

Master Thesis

StroCards: Simplified Presentation of Web Accessibility Test Results

Sophia Strobl

A thesis submitted in partial fulfillment of
the requirements for the degree of Master of Science.

Stuttgart Media University
Department of Computer Science and Media
Faculty of Print and Media

First examiner: Prof. Dr. Gottfried Zimmermann
Second examiner: Christophe Strobbe

January 13, 2021

Ehrenwörtliche Erklärung (Declaration on Honor)

Hiermit versichere ich, Sophia Strobl, ehrenwörtlich, dass ich die vorliegende Masterarbeit mit dem Titel:

„StroCards: Simplified Presentation of Web Accessibility Test Results “

selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in anderer Form als Prüfungsleistung vorgelegt worden.

Ich habe die Bedeutung der ehrenwörtlichen Versicherung und die prüfungsrechtlichen Folgen (§26 Abs. 2 Bachelor-SPO (6 Semester), § 24 Abs. 2 Bachelor-SPO (7 Semester), §23 Abs. 2 Master-SPO (3 Semester) bzw. § 19 Abs. 2 Master-SPO (4 Semester und berufsbegleitend) der HdM) einer unrichtigen oder unvollständigen ehrenwörtlichen Versicherung zur Kenntnis genommen.

Ort, Datum

Sophia Strobl

Abstract (English)

Web Accessibility is becoming increasingly important. Guidelines and according tests were created in order to ensure Web Accessibility for everyone. Detailed reports are created in order to advise content creators on this topic. However, these reports can be even more elaborate than the guidelines themselves with their very specific and technical vocabulary and their sheer length. This makes it hard, especially for non-experts, to understand what the results mean and to know where to start.

StroCards is a functional prototype developed to help viewers of Web Accessibility reports understand their contents easier. One way of doing this is by sorting and filtering identified accessibility issues. It can generate charts from the number of failed, passed and not applicable success criteria that highlight aspects that are not explained in the report itself. It can explain the user how well each of the tested website performs in terms of accessibility regarding different responsibilities. One of its key features is generating individual reports for individual responsibilities like e.g. visual design. With this functionality a designer like in this example, could receive a list of issues that are relevant to them without being overwhelmed by issues that they cannot solve. This creates a more efficient handling of the report. Besides displaying the report highlighting project roles, StroCards can have a more human-centered and empathetic approach by showing which user groups are affected and therefore excluded by accessibility issues on the website. This makes the long list of guidelines more tangible – especially for non-experts.

In the process of developing StroCards, some design decisions were made with a group of experts. The implemented functional prototype was tested in a qualitative and quantitative user study. It was perceived as easier to understand and better to work with.

A tool like this could wildly help people maintaining, creating, and developing websites to put these Web Accessibility guidelines into practice and consequently minimize exclusion of people from websites.

Keywords: Web Accessibility, Web Development, WCAG, BITV, Data Visualizations, User Tests

Abstract (German)

Die Barrierefreiheit im Web wird immer wichtiger. Es wurden Richtlinien und entsprechende Tests erstellt, um die Barrierefreiheit im Web für alle zu gewährleisten. Ausführliche Berichte können in Auftrag gegeben werden, um den Verantwortlichen der getesteten Webseite zu diesem Thema zu beraten und Anhaltspunkte zu liefern. Allerdings können diese Berichte mit ihrem sehr spezifischen und technischen Vokabular und ihrer schieren Länge viel ausführlicher sein, als die Richtlinien selbst. Das macht es vor allem Laien schwer zu verstehen, was die Ergebnisse bedeuten.

StroCards ist ein funktionaler Prototyp, der entwickelt wurde, um den Betrachtern von Barrierefreiheitsberichten zu helfen, deren Inhalt leichter zu verstehen. Eine Möglichkeit, dies zu tun, ist das Sortieren und Filtern der Prüfschritte. StroCards kann aus der Anzahl der nicht erfüllten, erfüllten, und nicht anwendbaren Erfolgskriterien Diagramme erzeugen, die Aspekte hervorheben, die im Bericht selbst nicht erklärt werden. Die Anwendung kann dem Benutzer darstellen, wie gut die getestete Website in Bezug auf die Barrierefreiheit in verschiedenen Verantwortlichkeiten abschneidet. Eine der wichtigsten Funktionen ist die Erstellung individueller Berichte für einzelne Verantwortungsbereiche wie z.B. die visuelle Gestaltung. Mit dieser Funktionalität kann beispielsweise ein Designer eine Liste von Problemen erhalten, die für seine Rolle relevant sind, ohne von Problemen überwältigt zu werden, für die er oder sie nicht verantwortlich ist. Dies schafft eine effizientere Handhabung des Berichts. Neben der Darstellung, die Projektrollen hervorhebt, verfolgt StroCards einen menschenzentrierteren und einfühlsameren Ansatz, indem gezeigt wird, welche Benutzergruppen von Barrieren auf der Website betroffen und daher ausgeschlossen sind. Dies macht die lange Liste von Richtlinien leichter zu handhaben - besonders für Laien.

Im Prozess der Entwicklung von StroCards wurden einige Designentscheidungen mit einer Gruppe von Experten getroffen. Der implementierte Prototyp wurde in einer qualitativen und quantitativen Nutzerstudie getestet. Er wurde als einfacher zu verstehen und besser zu handhaben empfunden.

Ein Werkzeug wie dieses könnte Personen, die Websites warten, erstellen und entwickeln, sehr dabei helfen die Richtlinien für Barrierefreiheit in die Praxis umzusetzen und somit den Ausschluss von Menschen von Websites zu minimieren.

Schlagwörter: Web-Barrierefreiheit, Web-Entwicklung, WCAG, BITV, Datenvisualisierungen, Benutzertests

Contents

1. INTRODUCTION	8
2. ISSUES WITH CURRENT WEB ACCESSIBILITY REPORTS	9
2.1. Web Accessibility	9
2.1.1. <i>Standardized Guidelines</i>	10
2.1.2. <i>Difficulties with Web Accessibility Testing</i>	13
2.2. Requirements for StroCards	13
3. RELATED WORK.....	14
3.1. General Approaches and Aspects	14
3.1.1. <i>The Importance of Interaction</i>	15
3.1.2. <i>Designing for the Target Audience</i>	15
3.2. Web Accessibility Monitoring Tools.....	16
3.3. WebAIR	17
3.4. Balanced Scorecards.....	17
3.4.1. <i>Characteristics and Usage</i>	17
3.4.2. <i>Using Scorecards for Web Accessibility</i>	19
3.4.3. <i>Example Scorecards and Aspects that could be adopted for StroCards</i>	19
4. DEFINING STROCARDS' CONTENT	23
4.1. Page Types	23
4.2. Dimensions	24
4.2.1. <i>User Groups/User Accessibility Needs</i>	26
4.2.2. <i>Project Roles</i>	30
4.3. Selected Content	39
5. DEFINING THE DESIGN: WIREFRAME DEVELOPMENT	40
5.1. Specific Iterations and Approaches	40
5.1.1. <i>Icons</i>	40
5.1.1. <i>Bar Charts</i>	46
5.1.2. <i>Report</i>	48
5.1.3. <i>Page Overview</i>	50
5.1.4. <i>Designs showing values over time</i>	52
5.2. Recapitulation: Designs presented to the experts.....	53

6. WIREFRAME REVIEW WITH A GROUP OF EXPERTS AND IMPLEMENTATION	57
6.1. Goals of the Review	57
6.2. Participants and Method.....	58
6.3. Results	58
6.4. Implementation.....	62
7. USER TESTING	65
7.1. Method	65
7.1.1. Setup and Data Acquisition	65
7.1.2. Procedure	66
7.1.3. Study Population	67
7.2. Results	67
7.2.1. Experience of the Participants	67
7.2.2. User Experience Questionnaire (UEQ)	69
7.2.3. Qualitative Evaluation	75
7.2.4. Limitations	92
7.3. Conclusion	94
8. FURTHER DEVELOPMENT	97
<u>APPENDIX</u>	<u>99</u>
A1. MAPPING OF SUCCESS CRITERIA: UANS	99
A2. MAPPING OF SUCCESS CRITERIA: ROLES.....	102
A4. USER TEST QUESTIONNAIRES (GERMAN ORIGINAL)	105
Fragebogen 1: Fragen zur Erfahrung	105
Fragebogen 2: zur BITV-Testdarstellung	105
Fragebogen 3: zu StroCards.....	106
A4. USER TEST QUESTIONNAIRES (ENGLISH TRANSLATION).....	106
Questionnaire 1: about prior Experience	106
Questionnaire 2: on the BITV presentation of results	107
Questionnaire 3: on StroCards	107
A5. UEQ EVALUATION BITV REPORT.....	109
BITV: Distribution of Answers per Item.....	109
BITV: Mean Value per Item	110
A6. UEQ EVALUATION STROCARDS	111
StroCards: Distribution of Answers per Item.....	111
StroCards: Mean Value per Item	113

A7. QUALITATIVE EVALUATION.....	114
Prior Experience (German).....	114
BITV report: negative comments.....	115
BITV Report: positive	117
BITV Report: wishes.....	118
Comparison: Which report was preferred?	119
Why was StroCards better?.....	119
StroCards: Preferred and probably most viewed Diagrams/Pages	121
StroCards: Wishes, Ideas and Recommendations	121
StroCards: Functions and Views to leave out and negatively perceived aspects.....	123
StroCards Report: List or Table?.....	123
<u>BIBLIOGRAPHY.....</u>	<u>125</u>
<u>LIST OF FIGURES</u>	<u>129</u>
<u>LIST OF TABLES</u>	<u>131</u>
<u>LIST OF SCREENSHOTS.....</u>	<u>132</u>
<u>ACRONYM</u>	

1. Introduction

Making websites more accessible is an important goal and is increasingly becoming legally binding. There are many people involved in developing and maintaining a website – each with a different level of an understanding for Web Accessibility. When receiving results from expert accessibility testers website creators can be overwhelmed by the long checklists with comments that are hard to understand for non-experts. Furthermore, if they are responsible for managing identified issues it is not clear how to prioritize them. It must be determined which person in the team is responsible for which accessibility issue.

A prototype for filling this gap – called StroCards – is intended to help to manage this issue: How can Web Accessibility test results be presented in a way that is easy to understand for non-experts? By moving the focus away from the number of success criteria, viewers can gain a better understanding of Web Accessibility with a more human-centered approach.

By making test results available as an input, StroCards shows the user in different views which user groups are most affected by accessibility issues in the evaluated website. Furthermore, by being able to interact with these results, users can access information and references on how to address and resolve identified issues. Learning during the process of development can ensure the improvement of tracking updates and projects long term. To facilitate and guide the team to a more accessible website, issues can be filtered by the relevant sample page, user group, or project responsibility.

This work will describe the current state and issues in Web Accessibility testing with a focus on the legally binding guidelines in Germany (2.). 3. will highlight work related to StroCards that could be helpful as a foundation in its development. Balanced scorecards, a more general way of displaying performance, will be discussed in 3.4. Chapter 4. Describes the process of defining the contents for StroCards: the different page types and filters with which success criteria can be viewed. Chapter 5. Discusses different ways of visually presenting that information. These designs were reviewed with a group of experts and their feedback was integrated into a prototype (6.), user tests were conducted (7.). Finally, chapter 8. concludes this work with a discussion of potential further developments.

2. Issues with current Web Accessibility Reports

This chapter highlights the issues that the simplified presentations of Web Accessibility test results – StroCards - aims to resolve. After discussing Web Accessibility in general and issues related to the testing procedure, requirements for StroCards are defined.

2.1. *Web Accessibility*

It may be argued that the efforts necessary to provide Web Accessibility are not important since they do not affect many people. E.g. the World Health Organization (WHO) reported that 0.3% of the people living in Europe are blind (WHO, 2010, p. 2). Additionally, 2.9% in Europe are visually impaired. These relative numbers may seem small, but they are also trending to rise in the WHO's projections. One reason for this is, for instance, age-related impairments which are increasingly prevalent in an aging population. Taking into account their absolute numbers, and that this is not the only type of disability, should be a convincing argument to create web content that is as inclusive as possible.

Beyond the act of inclusivity, a more accessible website results in a more usable website. Having enough contrast, for example, is also helpful for those without visual impairments when looking at their phone in bright sunlight: Web Accessibility is “essential for some and useful for all” as the Web Accessibility Initiative (WAI) tends to emphasize in its Web Accessibility perspectives videos (Initiative (WAI), 2020) among other things. These are meant for people new to the topic in order to convince them of the importance of accessible practices.

The internet has gained a great importance in people's everyday lives. Excluding anyone from internet access is considered to be an unethical practice. Especially during the COVID-19 pandemic many people rely on previously in-person services that have moved to the web. Ordering products online is safer, especially for people in high-risk groups, due to the lack of contact with other people. The elderly that are at high risk of serious symptoms are also more likely to have impairments that could hinder their use of online shops. Bridging that gap is a responsibility of web content creators regardless of whether it is a legal requirement or not. Nevertheless, there is an increasing number of countries that enforce and encourage more inclusivity towards people with disabilities. (Lopes et al., 2009)

2.1.1. Standardized Guidelines

There are several resources and guidelines for the development of accessible websites. Among the most prominent ones are the Web Content Accessibility Guidelines (WCAG) created by the World Wide Web Consortium (W3C). The current version is the WCAG 2.1 (W3C, 2018). It consists of 60 so-called success criteria that describe requirements that a website should fulfil in order to be accessible. An example for this is success criterion 1.1.1 Non-text Content. Each success criterion is assigned a *Level of Conformance*: A, AA or AAA. The minimum level of conformance is A. The WCAG are structured in four *principles*:

- Perceivable (i.e. making images perceivable for people who are blind with alternative text)
- Operable (i.e. making it possible for people with limited dexterity to click a target button by making it bigger)
- Understandable (i.e. avoiding overly complicated text in order not to exclude people with difficulty reading)
- Robust (i.e. developing a website that is compatible with assistive soft- and hardware that users might use)

The examples are merely given in order to illustrate these principles and are by no means complete. The European Standard EN 301 549 is compatible with the WCAG and describes further apps and services in addition to web content.

In Germany, the *Barrierefreie-Informationstechnik-Verordnung* (BITV) – roughly translated to „Accessible Information Technology Regulation“ – in its current version from May 2019, references both the WCAG 2.1 and the European Standard EN 301 549. German legislation following an EU guideline in 2016 stipulates that all websites should be accessible by September 2020. Validating this can be done with the BITV-Test that was first published in 2004. It used to include evaluations of a success criterion's conformity in five levels like, “compliant”, “partially compliant” and “not compliant”. Currently, compliance grades are no longer distinguished, and success criteria can only be passed or failed. This binary approach makes for a less differentiated result. A complete website's conformity is only assured with a compliancy of every single success criterion (BIK BITV-Test, 2021), (BIK BITV-Test, n.d.). A tool for creating a report through self-assessment is provided. However, it should be noted that this kind of report is not comparable to one created by an evaluation professional.

An example of what a BITV test report can look like was captured in Screenshot 1. This report was shortened drastically and is meant to illustrate the structure. The main headings include general

information about the report itself like, the testing standard, website address and the tester. A brief statement about the conformity of the number of compliant sample pages is given. In the body, the report gives details on each of these samples. Lastly, it lists all of the success criteria, each with an evaluation and sometimes including comments for each of the sample pages. The success criteria are divided into sections: “failed”, “passed” and “not applicable” success criteria. Each of these sections is introduced with the number of success criteria with this status. (This example report does not have a section for failed success criteria since it was created for a website that was evaluated as compliant.

This kind of report is static. It does not enable the user to view the results e.g. for one of the sample pages. If the viewer is unfamiliar with these guidelines, it might be hard for them to understand its content due to the use of technical terminology and simply the sheer amount of content may be overwhelming and potentially hard to work through. This reaction of non-experts was recorded in the user tests that were conducted in order to compare this report with the prototype of StroCards (chapter 7.). A challenge in the development of StroCards was to format this information in a way that makes it easier to work with. This kind of PDF that a project manager might receive is very detailed and high in quality. In order to fully make use of the potential, this document has to make a website more accessible. Equally, it should be more accessible to its reader.

Screenshot 1 Shortened example of a BITV report with all success criteria passed.

<https://testen.bitv-test.de/index.php?a=ti&sid=3088>

Beobachtungsstelle für gesellschaftspolitische Entwicklungen in Europa

■ Prüfung

Standard: BITV 2.0 / EN 301 549

Startadresse: [https://beobachtungsstf\[...\].gesellschaftspolitik.de/](https://beobachtungsstf[...].gesellschaftspolitik.de/)

Prüfstelle: BIK-Beratungsstelle

Prüfer/in: Brigitte Bornemann, bb@bit-informationsdesign.de

Prüfdatum: 30.04.2020

Anmerkungen zum Prüfgegenstand

PDFs werden nicht geprüft.

■ Ergebnis

3 von 3 Seiten BITV-konform

■ Geprüfte Seiten

Seite 1 (Startseite)

Titel: Beobachtungsstelle für gesellschaftspolitische Entwicklungen in Europa

URL: [https://beobachtungsstf\[...\].gesellschaftspolitik.de/](https://beobachtungsstf[...].gesellschaftspolitik.de/)

Ergebnis: BITV-konform, 0 von 50 Anforderungen nicht erfüllt

Seite 2 (Inhaltsseite)

Titel: Gleichbehandlung und Gleichstellung | Beobachtungsstelle für gesellschaftspolitische Entwicklungen in Europa

URL: [https://beobachtungsstf\[...\].gleichstellung.de/](https://beobachtungsstf[...].gleichstellung.de/)

Ergebnis: BITV-konform, 0 von 50 Anforderungen nicht erfüllt

Seite 3 (Suchergebnis)

Titel: Beobachtungsstelle für gesellschaftspolitische Entwicklungen in Europa

URL: [https://beobachtungsstf\[...\].gesellschaftspolitik.de/suche/](https://beobachtungsstf[...].gesellschaftspolitik.de/suche/)

Anmerkungen:

Seite erzeugen durch Eingabe des Suchworts "Frauen" im Suchfeld auf der Startseite.

Ergebnis: BITV-konform, 0 von 50 Anforderungen nicht erfüllt

■ Bewertung und Anmerkungen zu einzelnen Prüfschritten

60 Prüfschritte prüfen 50 Anforderungen der BITV.

Erfüllt sind 44 von 60 Prüfschritten:

- **Prüfschritt 1.1.1a - Alternativtexte für Bedienelemente**
Seite 1, 2, 3: BITV-konform
- **Prüfschritt 1.1.1b - Alternativtexte für Grafiken und Objekte**
Seite 1: BITV-konform
Seite 2, 3: nicht anwendbar
- **Prüfschritt 1.3.1a - HTML-Strukturelemente für Überschriften**
Seite 2: BITV-konform
Das Logo sollte besser nicht als Überschrift ausgezeichnet sein.
Seite 3: BITV-konform
Das Logo sollte besser nicht als Überschrift ausgezeichnet sein.
Seite 1: BITV-konform

■ ...

Nicht anwendbar sind 16 von 60 Prüfschritten:

- **Prüfschritt 1.1.1c - Leere alt-Attribute für Layoutgrafiken**
Seite 1: nicht anwendbar
Im Slider haben die Pfeile kein alt-Attribut.
Seite 2, 3: nicht anwendbar
- **Prüfschritt 1.1.1d - Alternativen für CAPTCHAs**
Seite 1, 2, 3: nicht anwendbar
- **Prüfschritt 1.2.1a - Alternativen für Audiodateien und stumme Videos**
Seite 1, 2, 3: nicht anwendbar

■ ...

2.1.2. Difficulties with Web Accessibility Testing

The moment a website has been tested for Web Accessibility not a single line of code has been changed to improve it. A crucial step is getting the results with its recommendations for conformity to the people who can implement them.

Adhering to guidelines for Web Accessibility can be confusing and overwhelming (This has also been demonstrated in the user test conducted for this project described in chapter 7.) Improving a website for accessibility can seem like an impossible task when confronted with a report that is presented as a long checklist with many discouraging errors and terms that are not necessarily understood by everyone. This makes improving Web Accessibility not very accessible in itself. As suggested by Milne et al. a list of guidelines to follow might not be enough for designers to work with (Milne et al., 2005).

This work intends to provide a way of making these reports more engaging and easier to process – especially for people who are new to the topic.

2.2. *Requirements for StroCards*

The aim of StroCards is to present an existing BITV report in a way that makes it easier to understand for non-experts. The intent is to make it easier for the viewer to navigate through the report. Existing BITV reports can oblige a viewer to scroll for a long time which is not very user friendly. StroCards aims to make it possible to filter success criteria. By enabling the sorting of results by the sample page to which they apply, a lot of redundant information can be left out of the report when doing work on a given page. Also, charts could show the status of each page. A way of visualizing the status of aspects that cannot be directly derived from the report should be possible. A non-expert might not understand the effect that a certain failed success criterion might have on a person that is blind, for example. Enriching the report with an idea of which people might be excluded from the website could make it easier to understand what effect solving the issues might have.

Needless to say, the prototype should be a good example and be accessible in itself too.

3. Related Work

This chapter provides an overview of related work. This can be used as a guide for the development of StroCards and account for its value.

3.1. *General Approaches and Aspects*

From all of our senses, vision happens to be the most dominant one. "It is estimated that up to eighty percent of our perception, learning, cognition and activities are mediated at least to some extent through vision." (Ripley & Politzer, 2010). This suggests that appealing to our sense of vision might facilitate the understanding of data. Enriching plain texts with graphics has the advantage of enhancing the effect they have on the viewer, especially through the use of colors. Furthermore, they can help overcome language barriers (Bounford & Campbell, 2001, p. 24). As Tufte puts it in *The visual display of quantitative information*: "Modern data graphics can do much more than simply substitute for small statistical tables. At their best, graphics are instruments for reasoning about quantitative information. Often the most effective way to describe, explore, and summarize a set of numbers – even a very large set – is to look at pictures of those numbers. Furthermore, of all methods for analyzing and communicating statistical information, well designed data graphics are usually the simplest and at the same time the most powerful." (Tufte, 2007, p. 12)

Bounford & Campbell contradict Tufte's opinion on graphics and explain that tables are a fitting way of presenting a bulk of information. The viewer is steered to what is important without being distracted by details that are less important. Lines that separate rows and/or columns make the data easier to read (Bounford & Campbell, 2001, p. 38). It should be considered, though, that neither define the amount of information they mean. A part of this work is to explore the best ways to display the information in Web Accessibility reports. Different approaches and aspects to consider will be explained in the following.

Even though there is an emphasis on visualization at this point in this paper, it should be noted that the visually impaired should be taken into consideration as well. Using colors with high contrast, not conveying information with color only and providing text alternatives are examples of aspects that should be respected.

3.1.1. The Importance of Interaction

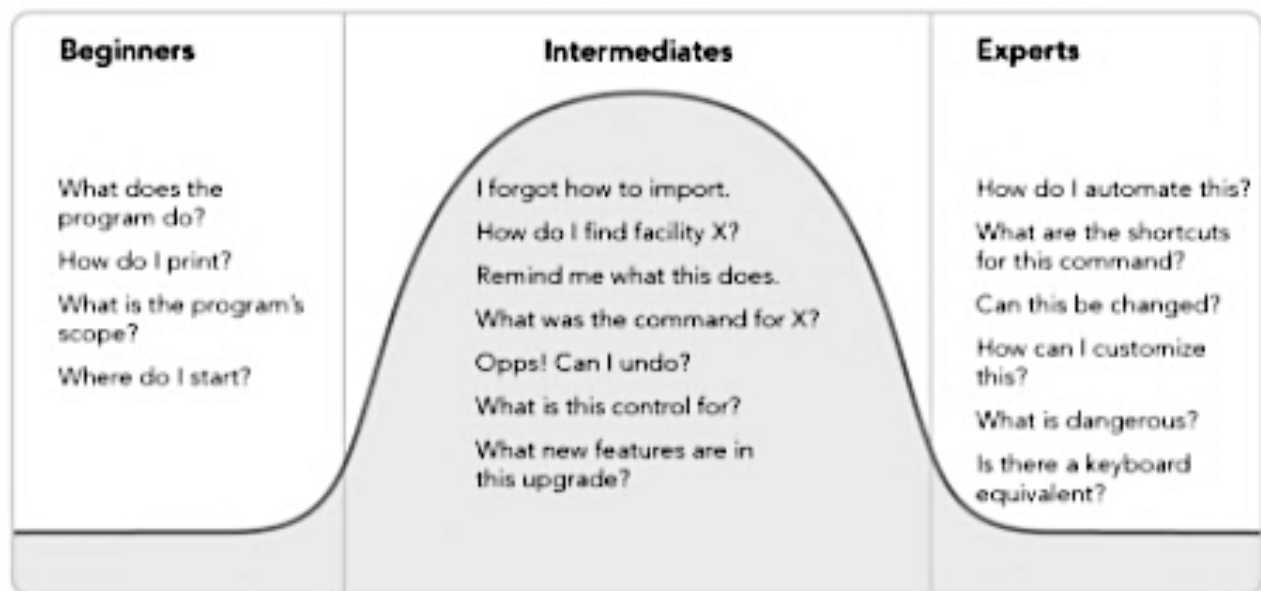
Providing interaction in the charts and infographics would use more of the potential that a website has to offer. It can support the mental model the user may have of the system and allow for more engagement with the matter.

Customizing the displayed content in the way that is more efficient for the user can make the report more valuable and easier to work with.

3.1.2. Designing for the Target Audience

One important aspect that should not be neglected during the design process is the target audience. One of StroCards' requirements is to make a Web Accessibility report as easy to understand as possible for non-experts. Cooper et al. note in *About Face 3: The Essentials of Interaction Design*: "Most users are neither beginners nor experts; instead, they are intermediates." (Cooper et al., 2012, p. 42). It is explained that people dislike how incompetent they are and therefore learn and get better to counteract this or let it be completely. According to Cooper et al., this is applicable for "almost any activity requiring knowledge or skill". Figure 1 shows the bell curve distribution of skill levels of users.

Figure 1 distribution of skill levels of users (Cooper et al., 2012, p. 42)



Whereas beginners are completely new and are trying to understand what a system does and where to start, intermediates know about functionalities and may benefit from some reminders about how to invoke them. They show interest, whereas some beginners might be too intimidated

and give up. Of course, this is not a desirable reaction and should be avoided. Methods that could prevent this may include avoiding vocabulary that is highly specific, providing an easy introduction to the topic, and designing an engaging and encouraging system without overwhelming the user. Cooper et al. point out, that a well-balanced user interface should accommodate the majority of users that are intermediates on the one hand and still provide ways of making the use of beginners and experts effective on the other. As for the knowledge of StroCards' users, it is safe to assume that they at least know what a Web Accessibility test and report are, for instance.

3.2. *Web Accessibility Monitoring Tools*

Automated Web Monitoring Tools play an important role in everyday development for web content creators. They offer immediate feedback for new web content and adapt to dynamic changes in a website. BITV reports are more static in their nature since they are not created entirely automatically and require careful testing and knowledge of an expert. They are not performed with the frequency of an automated test but have the capability of delivering evaluations of guidelines beyond the website's code. This makes their content even more elaborate and more likely to overwhelm the viewers. Design decisions have to be carefully considered and tested with users, ideally. Using existing Web Accessibility Monitoring Tools as a foundation does not only provide an approach of addressing the issue but it also leverages the advantage of familiarity. Using design patterns that are recognizable by the viewer improves the level of understanding and effectively navigating in a new application like StroCards.

In order to interconnect the report with the tested website, some monitoring tools (e.g. WAVE) work like an overlay on the website. This shows the user exactly where the source of an accessibility issue lies by highlighting the relevant HTML element directly. This improves the time efficiency of resolving the issue. StroCards cannot do this since the underlying BITV report was created manually: not every instance of an issue is recorded and finding exact elements with a human description of it is out of scope for this project.

