2019
The Journal on Technology and Persons with Disabilities

# LaTeX is NOT Easy: Creating Accessible Scientific Documents with R Markdown

JooYoung Seo, Sean McCurry

Learning, Design, and Technology, The Pennsylvania State University; IT Accessibility Team, The Pennsylvania State University

jooyoung@psu.edu, smm48@psu.edu

## Abstract

Although recent advancements in assistive technology has increasingly enabled people who are blind or visually impaired to challenge themselves to science, technology, engineering, and mathematics (STEM) subjects and careers, there is a lack of authoring tools available to independently produce scientific documents and materials which are inevitably necessary for a better communication in mainstream practices. While LaTeX, a plain-text-based document preparation system, has been considered an accessible full-fledged authoring and reference management tool, its steep learning curve and limited output type to PDF have made some blind people, who lack programming background and/or who would like to produce different accessible output formats, discouraged. This paper calls attention to the need of introducing an easy-to-write and accessible scientific document authoring tool by defining the scope of a scientific document, highlighting some issues of the conventional methods that the blind community has employed for document production, and suggesting an R Markdown system as a compelling solution. This research has developed and detailed the Accessible RMarkdown Online Writer (AROW) as a hands-on demonstration that proves its capability for highly accessible scientific document production that can be done by a blind individual in multiple formats including Word, RTF, PDF, MathML/MathJax enabled HTML, and presentations.

## Keywords

LaTeX, MathML, Markdown, RMarkdown, Scientific Document, Technical Writing

**Introduction**

The last few years have witnessed increased scholarly and practical discourses on how to engage people who are blind or visually impaired in science, technology, engineering, and math (STEM) subjects and careers. For example, the National Center for Blind Youth in Science (NCBYS), funded by the National Science Foundation, has actively held various annual and biennial summer camps and trainings tailored for blind children since 2004 to promote science opportunities for blind or low-vision youth in STEM fields. Behind the burgeoning attempt to adventure the new territory of STEM, which was long believed an untouchable area for blind people, lies vigorous work among researchers, developers, and diverse institutes seeking solutions through technologies that would otherwise be more challenging or partially impossible. For sure, advanced hardware such as high-resolution tactile embossers and 3D printers has offered us easier STEM material production in an efficient way than ever before (Ladner et al.; Wedler et al.), and increasingly accessible web technologies (e.g., HTML5; MathML; SVG) coupled with screen readers have continuously been allowing STEM-driven blind people to learn through a number of digital contents (Ausbrooks et al.; Rotard et al.).

Physically and digitally transforming STEM materials into a form that could be better consumed by a blind person is an important area of inquiry, and has made it possible for more blind and low-vision people to dive into the new challenge; however, remarkably few studies have been designed to enable visually impaired people as STEM content producers going beyond the consumer level. More specifically, relatively little is understood about how a blind individual can independently produce scientific and technical materials in an accessible way to communicate with others or for themselves.

This study addresses current challenges that people with visual impairments face in independently writing academic, technical, and scientific documents. The scope of "scientific documents" we define here involves as follows: (1) documents requiring Math formulas that are hard to type with plain/rich text editors; (2) documents that need to have embed image file(s) or visual graph(s); (3) documents that demand some table(s); and (4) documents that include citations and cross-references in accordance with a specific academic style (e.g., American Psychological Association 6th edition, Modern Language Association 8th edition, etc.). Having the definition and scope in our hands, in the following section we will discuss current issues of the conventional methods that the blind community use to employ for scientific document production primarily focusing on LaTeX. The rest of the paper will then be dedicated to detailing our development research of Accessible RMarkdown Online Writer (AROW) as a practical solution.

