 0.50

the chapter due to low contrast. Moreover, one student commented that it was difficult to know where he or she was in the reading.

The effects of accessibility levels on subjective experience ratings provided evidence to support the methods developed by Sun et al. [7] and Chan et al. [8] for determining accessibility of eBooks. Main effects of level were obtained on all subjective user experience measures, which suggest that both groups (students with visual impairments and students without visual impairments) were affected by the accessibility levels. The results suggest that usability and accessibility are related and may be similar in evaluation criteria. For example, one of the criteria in the accessibility evaluations was that users could navigate through various types of content by checking the structural markups of the materials. This is related to user control and freedom and the error prevention criterion used in heuristics evaluations for usability [13].

The findings from this study were consistent with a study conducted by Bayer and Pappas [14], who performed accessibility testing on JAWS and Microsoft Word using usability evaluation techniques. They found that almost all of the problems identified from the tests were either navigation or screen reading problems. Bayer and Pappas found that standard key combinations (e.g., activating screen reading) did not work for the participants. Participants were not able to access some of the books because they were using standard key combinations that they use daily for other computer activities.

Overall, the results suggested that the accessibility levels obtained by the method developed by Sun et al. and Chan et al. [7, 8] discriminated between eBooks with respect to user experiences. The current study showed that a book rated high in accessibility provides a better user experience for users with disabilities than a book rated low in accessibility. However, performance was unaffected by accessibility level, possibly due to the limited data collected.

Further research is needed using a larger sample to re-examine whether there is a difference in user performance by accessibility level. Inclusion of user groups beyond visually impaired and normal or corrected-to-normal vision students should be considered. Another possible direction for future research would be conducting user testing online. All of the eBooks evaluated by Sun et al. [7] were available online through the Cool4ed website. An online study would allow users to complete the study at home, which could broaden the pool of participants.

To improve the general accessibility of eBooks for the general population and for users with disability, publishers are encouraged to follow the accessibility standards while developing books and to evaluate their books using these standards after publication. Based on the findings from this study, a major redesign recommendation to assist students with visual impairments would be make the navigation of eBooks to match standard commands. Students with visual impairments were not able to use the standard combinations for commands to activate the assistive technologies; thus, they could not access content of these eBooks using the commands that they normally use. Coding the books to match standard combinations for commands would greatly improve accessibility of these eBooks for students with visual impairments.

References

1. Booth, C.: Why accessibility? Libr. Technol. Rep. **48**(7), 5 (2012)
2. Adobe. What is accessibility? (2016). http://www.adobe.com/accessibility/gettingstarted.html
3. W3C. Accessibility (2010). https://www.w3.org/standards/webdesign/accessibility
4. Cool4Ed. About this project (2014). http://www.cool4ed.org/about.html
5. Abascal, J., Arrue, M., Fajardo, I., Garay, N., Tomás, J.: The use of guidelines to automatically verify web accessibility. Univ. Access Inf. Soc. **3**(1), 71–79 (2004)
6. Web Accessibility Initiative. Introduction to web accessibility (2010). https://www.w3.org/WAI/intro/accessibility.php
7. Sun, Y., Manabat, A.K.M., Chan, M., Chong, I., Vu, K.: Accessibility evaluation: manual development and tool selection for evaluating accessibility of e-textbooks (2016). http://link.springer.com/chapter/10.1007/978-3-319-41691-5_28

8. Chan, M., Sun, Y., Tesoro, A., Vu, K.: Development of a scoring system for evaluating the accessibility of etextbooks (2016). https://www.researchgate.net/publication/304995934_Development_of_a_Scoring_System_for_Evaluating_the_Accessibility_of_eTextbooks
9. SkillsCommons. Accessibility and SkillsCommons. http://support.taaccct.org/accessibility/
10. Sauro, J.: Measuring usability with the System Usability Scale (SUS) (2011). http://www.measuringu.com/sus.php
11. Usability.gov. Heuristic evaluations and expert reviews (2014). http://www.usability.gov/how-to-and-tools/methods/heuristic-evaluation.html
12. Tullis, T., Albert, B.: Measuring the user experience: collecting, analyzing, and presenting usability metrics, 2nd edn. Morgan Kaufmann, Waltham (2013)
13. Neilson J.: Beyond accessibility: treating users with disabilities as people (2001). https://www.nngroup.com/articles/beyond-accessibility-treating-users-with-disabilities-as-people/
14. Bayer, N., Pappas, L.: Accessibility testing: case history of blind testers of enterprise software. Tech. Commun. **53**(1), 32–38 (2006)