Other facets of these monitoring tools can be applied in the development of StroCards. For instance, not only displaying errors to the viewer but also showing successful steps that have been made towards Web Accessibility could have a motivating effect. The use of scores can summarize the number of accessibility issues in one value and enable easy comparisons over time or with competitors for example.

3.3. *WebAIR*

Existing projects that address the issue of making Web Accessibility reports easier to understand for non-experts are fairly sparse.

Swallow et al. Published their evaluation of “WebAIR” ((Swallow et al., 2014), (Swallow et al., 2016)): Web Accessibility Information Resource. It is supposed to help with the problem of developers commonly being overwhelmed by all the guidelines. However, the target group is a different one from StroCards. WebAIR breaks information down to several layers of detail to avoid an information overload on a single page. This enables users to learn on-the-fly and choose for which success criteria they want to obtain more information. This nature of learning-by-doing would be a welcome side-effect of StroCards. WebAIR is structured by elements (e.g. links or forms) hoping to accommodate the developers’ workflow. This could be an interesting display option for StroCards since different people looking at evaluation results will have a different focus. Guidelines in WebAIR were stated as questions purposely avoiding language specific to Web Accessibility. This shifted the focus to a more functionality-centered approach, facilitating the work of developers with the tool. It might be worth rephrasing success criteria for responsibilities other than development as well, but this is out of scope for StroCards.

3.4. *Balanced Scorecards*

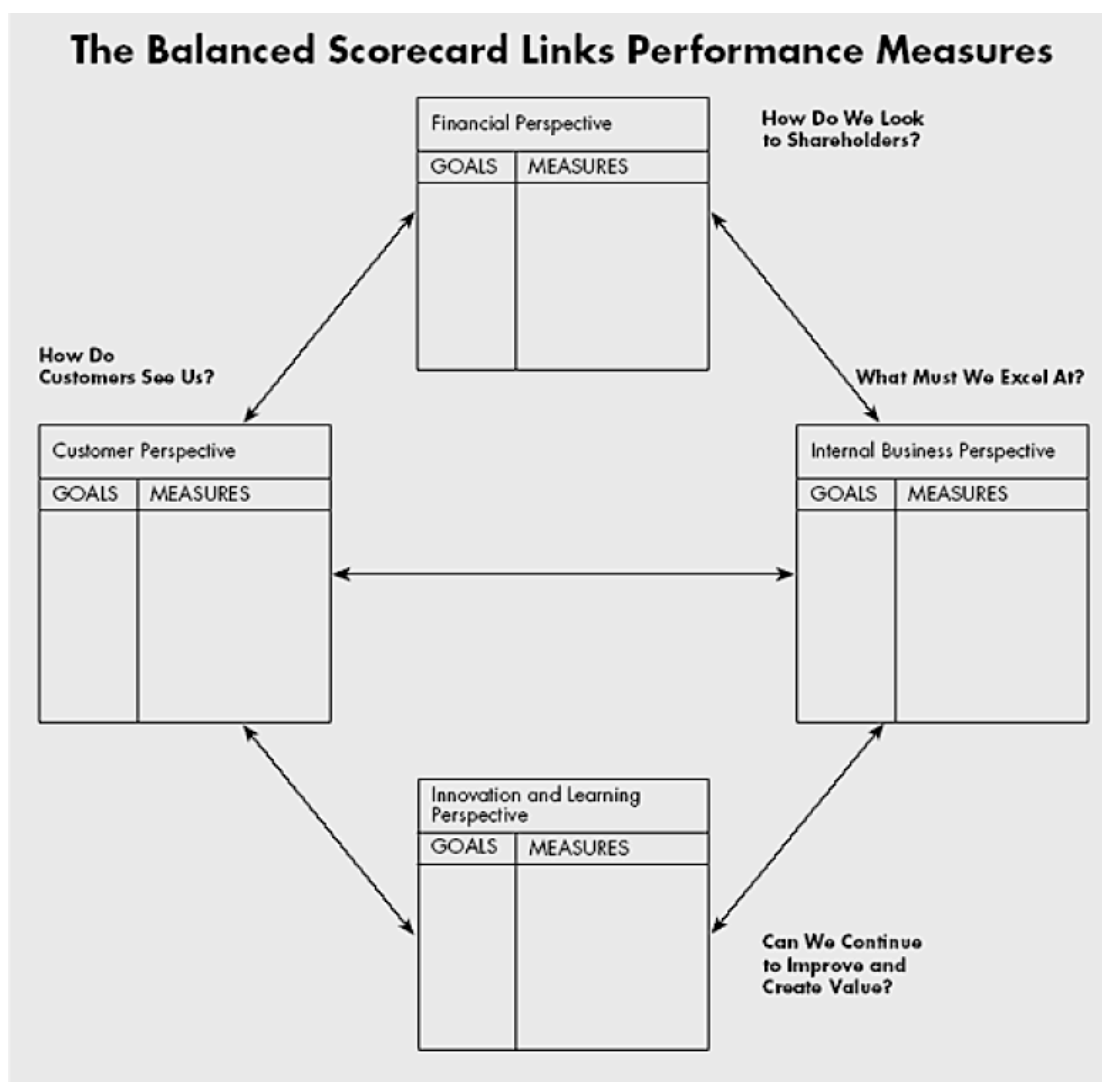
There have been several systems which extend financial progress of businesses with financial ones. One of the earliest was described by Schneiderman in 1987 (Schneiderman, 2006). However, it was Kaplan and Norton that popularized their system first in 1992 (Kaplan & Norton, 1992) followed by other articles ((Kaplan & Norton, 1993), (Kaplan & Norton, 1996b)) and an entire book (Kaplan & Norton, 1996a). The balanced Scorecard has since been modified for more specialized business purposes and can be applied more universally. One way it could be adapted is for the purpose of evaluating Web Accessibility

3.4.1. *Characteristics and Usage*

The main characteristic of the balanced Scorecard presented by Kaplan and Norton is the mix of measures that are both financial and non-financial ((Kaplan & Norton, 1992), (Kaplan & Norton, 1993), (Kaplan & Norton, 1996b)). These measures would be brought into four groups or so-called *perspectives*. First published (Kaplan & Norton, 1992), they were named “Financial”, “Customer”, “Internal Process” and “Innovation and Learning”. In 1996 the latter were renamed to “Internal

Business Process” and “Learning and Growth”. Measures and goals for each of these perspectives were chosen. The design that was first published can be seen in Figure 2. Kaplan and Norton focus their work on selecting, grouping and linking specific measures for businesses since they claimed this was an essential step for the success of the system. Balanced Scorecards could be a valuable inspiration for Web Accessibility test results and with it for StroCards not only restricted to its exact content but also looking at the design. In his book about business strategy (Ryan, 2020, Chapter 6), Ryan emphasized the advantages of using scorecards. They include the ability to see trends in performance: relative to an accepted standard or relative to other scorecards. These are desirable aspects also for someone monitoring a website’s performance in accessibility.

Figure 2 "The Balanced Scorecard Links Performance Measures" as depicted in the first article by Kaplan and Norton (Kaplan & Norton, 1992)



3.4.2. Using Scorecards for Web Accessibility

An example of a successful use of scorecards is described in a Prevos' 2015 Paper "Visualising water quality: A graphical index for drinking water system performance" (Prevos, 2015). Although this may seem far away from Web Accessibility, this paper shows some valuable take-aways for StroCards: having calculated scores made understanding the complex topic of water quality easier even for no-experts. Especially offering a graphical representation was helpful for viewers gaining an overall understanding of the matter.

Spontaneously, one could see the WCAGs four principles in the perspectives defined for the classic Balanced Scorecard. Also, the "goals" are comparable to the success criteria that are part of each of the principles. Lastly, the contents of the guideline could be seen as the "measures" since they make an evaluation fairly precise. Assessing each success criteria the way the BITV test used to work with its 5 levels of compliance (BIK BITV-Test, 2021), could have made for more quantifiable measures: quality noted as important by Kaplan and Norton (Kaplan & Norton, 1992). An issue with mapping the principles of Web Accessibility to the four perspectives is their very diverse number of success criteria assigned to each of them. Additional perspectives that were defined for StroCards are explained in chapter 4.2. Dimensions.

Similarly to what Prevos describes, scores could be calculated for Web Accessibility compliance. Examples of factors in this calculation include "how often barriers to accessibility occur in reviewed web page(s) and the likely seriousness of those barriers." (Hudson, 2011). Unfortunately, this information is either not included in a uniform way or not easily machine-readable in BITV reports. Nevertheless, they are factors that could be regarded in the future for similar score calculation. In this work, the score simply equals to the number of passed and not applicable success criteria over all success criteria assigned to a category.

3.4.3. Example Scorecards and Aspects that could be adopted for StroCards

Since the use of scorecards beyond the first published design is very diverse, this chapters shows some examples of recurring patterns focusing on the visualization and somewhat disregarding its contents.

Using traffic light colors seems like a common practice to underline values that are displayed. They can help categorize them at a glance in good and bad without having to specifically read the numbers. Besides using this way of color-coding, Figure 3 also shows a way of comparing different

Related Work


values over time by adding more columns. It also strays from the classic design (Figure 2) with its simple presentation as a table.

Figure 3 example scorecard : industry overview (<https://csipbl.com/software/scorecard-examples/industry-overview/>)

Objective	Target	Owner	Frequency For Review	January 2013	February 2013	March 2013	YTD FY 2013
Government & Public Works							
Customer Service							
Reduce average call wait time from x to y	2	Paul Campbell	Monthly	1.25	1.471	1.278	3.778
Reduce resolution time of complaints from x days to y days	3	Paul Campbell	Monthly	3	4	2	3
Improve average overall customer satisfaction rating to x%	85%	Paul Campbell	Monthly	79%	82%	90%	83.667%
People							
Increase the % of employees participating in community service projects from x% to y%	20%	Paul Campbell	Monthly	22%	13%	16%	17%
Increase employee satisfaction rating from x% to y%	85%	Paul Campbell	Monthly	72%	90%	86%	82.667%
x% of Training Verification Completed	100%	Paul Campbell	Monthly	100%	100%	100%	100%
Operations							
Reduce the # of key projects open from x% to y%	3%	Paul Campbell	Monthly	5%	3.6%	4.3%	4.3%
# of street/road signs not repaired	0	Paul Campbell	Monthly	0	0	1	1

Figure 4 shows the use of a scorecard for employee performance. It transparently shows the weighting used for each measure and provides a simple statement on the overall performance at the top of the scorecard ("18 points above target! Placement: #2nd")

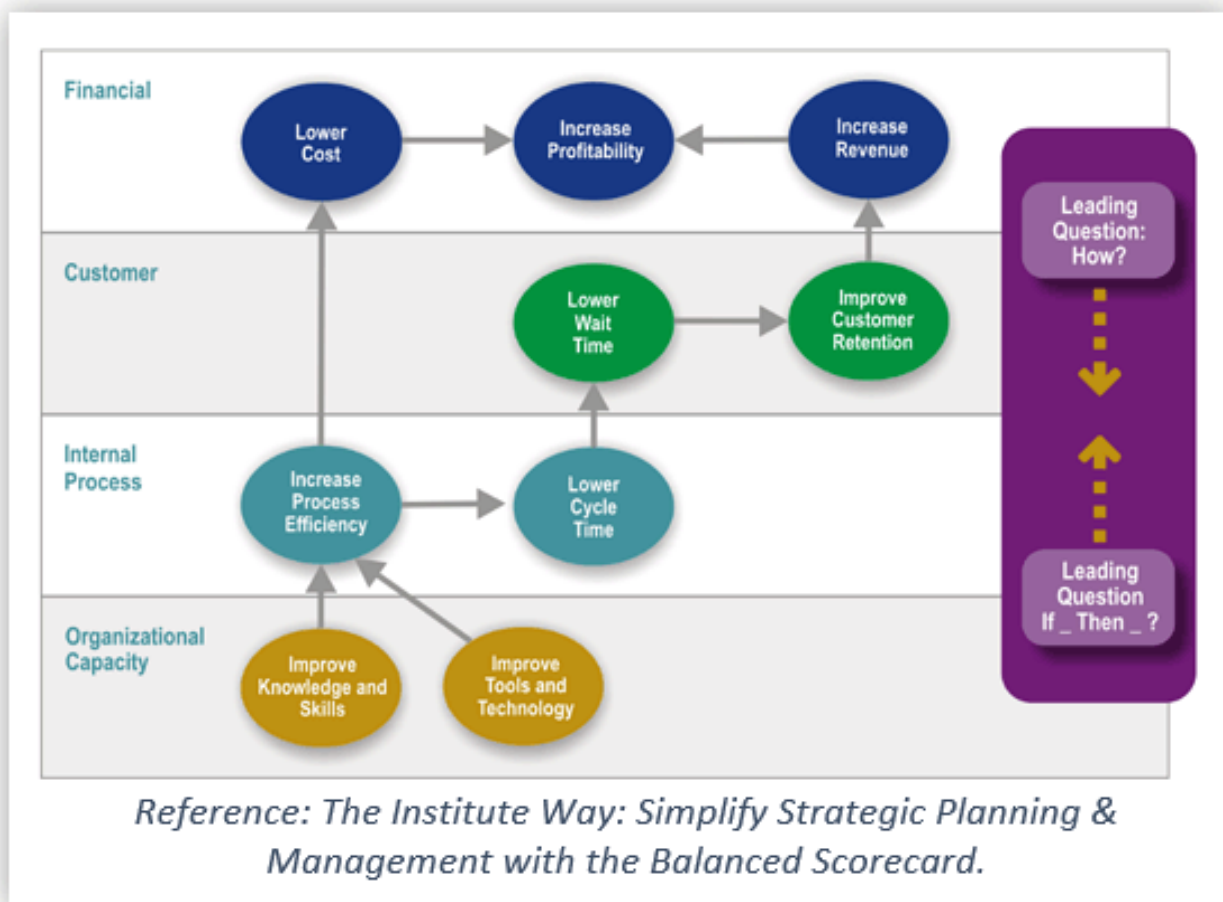
Figure 4 example scorecard for employee performance (<https://www.datapine.com/blog/kpi-scorecard-examples-templates-to-track-performance/>)

 SOPHIE MCADAMS EMPLOYEE ID: 071988 REGION: Texas (East) TIME PERIOD: 2018		GRADING SYSTEM -20+ -10 to -20 -10 to 0 0 to 10 10 to 20 20+			18 points above target! PLACEMENT: #2ND	
	ASPECT	ACTUAL	TARGET	OFF TARGET	SCORE	WEIGHTING
○ AVERAGES	CALLS PER HOUR	5	6	-1	12	30%
	ISSUES SOLVED PER DAY	88	73	+15		
	% RESPONSE TIME	79%	81%	-2		
○ CUSTOMER FOCUS	AVG. RESPONSE TIME (IN MINUTES) *30min = 1	180	120	-2	24	50%
	CUSTOMER RETENTION	89%	69%	+20		
	SATISFIED CUSTOMERS	71%	65%	+6		
○ WEEKLY CALL CYCLE (AVG)	FIRST CALL	110	115	-5	14	20%
	SECOND CALL	49	35	+14		
	THIRD CALL	30	25	+5		

Related Work

A strategy map like the one in Figure 5 can help to visualize dependencies and correlations. Applying this to Web Accessibility is possible but not with the scope of the BITV report only. A strategy like this would require looking at the bigger picture and including everyone working on the website, Web Accessibility testers, budgets, etc. Even though it is a common use of scorecards, it is barely applicable in the development of StroCards in its defined requirements.

Figure 5 example scorecard: strategy map (<https://balancedscorecard.org/bsc-basics/what-is-a-strategy-map/>)

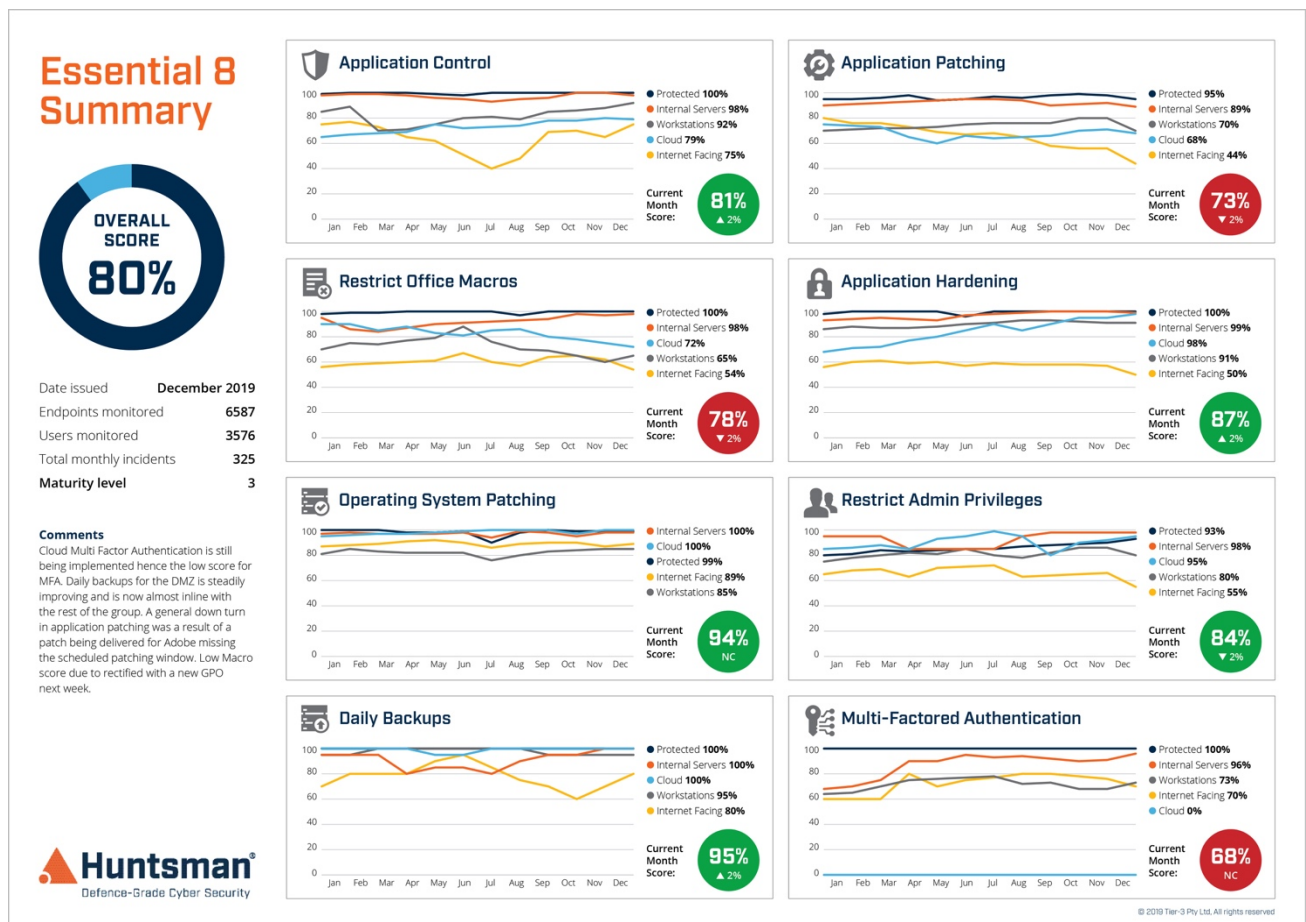


Related Work

Another field inspired by the use of scorecards is cyber security. The company SecurityScorecard (SecurityScorecard, n.d.) and Huntsman Security (Huntsman Security, n.d.) give examples for this. Figure 6 shows Huntsman's *Essential 8 Summary*. It provides an overall score and line charts and scores for each of the 8 categories.

Figure 6 Essential 8 Security Scorecard

(<https://www.huntsmansecurity.com/products/essential-8-security-scorecard/#:~:text=The%20Essential%20Scorecard%20is,mitigate%2085%25%20of%20cyber%20threats>)



4. Defining StroCards' Content

This chapter describes the development process of StroCards now that its requirements are defined (2.2.) and techniques for visualization were discussed (3.). The following presents the information that should be displayed with StroCards and how it is structured. It somewhat describes the backend of the final web prototype while 5. presents wireframes that were developed in order to visualize StroCards before actual implementation. This chapter focuses on how the success criteria can be categorized in order to highlight different aspects.

4.1. *Page Types*

The main advantage of using StroCards compared to the current BITV report is that it breaks down the long list of success criteria in order to make it easier to understand even for non-experts. At the same time, new diagrams and charts are used to add more value to the report. This should be done in a way that is less overwhelming for the user than the current report.

StroCards introduces 3 different page types to view the report:

1. Page overview
2. Dashboard
3. Report

This order is also the hierarchy that is being aimed for in development. Similar to the project WebAIR presented in chapter 3.3. , it is meant to lead the user from a general overview of each page (1.) to charts showing the numbers of passed and failed success criteria regarding different aspects (2.) to a detailed report (3.) similar to a current BITV report. The additional value added by StroCards to the detailed report (3.) is that it can be filtered to show e.g. success criteria that are relevant for the developer in the team that is creating the tested website. This can reduce the length of the report drastically and provide e.g. the developer only with issues that are in their area of responsibility. Not only can the report be filtered by aspects that will be explored in more detail in **Error! Reference source not found.** but also by page. The current BITV report is sorted by the type of success criterion (fail, pass, not applicable) and within these types the order is the same as in the guidelines. This results in mentioning different pages with each success criterion. Since the information is existent, filtering success criteria by the affected page seems like a useful functionality. Similarly, there should be an option to display the dashboard for all sample pages or for a particular one. This way, if a particular sample page consists mostly of third-party content, a

report and a dashboard can be generated that encompasses the most important information for said third party.

4.3. discusses the design of each of these pages. The content of the dashboard is more extensive and will be described in its own chapter that follows.

4.2. *Dimensions*

StroCards aims to give the user scores for different aspects of the report's content. An example for a possible dimension can be taken directly from the current report: the four principles. In its already occurring structure, BITV guidelines are divided into four sections:

1. Perceivable
2. Operable
3. Understandable
4. Robust

This categorization allows summaries for each of these principles to be made. E.g. 9 of the 17 success criteria that are relevant for how understandable the website is, are not met. Having statements like this next to each other could provide more insight for the viewer on what kind of accessibility issues users may encounter.

Another possible dimension is the conformance level, classified as A, AA and AAA. How meaningful a differentiation like this could be, was discussed with some experts. The main points of this conversation are described in 6.3. However, taking this a step further could mean creating new dimensions in addition to these two existing ones (principles and conformance levels).

Other dimensions that StroCards could cover, that cannot be interpreted alone from the list of success criteria, will be discussed in this chapter. These dimensions include:

- User Accessibility Needs (e.g. visually impaired, hard of hearing): chapter 4.2.1.
- different Roles and responsibilities in the development and maintenance of the website (e.g. visual designer, content writer): chapter 4.2.2.
- elements of the website (e.g. images, text)
- general aspects (e.g. keyboard accessibility, visual distinguishability)

Structuring the information this way, viewers of the test results can see at a glance e.g. their website is hard to use for people with motor disabilities or that their team responsible for the visual design is not paying enough attention to Web Accessibility.

For a developer viewing StroCards it might be interesting to see how many failed or passed criteria there are for a certain type of element. Swallow et al. described the workflow of web developers as

being structured by elements (Swallow et al., 2016, sec. 3.2.). Formatting the report like this might be less relevant for e.g. a public authority that has less knowledge of the programmatic structure of the website and might be more interested in how different user groups are affected by issues occurring in the website. In any case, it is helpful for the user to be able to switch between the information they are presented with and make the dashboard customizable. Depending on the situation and the focus when using StroCards, the user might not even navigate all the way through to the report page. If users have little understanding about how Web Accessibility is achieved in practice, viewing the dashboard might be enough for them just to see where the website stands in terms of accessibility.

The listed dimensions show possibilities of structuring results. Only if a correct assignments of success criteria to each category is possible does the resulting chart have value. For this reason, only two of the previously presented dimensions were chosen to be implemented in StroCards:

1. User Accessibility Needs (chapter 4.2.1.)
2. Roles (chapter 4.2.2.).

The definition of their sub-categories and success criteria assignment will be discussed in the following chapters. The other two considered dimensions were too difficult to design and could be addressed in further development.

The definition of elements of a website can be easily done e.g. by looking at available tags in HTML5 or the tags used to filter the success criteria in the WCAG Quick Reference (presented in more detail in 4.2.1.). However, their number is too overwhelming for a single chart. Furthermore, the practical value of such a chart is not clear. Having this structure in a report could be very valuable for a developer when coding each of the website components individually. Some success criteria may be assigned to the contents of a `<p>` tag for instance. Solving an issue concerning textual content is not typically a responsibility of a developer. Furthermore, many success criteria are not relevant for a single element but rather for the whole document structure. An example for this would be 2.4.3 Focus Order (W3C, 2018). Using the dimension of website elements was not implemented in StroCards. Similarly, general aspects were not regarded as a way of structuring reports either.

First of all, the interest for each of the general aspects can widely differ depending on the viewer. Secondly, the WCAG (and with it the BITV) have a structure on their own which covers one aspect after another and a mapping with it (e.g. 1.1 Text Alternatives, 1.2 Time-based Media, etc.) (W3C, 2018). This could result in a rather elaborate number of categories. Going a step further up in the hierarchy of the guidelines would reach the four principles the WCAG define: perceivable, operable,

understandable and robust. This categorization is valid and straightforward. Its implementation was considered, discussed with experts (6.) and used in the version that was tested with users.

4.2.1. User Groups/User Accessibility Needs

Displaying issues by user groups can show the user of StroCards which people are affected by accessibility issues in the website. The WAI describes the diversity of abilities. It names age-related impairments, health conditions and situational limitations to name a few (W3C Web Accessibility Initiative (WAI) et al., 2017). StroCards does not attempt to make this kind of differentiation. It rather aims to name compromised abilities regardless of the circumstances. Their list of diverse abilities and barriers serves figurative purposes and is not too be seen as complete. It does divide them into five categories however:

- Auditory
- Cognitive, learning, and neurological
- Physical
- Speech
- Visual

For each of these categories, the WCAG lists some examples of barriers that people might encounter. Since this source provides only examples and does not list specific success criteria it is rather unsuitable to use as a base for the mapping for StroCards. However, it plays a role when comparing different sources for this mapping.

Another way of categorizing disabilities can be taken from the University of Cambridge's Exclusion Calculator. This tool helps assess products and services in terms of accessibility. It calculates how many people would be excluded from a service or task based on population statistics from 1997 (University of Cambridge, n.d.) .

The categories used in the Exclusion Calculator are the following:

- Vision
- Hearing
- Thinking
- Dominant Hand
- Non-dominant Hand
- Mobility

Some differences are noticeable when comparing these to the categories listed by the WAI. The words used are not the same, but their meanings are comparable (i.e. "Hearing" corresponds to

"Auditory"). All of the WAI's categories are present except for "Speech". Furthermore "Physical" is split into 3 categories in the Exclusion calculator: "Dominant Hand", "Non-dominant Hand" and "Mobility". Regarding Web Accessibility, making this differentiation is not necessary. "Mobility" is further subdivided into "Walking", "Stair climbing" and "Standing and balancing". These are abilities not specifically mentioned by the WAI in their corresponding resource. This is also the case for differentiating between the "Dominant" and the "Non-dominant Hand": this is not mentioned by the WAI explicitly either. Regarding the categories used for StroCards, "Physical" should be sufficient enough as a category. Also, "Speech" should be included since this ability is necessary if the website relies on voice interaction or if the website owner provides nothing but a telephone number in order to contact them.