## Related Work

### *LaTeX*

LaTeX (Lamport), a document preparation system, has been widely used and recommended for blind people writing scientific documents due to the following reasons. First, the plain text-based LaTeX markup, as opposed to the formatted text, along with command-line executable environments guarantees trustful accessibility for screen reader users in principle. All the typesetting adjustments including formatting, inserting tables and figures, citations and cross-references, and writing mathematical notations (these satisfy our definition of "scientific document" above), can be fully controlled by a keyboard in a plain text; which is not quite accessible for blind people in other graphical user interface (GUI) based reference management tools (e.g., EndNote, Mendeley, and Zotero). Second, the rich applicability of LaTeX has also made it preferable among the blind community. Using a wide variety of LaTeX packages, any non-typical document forms including but not limited to presentations, diagrams, and music notations, can be expressed and produced as high-quality print materials without vision by a blind individual armed with LaTeX syntax. In sum, the strengths of both accessibility and applicability that LaTeX offers have been demonstrated by many blind or low-vision people so that LaTeX now has been positioned as one of the best must-have items among blind people working on any scientific documents.

However, there have been some trade-offs (Maneki and Jeans): (1) Getting used to LaTeX requires considerable efforts due to its steep learning curve; (2) The typical PDF output of LaTeX document is not fully accessible, in particular for math content; (3) The debugging process of compiling LaTeX file is overwhelming for those lacking programming experience; and (4) there still seems a lack of accessible TeX editors that offer screen reader users a satisfactory experience in Windows platform (Melfi et al.).

*Markdown*

While LaTeX demands a certain level of technical knowledge, there are lightweight markup languages (LML) that enable scientific writing with simpler and more unobtrusive syntax. Markdown is one of them that we would like to focus on. Markdown was originally created to allow web writers to produce HTML files using an easy-to-read and easy-to-write plain text formatting syntax. The primary design goal is readability, that Markdown-formatted document be readable as-is "without looking like it's been marked up with tags or formatting instructions" (Gruber). The early effort conversion from Markdown-formatted plain text into structurally valid XHTML (or HTML) was done by a Perl script (i.e., Markdown.pl) (Gruber); however, the recently developed tool Pandoc, an open-source document converter based upon an extended version of Markdown (MacFarlane), has replaced it with support for a wider variety of output formats including but not limited to MS Word, PDF, RTF, EPUB, and PowerPoint. The term "Markdown," hereafter, will refer to Pandoc's flavor of Markdown unless otherwise specified as it has been known to be the most active and comprehensive (Xie, JJ Allaire, et al.), among other variants such as MultiMarkdown, CommonMark, GitHub-Flavored Markdown, and Markdown Extra.

Some possible strong points that benefit from Markdown in relation to screen reader accessibility are as follows. First, Markdown is much easier to learn for a practical use than any markup languages. Table 1 shows some example comparisons between HTML, LaTeX, and Markdown. These comparisons show that Markdown can remove a considerable amount of effort for people without any programming background knowledge and allow them to produce high quality scientific documents. Second, Markdown permits LaTeX math equations and environments that can be turned into HTML/MathML with MathJax scripts enabled. Any LaTeX math expressions written between a pair of dollar ($) signs are automatically converted into accessible inline MathML when output is set to HTML. Because of this, blind people can interactively read through the Markdown-generated output using refreshable braille display device and/or screen reader that supports MathML/MathJax (e.g., NVDA 2015.2 or later; JAWS 16 later). For other output file formats, LaTeX math expressions are rendered into images with auto-transcribed alt text, which can be also detected by a screen reader. Last but not least, Markdown is highly extensible. As we mentioned above, one Markdown-formatted plain text source can be converted into multiple output formats such as HTML, DOCX, RTF, PDF, EPUB, PowerPoint Presentation and more. Additionally, users with higher levels of programming proficiencies can also extend more options for each output format. For example, people familiar with LaTeX can seamlessly utilize any TeX packages or environments within Markdown source for PDF output; LaTeX codes except for the math equations denoted with a pair of dollar signs are ignored when rendered for other formats; and HTML, CSS, and JavaScript syntax can be employed along with Markdown for HTML output. Therefore, Markdown allows blind people to produce scientific documents in the format that best meets their needs which would otherwise be only limited to PDF output in LaTeX system.