A first reference for mapping the success criteria to these categories was PersonaBrowser presented by Henka and Zimmermann in 2014 (Henka & Zimmermann, 2014). Its approach is to help developers understand the Web Accessibility guidelines by using personas that have individual stories, traits, and habits. The functionality of the PersonaBrowser includes a presentation of each of the developed personas and a list of success criteria that are relevant for them individually. The personas are assigned to one of four categories each: Visual, Auditory, Physical and Cognitive. These four categories happen to be the four categories that are used both by the WAI and the Exclusion calculator. Although this might seem like a good foundation for the mapping of success criteria needed for StroCards the nature of the PersonaBrowser makes it not suitable. Personas describe a theoretical person that could fit a target group, hence having their own individual needs and behaviors. It would be inaccurate to use this information to make general assumptions.

Furthermore, they would be incomplete, since the PersonaBrowser was developed for WCAG 2.0 and thus not offering a mapping of the success criteria that were added with WCAG 2.1. Besides, the PersonaBrowser is said to be not developed to completion.

In their "Equal Access Toolkit", IBM offers resources for their designers and developers to educate them about how to incorporate accessible design in their work. They also explain different aspects of it from a designer's point of view, more specifically user experience design, visual design and content design. These design practices are broken down further. For example, User experience design is defined by four ways how users can interact with an application. They are: navigation, keyboard, user feedback and the pointer. For each of these aspects, IBM listed requirements from the designer's perspective, dividing them into three "paces of completion" similar to the conformance levels used in the WCAG. Each of these requirements includes links to third-party resources and related topics in the IBM Equal Access Toolkit. Depending on the requirement, links to relevant items of the IBM accessibility checklist are included. This checklist is based on the

revised Section 508 (the accessibility regulations in the USA) and WCAG 2.1 standards and meets the requirements of the EN 301 549 V2.1.2 standard (*IBM Accessibility Checklist 7.1*, 2019). Even though the IBM Equal Access Toolkits provides a mapping of success criteria compatible with the BITV, the categories they are mapped to do not include user groups or user needs. Nevertheless, they do inform about different types of disabilities (IBM Accessibility, 2020) which are:

- Blind
- Low vision
- Color-blind
- Deaf / Hard of hearing
- Motor
- Cognitive

It is noteworthy that this list does not simply include "Vision" like the WAI did, but it splits it further into 3 categories: "Blind", "Low vision" and "Color-blind". This partition seems like a reasonable one. The way a person that is blind interacts with a website can be very different from how a person with low vision does. By using a screen reader, for example, a different set of guidelines apply, whereas guidelines for the use of color are less important for a blind person and all the more relevant for someone with limited vision.

As for the IBM "Equal Access Toolkit", it offers some guidance for StroCards regarding the categorization of disabilities, but it is not helpful when it comes to applying success criteria to them. It can be regarded as a resource for the dimensions of Roles which will be discussed in 4.2.2.

T-Systems Multimedia Solutions, a company belonging to the Deutsche Telekom AG, includes Web Accessibility checks and reports in their services. They provide statements (T-Systems Multimedia Solutions GmbH, 2019) about the tested website's accessibility for each of the user groups:

- Visually impaired users („sehbeeinträchtigte Benutzer“)
- Blind users ("blinde Benutzer")
- Impaired motor function users ("motorisch beeinträchtigte Benutzer")
- Hearing-impaired users ("hörgeschädigte Benutzer")
- Cognitively Impaired users ("kognitiv beeinträchtigte Benutzer")

In addition to the general result concerning BITV-conformity, there is a description of accessibility for each of these groups. These results are very transparent since they explain which success criteria are included to make that statement for a user group. Besides mapping the success criteria to each of these groups, they provide a rating for each of them describing their severity on a range from 1 to 3. This mapping from T-Systems is a reliable source of information regarding StroCards. The rating is an interesting way of providing a way of prioritizing issues and could be integrated in

score calculation in StroCards. This was also mentioned in the conversation conducted with experts which is evaluated in chapter 6.3.

The groups used in the reports by T-Systems are based on the User Accessibility Needs (UANs) described in the EN 301 549 EU standard (The British Standards Institution, 2020, Chapter 4). These are so-called "performance statements". Similar to the definition of the WAI it is also explicitly mentioned that these can be "permanent, temporary or situational". Table 1 shows these performance statements and where they are present in the grouping specified by T-Systems.

Table 1 UANs listed in EN 301 549 compared to the user groups used by T-Systems

EN 301 549	T-Systems
Usage with limited vision	Visually impaired users
Usage without perception of color	
Usage without vision	Blind users
Usage with limited manipulation or strength	Impaired motor function users
Usage with limited reach	
Usage without hearing	Hearing-impaired users
Usage with limited hearing	
Usage with no or limited vocal capability	
Minimize photosensitive seizure triggers	Cognitively Impaired users
Usage with limited cognition, language or learning	
(Privacy)	

T-Systems differentiates between visually impaired and blind but does not have a separate group for color blindness like the IBM Equal Access Toolkit does. Motor impairments are split into two groups in the EU standard: "limited manipulation or strength" and "limited reach". Making this differentiation is rather irrelevant regarding Web Accessibility. "Limited reach" can be applied to "ICT products [that] are free-standing or installed". Concerning hearing related disabilities, T-Systems includes no or limited vocal capability in this group. Similarly, they include photosensitive seizures in the group that they call cognitively impaired. The last performance statement in the EN

301 549 is Privacy. Although it is an important topic, it is not applicable in this case of defining user groups to display in StroCards.

The groups stated in the EN 301 549 can be used for the Dashboard in StroCards. As for the mapping of success criteria, T-Systems offers a reliable foundation. Furthermore, Fischer presented a related mapping in his presentation about simplified monitoring (Fischer, 2020). Besides assigning them to User Accessibility Needs, he highlighted to how many of them they were assigned. This way, conclusions about their importance could be made based on how many different types of disabilities they exclude. A goal of this presentation was to enable reports to make statements with practical relevance and to register the most common barriers. Assignments were made based on expert knowledge and estimation. This information together with what T-Systems uses was merged and became the main reference when developing the assignment of success criteria for StroCards. Decisions on which of the sub-groups were meant in the assignments by T-Systems could be made based on Fischer's presentation since he addressed all 9 of the groups listed in the EN 301 549. There were some success criteria were only one of the sources mapped them to a user accessibility need. Even then, the assignment was adopted for the prototype. The results of this can be viewed in detail in the appendix.

For the review a version of the resulting chart was created which combined some of the UANs in order to create a shorter chart. This was done with the groups defined by T-Systems. Regarding the group of users with hearing related limitations, the UANs they combine (without and with limited hearing) are mapped to the same success criteria. Combining other groups should be done with caution in order to prevent over-simplifying matters. This is discussed in chapter 6.3.

4.2.2. Project Roles

Creating the dimension of Roles or responsibilities in StroCards dashboard could be very helpful. If the viewer of the chart manages the team(s) responsible for the tested website, they can easily see how many failed success criteria are caused i.e. by the writers or by visual designers. Besides presenting the numbers of failed/passed/not applicable success criteria, StroCards can also provide the details of each of these success criteria. This way, role-specific reports can be created. For instance, a graphic designer could receive a different part of the BITV report than the writer containing information that is relevant for them. Which Roles or responsibilities this functionality should include and which success criteria could be assigned to each of them will be discussed as follows.

Defining StroCards' Content

The WAI-Engage community created an article (WAI-Engage: Web Accessibility Community Group, 2012) mapping each success criterion from WCAG 2.0 to 10 responsibilities. These Roles are listed in

Table 2.

This mapping is very detailed but only a large-scale project might have individuals responsible for each of these aspects. Regarding the resulting chart, the number of Roles could make it hard to read due to the elaborate length. Nevertheless, this article includes a mapping of each success criterion which could make it a quite valuable resource for StroCards. The number of used Roles can be reduced by combining them. An approach for this is proposed on the website A11YCLE created by Joshua Randall which supports the Cleveland Accessibility Meetup group (*Roles*, 2018).

Table 2 shows how each of the responsibilities by WAI-Engage map to the Roles listed on A11YCLE. The responsibilities they use for this mapping are claimed to be adopted by the International Association of Accessibility Professionals (IAAP). They concur with the WAI-Engage responsibilities except for an additional one named "Content Authoring". Therefore, an existing mapping of success criteria for this responsibility and the resulting role "Marketing" cannot be assumed. Similarly, the responsibility "Search Engine Optimization" listed by WAI-Engage is not included in any of the Roles by A11YCLE, suggesting a possibly incomplete resource.

A11YCLE links to an interactive tool created by the Design Agency Viget which allows the user to filter the WCAG 2.0 by the Roles Content, Design, General Development, Front-End Development, and User Experience but it is no longer maintained since November 2018 (*Vigetlabs/Interactive-Wcag*, 2020). Nevertheless, this current tool merely links to the WCAG Quick Reference (W3C, n.d.).

Table 2 responsibilities defined by WAI-Engage and how A11YCLE grouped them further

WAI-Engage	A11YCLE						
	Business Analyst	User Experience	Designer	Marketing	Front-End developer	Back-End Developer	Quality Control
Analysis	X	X					
Architecture	X						
Interaction Design / Usability	X	X					
Graphics Design			X				
Content Strategy	X	X					
Search Engine Optimization							
HTML/CSS Prototyping		X (“to lesser extent”)			X		
Front-End Development					X		
Back-End Development						X	
Quality control							X
				Content Authoring			

The WCAG Quick Reference (W3C, n.d.) was updated to meet version 2.1. Besides displaying the guidelines to the user with its structure and collapsible content like techniques and failures for each success criterion, it allows to filter the guidelines. Filters include tags, levels of conformance,

techniques and technologies. Tags are terms like "animation" or "carousels". These tags can be filtered with four checkboxes:

- Developing
- Interaction Design
- Content Creation
- Visual Design.

Screenshot 2 shows what these tags and checkboxes look like when all four are selected. These categories happen to be project Roles that could be adopted for the StroCards dashboard and by activating all tags for each of these categories, a mapping of success criteria can be extracted. Furthermore, these four categories are far less elaborate than the ones listed so far which could make for a simple chart that comes to the point. However, looking at the success criteria mapped to "Content Creation" especially, the very broad definition used for this role is noticeable. Success criteria that appear include e.g. 1.4.2 Audio Control and 2.2.2. Pause, Stop, Hide. These guidelines for the existence and functionality seem rather irrelevant for a person creating texts, videos and other content for a website. Having at least one role referring to responsibilities that have nothing to do with the code of the website could be an important distinction.

Screenshot 2 Filter options in the WCAG Quick Reference: the selectable tags change with the settings of the checkboxes

Tags [x Clear tags](#)

<input checked="" type="checkbox"/> Developing	only
<input checked="" type="checkbox"/> Interaction Design	only
<input checked="" type="checkbox"/> Content Creation	only
<input checked="" type="checkbox"/> Visual Design	only

animation	audio	
auto complete	autoplay	
blinking	buttons	captcha
captions	carousels	
changing content	color	
components		
consistent experience	content	
contrast	controls	errors
events	flashing	focus

A distinct role with responsibilities that have nothing to do with the code – an editor or content creator – is also a profile used by AnySurfer (*Les Directives AnySurfer - AnySurfer*, 2020). AnySurfer is a project that offers resources, assistance, trainings and awareness campaigns for Web Accessibility founded in Belgium. In July 2020 their website provided the functionality of filtering their own accessibility checklist (which they no longer use) by relevance for the following three Roles:

- Rédacteur (editor, writer)
- Développeur – Intégrateur (developer – integrator)
- Graphist[e] (graphic designer)

This categorization is supported by the WAI. They provide tips for getting started in

- Writing
- Designing
- Developing

for Web Accessibility (W3C Web Accessibility Initiative (WAI) & Lawton Henry, 2019). Each of the subpages with tips include references to the relevant success criteria.

Roles that they address (W3C Web Accessibility Initiative (WAI), 2020) can be even more detailed and include the following:

- Content Writers
- Designers
- Developers
- Evaluators, Testers
- Managers
- Policy Makers
- Trainers, Educators
- Web Users, People with Disabilities, Advocates

Not all resources for each of these Roles contain specific WCAG guidelines which makes them irrelevant to StroCards. The first three fulfill this criterion.

The IBM Equal Access Toolkit (introduced in chapter 4.2.1.) give some more detail on the success criteria relevant for Designers and Developers. The groups they used were less relevant for the different disabilities that users might have but could be more of importance regarding these Roles. During the work on this project the proposed structure by IBM changed slightly but the current one is as follows:

- **User experience design:**
 - Navigation
 - Keyboard
 - User feedback
 - Pointer
- **Visual design**
 - Color and contrast
 - Text style and layout
 - Transformation
 - Flashing and motion
- **Content design**
 - Text equivalents
 - User input
 - User feedback
 - Text meaning

These "Aspects of Design" used to be in a slightly different order and with slightly different titles. "Feedback" became "User feedback" for user Experience Design and in the aspect of Content design, "Text updates" became "user feedback" this might show a current shift of focus of IBM towards a more user-centered approach. Nevertheless, IBM defines 3 categories of design that could be used for the Roles diagram in StroCards: User Experience designer, Visual designer and Content Designer. As previously described in 4.2.1. each of these listed points contain references to the IBM accessibility checklist and as a consequence this also references the WCAG and the BITV. The mapping of the success criteria mentioned by IBM could be adopted. Furthermore, these categories are related to the filter option available in the WCAG Quick Reference with slightly different wording: Interaction Design, Visual Design and Content Creation. This somewhat solidifies the use of these Roles. Also the last one, development, is picked up by IBM.

The categories with their respective sub-categories are as follows:

- **Core considerations**
 - Reducing effort
 - Unit testing
- **Coding interactions**
 - Keyboard interactions
 - Dynamic content
- **Coding info and relationships**
 - Non-text content
 - Text meaning
 - User input

This covers a number of resources that could be used as a reference to create this chart by project Roles in StroCards. The defined Roles are:

1. User Experience designer
2. Visual designer
3. Content designer
4. Developer
5. Writer

The first four have an explicit mapping of success criteria in the WCAG Quick Reference and the IBM Equal Access Toolkit. The fifth – Writer – is chosen in order to include a role for responsibilities not related to the code of the website. A list of relevant success criteria can be read in the tips given for writers by the WAI (Initiative (WAI) et al., 2020). A detailed table containing all assignments can be found in the appendix.

In the prototype and consequently in user tests only the assignments from the WCAG Quick Reference and the tips by the WAI were used. An issue that was further discussed reviewing the wireframes with a group of experts in chapter 6. is that the WAG Quick Reference leaves out only a small number of success criteria for each of the Roles. This means, that the number of passed/failed/not applicable number of success criteria do not differ much between Roles. The assignments could be refined further by taking the IBM Equal Access Toolkit into account. On the other hand, this might lead to role-specific reports that are not very universal. By including many success criteria as the WCAG Quick Reference does, e.g. a content creator that might be also responsible for controls will have guidelines 4.2 Audio Control and 2.2.2. Pause, Stop, Hide listed in their report. At the same time, the developer will have it too. Finding a meaningful consideration of this redundancy while ensuring that relevant success criteria are listed for the right Roles is not

trivial. This problem could be eliminated by providing customization options for users of StroCards in order to define the real Roles that they have in their team. It is possible to use the same tags like in the WCAG Quick Reference. This way, the user is offered a level of abstraction, eliminating the need of them looking through each and every success criterion and deciding which of the Roles they apply to. Due to time constraints this was not implemented but could be considered in future development.

4.3. *Selected Content*

Besides displaying the original report, StroCards allows to filter the listed issues by sub-categories of the two dimensions: User Accessibility Needs and Roles.

The content of the StroCards dashboard that was discussed with the group of experts in chapter 6. includes charts displaying the number of passed/failed/not applicable success criteria for each of these sub-categories. Similar charts can be implemented for the four principles and for the three levels of conformance. In order to make values more comparable, StroCards will display scores for each of these values. They are the percentage of passed and not applicable success criteria of all relevant success criteria. A separate chart simply displaying these scores without the more detailed information of how many success criteria are passed/failed/not applicable can be displayed as well. Additional summaries of these reports could be created that simply show the sub-category with the maximum or minimum score. The dimension of User Accessibility Needs specifically will have an alternative version with combined groups in order to provide a shorter and less overwhelming chart.

The wireframe review with experts resulted in some changes of the wireframes concerning their content. The changes are explained in more detail in chapter 6.3. There was no agreement on including a chart displaying the success criteria by level of conformance. The chart with success criteria by principle, the chart with the scores only, and the minimum/maximum summaries were deemed rather unnecessary. The use and structure of dimensions was regarded as fitting. It was recommended to provide a chart with the numbers of all success criteria and to make the differentiation very clear between success criteria that are passed and those that are not applicable. Furthermore, it was suggested to provide an overview of the status of each page of the website that was tested. This final structure was already described in chapter 4.1.

5. Defining the Design: Wireframe development

The design of StroCards will be described in this chapter considering the decisions made in the previous one,

5.1. *Specific Iterations and Approaches*

Chapter 3.1. explained some rather general aspects of information design and user interface design. This chapter intends to apply them to design decisions made for specific aspects of StroCards.

5.1.1. Icons

Unambiguous icons can wildly improve the clarity of a website and displayed data and are thus a great addition for StroCards' dashboard. Instead of needing to read text, the user can rely on icons that have the potential to convey the same information. This chapter describes the process of icon design for each of the UANs and Roles, since they were a focus of this project. Icons for the two dimensions principles and conformance levels were not created since they were evaluated as not essential in the expert interview (chapter 6.3.).

5.1.1.1. User Accessibility Needs

The chart for success criteria by user accessibility need require icons for the following meanings:

- Usage without vision
- Usage with limited vision
- Usage without perception of color
- Usage without hearing
- Usage with limited hearing
- Usage with limited manipulation or strength
- Usage with no or limited vocal capability
- Minimize photosensitive seizure triggers
- Usage with limited cognition, language or learning

Besides using this categorization an aim is to combine some of them in order to make the chart less elaborate. The icons needed for the combined groups are the following:

- Usage with limitations related to vision
- Usage with limitations related to hearing
- Usage with limited manipulation or strength
- Minimize photosensitive seizure triggers
- Usage with limited cognition, language or learning

The EU standard listing these User Accessibility Needs does not provide icons for each, but other resources that were described in more detail in chapter 4.2.1. do. Used icons are compared in Table 3. The icons proposed by the W3C Web Accessibility Initiative are still in development as of September 2020 (W3C, 2020). What stands out with the icons by T-Systems are the strikethroughs of the eye, hand and ear. This could be problematic due to its negative connotation.





























The icon set used in the Persona Browser is inconsistent in that sense since it uses this symbolism of the strike only for the meaning of showing limits of physical strength. It is also the only resource to display the whole arm for this meaning and not simply just the hand. Furthermore, compared to T-Systems, it uses a rectangular border around the icon. This could help the icon be more distinguishable from the page layout it is used in. It also uses four different colors for each icon. This design concept is followed even further in the PersonaBrowser itself, using these colors in other layout elements. The University of Cambridge Exclusion Calculator (University of Cambridge, n.d.) uses very similar ideas in their icons. Since it was developed for services in general and not only websites, it provides separate icons for the dominant, the non-dominant hand and mobility. This should not be necessary which is why simply one hand is sufficient. The pointing hand with a button used by T-Systems is more specific than a spread-out hand only. This might be the intention the WAI had with their icon for the same purpose. They also refrain from crossing out the main element and provide an extra icon for speech impairment.

Having the icon on a dark round background could make it stand out even more in a web layout since it provides a strong contrast in color to a potentially white background and a contrast in shape with the circle in a typically rectangular web-aesthetic.

The icons for cognitive abilities differ the most in their content. T-Systems use a speech bubble in a stylized head whereas the PersonaBrowser and the WAI use a structure comparable to a speech bubble but also the brain itself that is located at precisely that orientation. The Exclusion Calculator on the other hand simply uses a thought bubble. For StroCards, the idea of the head was adopted, but since this icon should represent more processes in the brain than cognition, a gear seemed fitting to symbolize all of them and not simply e.g. speech.

StroCards' icons hold on to the general ideas of the other icon sets and fill out the gap of color perception and photosensitive seizure triggers. Since the latter is mapped to the success criteria describing flashes on the website amongst others, this symbol was used. Perception of color is symbolized by an abstraction of a graphic used to explain the mixing of three colors. StroCards refrains from using borders or background with another shape since their use in a chart is structured enough. Also, they are all filled with the same color: black. Using several different ones like the Persona Browser or the Exclusion Calculator could lead to more low-contrast issues with the background more easily. It is assumed that the icons and the text displayed next to them provide enough ways of differentiation.

Table 3 comparison of icons used in different resources for the presented User Accessibility Needs and the icons developed for StroCards

User Accessibility Needs	Icon resources				StroCards
	T-Systems	Persona Browser	Exclusion calculator	W3C WAI	
without vision					
limited vision perception of color					
manipulation or strength			  		
Without hearing					
Limited hearing					
cognition, language or learning					
speech					
photosensitive seizure triggers					

The issue of making different icons e.g. for no vision and limited vision is still existent. T-Systems solved this by altering the number of strikethroughs. Figure 7 shows different approaches to achieve this without the problematic crossing through. They were further discussed with the group of experts (6.). Possibilities of visualizing this difference include relying on the outlines only for one version, or modifying the icon by making it slightly transparent, using an overlay with a gradient or even blurring it. These solutions except for the one using the outlines only were not accepted by the experts since differentiating them requires good eyesight which makes them inaccessible. The only other icon that needs this kind of modification is the ear. Using the outlines only makes for many lines next to each other due to the nature of the icon. It was not developed for user tests due to time constraints. Only one version of each icon was used. Not having this visual differentiation was not a mentioned point during user testing (chapter 7.).






Figure 7 possible variations of the exemplary vision icon to differentiate between no vision and limited vision



5.1.1.2. Roles

The Roles described in chapter 4.2.2. and the icons designed for StroCards are shown in Table 4. The resources that provided these categories do not provide adoptable icon sets so they were created individually. Development is symbolized by HTML-typical brackets in a browser window and visual design is represented by an artistic paintbrush. Using an eye for this icon was considered but this could lead to confusion with the icon for visual impairments. Interaction Design is not visualized with a browser window like the others but rather by a pointing hand tapping. This icon is comparable to the one used by T-Systems for motor disabilities. Making a visual difference between the icons for Content Creation and Writing is not trivial. Both were represented with a pencil.

Table 4 icon for each role

Role	StroCards icon
Developing	
Interaction Design	
Content Creation	
Visual Design	
Writer	

5.1.1. Bar Charts

The development of the design of the charts used for the dashboard will be described in the following. Their purpose is to visualize the number of passed/failed/not applicable for each category. A requirement for them is to be as accessible as possible. The colors used – red and green are widely used to symbolize good and bad. At the same time, they are impossible to tell apart for people with red-green color blindness. This is why the accessibility guidelines describe using color-independent patterns or text as well.

Figure 8 shows the first generated design for a chart displaying success criteria for each of the four principles. Charts were partially generated with the Python library *pygal*. The idea was to show the proportions of those principles at the same time which is why everything is presented in a single chart. The outer ring shows the proportions of success criteria within the principles. This chart was presented to the experts and did not convince them. Results are explained in 6.3. Besides the reasons the experts mentioned, this chart is unsuitable for the other dimensions like User Accessibility Needs and Roles. The problem here is that one success criterion can be mapped to several categories devaluating the pie analogy.

Figure 8 first pie chart by the four principles created with pygal

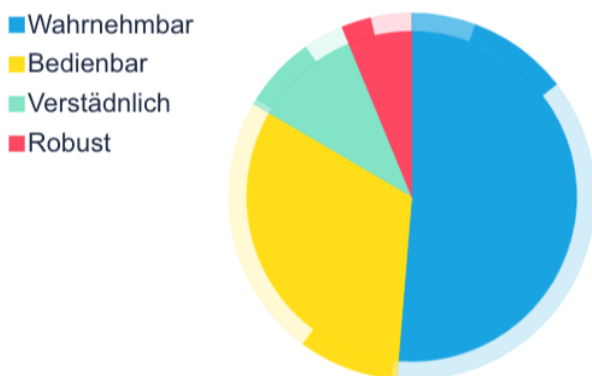
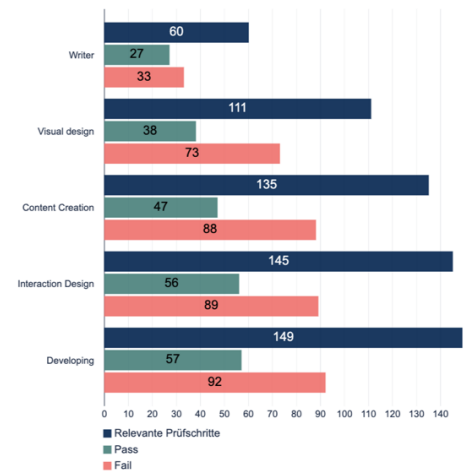


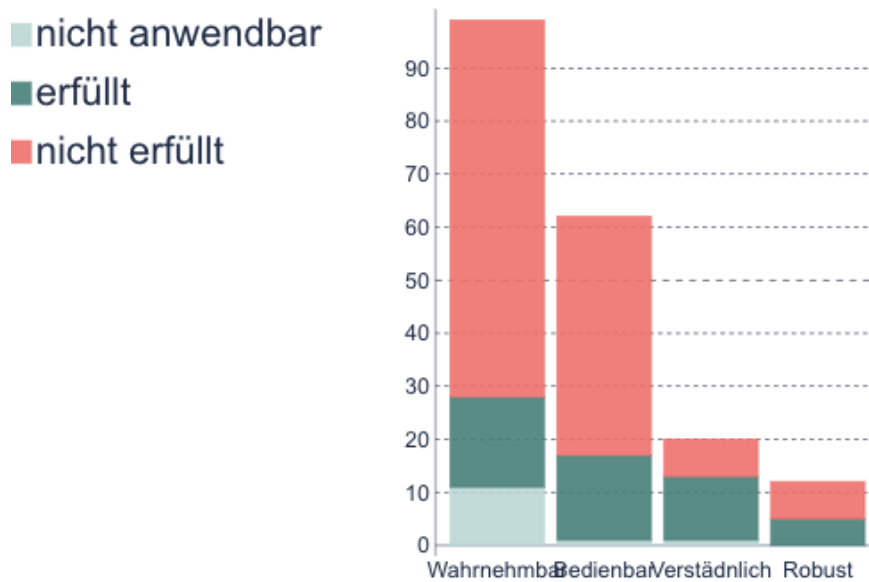
Figure 9 bar chart created with pygal showing success criteria by role



Further possibilities with the *pygal* library were explored resulting in Figure 9. Colors were modified to transmit the meaning of passed/failed/not applicable better. The bar chart has three bars per category. One is dedicated to showing the number of success criteria relevant for that category – the maximum number of passed or failed criteria. In this iteration, passed and not applicable success criteria were put together.

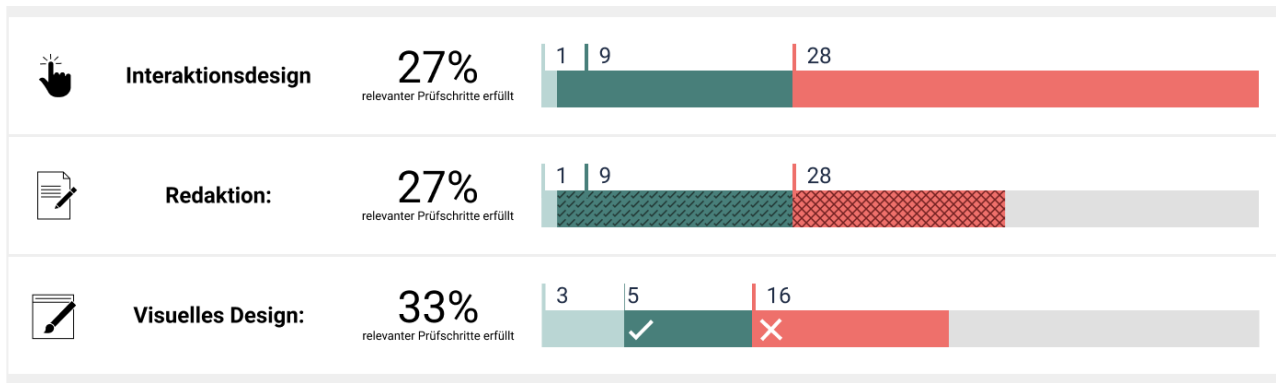
A way not to show a separate bar for the sum of success criteria like Figure 9 does, is to use a stacked bar chart. This was implemented in Figure 10. A lighter green was chosen for the success criteria that are not applicable. This symbolizes their relations to the passed ones.

Figure 10 a stacked bar chart created with pygal



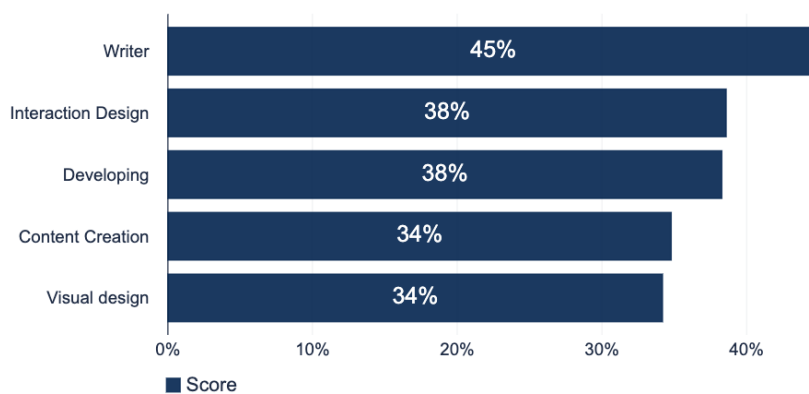
Some limitations of the library had to be overcome with a custom bar chart design. Besides preventing an overlap of the labels like in Figure 10, this enables the use of patterns or symbols in the bars themselves, inserting icons and the placement of further text or information within the chart layout. Different iterations of this custom design can be seen in Figure 11. The second row shows the first use of patterns. They were too small and did not have enough contrast to the background, so the checks and Xs were colored white instead and increased in size. This can be seen in the third row. As for the review with experts, the stage in the first row was discussed (chapter 6.)

Figure 11 development of the custom bar chart design including icons



For displaying the scores only for each category of a dimension, a simple out-of-the-box bar chart from pygal was used. It can be seen in Figure 12. Further development could introduce a variation of Figure 11 to accommodate the scores to make for a more uniform overall design.

Figure 12 bar chart with scores only, created with pygal



5.1.2. Report

In the StroCards prototype, the report page shows reports that do not differ much from the original BITV report. The same three sections are used: passed, failed and not applicable, listing every success criteria with the comments made during evaluation. Since these categories are used throughout the dashboard as well, the same colors are used in the headings (Screenshot 3 and Screenshot 4). This helps to fortify the connection of the color with its meaning and thus makes these headings in the report and the bars on the other pages easier to understand. Screenshot 3 shows an excerpt of a report of what was later called "list view". Its design is very close to the one in the BITV report. Some changes include more apparent differences in font sizes in order to emphasize the hierarchy and the use of a different color and font for code-snippets. Screenshot 4 shows an example of the table view. Which uses different columns instead of a visual hierarchy.

This page with the report was not discussed with the experts, but curiosity as for which of these views is preferred led to this being explicitly asked during the user tests. Results are discussed in chapter 7.2.

Screenshot 3 an excerpt from a StroCards report showing some success criteria in list view

Erfüllte Prüfschritte

✓

- 1.3.1a [HTML-Strukturelemente für Überschriften](#)

Seite 1:

Unter dem Logo steht ein leeres h1-Element (siehe Prüfschritt 2.4.6a, wo dies strenger bewertet wird). Man sollte dieses Element entweder entfernen (und eventuell "Herzlich Willkommen ..." mit h1 statt h2 auszeichnen) oder mit einem sinnvollen Inhalt versehen.

- 1.3.1a [HTML-Strukturelemente für Überschriften](#)
- 1.3.1d [Inhalte gegliedert](#)

Seite 1:

Leerer Absatz (p-Element enthält lediglich nbsp-Zeichen-Entity-Referenz): Löschen und eventuell ersetzen durch CSS-Regel die der h2-Überschrift einem größeren margin-bottom gibt.
(Das leere p-Element wird in Firefox von NVDA tatsächlich "gelesen": NVDA sagt "leer". Auf Seiten mit einer Zeitbegrenzung, wird das Vorlesen von leeren p-Elementen den Screenreader-Benutzer extra benachteiligen.)

- 1.3.1d [Inhalte gegliedert](#)

Seite 2:

Es empfiehlt sich, das leere p-Element zwischen der Liste von Assessments und dem footer-Element durch `margin-bottom` oder `margin-top` im CSS zu ersetzen.
Da die gefetteten Text wichtig sind, sollte man die b-Element durch strong-Elemente ersetzen.

Screenshot 4 an excerpt from a StroCards report showing some success criteria in table view

Erfüllte Prüfschritte			
Prüfschritt	Beschreibung und Link zur Referenz	Seiten	Kommentar
1.3.1a	HTML-Strukturelemente für Überschriften	1	Seite 1: Unter dem Logo steht ein leeres h1-Element (siehe Prüfschritt 2.4.6a, wo dies strenger bewertet wird). Man sollte dieses Element entweder entfernen (und eventuell "Herzlich Willkommen ..." mit h1 statt h2 auszeichnen) oder mit einem sinnvollen Inhalt versehen.
1.3.1a	HTML-Strukturelemente für Überschriften	2, 3, 4, 5, 6, 7, 8	
1.3.1d	Inhalte gegliedert	1	Seite 1: Leerer Absatz (p-Element enthält lediglich nbsp-Zeichen-Entity-Referenz): Löschen und eventuell ersetzen durch CSS-Regel die der h2-Überschrift einem größeren margin-bottom gibt. (Das leere p-Element wird in Firefox von NVDA tatsächlich "gelesen": NVDA sagt "leer". Auf Seiten mit einer Zeitbegrenzung, wird das Vorlesen von leeren p-Elementen den Screenreader-Benutzer extra benachteiligen.)
1.3.1d	Inhalte gegliedert	2	Seite 2: Es empfiehlt sich, das leere p-Element zwischen der Liste von Assessments und dem footer-Element durch <code>margin-bottom</code> oder <code>margin-top</code> im CSS zu ersetzen. Da die gefetteten Text wichtig sind, sollte man die b-Element durch strong-Elemente

5.1.3. Page Overview

The final page that needed to be designed is the page overview (Figure 13). It provides a bar chart visualizing passed/failed/not applicable success criteria for every page. Each of these sections provides general information about that page like its URL and a brief statement about its conformity. Furthermore, a link to the dashboard specifically for that page is provided. The design in Figure 13 used borders to visually separate these tiles from one another. The colors for these borders are intended to emphasize the conformity of that specific page. One could argue that the use of colors that are the same as those used in the bar charts should not be used for another purpose like this. This could cause confusion for the viewer. Further confusion can occur with an apparition of this green border. When viewing this kind of report, it is more likely that none of the pages are conform, since all success criteria have to be passed without exception. This results in the user being familiarized with all red borders only leading to the user never learning the meaning of the colors. Additionally, the coloring of the border may seem redundant since the information it is supposed to convey already is provided in text form. For these reasons, the borders were eliminated in the implemented and user tested version.

This design was not discussed by the group of experts in chapter 6. It is a result of input provided by them. Similarly, the single bar chart at the top of the page was one of their recommendations. It is supposed to provide an executive summary that contains information of all the success criteria for all pages.

Figure 13 Page overview initial design



5.1.4. Designs showing values over time

Since several Web Accessibility reports can be created over time, allowing users to see the development of the results could be a desirable feature of StroCards. This was also mentioned by experts (described in more detail in chapter 6.). Providing a presentation of the development can be encouraging for the viewer if they have a direct comparison to what their website used to be like. Figure 14 shows what a row of data could look like in the dashboard with a line chart instead of a bar chart. This visualization requires a lot of detail. As Tufte suggests: “Why waste the power of data graphics on simple linear changes, which can usually be better summarized in one or two numbers?” (Tufte, 2007, p. 30). Especially if the dataset consists of merely two values, providing a tendency or those two values could make for a clearer and more precise design.

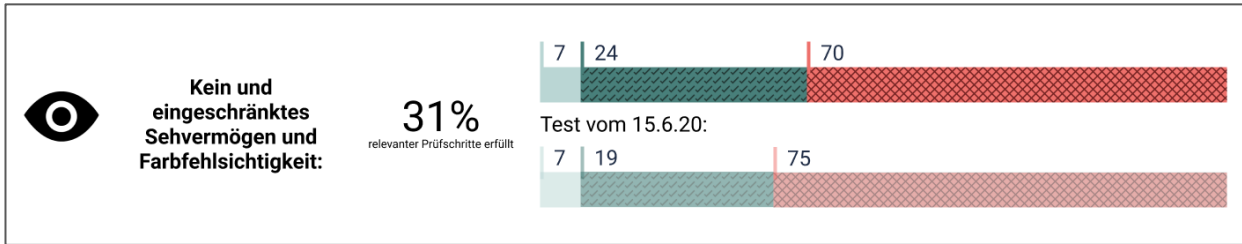
Figure 14 possible design for the data for the role of writer in the dashboard with a line chart instead of a bar chart



Of course, if more data points are available, meaning more reports for the same page, a line chart could be more fitting. These data points could also be achieved if resolved issues could be documented directly with StroCards. Since this is not the case, the idea of visualizing data over time was not followed through any further. Comparing reports this way – that tested a different set of pages – is not possible and there were not several comparable example reports available for this project’s purpose.

Another way of presenting this data can be seen in Figure 15. It displays the bar chart of the last report with the current one. Making this one slightly transparent is supposed to convey that it is old and out of date. On the other hand, this modification is not accessible due to a lack of contrast with the background. Using this design would require decisions like how many bar charts (and consequently reports) can be presented simultaneously. If the decision is two, which report should be compared to: the very first or the second last next to the newest?

Figure 15 possible design for the data for impairments related to vision in the dashboard with two line charts



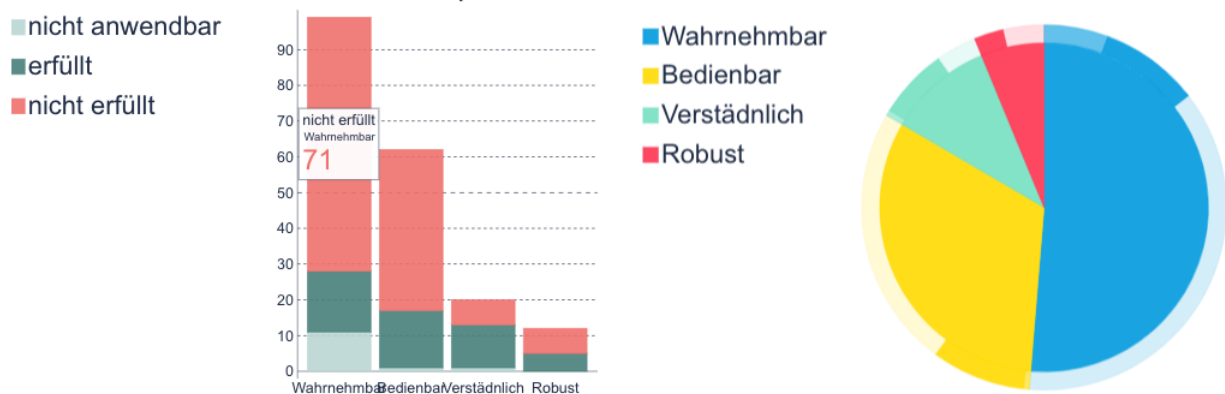
5.2. Recapitulation: Designs presented to the experts

5.1. Specific Iterations and Approaches described design changes considering the recommendations by the experts presented in the next chapter. This chapter's purpose is to provide an overview of the status of the designs that were discussed with the experts.

This discussion included content for the dashboard only. The concept of the page overview and of the report were developed later.

The presented elements are the following:

Figure 16 two variations for a chart by principle. Left: vertical stacked bar chart with a tool tip. Right: pie chart



Defining the Design: Wireframe development

Figure 17 horizontal bar chart with separate bars for passed/failed success criteria and a separate bar for the sum of success criteria per role

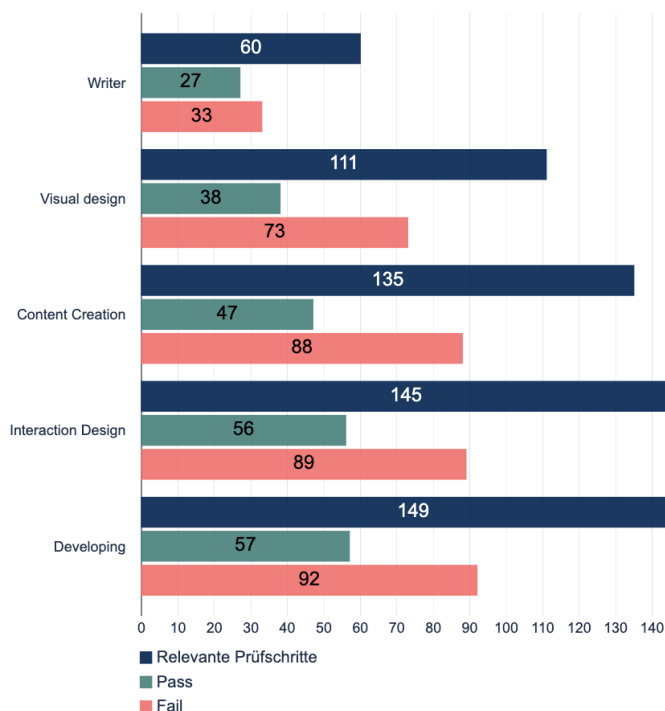


Figure 18 horizontal bar chart for scores for each role

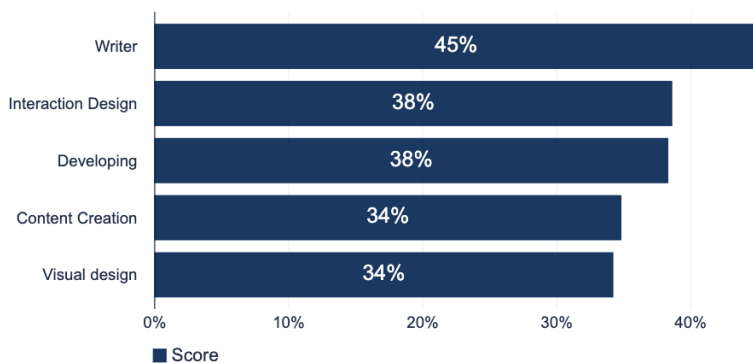


Figure 19 example of a tool tip in a custom bar chart



Figure 20 design for a legend



Figure 21 summary with the highest and lowest scores for the UAN dimension

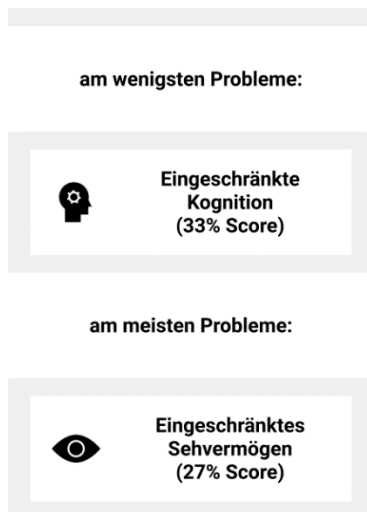


Figure 22 chart presenting success criteria by level of conformance. The typical order (A, AA, AAA) is changed and sorted by value. Keeping the natural order would be in accordance with Nielsen's Usability Heuristic #2 (Nielsen, 1994)

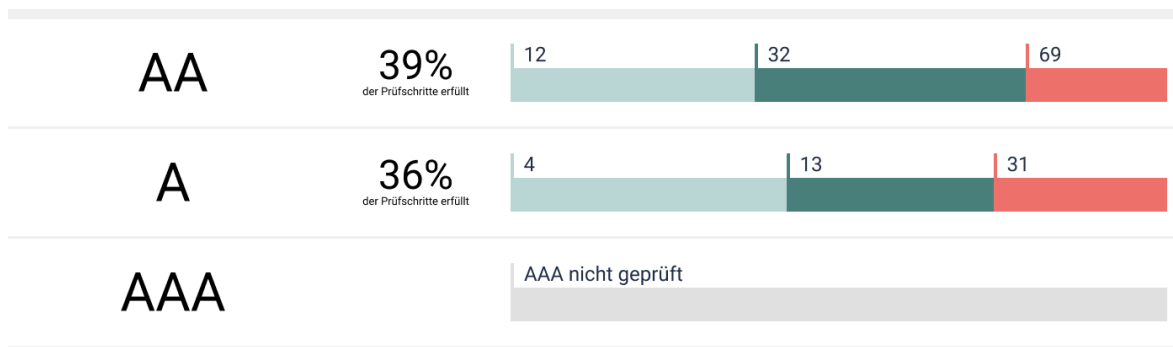
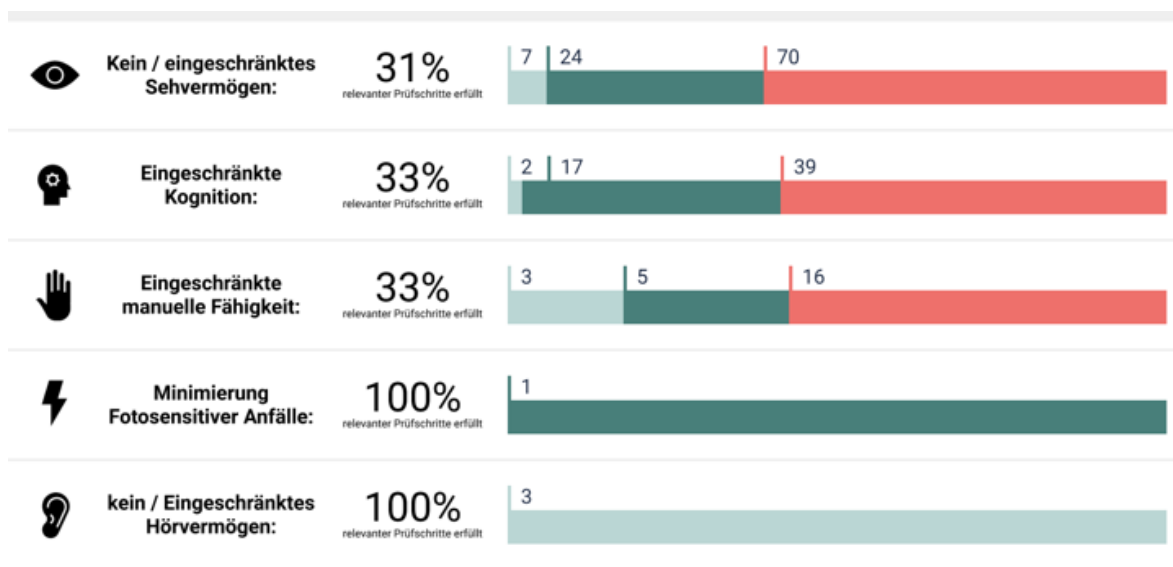


Figure 23 example for a chart using combined UAN



Defining the Design: Wireframe development

Figure 24 example of generating icon variations in order to differentiate between limited vision and no vision



Figure 25 two variations of a chart showing success criteria by UAN. Left: passed and not applicable success criteria are treated separately. Right: not applicable success criteria are summed with the passed ones



6. Wireframe Review with a Group of Experts and Implementation

In order to improve created wireframes and decide on different design approaches, several experts with different backgrounds were gathered in a virtual meeting. A document with several sketches for different aspects of StroCards (chapter 5.2.) were shared with the participants in advance and were later discussed in a moderated fashion similar to a focus group. This chapter presents the goals of the expert review, summarizes said meeting, and presents its results and what was decided to be implemented in the StroCards prototype.

6.1. *Goals of the Review*

The elements and questions that were to be discussed in the expert interview are the following:

1. What type of chart is preferred when showing the amount of success criteria for each dimension? Bar Charts, Stacked Bar Charts or Pie Charts? (Figure 16, Figure 17, Figure 18)
2. Which charts/dimensions seem necessary to the consulted experts?
 - a. User Accessibility Needs
 - b. Roles
 - c. Levels of Compliance
 - d. Principle
 - e. Scores for each of these aspects
3. If hover-tooltips (Figure 19) seem necessary: how could they be implemented? Or is a Legend (Figure 20) for each chart a better option?
4. Does it make sense to show a summary with the highest/lowest score for a dimension? (Figure 21)
5. Should categories in a dimension be sorted by their values or kept in the same order throughout the application? (Figure 22)
6. Do the experts approve of combining groups with comparable User Accessibility Needs to allow the final report to be shorter? (Figure 23)
7. What kind of icons are preferred? How are modifications received that convey the difference between e. g. "Visually impaired" and "blind" with the same icon as a base? (Figure 24)

6.2. *Participants and Method*

An advantage of this type of feedback is the dynamic between participants. By sharing their thoughts, they can influence each other's ideas and develop these further. This is a dimension lost in an interview with one single person.

Participants were recruited by the Competence Center on Digital Accessibility of the Stuttgart Media University, since it has many contacts in the field. Besides members of said Competence Center, two people directly involved in the maintenance of the BITV test, a Web Accessibility evaluator currently working on how to facilitate that job, and a Computer Science student were part of the meeting.

Participants were given the wireframes and designs (chapter 5.2.) a week in advance to enable them to prepare for the video conference.

6.3. *Results*

Summarized answers to the questions listed in 6.1. are as follows:

1. As for the **type of chart**, pie charts were mostly disapproved of. Possibly the labelling used in the version that was discussed tended to be more misleading than helpful. Rather than having this type of pie chart with this hierarchical data, bar charts were favored by all experts. Even though proportions of values relative to each other are easily understood in a pie chart, they do not convey absolute numbers very well. Furthermore, a full pie would symbolize accessibility as a whole, not leaving room for encouraging further improvements. There also was accordance in the group agreeing on a stacked bar chart over a regular one since it makes it easier for the viewer to compare data. There was a noteworthy remark in changing the diagrams for each dimension: the colored parts of the charts should not be the same length. This implies that e.g. every UAN has the same number of success criteria mapped to it, which simply is not the case. The length of these bars should be put into correct proportion to each other. It was important to the experts to have an accessible application for showing another website's accessibility. A recommendation was to not solely rely on color coding for each type of success criteria but to use color-independent symbols or patterns as well. This is even more important since the chosen colors are red and green. These colors are impossible to differentiate for people with red-green color blindness. One of the experts recommended to leave out not applicable and passed success criteria since these are not of much interest for the viewer. It was contradicted that seeing how

much has been achieved in terms of accessibility can be very motivating. A solution for this is to make which type of success criteria to display a customizable setting.

2. The reaction to each of the charts was as follows:

a. **User Accessibility Needs**

This chart was described as a “nice to have”, redundant for several use cases, with a lower priority since a website that is BITV-conform is accessible to all of these groups to the same extent. Differentiating these groups might be problematic when it leads to favoring some of them just because a certain group consists of more people. This would be against the ideology “Accessibility is for everyone”. It was noted that viewing this chart could make the user more aware of the relationship between success criteria and how they actually affect people. This tendency is reflected in drafts for the upcoming version of WCAG where this structure or hierarchy becomes important. Fixing issues for a certain group could feel less abstract and more motivating for developers. A compromise in this discussion was to change the order of graphs to show the chart for the Roles first, since this is more important for many viewers of the report. This led to the statement that providing alternatives for the presentation of information and personalization is important for the user.

b. **Roles**

The idea of this diagram seemed very useful to the experts. They recognized its value because it helps understand which team is responsible for which issues. The names used for each role could be changed since the difference between “Content Creator” and “Writer” is not obvious. It was noted that not every client looking at the report might have teams for each of the listed Roles. “Interaction design” was given as an example for a rather rare role. This could also diminish general comprehension of the chart and/or cause confusion.

c. **Levels of Compliance**

This diagram was deemed not relevant in all of the experts’ opinions. BITV implies conformity with level AA.

d. **Principles**

It was mentioned that the distribution stays the same, since the number of success criteria for each principle stay the same. Just seeing how many of them are passed/not passed is a “goodie”, maybe at the end of the report. It elongates the report with data that is not of much value for a user who is not very interested in this chart.

e. **Scores** for each of these aspects

Calculating and displaying scores was heavily discussed among the experts. The necessity of making it as clear as possible how the scores were calculated was discussed. SiteImprove was mentioned as an example. This monitoring tool provides a score and a comparison to the industrial standard. Especially when being used consistently, scores can be used in a meaningful way. It was criticized, though, that the same scores were already displayed within the diagram with the bar charts. Regarding this, they seem to be redundant. There was a consensus though, that scores should not be calculated with an absolute number of affected people. Counts of people with certain disabilities are not reliable and favoring certain groups because they are simply bigger should not be encouraged. Furthermore, it is not only the disabled that benefit from Web Accessibility.

f. **Recommendations:**

- **“Executive Summary”**

It was mentioned that a kind of summary was missing in the distributed sketches. There was nothing describing all success criteria as whole. This was missed by one expert particularly who expected this at the beginning of the report to instantly have a good overview of the status of the website.

- **Pages Overview**

Developing a chart presenting passed and failed success criteria for each tested page was recommended. Having a page layout with a tile for each page enables the user to have an overview of the status of each page. Relevant actions can be taken if one page has an exceptional amount of third-party content and more failed success criteria.

- **Price tag**

Which team is responsible for which accessibility issues? Does it affect an internal team or one that is in-house?

Ideally, each issue would be displayed with a price tag to give the viewer a feeling of the effort that is needed to fix the issue. This was admitted to be an unrealistic wish by the expert, but helpful, nonetheless.

- **Criticality Flags**

Slightly more realistic than calculating a price for resolving each issue would be to emphasize especially critical issues. There used to be a weight for each success criteria, but criticality is not described in current standards. It is possible for the

person testing the website to provide information on criticality and to make general assumptions: e.g. missing alternative text are always critical. In turn, however, this is always depending on the context of the tested website.

Even though this kind of information might still be available in the BITV database, making the effort to accommodate it in StroCards is not efficient since it would not be available for all the reports.

Nevertheless, this way of prioritizing issues would be extremely valuable for users and a system for scoring is discussed for future versions of the WCAG.

- **Development over time**

If more than one BITV tests were conducted for the same page, it would be desirable to be able to see results next to each other in order to see the development over time. This could be e.g. by placing bar charts over each other.

This could have a motivating effect on viewers since they can easily see the progress they have already made. It was mentioned that monitoring tools mostly rely on line charts for a timeline. This is an efficient way of displaying lots of data over time.

However, in the case of BIVT tests, they are simply a snapshot for a specific moment and not conducted frequently – at least not at the time this discussion took place.

3. The document that was sent to the experts in advance showed the state of a chart with and without a **hover-tooltip** (Figure 19). The experts immediately expressed disapproval of using tooltips for a few reasons: when printing a website or specifically this report, these tooltips would be missing or all of them would be displayed. This is problematic in both scenarios. Furthermore, there is no “hover” when using the website on a mobile device, rendering these tooltips not usable. In conclusion, they should be avoided or not be the only source of describing the information visualized. Having all information visible at once makes it easier to include more people in the viewing and understanding of the report. Having a legend could be a better option. In general, it was encouraged to give users several ways of accessing and displaying information. It was praised how the labelling with the numbers was designed. Connecting the start of a bar to the number above with a “nose” instead of displaying text in the bar itself is an elegant solution compared to having cut-off text when the values are very small.
4. A **summary** with the highest/lowest score for a dimension was seen as rather redundant by the experts. An application like StroCards should be focusing on making a report as short as possible. This element would make it more extensive by providing information that is easily

extracted from other diagrams, assuming it is displayed next to the detailed chart it summarizes.