Table 1. Table of Example Comparisons Between HTML, LaTeX, and Markdown.

| HTML | LaTeX | Markdown | Result |
|---|---|---|---|
| `<h1>HEADING</h1>` | `\section{HEADING}` | `# HEADING` | **HEADING** |
| `<b>BOLD</b>` | `\textbf{BOLD}` | `**BOLD**` | **BOLD** |
| `<i>ITALIC</i>` | `\textit{ITALIC}` | `*ITALIC*` | *ITALIC* |
| - | `\textcite{BibTeXkey} have developed AROW.` | `@BibTeXkey have developed AROW.` | Seo and McCurry have developed AROW. |
| - | `AROW is useful \parencite{BibTeXkey}.` | `AROW is useful [@BibTeXkey].` | AROW is useful (Seo and McCurry). |
| <math><mrow><mi>E</mi><mo>=</mo><mi>m</mi><msup><mi>c</mi><mn>2</mn></msup></mrow></math> | `$E=mc^2$` | `$E=mc^2$` | $E = mc^2$ |

Despite the possible convenience that Markdown can offer blind people over LaTeX, a search of literature revealed that little research has explored this subject. Voegler et al. suggested Markdown as a transcription notation system for creating accessible study materials for students with print disabilities. The authors demonstrated that the utilization of Markdown could not only promote the technical quality on transcribers' end, but also improve the accessibility of the output documents on the students' end. This study is valuable in the sense that it has shed light on the power of Markdown; however, the usage of the Markdown that the research addressed was limited to sighted transcribers as the producer and blind learners as the consumer. Thus, the study has not resulted in proposing an accessible Markdown authoring tool for blind individuals.

The recent discussion on Markdown for blind people, among the very few studies, has been proposed by Jonathan Godfrey, who is blind statistician in New Zealand. Utilizing wxPython cross-platform GUI library (Dunn and Pasanen), Godfrey and James developed an accessible text editor called "WriteR" supporting RMarkdown, the combination of statistical language R and the Pandoc version of Markdown (Xie, J.J. Allaire, et al.). This in-progress work is worthy of notice since it has been the first attempt to enable people with visual impairments to be the content creator of scientific documents using RMarkdown. As of this writing, however, there are some challenges in using WriteR. First, the required local installation of the application requires basic knowledge about the R language and Python package management due to its dependencies (e.g., R, Python 2.x, wxPython, and Pandoc). Although BrailleR (Godfrey et al.), an R package for improving accessibility of Statistics for blind people, has a function that automates the local installation of Python and its required library for the WriteR editor, those who do not know how to invoke this function in R will find the installation process difficult. Second, WriteR currently lacks some functionalities for scientific document production. For example, the current version of WriteR does not provide screen reader users with improved accessibility for citations or cross-references which are essential to academic paper composition. WriteR requires users to manually find and insert each citation key from a BibTeX file for citation reference; this is opposed to other reference management tools such as EndNote, Zotero, etc. which have included their own convenient workflows for inserting citations from customized reference libraries. In other words, WriteR lacks ease of use for scientific features, and having adequate R knowledge precedes the actual benefit of WriteR application. This could discourage some blind individuals who would otherwise like to use Markdown for document creation.

**Development Procedures**

Given that any academic work demands the elements of scientific writing to some degree, the need for creating technical documents is not limited to persons pursuing STEM subjects. This has motivated our research to develop a more inclusive, accessible, and easier authoring tool for scientific document production that includes all the benefits of LaTeX. Thus, we have developed the Accessible RMarkdown Online Writer (hereafter, AROW) aiming at the following two goals. First, the authoring tool should provide blind people with a varying range of technical ability with equal accessibility for writing and rendering processes to the greatest extent possible. Second, the tool should be seamlessly extensible for advanced users to employ other markup and/or programming languages for their work while benefiting from Markdown capability. In what follows, we will detail our development work towards these ends.