5. The experts could agree on always **displaying the categories of a dimension in the same order**. This should prevent confusion. Sorting them by their values should not be necessary since values are represented in the bar charts. This coincides with WCAG 2.1. Success criterion 3.2.3 Consistent Navigation (W3C World Wide Web Consortium, 2018).
6. Some of the experts liked the idea of **combining the User Accessibility Needs** in order to make the report less overwhelming. It was discussed how far one can go when combining them: Both groups “not sighted” and “color blindness” are related to eyes but the relevant success criteria differ greatly. Furthermore, combining “photosensitive epilepsy” with another group is problematic since it is caused by the eyes but takes place in the brain which can make it related to cognition. In any case, combinations should be made transparent enough in StroCards e.g. with a paragraph or a tooltip. Alternatively, each combined group could be made expandable to display the groups it consists of.
7. In order to provide **icons** for combined user groups, it was recommended to place the respective icons next to each other.

The presented modifications that were made in order to differentiate between e.g. “Visually impaired” and “blind” were not supported. These changes – making the icon more transparent or blurry – would be extremely hard to tell apart from the original for people with visual impairments.

In conclusion, discussing with these experts revealed many important points. The conversation went astray at times, debating adjacent topics, nevertheless providing good recommendations for the wireframes. The fact that some of them were contradictory (e.g. whether or not passed and not applicable success criteria should be presented) proves how individual the use of StroCards can be. This encourages the possibility of providing customizable settings in StroCards in order to accommodate as many users as possible.

Any occurring uncertainties are meant to be evaluated in user tests that are described in chapter 7.

6.4. *Implementation*

StroCards should generate pages dynamically since there are several combinations a user can choose: dashboard or report, for all pages or for a specific one, for a certain user Accessibility Need or for a certain role.

This was implemented with Python and the web Framework Flask. The Jinja2 templates can be filled with data that is formatted with Python. Flask is easily extendable to include user management for instance. An SQLite database is stored on the same server as the website itself and contains data for user management, their uploaded reports and their personal settings for the dashboard. Flask integrates this nicely with the Python toolkit SQL-Alchemy. The settings that were stored were the elements of the dashboard the user choses to be displayed. Which type of success criteria (passed/failed/not applicable) are to be displayed is not stored and reset every time the report is loaded.

On the subject of database, accessing existing BITV reports was not straightforward. The reports that are public might not be the most interesting to see in an application like StroCards since they are BITV compliant. In order to count e.g. failed success criteria, the data was stored in a JSON object. The Python library Pandas allows importing data as JSON to create a dataframe. These dataframes enable calculations and modifications of values within itself very efficiently. This is useful for calculating scores for e.g. every project role.

Since accessing realistic report data was difficult at the beginning of development, this data was extracted from HTML reports. This was done using the tool Beautiful Soup and to be regarded as a temporary solution in order to develop the prototype. Scraping HTML documents is not very reliable since any slight deviation in structure can produce an erroneous conversion. Testing of this feature was not very extensive since accessing data directly was planned to be a next step. Since this was implemented towards the end of this project, the user tests (chapter 7.) were conducted with a version that relied on HTML scraping.

The mappings of the success criteria to each role and UANs respectively are saved in a CSV format. This is easy to edit with Applications such as Microsoft Excel. The Python script csvConverter.py can create a Python dictionary from the csv – assigning each role a list of success criteria.

For creating dynamic charts as SVG several existing libraries were compared. They proved to be not flexible enough in terms of customization. E.g. “pygal” can create stacked horizontal bar charts but inserting an icon for each of the charts is impossible. Also changing charts without reloading the whole page was near to impossible. This function is needed to update the charts depending on which success criteria the user wants to see (passed, not passed, not applicable). Also, tested libraries relied heavily on tooltips which was discussed in the interview with experts in the design phase (6.). An accessible legend for a bar chart for instance cannot only rely on differentiating the information by colors. In order to include people who are color blind, information should be coded

not only by color only but also with symbols or patterns. Creating bar charts from scratch seemed like a more feasible task considering these requirements and the desired design.

As for the frontend, some UI frameworks were compared. Essential requirements for accessibility were that they display elements with enough contrast ratio. In the course of development – starting with the simple markup – it seemed too overengineered to use a UI framework since StroCards has fairly simple interfaces.

7. User Testing

The meeting with the experts was comparable to a focus group. The Nielsen Norman Group discourages them being the only source of user feedback when developing interactive systems. They say the only way to test such a system for its user experience is to have individuals test it and try it out on their own without a group influencing their actions. Furthermore, actively using a product is very different from being passively exposed to a demonstration – as was the case in the previously presented approach. (Nielsen, 1996) Since these results can be considered like an iteration in the design process, individual user tests were conducted with the functional prototype of StroCards. User tests were facilitated, with quantitative and qualitative data retrieval.

User tests were conducted in order to validate the efficiency and benefit added to BITV Web Accessibility Reports when using StroCards. In order to make this comparison, both the current BITV report and StroCards were tested in a similar way.

The goals of the study and questions to be answered are as follows:

1. How easy/hard to understand is the current presentation of BITV reports for the testers?
2. Could the prototype convey the same information better – especially for non-experts?
3. Are there issues concerning comprehension and usability of the prototype?
4. Which customization settings are preferred?

7.1. Method

After describing the technical setup of the user tests and which data was to be acquired in the following, the exact procedure is described (7.1.2.). Finally, the population of the study is outlined (7.1.3.).

7.1.1. Setup and Data Acquisition

Due to safety measures during the COVID-19 pandemic and with the possibility of recruiting more participants for the study, user tests were performed **remotely**. With the application TeamViewer, testers were able to access the prototype that was running on a local machine at the time of testing. They were moderated and were meant to provide both **quantitative** and **qualitative** data. The screen and the voice chat were recorded, and an additional transcript writer provided a backup for the recording. Including a video chat and recording it would have been helpful to

understand the testers' reactions more clearly, but it was omitted due to performance and bandwidth issues.

All texts in the prototype are written in German to enable the exemplary report blend into the application. For this chapter, all texts were translated into English. The original forms that the participants filled out can be looked up in the appendix. Since German was the native language of all participants, this was the language spoken during the user tests.

An advantage of quantitative results is how expressive the results in the form of numerical values can be. At the same time it might result in not well-founded statements when interpreting and explaining correlations (Nielsen, 2004). This is why qualitative research is important. Furthermore, it makes it possible to find answers that are not listed in a form with checkboxes making them more specific to the testers without them having to compromise with a given option.

Quantitative Data was retrieved with the User Experience Questionnaire (UEQ) (Laugwitz et al., 2008). Testers are presented with 26 pairs of opposite adjectives. For each of these items they check a value in a 1 to 7 Likert scale to describe their experience or impression of the system. Items are mapped to factors or scales: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty. The UEQ offers benchmark data and tools to help evaluate results. The System Usability Scale (SUS) was considered as an alternative to the UEQ. Having only 10 questions it is "quick and dirty" as it was called in its introduction (Jordan et al., 1996, Chapter 21). It is very focused on usability though and not the user experience and heuristics. Also, since it was developed to evaluate systems, the questions asked might be hard to transfer to the static report that is being tested.

7.1.2. Procedure

The procedure of each test was as follows: in order to classify the results of each participant, they were asked to answer a few questions about their experience with Web Accessibility (A3. User Test Questionnaire) first. It was important to have values to compare the results of StroCards to, so testers were invited to have a look at a BITV example report. This was conducted in an explorative manner, meaning they were not given tasks. Testers stated their impressions and scrolled through the report, reading parts of it that caught their attention. Afterwards they filled out the UEQ for this report and answered a few open questions in a form (**Error! Reference source not found.**). In the next step they interacted with StroCards in a similar manner, clicking through available the views. Again, they were asked to fill out the UEQ and similar questions to the ones in the last form

(**Error! Reference source not found.**). These questions were more specific though and encouraged the testers to compare the two systems.

7.1.3. Study Population

In order to have the most representative results as possible, recruiting was focused in institutions with an affinity to computer programming, digital product management and Web Accessibility. Universities, that require accessible (web) content by law and monitoring centers were contacted but unfortunately could not recommend voluntary participants.

The study population consisted of **10** people with the following occupations. Each participant was given an ID to enable anonymized documentation which is written in brackets:

- **5** Computer Science and Media Students (TP1, TP4, TP5, TP6, TP10)
- **2** Product Managers for Digital Products (TP8, TP9)
- **1** App Developer (TP2)
- **1** Web Administrator (TP7)
- **1** Research Assistant for Digital Accessibility (TP3)

7.2. Results

7.2.1. Experience of the Participants

Looking at the results of the first questionnaire (Figure 26) it is safe to say that participants should be pretty confident using a computer. This question was posed in order to detect outliers in the questionnaires to come.

How experienced testers were with Web Accessibility was important to learn since the prototype is supposed to be validated with non-experts. Also, if there had been a participant with no prior knowledge to Web Accessibility in any way, combining their answers with those of the other participants could be problematic since a requirement for looking at Web Accessibility tests at all would be knowing what Web Accessibility is. Figure 27 shows that none of the participants checked 0: no experience. Also, according to this result, none of them deals with it every day which could define them as an expert.

When it comes to testing, none of the participants have commissioned a test to be done (Figure 28). Half of them have never conducted one themselves and the other half has conducted a test

themselves (at least partially). This shows that at least half of the study population is familiar with the structure of a Web Accessibility report.

Figure 26 Distribution of Answers: How experienced are you with the use of computers?

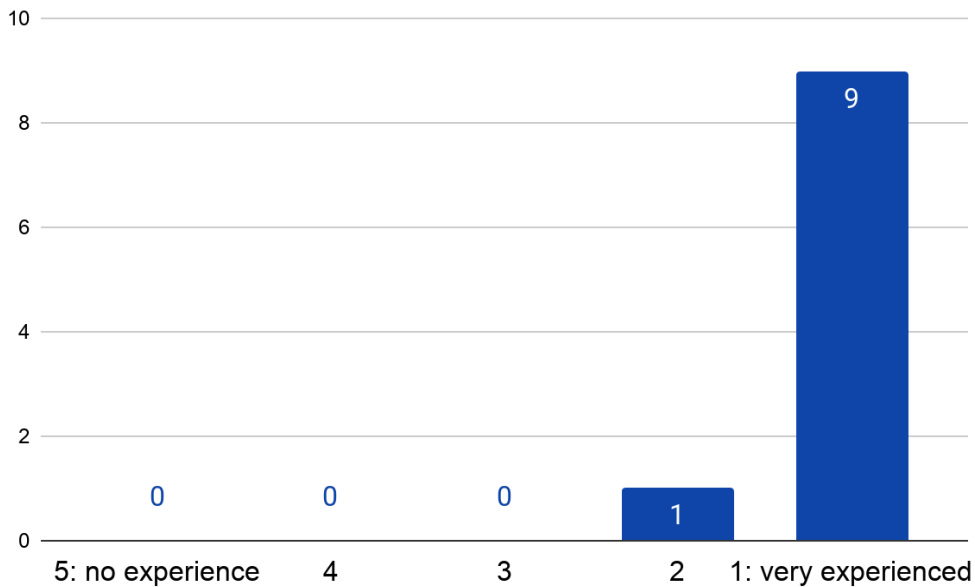


Figure 27 Distribution of Answers: How experienced are you with Web Accessibility?

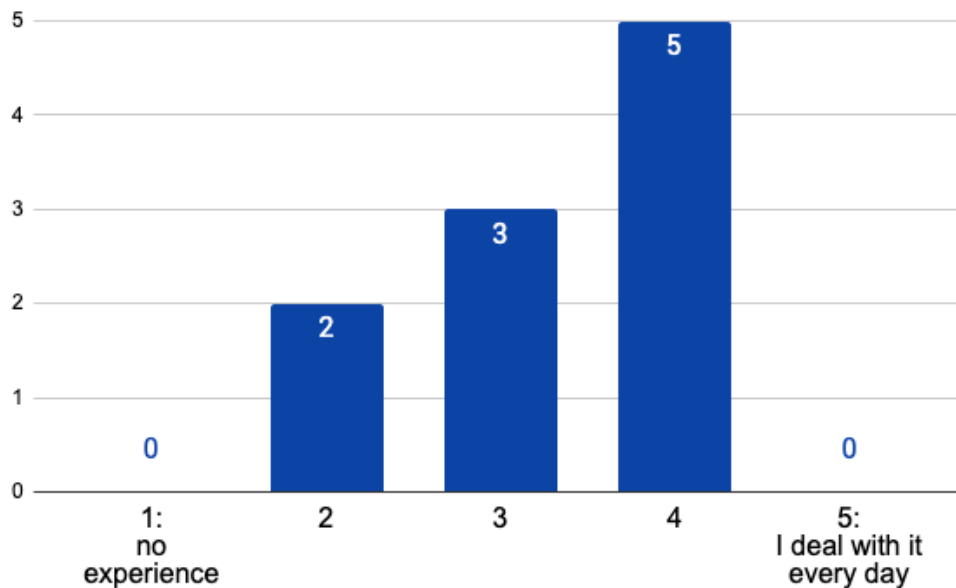
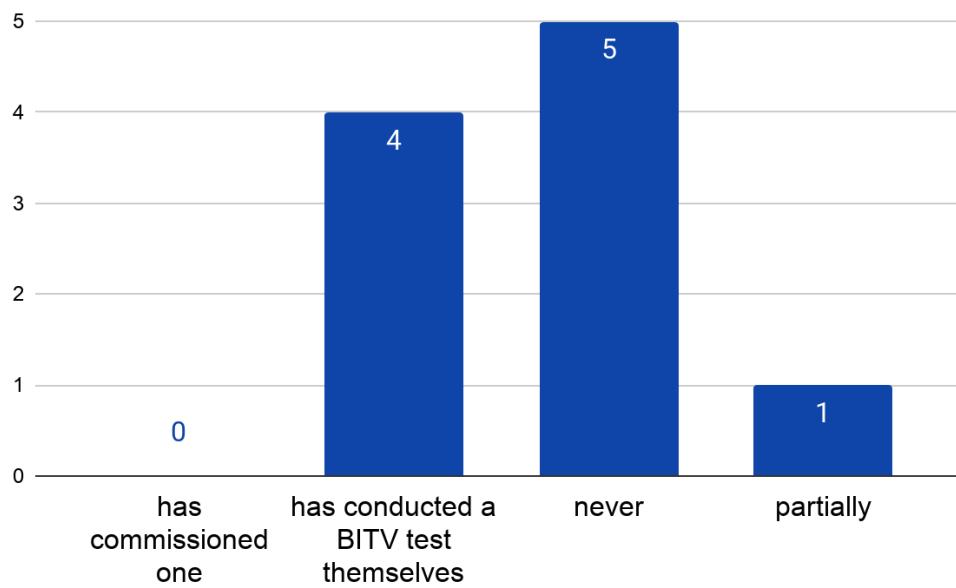


Figure 28 Distribution of Answers: Have you ever conducted or commissioned a Web Accessibility test?



7.2.2. User Experience Questionnaire (UEQ)

The results of the UEQ showed how StroCards has a better user experience in every aspect that the questionnaire addresses.

Figure 29 and Figure 30 show the distribution of the values that participants checked for each of the items between -3 (horribly bad) and +3 (extremely good).

Figure 29 shows the distribution for the BITV report and it is easily noted that there are a lot more negative (scores from 1 to 3) responses compared to the distribution for StroCards (Figure 30). The most negative response – a 2 – was found in the item “usual/leading edge”. A larger depiction of these charts is available in the appendix.

Means for each of these items are displayed in Figure 31 and Figure 32. In comparison, values were positive for StroCards without any exception. Each of these items is assigned a quality which is visualized by the use of different colors. Figure 31 shows for instance that the 3 items with the highest values describe the same quality: Dependability. The 3 items with the lowest values describe Novelty. These qualities are further explored in Figure 33 and Figure 34. Each quality for both of the reports can be compared and all of them show an increase in StroCards without exception. Moreover, the variance was smaller for each quality in StroCards which supports the reliability of the retrieved data. The diagrams depicting the qualities (Figure 35 and Figure 36) profit from this security respectively since they show averages of Perspicuity, Efficiency and Dependability for the pragmatic quality and Stimulation and Originality for the hedonic quality.

When compared to UEQ benchmark values, the BITV report’s scales are all “below average” or “bad” (Figure 37). Whereas for StroCards they are all “good” or “excellent”, except for the scale Novelty that has an “above average” result in accordance to the benchmark data. This fact is reflected in the example given before, that the item “usual/leading edge” scored rather poorly.

Novelty was the poorest scale in both report representations. A comment that might reflect this fact is: “At the end of the day it’s still a report” (TP3). Looking through guidelines “isn’t fun” (TP6) but if an application like StroCards can influence the Attractiveness or Stimulation of such a report, it is definitely an advantage. The next chapter will cover more on what the testers said and did while testing.

Figure 29 UEQ Results BITV report: Distribution of Answers per Item

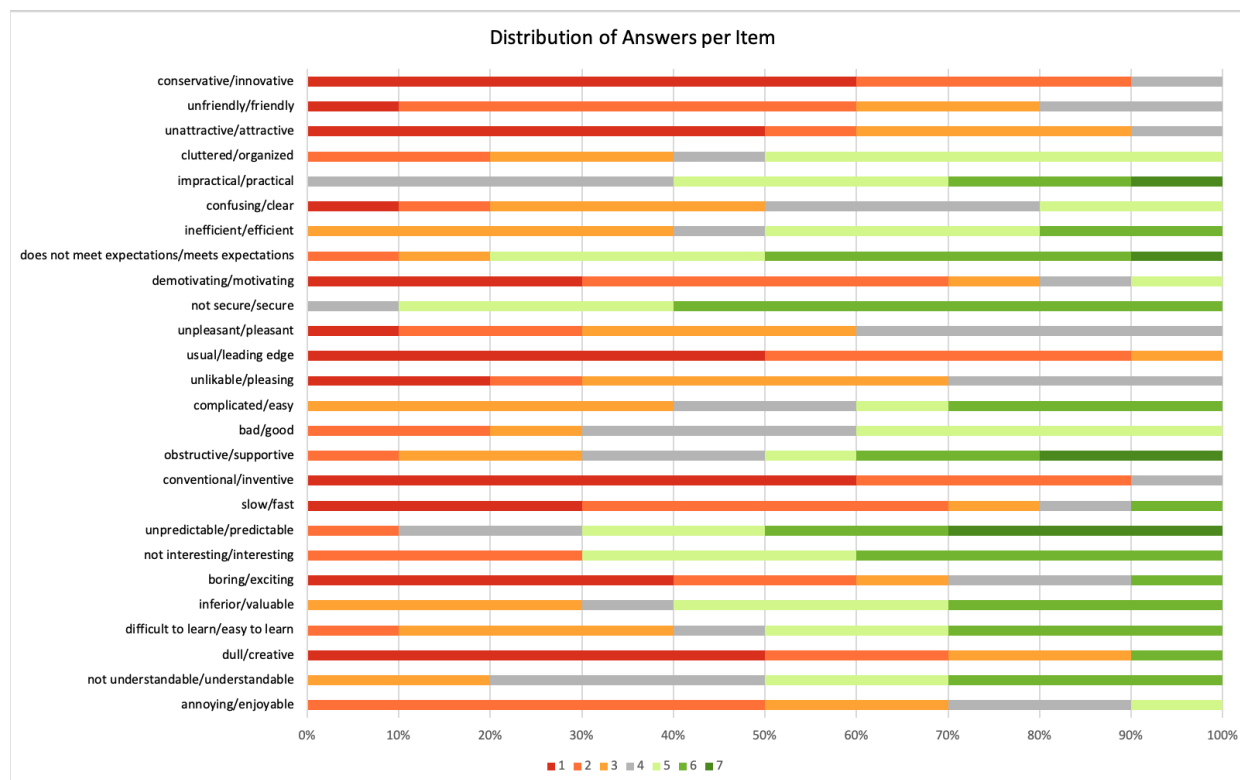
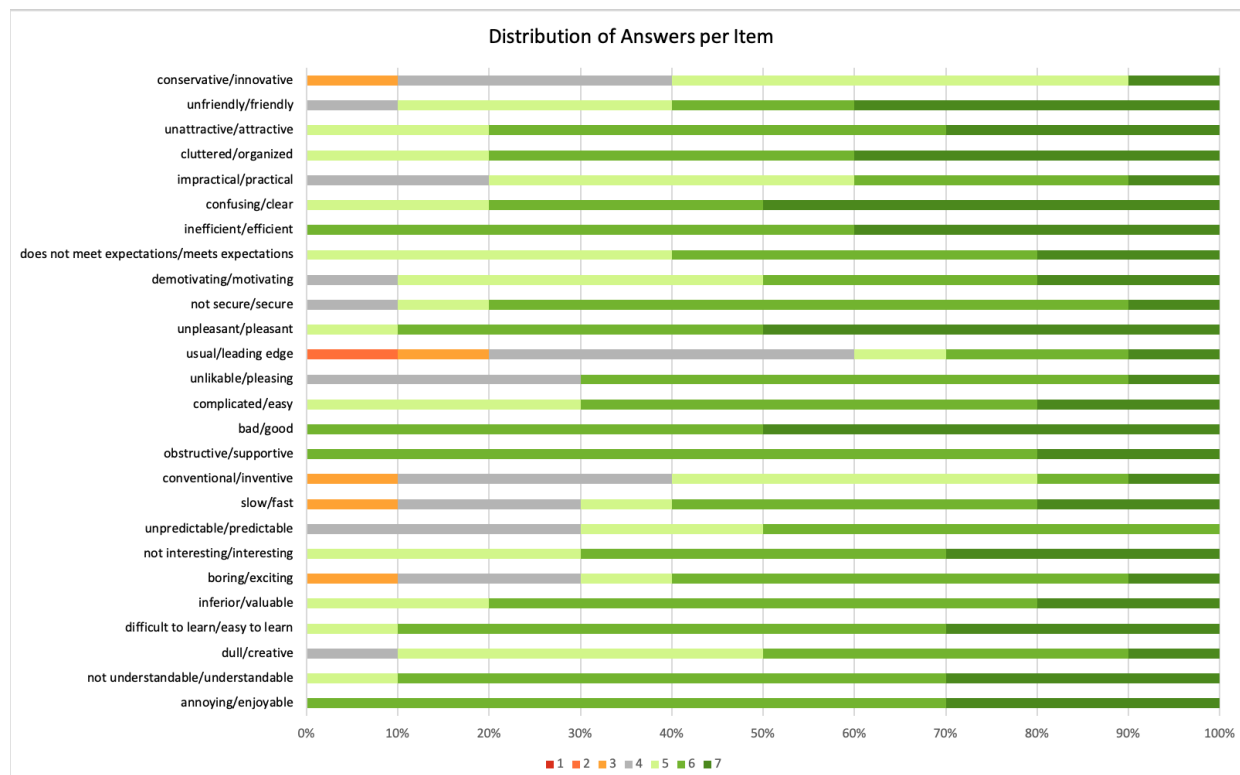


Figure 30 UEQ Results StroCards: Distribution of Answers per Item



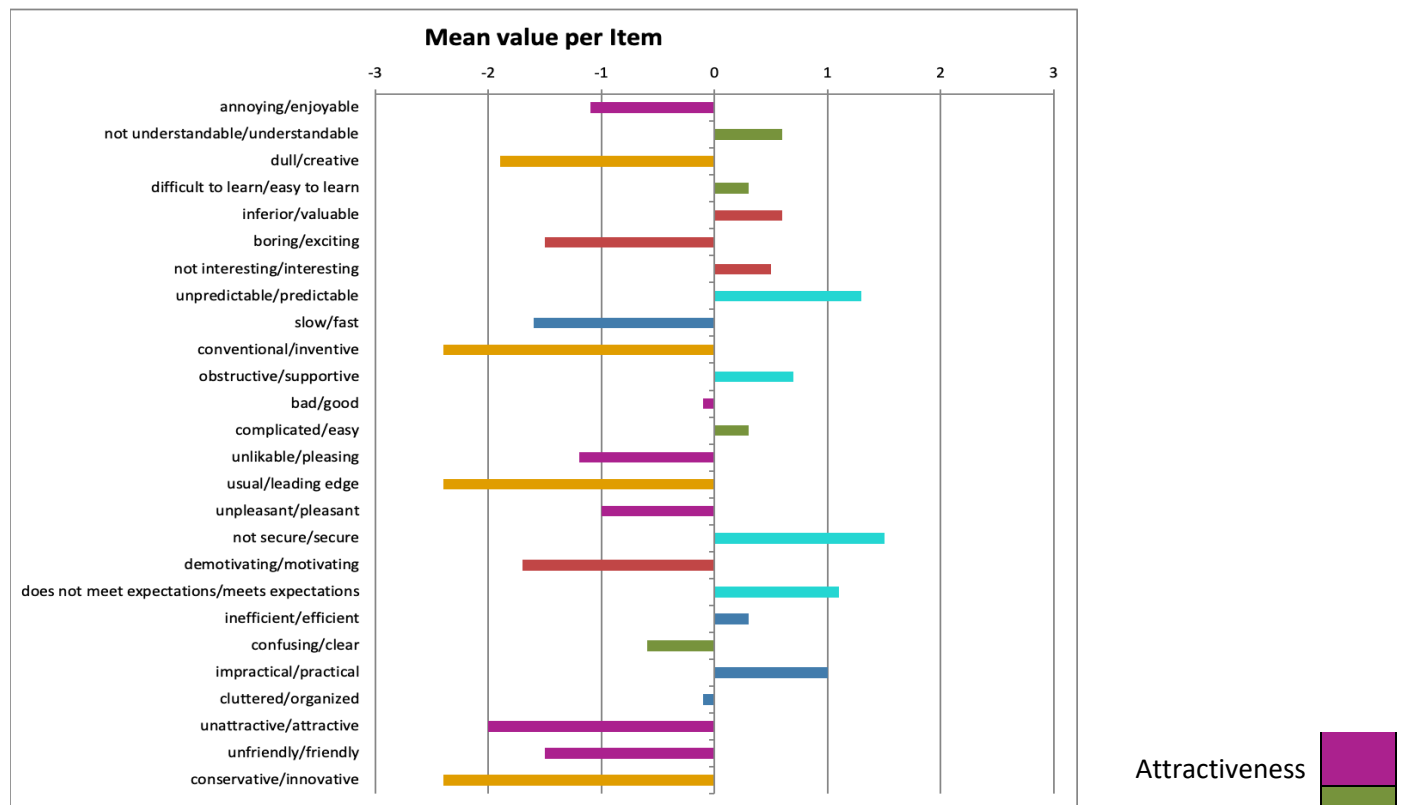


Figure 31 UEQ Results BITV report: Mean value per Item

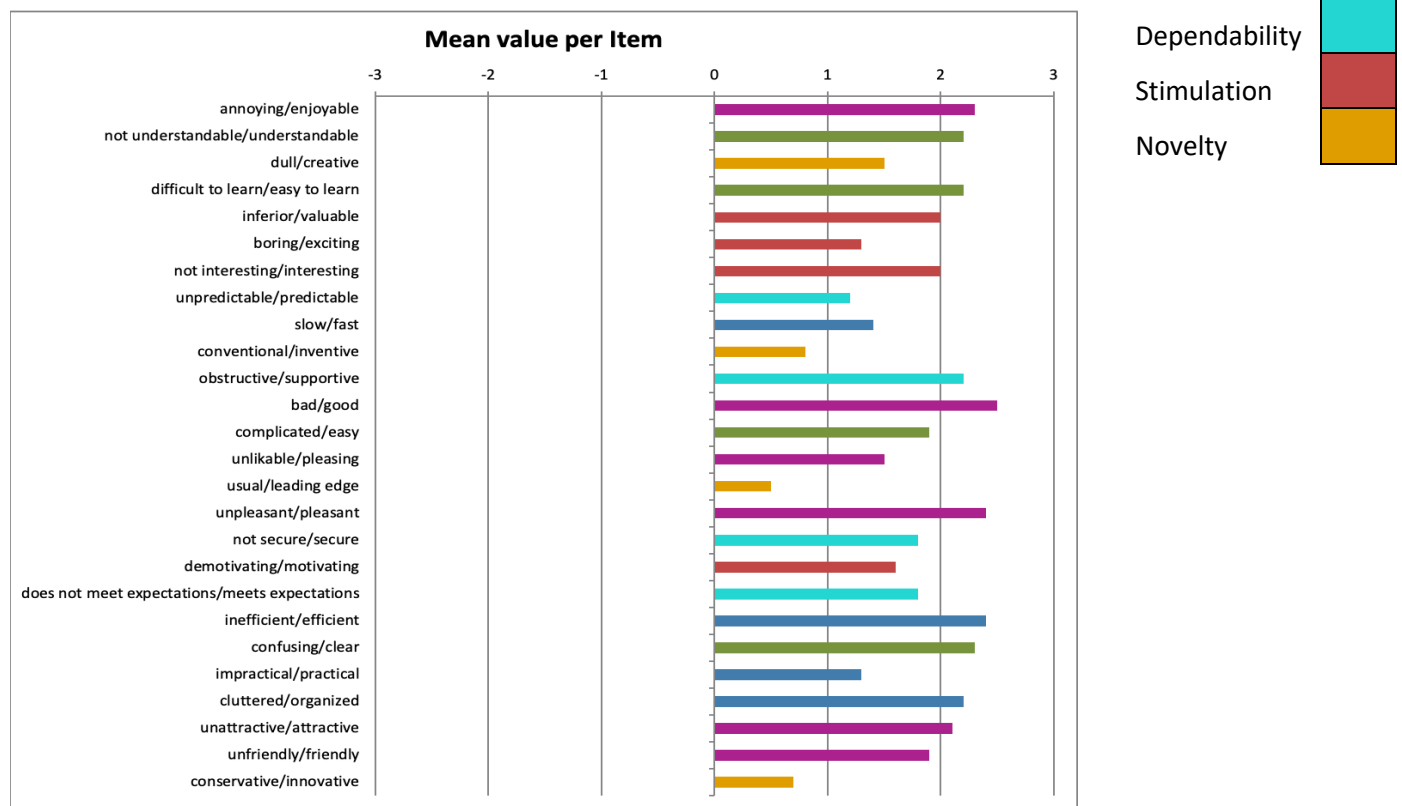


Figure 32 UEQ Results StroCards: Mean value per Item

User Testing

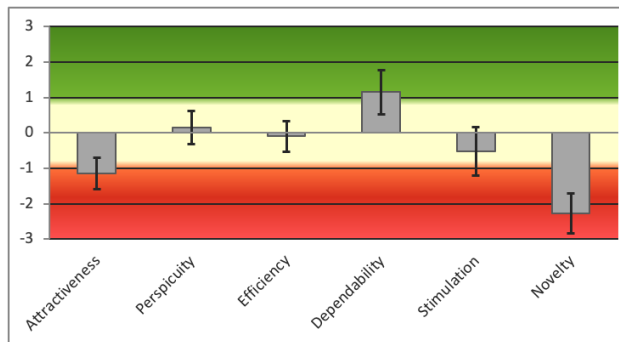


Figure 33 UEQ Results BITV report: Qualities

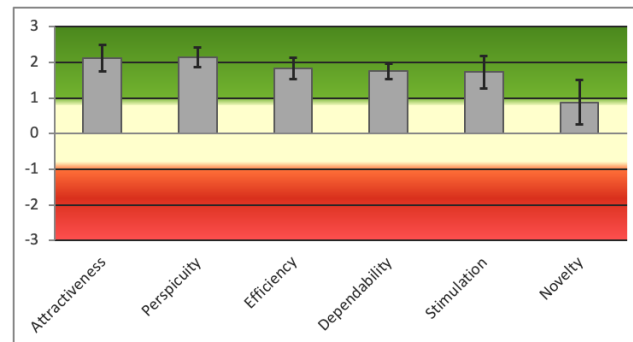


Figure 34 UEQ Results StroCards: Qualities

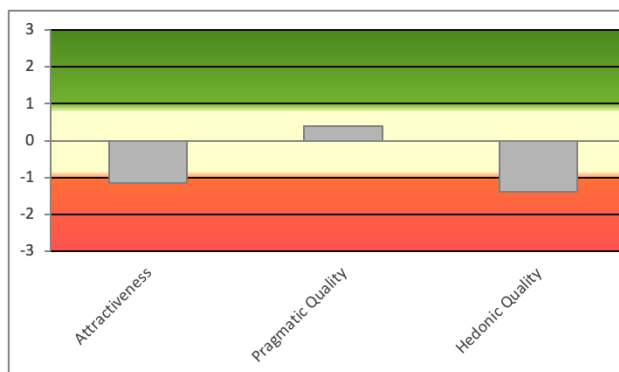


Figure 35 UEQ Results BITV report: Scales

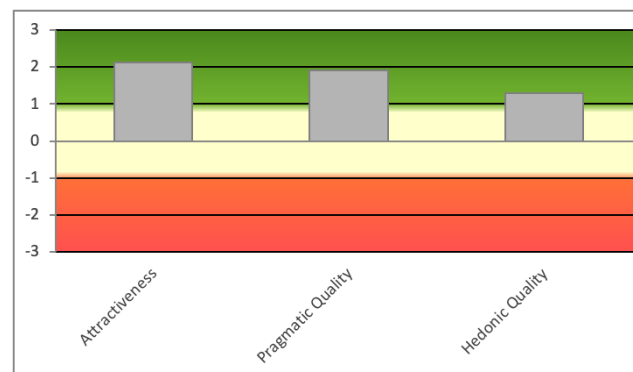


Figure 36 UEQ Results StroCards: Scales

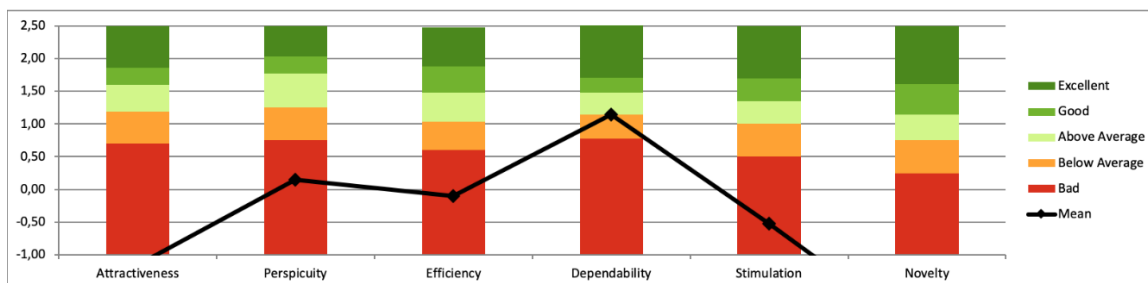


Figure 37 UEQ Results BITV report: Benchmark Comparison

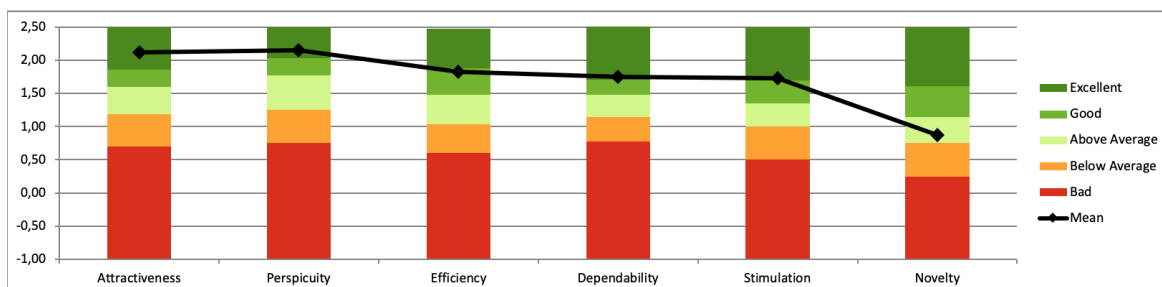


Figure 38 UEQ Results StroCards: Benchmark Comparison

7.2.3. Qualitative Evaluation

For this study there were two ways of retrieving qualitative data. For one, users were asked to answer non-leading questions after testing each report presentation. These questions were part of the questionnaire they filled out that included the UEQ. Answers were very different in how detailed they were, containing a single word to several sentence. This could have been influenced by how motivated participants or how attentive they were.

Next to the participants' questionnaire entries, notes were taken throughout the test by the facilitator and an additional transcript writer. The recording was rewatched after the test to make sure notes were as complete as possible.

To evaluate and combine both of these data sources, a table was created for each of the questions in the form. Occurrences of an answer or reaction were checked for every participant separately. If answers or reactions seemed similar enough, their counts were merged. An example of similar enough statements would be if one person said, "the information seems hidden" and another one "I really had to dig to find the information I'm looking for". This may have been slightly subjective in some cases, but it delivers more common themes than only comparing the use of the exact same words. Each of these statements received an occurrence count from the forms and an occurrence count through observation. It was necessary to display these counts in separate columns in the following tables in order to make possible inconsistencies transparent (e.g. 2 participants wrote they liked nothing but 9 said they value the information: these are opposing statements that should be coming from 11 people in total, following their contents' logic. The study however had only 10 participants.). Nevertheless, a total count will be displayed for each statement which is the number of participants that made that statement regardless whether this was in written or oral form. Responses of each individual participant are listed in the appendix.

For the first question after filling out the UEQ for the BITV report, testers were asked what they liked about it. Table 5 shows the results for that question in the previously stated structure. Sorted by the total number of occurrences (observed plus through the questionnaire) the most common positive feedback was for the contents of the report itself. Participants had the impression of reading a professional and thorough report. They also liked the fact of having a quick summary of the report with the number of passed and failed success criteria right at the beginning of the document. 3 testers said that the report became clearer over time – it is arguable whether this statement can be regarded as a positive one. On the other hand, it demonstrates that the report is not impossible to learn. Of course, this is a head start for the testers when being exposed to StroCards.

Since many of these statements refer to the content of the report, it is to be expected that testers would make the same statements about StroCards since these texts are not modified in any way. Simply, their structure is altered. Nevertheless, 2 participants replied that they did not like anything about the BITV report. This confirms a very negative experience with the report.

Table 5 Statement count: what did testers like about the BITV report?

Statement	Observations	Form Answer	Total
lots of (helpful information), very detailed	9	2	9
the result/general overview was at the beginning of the document	5	4	8
recommendations to improve website are included	6	1	6
the information feels correct, professional, to be taken seriously ("textbook-feeling")	6	0	6
Information becomes clearer with time	5	0	5
link to resources	3	1	4
examples are helpful	4	0	4
differentiation passed/not passed is well understood	4	0	4
the number of passed/not passed success criteria is clearly displayed	0	3	3
notes („Anmerkungen“) at the beginning of the page are large enough: reads them and understands why URLs are the same	3	0	3
each SC is structured	2	0	2
nothing.	0	2	2
everything is displayed at once	1	1	2
minimal design	2	0	2
seems manageable due to the titles of the success criteria	1	0	1
predictable	1	0	1
because of the general notes ("allgemeine Anmerkung") text is not repeated unnecessarily	1	0	1
seeing passed success criteria feels motivating	1	0	1
It is helpful to have the tester's contact data	1	0	1
having to URL is helpful to be able to test right away	1	0	1
at least the bold headings provided some structure	1	0	1

Table 6 lists statements that are answer to the question "What aspects of the report didn't you like?". It is visible at a glance that this table is more extensive. The most common observations include that the tester perceived the report to be a lot of text, frequently calling it plain, unformatted and very long. A comment by TP4 summarized the design like this: "This looks like Google results in 2003".

Most likely due to missing experience, half of the participants expressed incomprehension regarding the categories "fail", "pass" and not applicable. The answers collected in the questionnaire were not quite as explicit. The statement with the most occurrences was describing the BITV report as "just plain unformatted text". The missing explanations for each category were mentioned by two participants and the inconsistent order of described pages in the report as well. In this table specifically there are a number of statements that are solely observations. These might not have been perceived consciously by the tester like "misses a heading due to fast scrolling". It is in the nature of this fact that it could not have been actively mentioned by the tester in the questionnaire.

Table 6 Statement count: What testers did not like about the BITV report

Statement	Observations	form answer	total
"this is a lot of text"	7	0	7
just plain unformatted text	6	3	7
very long	7	1	7
calls the report "ugly", not attractive	7	0	7
missing explanations to the types, does not understand "not applicable" ("nicht anwendbar")	5	2	5
Not clearly laid out	4	1	5
does not understand "general notes" („allgemeine Anmerkungen")	5	1	5
presentation is boring, not interesting	5	0	5
no colors were used to code the type of SC	3	1	4
it could not be more unimaginative	4	0	4
not very demonstrative	4	0	4
First impression is confusing	3	1	3
not much helpful structure	2	1	3
the order of the tested pages (1, 2, 3...) is not consistent	2	2	3
no visual hierarchy	3	0	3
sorted, but not clear enough	2	1	3

Statement	Observations	form answer	total
text is too small	2	1	3
no Screenshots were used in the page overview	3	1	3
it is hard to tell which SC do not pass on which page	3	0	3
some information seems hidden	3	0	3
does not know where to begin	3	0	3
misses a heading due to fast scrolling	3	0	3
trouble with orientation when scrolling	3	0	3
everything is displayed at once	1	1	2
notices Grammar error	2	0	2
hard to learn	2	0	2
very technical, not everyone could understand it	2	0	2
confused how "general notes" ("allgemeine Anmerkungen") are listed inconsistently with numbers, letters or they are unordered	2	0	2
success criteria are sorted by category and not by their order in the Guidelines.	0	1	1
recommendations and pages were separated	0	1	1
the order of SC did not feel right	1	1	1
missing priorities	1	1	1
did not understand that each page was tested the same way	1	0	1
does not understand difference between "Prüfschritt" and "Anforderung"	1	0	1
It is not visually emphasized if a SC is passed	1	0	1
Feels like list of tested pages does not give more information	1	0	1
knows due to experience what "ET" means, but not explained any farther	1	0	1
page Overview: no number how many SC are not applicable	1	0	1
not according to expectation: thought there would be more levels to passed/not passed	1	0	1
It is not fun	1	0	1
misunderstands note at beginning as something legal	1	0	1
does not understand why it is the same URL everywhere	1	0	1
expects an annex	1	0	1
not many general numbers	1	0	1
Does not understand testing standard	1	0	1
not only ideal code is displayed but also erroneous code	1	0	1

Table 7 lists observed comments and written statements in the questionnaire answering the question “Which additional features would you like to see in this report?”. There is an extra column stating whether that idea or wish was attempted in StroCards.

Many testers orally commented on wanting more color (5 out of 10), more diagrams (4 out of 10) and a visual hierarchy (4 out of 10). A possibility these were not reflected as strongly through the form could be because participants did not call these wishes “features”. The question could be rephrased in future user tests to encourage testers to include this kind of feedback. The most common answer (5 out of 10) in the questionnaire was including a way to filter or sort the success criteria by type, page or their order in the reference guidelines. This is a predominant feature of StroCards. This is also the case for the mentioned enhanced visuals. A vast majority of the recommendations by the testers (with more than one occurrence in total), were attempted in StroCards. Recommendations that were not attempted and with more than one occurrence were to include some explanations for each type (5 occurrences). Another wish worth considering is the use of screenshots (mentioned by 3 testers). Using screenshots could make it easier to understand which page is being analyzed in the report. Having this quick visual should be more tangible than the currently used page ids. This caused confusion at times since it was not easily learned which page is assigned which id. Furthermore, testers were not familiar with the website that the report was about. This might make it even harder to understand what a page is about based solely on its title.

Table 7 Statement count: What additional features testers would like to see in the report and whether it was attempted in StroCards

Statement	Observations	Form answer	Total	attempted in StroCards
more colors	5	3	6	Yes
Filtering (by type) / Sorting (by type, page or guidelines)	3	5	5	Yes
more diagrams	4	4	5	Yes
more layout-design	3	2	5	Yes
explain the types, does not understand “not applicable” (“nicht anwendbar”)	5	2	5	No
use colors to code the type of SC	3	1	4	Yes
a visual hierarchie	4	0	4	Yes
grouping pages and recommendations	2	1	3	Yes
larger text	2	1	3	Yes
maybe show information as a table	2	2	3	Yes

Statement	Observations	Form answer	Total	attempted in StroCards
(A sidebar) Table of Contents	2	2	3	Yes (except for sidebar display)
Screenshots could be used	3	2	3	No
more interaction	1	1	2	Yes
not everything should be displayed at once	1	1	2	Yes
Further Types such as “rather passed”	2	1	2	No
information: how (in)accessible is the website for certain groups of people?	1	0	1	Yes
emphasize code-snippets better	1	1	1	Yes
Page Overview: show how many SC are not applicable	1	0	1	Yes
more general numbers	1	0	1	No
Collapse/Expand	1	1	1	No
add priorities	1	1	1	No
No need to explain more to SC that are passed (because they are passed)	0	1	1	No
General statement from the Tester	1	0	1	No
reasoning for a recommended change would be good	1	0	1	No
explain “ET” (Entwicklungsbegleitender Test - Development accompanying test)	1	0	1	No
link to “general notes” („allgemeine Anmerkungen“)	0	1	1	No

Testers were able to use StroCards which included many of the recommendations and wishes as listed in Table 7. The first question testers were asked after having filled out the UEQ for StroCards was which report they understood better. All of the 10 testers responded with having preferred StroCards. Table 8 lists all the reasons for this response that were part of the answers given in the questionnaire and observed positive comments.

The most common written comment was that StroCards is easier to understand (7 of 10 participants). The second most common one was that the diagrams make the data more demonstrative (6 of 10). Half of the participants wrote about the use of colors and the filtering function making it easier to work with the report. One of the most common spoken comments (9 out of 10) was excitement about the use of colors and usually made within the first seconds of viewing StroCards. 9 out of 10 participants also valued the use of the possibilities of customization using StroCards. Further notable praise was about how fast the general status of the tested website

is conveyed (6 out of 10). The same number of participants noted the way code was formatted in the report. They appreciated the use of an eye-catching color for this purpose. There were 3 occurrences of a comment about the mappings of the success criteria to development Roles and user accessibility need. This may not seem convincing since these diagrams were supposed to be a fairly strong novelty in StroCards.

When asked precisely which diagram participants especially liked, most of them (8 out of 10) answered with the chart for each user accessibility need (see Table 9) and slightly less with the chart for each role (7 out of 10). These are fairly evidentiary results, but it should be noted that answers might be different in a scenario that is not a user test like this one. In some of the tests, participants expressed which variation of the UANs chart they preferred. There is a version with all 9 of them and another one where similar ones are combined (e.g. hard of hearing and hearing loss are combined). 3 out of 5 preferred the combined ones and 2 the complete ones. (The other 5 did not make a statement.) In this case, making it possible for the user to decide which chart to display is recognized as a good feature by 1 of those 5 participants, saying that they would switch between the charts depending on the scenario.

The chart displaying the scores was actively mentioned as a good diagram by 4 of the 10 participants. Contrarily to this, 3 of the 10 participants stated they would completely leave out this diagram (Table 10). The chart displaying success criteria for each principle had 2 mentions as an answer to this question. 3 testers would leave out the legend. These participants felt like the colors and symbols were self-explanatory enough. Furthermore, having the checkboxes which success criteria to display at the beginning of the page, acts like a legend in a sense. 6 of the participants would not leave anything out, which proves how for them StroCards does not feel too overwhelming. This was not the case for 1 tester who was overpowered by the "bar chart orgy" ("Balkenorgie") in the dashboard. Testing circumstances were not the same for them, since the dashboard was still set to showing every single possible diagram.

Testers were not explicitly asked what they did not like about StroCards in the questionnaire which could be regarded as a flaw in the test setup. However, negative comments not related to a specific chart were noted and are listed in

Table 11. The most common statement made by half of the participants was about the wording. Testers were not sure about what happens when clicking buttons and links labelled with "Dashboard" and "Report". This was somewhat reflected in the statement that StroCards was not the easiest application to learn (4 observations). The advanced settings (Screenshot 6) were slightly problematic sometimes. 4 participants missed them entirely while scrolling which suggests making them stand out more in further development. Some testers were not very confident using these

settings. 3 participants did not realize they had to reload the entire page to see the changes they made. Both of these issues could be resolved with an idea from TP1 and TP10: making each section/chart in the dashboard collapsible with its own toggle button. All written and observed answers to which new functionalities or improvements testers wanted to see are listed in Table 12.

Besides interacting with the advanced settings, the descriptions for each section were not obvious to 3 testers. There were several observed uncertainties while testers were interacting with the dashboard: e.g. not understanding the diagram for each principle (Screenshot 12), the meaning of scores (Screenshot 11 Dashboard: Scores for each Role) or varying lengths of bar charts (like in Screenshot 8). These aspects that are both general BITV facts and specific for StroCards could be explained better through brief notes. Having missed some explanations seems like a theme when looking at Table 12.

The second most common idea (3 mentions) was to include screenshots of the described pages in the report. These could be more helpful than the currently used URL and page description.

Further ideas that exceeded the scope of this project were to provide an export as PDF (1 mention) or provide an issue tracker (2 mentions). 2 testers stated they were not sure where to begin and wished for more calls to action and clear priorities. This topic was discussed with the group of experts in 6.3. and was proven to be too complex to be implemented in this prototype.

Many more suggestions were documented. If they were only stated by one tester, which is the case for many of them, considering whether to implement them in the next iteration has to be even more careful. These ideas are possibly very spontaneous and therefore not precisely thought through. Comparing a suggestion to issues in

Table 11 might lead to some evidence of its importance. This is the case for the recommendation for providing more explanations for each graph or making each chart collapsible to name two examples for this. Another one would be to pin the settings (Screenshot 5) in a sidebar in order to provide a legend at any given point when scrolling through the dashboard. Despite the fairly intuitive choice of color and symbols, two testers forgot the meaning of them and had to scroll up again to refresh their memory. Building on this idea, this element could be expanded with a table of contents with anchor links to each section. This received positive feedback on the StroCards report page. This also was a wish after testing the current BITV report. It can be concluded that providing a table of contents together with the settings in a sidebar could yield positive user test results.

3 testers did not write down any idea for further improvement in the questionnaire but 2 of them mentioned some while testing. Having testers reply in this way could mean that they did not regard their idea as noteworthy enough, maybe because their problem was resolved on its own. It is also possible that they just simply forgot their idea.

Some bugs in the code were found through these user tests. One of them was the erroneous generation of breadcrumbs. This could have induced 4 testers to answer “the navigation” upon being asked what they did not like about StroCards (

Table 11). It is possible that testers said this due to the misunderstood use of the words “dashboard” and “report”. On the other hand, adding a copy of the breadcrumbs to the bottom of the page and providing a “back to top”-button could facilitate navigation. Even though this was recommended by only one participant (Table 12) it might be worth implementing.

Table 8 Why testers preferred StroCards and their positive comments

Statement	Observations	Form answer	Total	Page
use of colors	9	5	9	general
customizable	9	3	9	general
easier to understand	6	7	8	general
diagrams (were demonstrative)	7	6	8	dashboard/page overview
filtering (makes working with it easier)	6	5	7	general
you get a good feeling for the status of the website very fast	6	1	6	general
code is highlighted in the text	6	0	6	report
color-coding	4	3	5	general
icons / symbols	5	0	5	general
linking to different parts in a page, interactive Table of Contents	5	0	5	report
more appealing	3	2	4	general
hierarchies	3	1	4	general
more motivated to work on accessibility of website	4	0	4	general
better structure	4	0	4	general
the same color coding is used throughout the application	3	0	3	general
not everything at once	2	2	3	general
easy to see who has to do what (Roles)	1	2	3	dashboard
user groups are interesting	1	2	3	dashboard
more inviting to read	3	0	3	general
so much more value is added with sorting SC by role or UAN	3	0	3	dashboard
so much more value is added with sorting SC by role or UAN	3	0	3	dashboard
proportions of the 3 cases	1	3	3	dashboard/page overview
symbols make colors accessible	0	2	2	general

Statement	Obeser- vations	Form answer	Total	Page
interaction	1	2	2	general
larger Text	1	1	2	general
it feels more positive: It does not say “no passed” everywhere	2	0	2	general
color coding and meaning of symbols was easy to learn	2	0	2	general
layout feels more modern	2	0	2	general
accessible through symbols	2	0	2	general
It is easier to compare pages	2	0	2	page overview
default settings are well chosen	2	0	2	dashboard
report for a single page	2	1	2	report
Breadcrumbs	1	1	1	general
differentiating between not applicable and pass	1	1	1	general
table view	1	1	1	report
is presents everything necessary	1	0	1	general
familiar layout	1	0	1	general
create relevant reports	1	0	1	report
appreciates note about the unusual number of SC	1	0	1	dashboard/page overview
tester imagines it to be motivating to see red bars become smaller over time	1	0	1	dashboard/page overview
reports for a single role are helpful	1	0	1	report

Table 9 Diagrams that testers liked especially or would use the most

Statement	Example Screenshot	Obeser-vations	Form answer	Total
diagrams / Dashboard		0	2	2
diagram: executive summary	Screenshot 7	2	3	3
diagram: Roles	Screenshot 8	4	6	7
diagram: user groups (UANs)		7	7	8
diagram: UANs rather than combined	Screenshot 10	1	2	2
diagram: combined UANs rather than all	Screenshot 9	3	1	3
diagram: which UAN chart depends on scenario		1	0	1
diagram: Scores	Screenshot 11	2	2	4
depends on situation		2	1	3

Table 10 Questionnaire answers and observations which charts testers would leave out


Statement	Example Screenshot	Obeser-vations	Form answer	Total
nothing		0	6	6
Scores	Screenshot 11	1	2	3
Dashboard		0	1	1
diagram: principles	Screenshot 12	1	1	2
Legend	Screenshot 9	3	0	3


Screenshot 5 Dashboard: Header and settings


Dashboard acme-test ET

[alle Berichte](#) > [acme-test ET Seitenübersicht](#) > Dashboard alle Seiten

Einstellungen

☒ Nicht anwendbare Prüfschritte anzeigen


☒ Erfüllte Prüfschritte anzeigen



☒ Nicht erfüllte Prüfschritte anzeigen



Erweiterte Einstellungen einblenden +


Bericht für alle Seiten anzeigen

Screenshot 6 advanced settings expanded, default charts are checked

Einstellungen

☒ Nicht anwendbare Prüfschritte anzeigen


☒ Erfüllte Prüfschritte anzeigen


☒ Nicht erfüllte Prüfschritte anzeigen


Erweiterte Einstellungen einblenden –

Erweiterte Einstellungen

☒ Balkendiagramm mit **Zusammenfassung** anzeigen

☒ Grafik nach **Rollen** anzeigen

☐ Grafik nach mit **Scores** nach **Rollen** anzeigen

☒ Grafik nach **Nutzergruppe** anzeigen

☐ Grafik mit **Scores** nach **Nutzergruppe** anzeigen

☐ Grafik nach **kombinierten Nutzergruppe** anzeigen
(Bsp. "kein Hörvermögen" und "eingeschränktes Hörvermögen" werden zusammengefasst)