*Application Overview*

In response to our two goals (i.e., accessibility and extensibility), we have developed and hosted a web-based application where anyone with a modern browser and desired assistive technologies can write and render scientific documents in multiple formats without any additional installation. We particularly chose RMarkdown (RMD) for our core Markdown system due to its added rich capabilities for supporting many programming languages (e.g., R, LaTeX, Python, HTML/CSS/JS, C/CPP, and more) on top of the basic Pandoc flavored Markdown (Xie, JJ Allaire, et al.). The use of RMarkdown, however, does not mean that more complexity is required; advanced users can take advantage of these features while basic users can simply use Markdown as discussed thus far. There exists a great on- and off-line RMD editor developed by RStudio; unfortunately, however, it is not yet accessible for screen reader users (Godfrey and James). Hence, our development work would be instrumental for those seeking alternative solutions.

We acknowledge that we have borrowed some menu layouts for inserting scientific notations (e.g., LaTeX Math expression and R code chunk) from the work of WriteR (Godfrey); and we have added more functionalities for scientific writing such as citations, cross-references, customizable YAML headers, etc. that will be further described below.

*Interface and Function*

The AROW is a simple web application that renders RMD text into various document formats. Users enter plain text written in the Pandoc Markdown syntax (when necessary, statistical computation can be further employed using advanced R blocks) into a text field, select from the available formats, and submit to let the system render their files where they are displayed for download (see Figure 1).

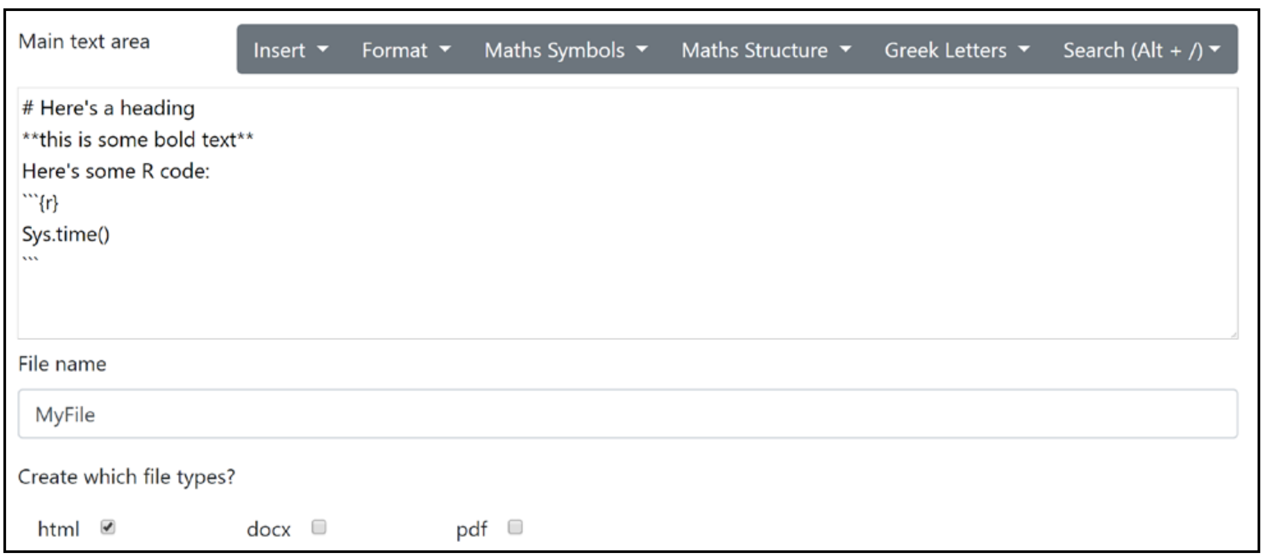



Fig. 1. The Main Interface for the Application

The main text area in which users compose their scientific documents has six assisting menu options with corresponding hotkeys: Insert; Format; Maths Symbols; Maths Structure; Greek Letters; and Search. Advanced users familiar with Markdown and LaTeX can directly type according to each syntax without going through the menu; however, beginners can benefit from these available items. Since all available options including sub-menu items are searchable using a

Google Doc inspired search menu, users can invoke any desired functionality in a timely manner without the cognitive load of remembering every single hotkey.

To allow users to create scientifically rigorous documents, advanced options such as reference management and field customization for YAML header are also provided (Figure 2). For instance, users can either upload BibTeX files or paste its entries into the AROW system, and conveniently insert cross-reference citations in their document text through the citation insert system. Furthermore, users can customize the YAML headers the document will be created with; commonly headers such as “title,” “author,” and “date.”

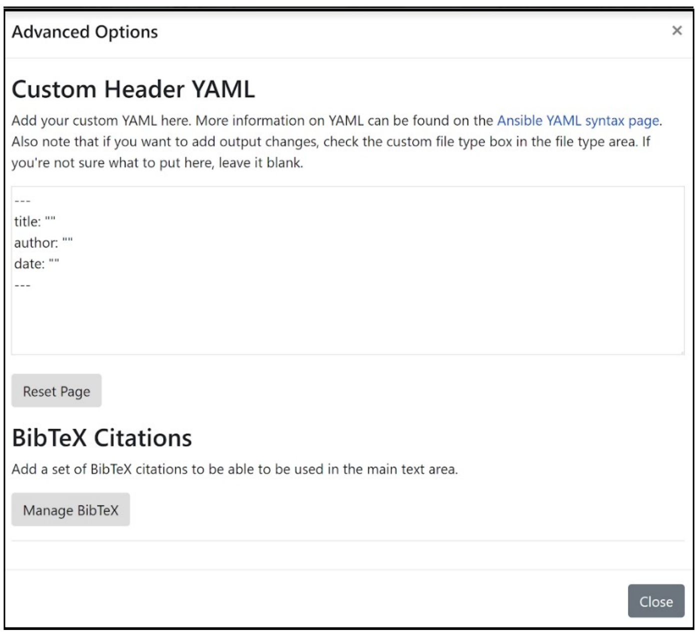



Fig. 2. An Interface for the Advanced Interface: Adding Custom YAML Headers and Citations.

*Document Conversion Methodology*

This application uses (1) a JQuery + Bootstrap frontend, (2) LAMP backend, and (3) R and Pandoc document converter for document rendering. On submission, RMarkdown text is sent from the frontend form via ajax, and is caught by PHP, which saves the text as a file and triggers the R render process. An R file is run via a PHP command line execution, and passes flags for the saved text file along with desired file types. Pandoc is called from the R process, which creates the output document files, and passes the success back through PHP to the frontend where the rendered files are displayed. Non private user data (desired file types, IP address, and timestamp) are stored in SQL for usage tracking, and to avoid collisions in the file system, users are assigned an IP based unique ID when initially loading the webpage. Figure 3 illustrates this rendering process as a diagram.