☐ Grafik mit **Scores** nach **kombinierten Nutzergruppen** anzeigen

☐ Grafik nach **Prinzipien** anzeigen

☐ **Legende** nach jeder Überschrift zeigen

Einstellungen speichern und Seite neu laden

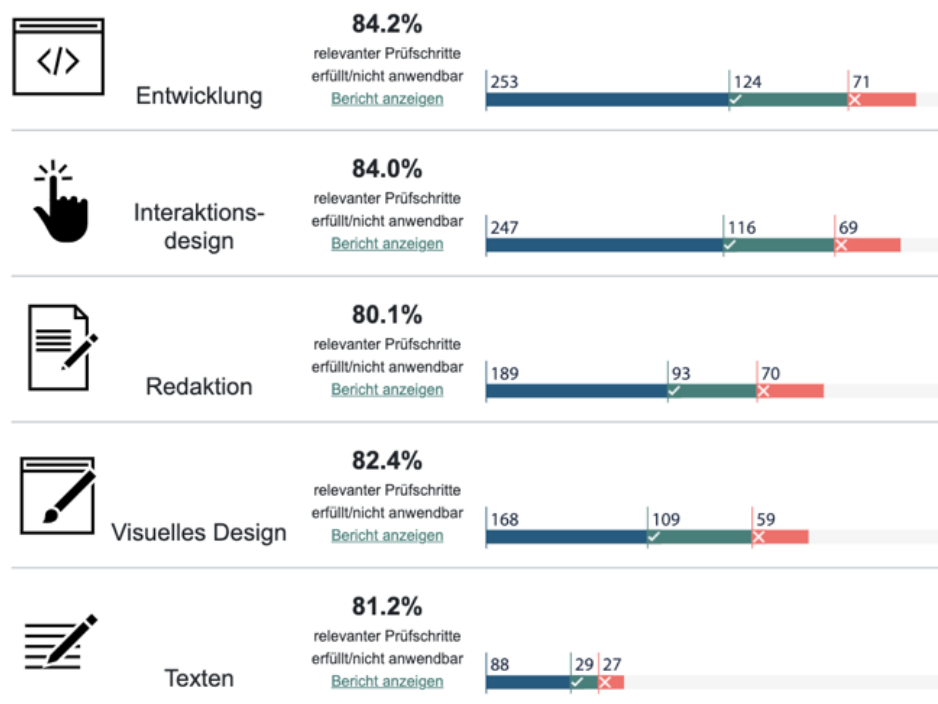
Screenshot 7 Dashboard: executive summary

Prüfschritte über alle Seiten



Screenshot 8 Dashboard: success criteria by Role

Prüfschritte nach Rolle

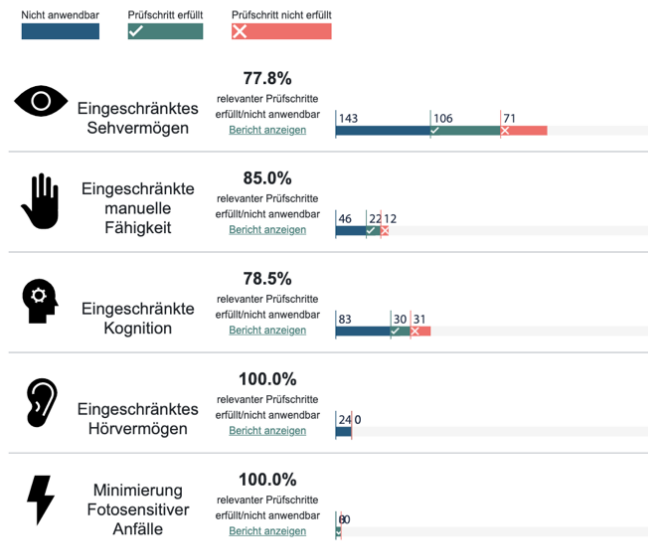


User Testing

Screenshot 9 Dashboard: combined User accessibility Needs, example for placement of legend

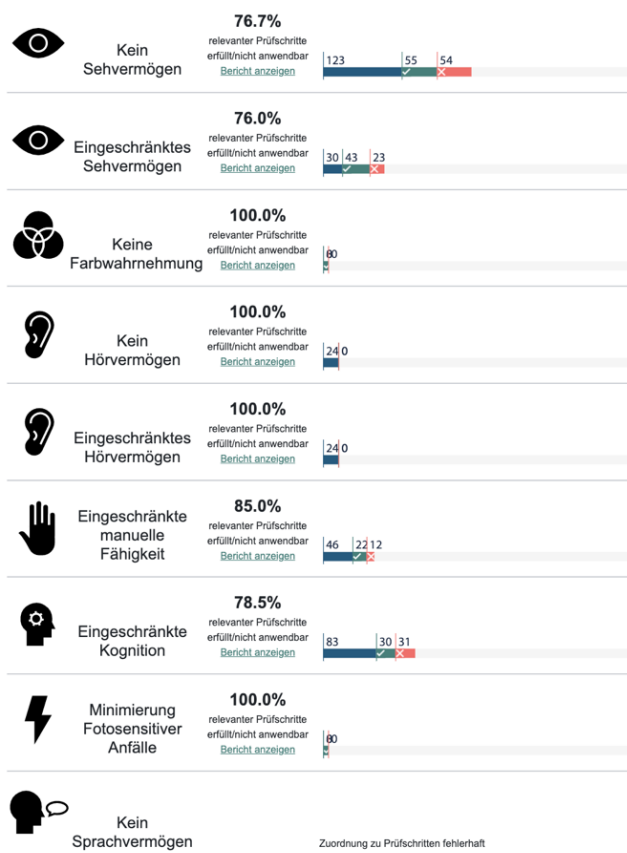
Prüfschritte nach kombinierten Nutzergruppen

Legende



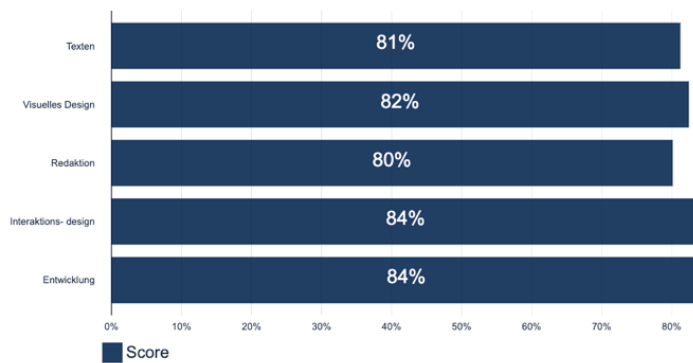
Screenshot 10 Dashboard: all User Accessibility Needs (without a legend)

Prüfschritte nach Nutzergruppen



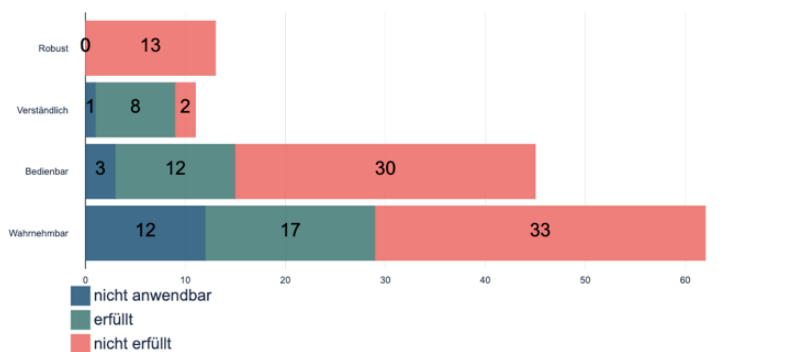
Screenshot 11 Dashboard: Scores for each Role

Scores für jede Rolle



Screenshot 12 Dashboard: success criteria by Principle

Prüfschritte nach Prinzip



Screenshot 13 Report: Display Settings and table of Contents



*Screenshot 14 Report: Table with tested pages and links to each dashboard and report***Getestete Seiten**

ID	Seitenname	URL	Link zum Dashboard	Link zum Bericht
1	Login	https://etest.acme-test.de/	Dashboard für Seite 1	Bericht für Seite 1
2	Anleitung	https://etest.acme-test.de/	Dashboard für Seite 2	Bericht für Seite 2
3	Numerisch logisches Denken	https://etest.acme-test.de/	Dashboard für Seite 3	Bericht für Seite 3
4	Deutsch	https://etest.acme-test.de/	Dashboard für Seite 4	Bericht für Seite 4
5	Rechtschreibung	https://etest.acme-test.de/	Dashboard für Seite 5	Bericht für Seite 5
6	Office Management	https://etest.acme-test.de/	Dashboard für Seite 6	Bericht für Seite 6
7	Endeseite	https://etest.acme-test.de/	Dashboard für Seite 7	Bericht für Seite 7
8	Datenschutz	https://privacy.acme-test.de/?l=de_DE	Dashboard für Seite 8	Bericht für Seite 8
			Dashboard für alle Seiten	Aktuelle Ansicht: Bericht für alle Seiten

Table 11 observed negative comments and misunderstandings

Statement	Observations	Page
wording: does not understand “Dashboard”	5	general
not the easiest to learn	4	general
missed advanced settings when scrolling	4	dashboard/page overview
advanced settings: did not expect having to reload page	3	dashboard
tester does not understand principles	3	dashboard
does not understand description of each chart in the advanced settings	3	dashboard
pretty long loading times	3	general
breadcrumbs are broken	2	general
Roles: what is the difference between “editorial” and “text”, does not understand Roles	2	dashboard
report: code bug: one SC is listed several times	2	report
overwhelming number of charts	2	dashboard
uncertain what score means	2	dashboard
report: still a lot of text	2	report
tester briefly forgets meaning of each color in the bar charts and has to scroll up again	2	dashboard/page overview
tester does not understand „left out SC“ („ausgelassene Prüfschritte“)	2	report
shorter bars in groups/Roles: Why?	1	dashboard
display categories checkboxes: expected this to be a diagram	1	dashboard/page overview

Statement	Observations	Page
Code is highlighted but acts too overpowering as an eyecatcher	1	report
Score graphs have a tooltip, but other diagrams do not display more valuable information	1	dashboard/page overview
bug: advanced settings: when there is no box checked, reverts to default	1	dashboard
some information is quite hard to reach, many clicks	1	general
did not see the effect of a change made in “advanced settings”	1	dashboard
being a report it is not the most exciting thing	1	general
tester had problems differentiating between text and interactive buttons or links	1	general
report: expects the decorative bar to be a chart	1	report
tester expects each bar to be clickable and navigate to the according report	1	dashboard/page overview
no change is obvious when switching between list and table view	1	report
report: tester did not understand the navigation purposes of the very first table	1	report
report: tables have columns of different widths, don't look uniform, seem like they present different information	1	report
does not understand shorter bars in charts for Roles	1	dashboard
overwhelmed with possibilities in “advanced settings”	1	dashboard
does not understand UANs, blames on the fact that the tested website is not their own	1	dashboard
diagram UANs: tester thinks these groups can be defined by themselves	1	dashboard
scores: same color as “not applicable”: confusing	1	dashboard
name of tester was available in BITV report: better	1	general
tester forgets what page an ID stands for	1	general
Diagram: does not see difference between the two UANs charts	1	dashboard
tester does not see settings for report display at first	1	report

Table 12 improvement ideas from testers (through questionnaire and observed)

Statement	Observations	Form answer	Total	Page
explanation for principles diagram	4	1	4	dashboard
incorporate screenshots	2	2	3	general
no farther ideas	0	3	3	general
explain why bars are shorter in Roles/group diagrams	2	1	2	dashboard
pin checkboxes (display passed/not passed/not applicable) and settings in a sidebar	2	1	2	dashboard/page overview

Statement	Observations	Form answer	Total	Page
toggle display button for each chart right next to it	2	1	2	dashboard
Issue tracker	2	0	2	general
anchor link to the “general notes” (“Allgemeine Anmerkungen”)	2	0	2	report
score bar chart: different colors (depending on score value)	2	0	2	dashboard
calls to action, criticality flags, priorities	2	2	2	general
bars: hover tooltip	1	1	1	dashboard/page overview
less overwhelming settings for first visit	0	1	1	dashboard
print/export	0	1	1	general
make order of charts on DB customizable	1	1	1	dashboard
reports are on a new page, could have a hover-over preview	1	0	1	dashboard/page overview
notes on improvement could be defecated from the rest of the text	1	0	1	report
advanced setting: emphasize the button when a change has been made	1	0	1	dashboard
reload changes in advanced settings automatically	1	0	1	dashboard
a link to the WCAG reference for the Roles	1	0	1	dashboard
report: order of categories should be different: not passed first because most important	1	0	1	report
it is not obvious that a report for a single role is a report for all pages: should be explained	1	0	1	report
if “not passed” SC are not checked, there might be a possibility of a user thinking that all SC are passed	1	0	1	general
report: show number of SC with each heading	1	0	1	report
more navigation elements at the end of a page (like “back to top” button or breadcrumbs)	1	0	1	general
a Table with contents with all possible generated pages linked, showing which ones have been viewed already	1	0	1	general
format code snippets more so they are differentiated more clearly from the rest of the text	1	0	1	report
button to reset to default settings	1	0	1	dashboard

7.2.4. Limitations

These results are extremely clear in showing the improvement of the user experience of a BITV report using StroCards. Nevertheless, some aspects have to be taken into consideration.

8 of 10 participants were recruited by the facilitator due to a personal acquaintance. Some of them had expectations regarding the project and could guess what the study was aiming for. Even though they were told to give spontaneous feedback, it is possible that they were more accommodating with their feedback than strangers would have been. Regardless, this does not take away all of the retrieved data's value since results were highly unequivocal. The other 2 participants did not give very different feedback than the others. It is to be expected that results could be similar with a different study population but not completely contrary.

Some of the unreliable data can be uncovered when comparing the participants' answers given in the forms with their spontaneous reactions. An example are the answers to the question "What aspects did you like about the BITV report.". Two participants (TP6 and TP8) answered "nothing" in written form. Contrarily, they made comments in oral form about how reliable and detailed the information seems or how helpful examples were. Possible explanations for this include that by the time they got to filling out the form they had forgotten about those rather positive aspects. Another one could be that they did not see them as positive enough to mention or possibly the negative aspects were too overpowering in their perception.

It is possible that changing the order of the tested systems could have an effect on the results. Exposing the tester to StroCards first could also lead to comments on being overwhelmed. Testers had the opportunity to learn the structure of this kind of report before using StroCards and had certain expectations. Having half of the participants test the systems in opposite sequence could reveal more reliable data.

If farther testing is conducted there should be a bigger focus on recruiting more participants that actively work with BITV reports. This study had only one participant who actively writes these reports due to their occupation. Other testers knew of the BITV test but never dealt with it in a practical scenario. Having more results from the actual target group could deliver more relevant insights.

It could be interesting to see how testers would react to StroCards if it were to display a report for their own website. They might already be aware of some possible issues or more understanding regarding the feedback. Furthermore, testers could easily visualize for themselves which tested page is being criticized in the report. Relying only on their titles and assigning each page an id seems very minimal and it was confusing for testers sometimes. Not having this mapping and seeing right away which page is concerned could make the report faster to understand. The use of screenshots has also been recommended by participants.

It is mentioned in resources offered for the UEQ, but it was also observed during the user tests that participants interpreted the presented sets of adjectives differently. E.g. "slow/fast" was

sometimes interpreted as the time it takes to work through the report or as the actual loading speed of the application. The latter was intensified due to the technical setup. The remote machine was used to its full capacity running TeamViewer, recording the session and running the local server at the same time. This also resulted in a delay of a few seconds in the voice chat leading to misunderstandings or simultaneous talking that had to be repeated to enable good communication between the facilitator and the testers. Another disadvantage of the remote test setup was not being able to hear whether a tester was clicking. This is important to understand if there are elements that look like they are interactive to users but are not in reality. In order to retrieve this kind of information, testers were asked by the facilitator whether they clicked or not. The problem with this is that a probable moment for clicking has to be noted and reacted to very fast and another problem being that this could disrupt the chain of thought of the tester.

Testers were told the study would take around 30 to 45 minutes. This timeframe was met in some of the conducted tests. It should be noted though, that several lasted a lot longer (maximum 111 minutes) and it is to be expected that the participants' focus and motivation decrease towards the end of the test. Communication was not as smooth as it could have been like for an in-person user test due to the delay. Having technical issues like disconnecting from the TeamViewer meeting (TP3 and TP9) also resulted in a longer duration of the user test.

7.3. Conclusion

The goal of this study was to answer the questions posed in chapter 7. Regarding the results presented in 7.2. the following statements can be made:

1. How easy/hard to understand is the current presentation of BITV reports for the testers?

The UEQ quality Perspicuity which describes how easy to understand a system is, performs very poorly for the BITV report and a lot better for StroCards. Comments like "This is a wall of text" (TP1) or "I don't know where to begin" (TP2) support this. Showing all the information at once was overwhelming for the testers. Presenting it in smaller portions seemed to be a good improvement.

2. Could the prototype convey the same information better – especially for non-experts?

All of the participants said that StroCards made the report easier to understand. This is supported by the fact that the scale Perspicuity performed a lot better.

Since half of the participants deal with Web Accessibility only in rare occasions, they could be considered non-experts. At the same time their judgment might not be very reliable due to their lack of experience with Web Accessibility reports.

3. Are there issues concerning comprehension and usability of the prototype?

The user tests resulted in a substantial list of desired improvements. Some of them concerned usability, others were issues regarding the content of the report. The actual content was untouched by StroCards, it simply rearranged it and generated diagrams according to its structure. As for the validation of the prototype these issues can be neglected but a few examples of usability issues and possible improvements are as follows:

- In observations, some testers had issues with navigation and mentioned this. It should be considered to redesign the graphical user interface for navigation. Since StroCards introduces wording like “Dashboard”, the content that is to be expected could be displayed in a little icon to support the learning process of these terms. Another option would be to give the user a preview or an introductory sentence of the dashboard.
- Several issues could be fixed with short explanations. E.g. shorter bars in UANs-chart or “left out” success criteria in a filtered report. Some of these explanations would also be able to fill in lack of knowledge e.g. describing the “principles” that some testers were not familiar with.
- Two testers suggested using Screenshots. These could make the report easy to understand at a glance. Instead of remembering the ID of a tested page, a screenshot could simply show the user the page.
- It was noted that the settings could be more responsive, meaning not making it necessary to reload the whole page. A suggestion was to make each graph collapsible (TP1) which would give the user the possibility to see right away which graph they are activating or deactivating. This could be easier to understand than the long list of descriptions with checkboxes that was implemented at the time of testing.
- Another suggestion was to pin a table of contents and the settings to a sidebar. This could prevent a lot of scrolling when going back if the user forgot what each color and symbol in the graphs stands for or when they want to go back to the settings.

4. Which customization settings are preferred?

In general, testers liked having the chart for each role and for each UAN. There was no clear consensus, however, e.g. if they preferred the combined User Accessibility Needs or the detailed ones that are found in the EN 301 549 In that form.

As for the report, 8 of 10 testers said they prefer the table. 2 of 10 saw themselves switching between views and using both.

Customization makes it possible to accommodate the preference of as many users as possible. This feature was appreciated by many participants.

8. Further Development

Some usability issues were found through user testing which is described in detail in 7.2. Addressing and improving these issues would mainly mean a redesign of the current navigation interface and of the settings (mainly for the dashboard). Other elements could improve the user experience, such as providing a table of contents in a constantly available sidebar and more explanations throughout the application to ensure that non-experts are fully informed as well.

Looking at the prototype from a technical standpoint, loading times could be shortened by optimizing the generation of each page. (Some testers even mentioned waiting longer than expected for a page to load). Working with the report could be made easier by integrating it with issue management services such as Jira or GitLab. This could make the report more dynamic without having to implement an issue managing system. Simply creating an “export to PDF” function could help distribute issues in a team.

Currently it is necessary to upload a report as an HTML in order to display it in StroCards. Since data extraction through HTML scraping is extremely error prone, this is simply a functionality of the prototype and is not intended to be a permanent one. A way to connect to the BITV MySQL database would also mean deciding on a common JSON format or updating StroCards to use the SQL interface. This implicates additional security issues since BITV reports are kept password protected if they are not public. It should be possible to use the same login data for the BITV reports in StroCards. Handling two different identifications for the BITV report and StroCards seems redundant.

It is possible to include more customization settings in future development. Users could define their own Roles, for example. The current mapping of Roles to success criteria is taken from the WCAG Quick reference which might be too generic for some real-life teams. The very similar scores that are calculated in StroCards make it noticeable that many success criteria are assigned to most Roles. The chart by Roles could be more precise if users could define the responsibilities of the teams that work on the website or even if they could define the skills of each team member. This way more efficient reports could be generated.

Another missing feature in StroCards is taking possible impairments of team member into consideration. Another way to filter issues could be by defining some biological requirements. E.g. having an issue that calls for providing alternative text for images should be solved by a sighted person.

Further Development

Current development is focused on providing a different presentation of BITV test results. It should be compatible with other Web Accessibility testing standards in the future that are not limited to Germany. This could enable more users of StroCards that do not rely solely on BITV.

Appendix

A1. Mapping of Success Criteria: UANs

Meaning of abbreviations:

- F: (Fischer, 2020)
- T and O: (T-Systems Multimedia Solutions GmbH, 2019)
- P: PersonaBrowser (Henka & Zimmermann, 2014)

Guideline		Without vision	Limited vision	Limited perception of color	Without hearing	Limited hearing	Limited or no motor capabilities	Limited cognition	Minimize photo-sensitive seizure triggers	no or limited vocal capability
1.1.1	Non-text Content		FTP	TP		TP	TP	T	T	
1.2.1	Audio-only and Video-only (Prerecorded)		TP	TP		FTP	FTP		T	
1.2.2	Captions (Prerecorded)					FTP	FTP		T	
1.2.3	Audio Description or Media Alternative		FTP	TP					T	
1.2.4	Captions (Live)					FTP	FTP		T	
1.2.5	Audio Description (Prerecorded)		TP	TP					T	
1.2.6	Sign Language (Prerecorded)					OP	O			
1.2.7	Extended Audio Description (Prerecorded)		OP	OP					O	
1.2.8	Media Alternative (Prerecorded)		OP	O		O	OP		O	
1.2.9	Audio-only (Live)					OP	OP		O	
1.3.1	Info and Relationships		FTP	T				TP	T	
1.3.2	Meaningful Sequence		FTP	T					T	
1.3.3	Sensory Characteristics		FTP	TP		TP	TP		T	
1.3.4	Orientation		P	FTP	P			T		
1.3.5	Identify Input Purpose		P					T	FT	
1.3.6	Identify Purpose		P	OP	P				O	

A1. Mapping of Success Criteria: UANs

Guideline		Without vision	Limited vision	Limited perception of color	Without hearing	Limited hearing	Limited or no motor capabilities	Limited cognition	Minimize photo-sensitive seizure triggers	no or limited vocal capability
1.4.1	Use of Color		TP	TP	F				T	
1.4.2	Audio Control		FTP	P	P	T	T		FTP	
1.4.3	Contrast (Minimum)		P	FTP	P				FT	
1.4.4	Resize text			FTP			P		T	
1.4.5	Images of Text			FT	P				TP	
1.4.6	Contrast (Enhanced)		P	OP	P				OP	
1.4.7	Low or No Background Audio		P	P		O	O	P	O	
1.4.8	Visual Presentation		P	O				P	O	
1.4.9	Images of Text (No Exception)		P	O				P	O	
1.4.10	Reflow		P	FTP		P		TP	FTP	
1.4.11	Image Contrast		P	FTP				P	TP	
1.4.12	Text Spacing		P	FTP		P		P	T	
1.4.13	Hover/Focus		P	FTP				TP	TP	
2.1.1	Keyboard		FTP	P		P		FTP	P	
2.1.2	No Keyboard Trap		FT					FT	P	P
2.1.3	Keyboard (No Exception)		O					O	P	P
2.1.4	Character Key Shortcuts		TP	P				FTP		
2.2.1	Timing Adjustable		TP	TP		T	T	T	FTP	
2.2.2	Pause, Stop, Hide		FTP	FTP		T	T	TP	FTP	
2.2.3	No Timing		OP	OP		O	O	OP	OP	
2.2.4	Interruptions		OP	OP				P	OP	
2.2.5	Re-authenticating		OP	OP		O	O	OP	OP	
2.2.6	Timeouts		OP	P				P	O	
2.3.1	Three Flashes or Below Threshold			P					TP	F
2.3.2	Three Flashes		P	P				P	O	F
2.3.3	Animations from Interactions		P					P	OP	
2.4.1	Bypass Blocks		FTP			P	P	FT	T	
2.4.2	Page Titled		FTP	T		P	P		FT	
2.4.3	Focus Order		FT	F				FT	TP	

A1. Mapping of Success Criteria: UANs

Guideline		Without vision	Limited vision	Limited perception of color	Without hearing	Limited hearing	Limited or no motor capabilities	Limited cognition	Minimize photo-sensitive seizure triggers	no or limited vocal capability
2.4.4	Link Purpose (In Context)		FT	T					FTP	
2.4.5	Multiple Ways		T	T				T	FTP	
2.4.6	Headings and Labels		TP	T					FTP	
2.4.7	Focus Visible		P	FP				FTP	T	
2.4.8	Location		OP	OP					O	
2.4.9	Link Purpose (Link Only)		OP	OP					O	
2.4.10	Section Headings		OP	OP					OP	
2.5.1	Pointer Gesture		P	P				FTP	T	
2.5.2	Pointer Cancellation		P	TP	P			FT	TP	
2.5.3	Lable in Name		P	P				FTP	TP	
2.5.4	Motion Actuation		P	P				FTP	P	
2.5.5	Target Size		P	OP		P	P	OP	P	
2.5.6	Concurrent Input Mechanisms		P	P				O	OP	
3.1.1	Language of Page		FTP	P				P	TP	
3.1.2	Language of Parts		FTP					P	T	
3.1.3	Unusual Words		P			O	O	P	O	
3.1.4	Abbreviations		O			O	O		O	
3.1.5	Reading Level					O	O		O	
3.1.6	Pronunciation		O			O	O		O	
3.2.1	On Focus		T	T				T	FT	
3.2.2	On Input		T	T				T	FT	
3.2.3	Consistent Navigation		T	T					FT	
3.2.4	Consistent Identification		T						FT	
3.2.5	Change on Request		O	O		O	O	O	O	
3.3.1	Error Identification		FT	T		T	T		FT	
3.3.2	Labels or Instructions		T	FT				T	FT	
3.3.3	Error Suggestion		T	T				T	FT	
3.3.4	Error Prevention (Legal, Financial, Data)		T	T					FT	

A2. Mapping of Success Criteria: Roles

Guideline		Without vision	Limited vision	Limited perception of color	Without hearing	Limited hearing	Limited or no motor capabilities	Limited cognition	Minimize photo-sensitive seizure triggers	no or limited vocal capability
3.3.5	Help					O	O		O	
3.3.6	Error Prevention (All)		O	O		O	O	O	O	
4.1.1	Parsing		F			T	T		T	
4.1.2	Name, Role, Value		F			T	T			
4.1.3	Status Messages		FT	T		T	T	T	T	
										Only mapping: EN 5.4 (Fischer)

A2. Mapping of Success Criteria: Roles

Guideline		Development	Interaction Design	Content Creation	Visual Design	Writing
1.1.1	Non-text Content	X	X	X	X	X
1.2.1	Audio-only and Video-only (Prerecorded)	X	X	X	X	
1.2.2	Captions (Prerecorded)	X	X	X	X	X
1.2.3	Audio Description or Media Alternative	X	X	X	X	X
1.2.4	Captions (Live)	X	X	X	X	
1.2.5	Audio Description (Prerecorded)	X	X	X	X	
1.2.6	Sign Language (Prerecorded)	X	X	X	X	
1.2.7	Extended Audio Description (Prerecorded)	X	X	X	X	
1.2.8	Media Alternative (Prerecorded)	X	X	X	X	
1.2.9	Audio-only (Live)	X	X	X	X	
1.3.1	Info and Relationships	X	X	X	X	X
1.3.2	Meaningful Sequence	X	X	X	X	
1.3.3	Sensory Characteristics	X	X	X	X	
1.3.4	Orientation	X	X		X	
1.3.5	Identify Input Purpose	X	X			
1.3.6	Identify Purpose	X	X	X	X	

A2. Mapping of Success Criteria: Roles

Guideline		Development	Interaction Design	Content Creation	Visual Design	Writing
1.4.1	Use of Color		X	X	X	
1.4.2	Audio Control	X	X	X		
1.4.3	Contrast (Minimum)		X	X	X	
1.4.4	Resize text	X			X	
1.4.5	Images of Text	X	X	X	X	
1.4.6	Contrast (Enhanced)		X	X	X	
1.4.7	Low or No Background Audio	X	X			
1.4.8	Visual Presentation	X	X	X	X	
1.4.9	Images of Text (No Exception)	X	X	X	X	
1.4.10	Reflow	X	X	X	X	
1.4.11	Image Contrast		X	X	X	
1.4.12	Text Spacing	X	X	X	X	
1.4.13	Hover/Focus	X	X		X	
2.1.1	Keyboard	X	X	X		
2.1.2	No Keyboard Trap	X	X			
2.1.3	Keyboard (No Exception)	X	X			
2.1.4	Character Key Shortcuts	x	X			
2.2.1	Timing Adjustable	X	X			
2.2.2	Pause, Stop, Hide	X	X	X	X	
2.2.3	No Timing	X	X			
2.2.4	Interruptions	X	X			
2.2.5	Re-authenticating		X		X	
2.2.6	Timeouts	X	X		X	
2.3.1	Three Flashes or Below Threshold			X	X	
2.3.2	Three Flashes			X	X	
2.3.3	Animations from Interactions	X	X		X	
2.4.1	Bypass Blocks	X	X	X	X	
2.4.2	Page Titled	X	X	X	X	X
2.4.3	Focus Order	X	X	X	X	
2.4.4	Link Purpose (In Context)	X	X	X		X
2.4.5	Multiple Ways	X	X		X	
2.4.6	Headings and Labels	X	X	X	X	X

A2. Mapping of Success Criteria: Roles

Guideline		Development	Interaction Design	Content Creation	Visual Design	Writing
2.4.7	Focus Visible	X	X	X	X	
2.4.8	Location	X	X			
2.4.9	Link Purpose (Link Only)	X	X	X		X
2.4.10	Section Headings	X		X	X	
2.5.1	Pointer Gesture	X	X			
2.5.2	Pointer Cancellation	X	X			
2.5.3	Label in Name	X	X	X	X	
2.5.4	Motion Actuation	X	X			
2.5.5	Target Size		X		X	
2.5.6	Concurrent Input Mechanisms	X				
3.1.1	Language of Page	X			X	
3.1.2	Language of Parts	X			X	
3.1.3	Unusual Words	X		X	X	X
3.1.4	Abbreviations	X		X	X	X
3.1.5	Reading Level			X	X	X
3.1.6	Pronunciation			X	X	
3.2.1	On Focus	X	X			
3.2.2	On Input	X	X			
3.2.3	Consistent Navigation	X	X	X	X	
3.2.4	Consistent Identification	X	X	X	X	
3.2.5	Change on Request	X	X			
3.3.1	Error Identification	X	X	X		
3.3.2	Labels or Instructions	X	X	X		X
3.3.3	Error Suggestion	X	X	X		
3.3.4	Error Prevention (Legal, Financial, Data)	X	X	X		
3.3.5	Help	X	X	X	X	
3.3.6	Error Prevention (All)	X	X	X		
4.1.1	Parsing	X				
4.1.2	Name, Role, Value	X		X		
4.1.3	Status Messages	X	X	X	X	

A3. User Test Questionnaires (German Original)

Fragebogen 1: Fragen zur Erfahrung

Wie sicher sind Sie mit dem Umgang mit Computern?