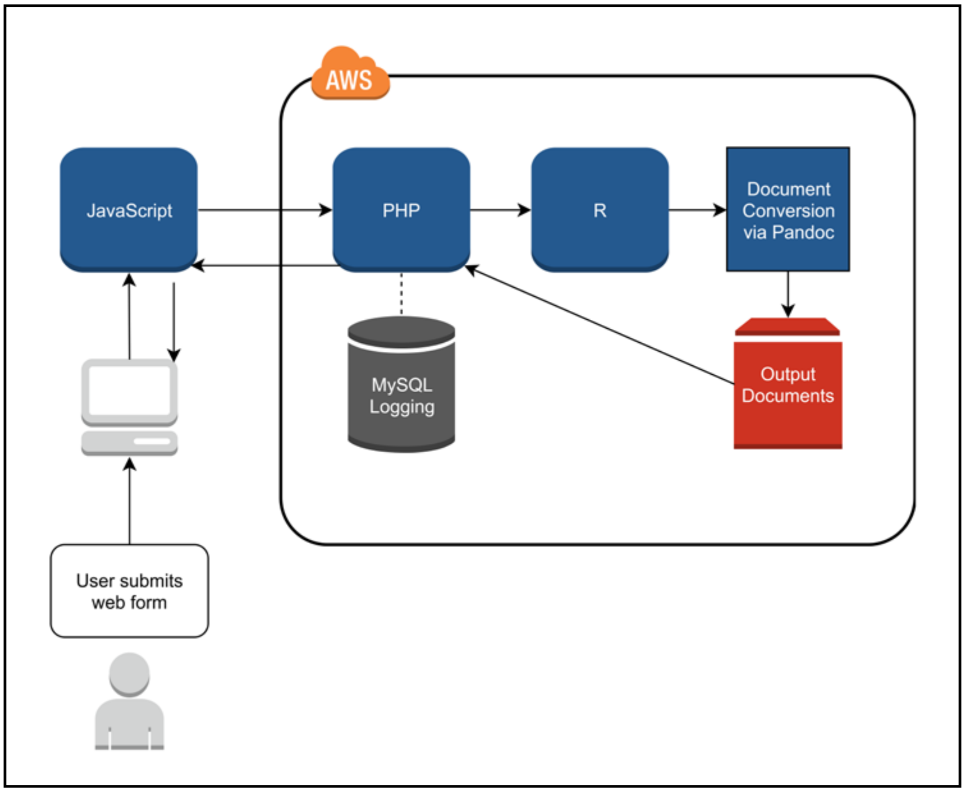
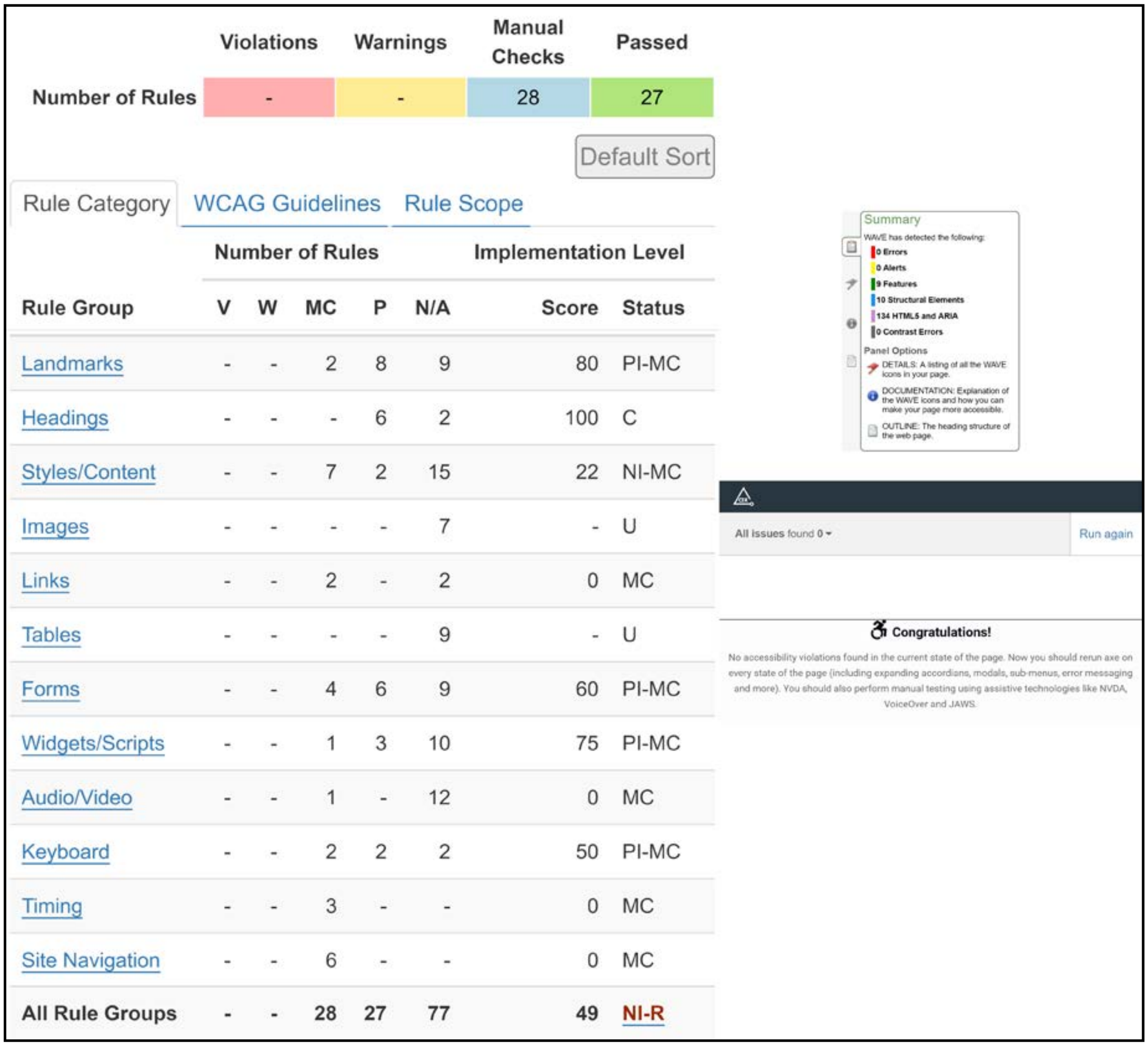



Fig. 3. A Computer Science Flowchart Describing the Process Flow of the Application.

*Accessibility Test*

To ensure accessibility and compatibility of our solution with Web Content Accessibility Guidelines 2.0 AA level (Web Accessibility Initiative), all auto-detectable HTML/CSS violations and warnings have been carefully addressed using WAVE, FAE, Axe, W3C HTML and CSS validators (see Figure 4 for test results). Manual testing was undertaken with four experienced accessibility testers (two of them were blind; the others were sighted) using different combinations of screen readers and browsers. On Windows, JAWS and NVDA with Firefox, Chrome, and Internet Explorer were tested; on Mac OSX, VoiceOver with Safari. In the lab trial, although all combinations were satisfactory, both JAWS and NVDA with Firefox were found to be optimal.



| | Violations | Warnings | Manual Checks | Passed |
|---|---|---|---|---|
| Number of Rules | - | - | 28 | 27 |

| Rule Group | V | W | MC | P | N/A | Score | Status |
|---|---|---|---|---|---|---|---|
| Landmarks | - | - | 2 | 8 | 9 | 80 | PI-MC |
| Headings | - | - | - | 6 | 2 | 100 | C |
| Styles/Content | - | - | 7 | 2 | 15 | 22 | NI-MC |
| Images | - | - | - | - | 7 | - | U |
| Links | - | - | 2 | - | 2 | 0 | MC |
| Tables | - | - | - | - | 9 | - | U |
| Forms | - | - | 4 | 6 | 9 | 60 | PI-MC |
| Widgets/Scripts | - | - | 1 | 3 | 10 | 75 | PI-MC |
| Audio/Video | - | - | 1 | - | 12 | 0 | MC |
| Keyboard | - | - | 2 | 2 | 2 | 50 | PI-MC |
| Timing | - | - | 3 | - | - | 0 | MC |
| Site Navigation | - | - | 6 | - | - | 0 | MC |
| All Rule Groups | - | - | 28 | 27 | 77 | 49 | NI-R |



Fig. 4. The Automated Compatibility Checker Results for FAE (Functional Accessibility Evaluator by The University of Illinois), WAVE (Web Accessibility Evaluation Tool by WebAIM), and aXe (by Deque).

## Conclusions

This paper thus far has discussed the current challenges of blind people writing scientific documents using LaTeX. The technically steep learning curve, lack of accessible editors, and limited output format to PDF have been pointed out as the trade-offs with the benefits of LaTeX. Highlighting the extensible simpler syntax and multiple formats of Markdown, we have introduced our development of Accessible RMarkdown Online Writer (AROW) as an alternative solution for blind people to independently produce scientific documents. Since we have recently prototyped our concepts, the accessibility and usability test has been limited to the lab trial. Further research will need to be done to investigate the general usability field test of our product with a broader population of blind people along with appropriate training for Markdown. It is our hope that this work can contribute to further discussion on how to engage blind people in writing scientific documents in inclusive, accessible, and extensible ways.

**Works Cited**

Ausbrooks, Ron, et al. *Mathematical Markup Language (Mathml) Version 3.0*. Tech. rep. World Wide Web Consortium (W3C), 2010. url: http://www.w3.org/TR/MathML3, 2010.

Dunn, Robin, and Harri Pasanen. *Welcome to wxPython!* The wxPython Team, https://wxpython.org/. Accessed 26 Nov. 2018.

Godfrey, A. Jonathan R. *WriteR: An Application for Editing R Markdown Documents That Is Accessible for Blind People Using Screen Reading Software*. 6 Feb. 2016, https://github.com/ajrgodfrey/WriteR.

Godfrey, A. Jonathan R., and M. P. Curtis James. "Simple Authoring of Statistical Analyses by and for Blind People." *Proceedings of the International Workshop on Digitization and E-Inclusion in Mathematics and Science 2016*, edited by Katsuhito Yamaguchi and Masakazu Suzuki, 2016, pp. 47–54.

Godfrey, A. Jonathan R., et al. *BrailleR: Improved Access for Blind Users*. 2018, https://CRAN.R-project.org/package=BrailleR.

Gruber, John. *Markdown*. Daring Fireball, https://daringfireball.net/projects/markdown/. Accessed 26 Nov. 2018.

Ladner, Richard E., et al. "Automating Tactile Graphics Translation." *Proceedings of the 7th International Acm Sigaccess Conference on Computers and Accessibility*, ACM, 2005, pp. 150–57, doi: 10.1145/1090785.1090814.

Lamport, Leslie. *LATEX: A Document Preparation System: User's Guide and Reference Manual*. Addison-wesley, 1994.

MacFarlane, John. *About Pandoc*. http://pandoc.org/. Accessed 26 Nov. 2018.

Maneki, Alfred P., and Alysha Jeans. "LaTeX: What Is It and Why Do We Need It?" *Braille Monitor*, edited by Gary Wunder, vol. 55, no. 7, The National Federation of the Blind, 2012, https://nfb.org/images/nfb/publications/bm/bm12/bm1207/bm120704.htm.

Melfi, Giuseppe, et al. "An Inclusive and Accessible Latex Editor." *International Conference on Computers Helping People with Special Needs*, Springer, 2018, pp. 579–82.

Rotard, Martin, et al. "Exploring Scalable Vector Graphics for Visually Impaired Users." *International Conference on Computers for Handicapped Persons*, Springer, 2004, pp. 725–30.

The National Center for Blind Youth in Science. *About the Ncbys*. The National Federation of the Blind Jernigan Institute, https://blindscience.org/about-ncbys. Accessed 26 Nov. 2018.

Voegler, Jens, et al. "Markdown–a Simple Syntax for Transcription of Accessible Study Materials." *International Conference on Computers for Handicapped Persons*, Springer, 2014, pp. 545–48.

Web Accessibility Initiative. *Web Content Accessibility Guidelines (Wcag) 2.0*. Edited by Ben Caldwell et al., World Wide Web Consortium, 11 Dec. 2008, https://www.w3.org/WAI/WCAG20/versions/guidelines/wcag20-guidelines-20081211-letter.pdf.

Wedler, Henry B., et al. "Applied Computational Chemistry for the Blind and Visually Impaired." *Journal of Chemical Education*, vol. 89, no. 11, ACS Publications, 2012, pp. 1400–04.

Xie, Yihui, et al. *R Markdown: The Definitive Guide*. CRC Press, 2018.

---. *R Markdown: The Definitive Guide*. Chapman; Hall/CRC, 2018, https://bookdown.org/yihui/rmarkdown.