Likert Skala von 1 bis 5: „sehr sicher“ bis „gar nicht sicher“

Kommentar

Freitext

Wie erfahren sind Sie im Bereich Web Accessibility?

Likert Skala von 1 bis 5: „überhaupt keine Erfahrung“ bis „ich habe jeden Tag damit zu tun“

Haben sie schon einmal einen Web Accessibility Test selbst durchgeführt oder durchführen lassen?

Optionen (nur eine ist auswählbar):

- *selbst durchgeführt*
- *durchführen lassen*
- *noch nie*
- *sonstiges*

Kommentar zu Ihrer Erfahrung mit Web Accessibility

Freitext

Beruf / Tätigkeit:

Freitext

Fragebogen 2: zur BITV-Testdarstellung

[UEQ Fragebogen]

Fanden Sie bestimmte Aspekte bei dieser Ansicht besonders gut? Wenn ja, welche?

Freitext

Fanden Sie bestimmte Aspekte besonders verwirrend oder unverständlich? Wenn ja, welche?

Freitext

Was für Funktionen würden Sie bei so einer Ansicht gerne sehen wollen?

Freitext

Fragebogen 3: zu StroCards

[UEQ Fragebogen]

Welchen Beispielbericht konnten Sie besser verstehen? Können Sie beschreiben, was Ihnen dabei besonders geholfen hat?

Freitext

Fanden Sie bestimmte Ansichten und Diagramme bei StroCards besonders gut? Wenn ja, welche?

Freitext

Fanden Sie bestimmte Ansichten und Diagramme besonders verwirrend oder unverständlich? Wenn ja, welche?

Freitext

Welche Darstellungen würden Sie am meisten verwenden? Wieso?

Freitext

Würden Sie noch weitere Funktionen sehen wollen? Wenn ja, welche?

Freitext

Würden Sie Funktionen auslassen bzw. Gar nicht verwenden?

Freitext

Welche Darstellung der einzelnen Prüfschritte hat Ihnen mehr gefallen? Die Tabelle oder die Liste?

Freitext

A4. User Test Questionnaires (English Translation)

Questionnaire 1: about prior Experience

How experienced are you with the use of computers?

Likert scale from 1 to 5: „no experience“ to „very experienced“

Comment

Text field

How experienced are you with Web Accessibility?

Likert scale from 1 to 5: „no experience“ to „I deal with it every day“

Have you ever conducted or commissioned a Web Accessibility test?

A4. User Test Questionnaires (English Translation)

Options (only one can be selected):

- *I have conducted a BITV test myself*
- *I have commissioned one*
- *never*
- *other*

Comment on experience with Web Accessibility

Text field

Profession / Occupation:

Text field

Questionnaire 2: on the BITV presentation of results

[UEQ]

Did you find that certain aspects felt particularly good in this view? If so, which ones?

Text field

Did you find certain aspects particularly confusing or incomprehensible? If so, which ones?

Text field

What features would you like to see in a view like this?

Text field

Questionnaire 3: on StroCards

[UEQ]

Which sample report were you able to understand better? Can you describe what helped you in particular?

Text field

Did you find certain views and diagrams in StroCards particularly good? If yes, which ones?

Text field

Did you find certain views and diagrams particularly confusing or incomprehensible? If yes, which ones?

Text field

A4. User Test Questionnaires (English Translation)

Which representations would you use the most? Why?

Text field

Would you like to see any other features? If so, which ones?

Text field

Would you omit functions or not use them at all?

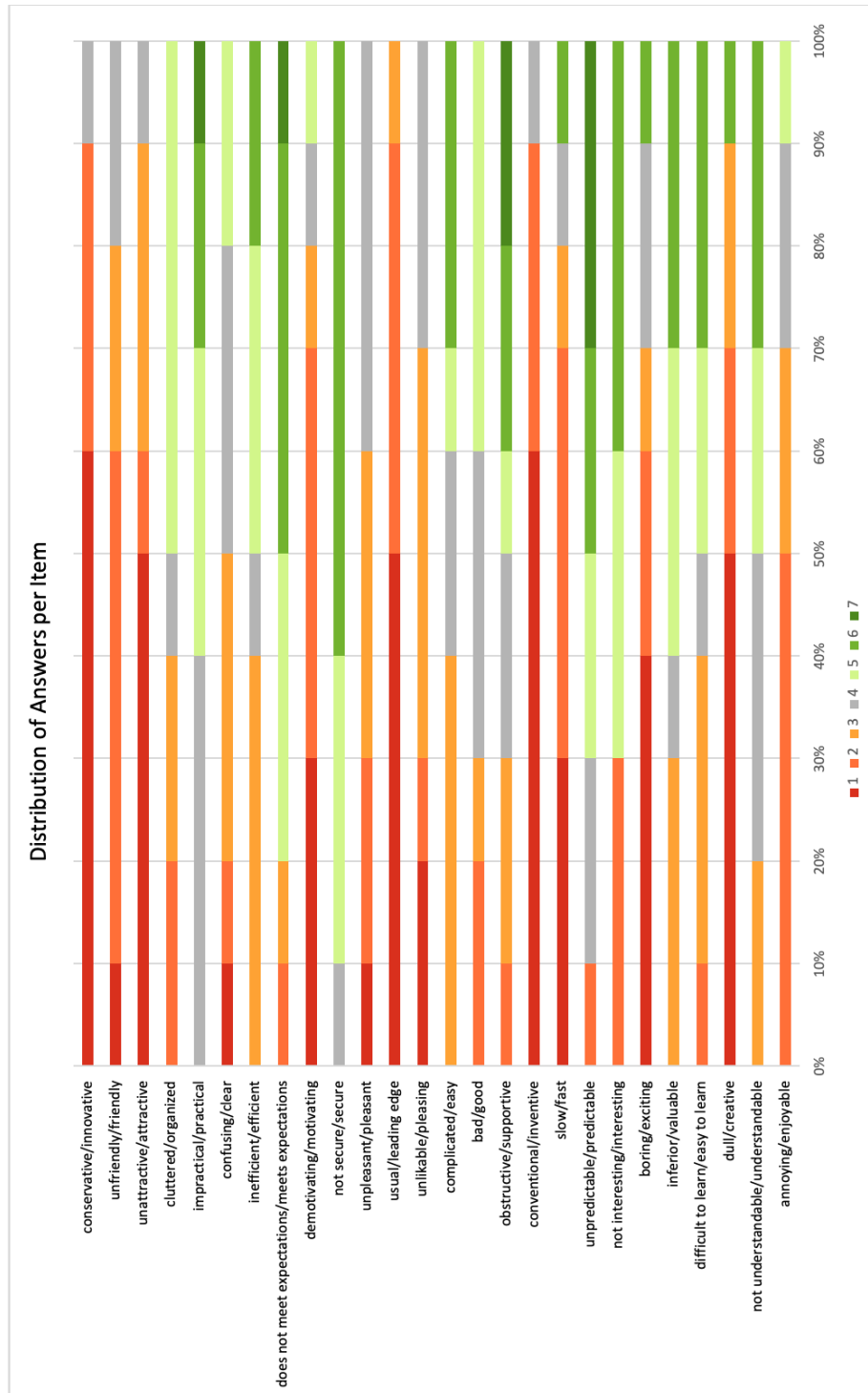
Text field

Which presentation of the individual test steps did you like more? The table or the list?

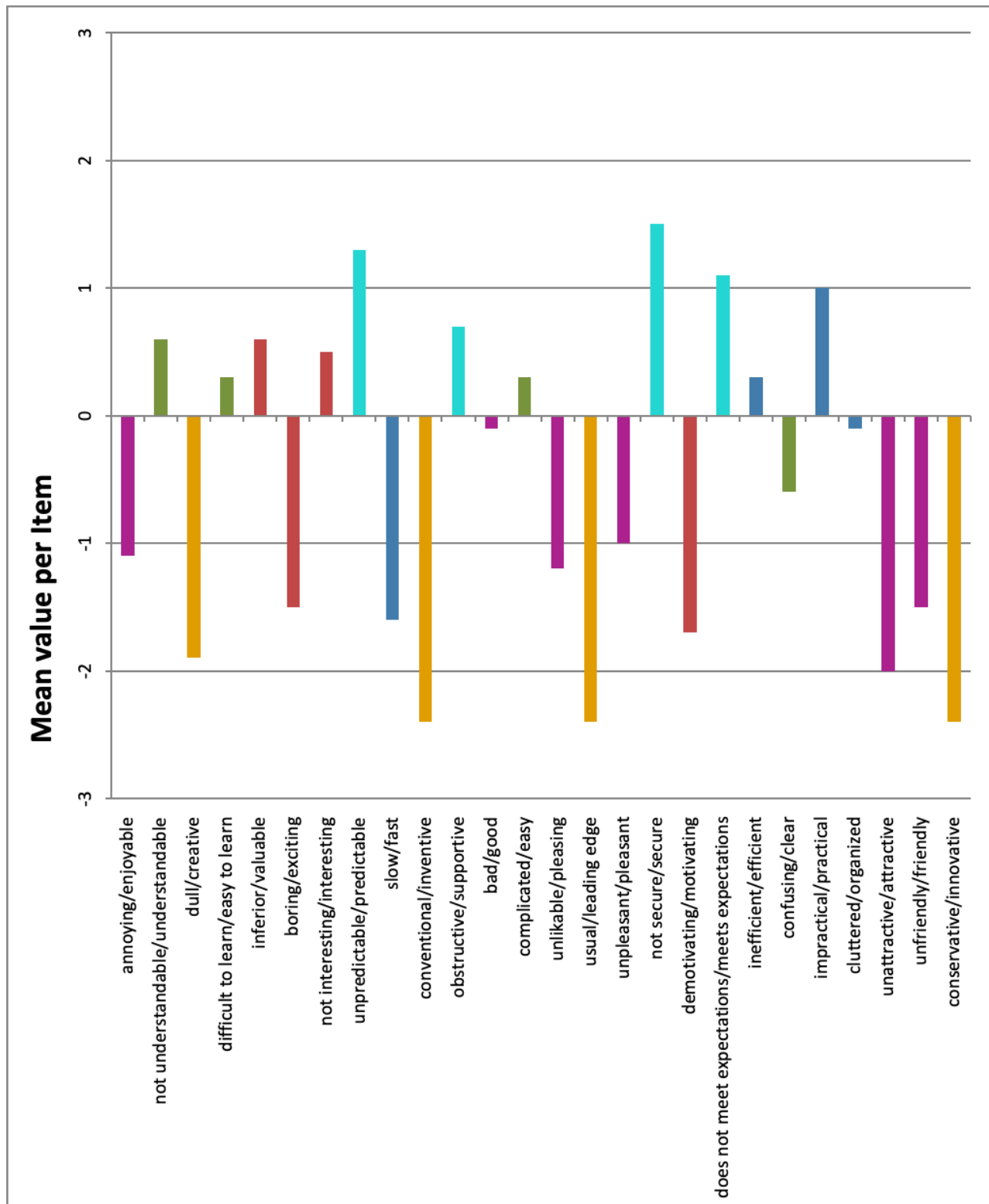
Text field

A5. UEQ evaluation BITV report

BITV: Distribution of Answers per Item



BITV: Mean Value per Item

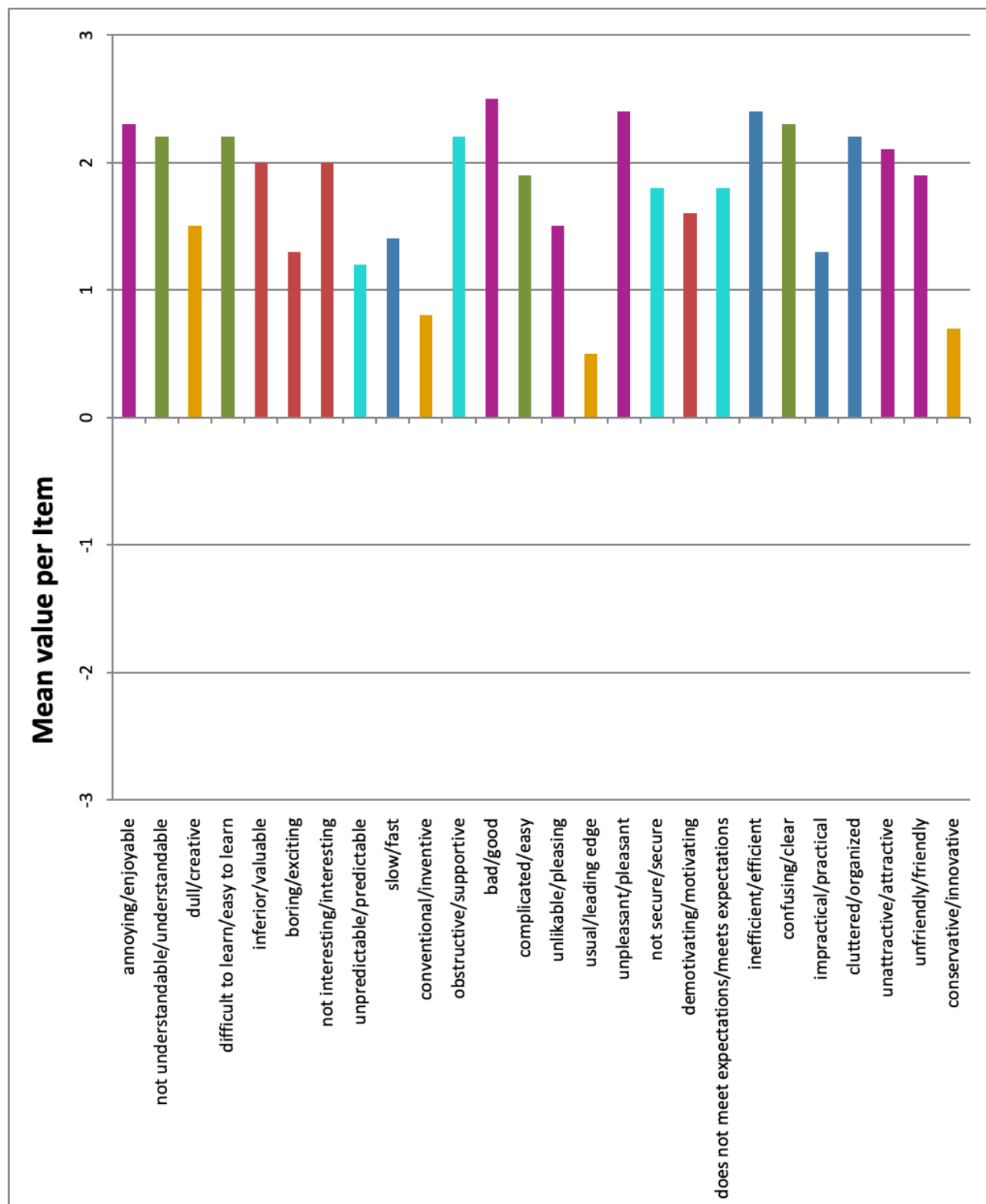


A6. UEQ evaluation StroCards

StroCards: Distribution of Answers per Item



StroCards: Mean Value per Item



A7. Qualitative Evaluation

Abbreviations:

- F: Form answer
- O: Observation

Prior Experience (German)

ID	Wie sicher sind Sie mit dem Umgang mit Computern?	Kommentar	Wie erfahren sind Sie im Bereich Web Accessibility?	Haben sie schon einmal einen Web Accessibility Test selbst durchgeführt oder durchführen lassen?	Kommentar zu Ihrer Erfahrung mit Web Accessibility	Beruf / Tätigkeit:
TP1	1		4	selbst durchgeführt	Nur im schulischen Umfeld	App Designer / Student
TP2	1		3	noch nie		Programmierer
TP3	1	Ist ein großer Teil meines Berufes.	4	selbst durchgeführt	Ich arbeite in diesem Bereich	Wissenschaftlicher Mitarbeiter am Kompetenzzentrum Digitale Barrierefreiheit
TP4	1		4	selbst durchgeführt	Vorlesungen an der HS	Student*in
TP5	1	Durch mein Studium habe ich das Gefühl mich damit gut auszukennen.	4	noch nie	Ich beschäftige mich im Rahmen meiner Masterarbeit mit dem Thema Web Accessibility	Student (Computer Science and Media)
TP6	1	Ich studiere Computer Science	2	Teilweise	Viele Richtlinien die einen etwas überfordern können	Student (Computer Science and Media - Master)
TP7	1		4	noch nie		MEdien- und KOMmunikationswirtin / Administration Web und CMS / PR & Marketing / Grafik- und Mediendesign
TP8	1		2	noch nie		Projektmanager Digitale Spielware (Spiele Begleit-Apps)
TP9	1		3	noch nie	Erfahrung mit Webentwicklung	Produktmanager / Produktentwickler
TP10	2	Als Entwickler einige Kenntnisse im Pc-Bereich.	3	selbst durchgeführt	Die Accessibility Veranstaltungen der HdM besucht, dort auch gelernt mich stark an den Guides zu orientieren und daran zu halten. Natürlich die Gesetzlichen.	Student - CSM

BITV report: negative comments

Statement	Observations	Form answers	Total	Page	TP1	TP2	TP3	TP4	TP5	TP6	TP7	TP8	TP9	TP10
wording: doesn't understand "Dashboard"	5	0	5	general		O				O		O	O	O
principles	2	2	4	dashboard		F	O				O	F		
none	0	4	4	general			F		F				F	F
Navigation	4	1	4	general				FO		O		O	O	
Scores	3	2	4	dashboard		O	O			F		FO		
not the easiest to learn	4	0	4	general		O	O			O				O
missed advanced settings when scrolling	4	0	4	dashboard/page overview				O				O	O	O
advanced settings: didn't expect having to reload page	3	0	3	dashboard	O								O	O
tester doesn't understand principles	3	0	3	dashboard						O	O	O		
doesn't understand description of each chart in the advanced settings	3	0	3	dashboard						O			O	O
pretty long loading times	3	0	3	general				O			O	O		
breadcrumbs are broken	2	0	2	general	O					O				
roles: what is the difference between "editorial" and "text", doesn't understand roles	2	0	2	dashboard	O								O	
report: code bug: one SC is listed several times	2	0	2	report		O	O							
overwhelming amount of charts	2	0	2	dashboard		O								O
uncertain what score means	2	0	2	dashboard			O					O		
report: still a lot of text	2	0	2	report	O			O						
tester briefly forgets meaning of each color in the bar charts and has to scroll up again	2	0	2	dashboard/page overview								O		O
tester doesn't understand "left out SC" ("ausgelassene Prüfschritte")	2	0	2	report						O				O
List View	0	1	1	report	F									
bars: more info on hover	0	1	1	dashboard	F									
many bar charts	0	1	1	dashboard		F								
shorter bars in groups/roles: Why?	0	1	1	dashboard							F			
display categories checkboxes: expected this to be a diagram	1	0	1	dashboard/page overview	O									
Code is highlighted but acts too overpowering as an eyecatcher	1	0	1	report	O									
Score graphs have a tooltip but other diagrams don't	1	0	1	dashboard/page overview	O									

A7. Qualitative Evaluation

display more valuable information															
bug: advanced settings: when there's no box checked, reverts to default	1	0	1	dashboard										0	
some information is quite hard to reach, many clicks	1	0	1	general		0									
didn't see the effect of a change made in "advanced settings"	1	0	1	dashboard			0								
being a report it's not the most exciting thing	1	0	1	general			0								
tester had problems differentiating between text and interactive buttons or links	1	0	1	general				0							
report: expects the decorative bar to be a chart	1	0	1	report				0							
tester expects each bar to be clickable and navigate to the according report	1	0	1	dashboard/page overview				0							
no change is obvious when switching between list and table view	1	0	1	report					0						
report: tester didn't understand the navigation purposes of the very first table	1	0	1	report						0					
report: tables have columns of different widths, don't look uniform, seem like they present different information	1	0	1	report						0					
doesn't understand shorter bars in charts for roles	1	0	1	dashboard					0						
overwhelmed with possibilities in "advanced settings"	1	0	1	dashboard						0					
doesn't understand UANs, blames on the fact that the tested website isn't their own	1	0	1	dashboard						0					
diagram UANs: tester thinks these groups can be defined by themselves	1	0	1	dashboard						0					
scores: same colour as "not applicable": confusing	1	0	1	dashboard							0				
name of tester was available in BITV report: better	1	0	1	general										0	
tester forgets what page an ID stands for	1	0	1	general										0	
Diagram: doesn't see difference between the two UANs charts	1	0	1	dashboard								0			
tester doesn't see settings for report display at first	1	0	1	report									0		

BITV Report: positive

Statement	Observations	form answer	total	TP1	TP2	TP3	TP4	TP5	TP6	TP7	TP8	TP9	TP10	Page
lots of (helpful information), very detailed	9	2	9	O	O	O		O	O	FO	O	FO	O	information
the result/general overview was at the beginning of the document	5	4	8	F	FO	O	O	FO		O		O	F	structure
recommendations to improve website are included	6	1	6	O	O				O	O		O	FO	information
the information feels correct, professional, to be taken seriously ("textbook-feeling")	6	0	6	O	O				O		O	O	O	information
Information becomes clearer with time	5	0	5		O			O	O		O		O	information
link to resources	3	1	4		O					O		O	F	information
examples are helpful	4	0	4		O						O	O	O	information
differentiation passed/not passed is well understood	4	0	4					O	O	O		O		information
The number of passed/not passed success criteria is clearly displayed	0	3	3	F		F		F						structure
"Anmerkungen"/notes at the beginning of the page are large enough: understands why URLs are the same	3	0	3	O		O						O		visuals
each SC is structured	2	0	2	O	O									structure
Nothing.	0	2	2						F		F			other
Everything is displayed at once	1	1	2				F						O	structure
minimal design	2	0	2		O								O	visuals
seems manageable due to the titles of the SC	1	0	1		O									information
predictable	1	0	1			O								information
because of the "allgemeine Anmerkung" Text isn't repeated unnecessarily	1	0	1					O						structure
seeing passed SC feels motivating	1	0	1						O					information
it's helpful to have the tester's contact data	1	0	1							O				information
having to URL is helpful to be able to test right away	1	0	1										O	information
at least the bold headings provided some structure	1	0	1										O	visuals

BITV Report: wishes

Statement (green: attempted in StroCards, red: no)	Obeser- vations	form answe r	total	attempted in StroCards	TP 1	TP 2	TP 3	TP 4	TP 5	TP 6	TP 7	TP 8	TP 9	TP1 0	area
more colors	5	3	6	1	O				FO		FO	O	O	F	visuals
Filtering (by type) / Sorting (by type, page or guidelines)	3	5	5	1		F	FO	F				FO	FO		structure/interaction
more diagrammes	4	4	5	1				O	F		FO		FO	FO	visuals
more layout-design	3	2	5	1	O				F			O	O	F	visuals
explain the types, doesn't understand "nicht anwendbar"	5	2	5	0						FO	O	O	FO	O	information
use colors to code the type of SC	3	1	4	1					F	O	O		O		visuals
a visual hierarchie	4	0	4	1	O							O	O	O	visuals
an explanation what each category means would be helpful	4	0	4	0						O		O	O	O	structure
grouping pages and recommendations	2	1	3	1	F							O	O		structure
larger text	2	1	3	1				F					O	O	visuals
maybe show information as a table	2	2	3	1					O				FO	F	structure
(A sidebar) Table of Contents	2	2	3	0.5 (not on every page)	FO					F		O			structure
Screenshots could be used	3	2	3	0					FO				FO	O	visuals
more interaction	1	1	2	1					F				O		interaction
not everything should be displayed at once	1	1	2	1				F						O	visuals
Further Types such as "rather passed"	2	1	2	0			FO			O					information
information: how (in)accessible is the website for certain groups of people?	1	0	1	1					O						information
emphasize code- snippets better	1	1	1	1										FO	visuals
Page Overview: show how many SC are not applicable	1	0	1	1			O								structure
more general numbers	1	0	1	0.5							O				structure
Collapse/Expand	1	1	1	0	FO										interaction
add priorities	1	1	1	0						FO					information
No need to explain more to SC that are passed (bc they are passed)	0	1	1	0										F	structure

A7. Qualitative Evaluation

Statement (green: attempted in StroCards, red: no)	Obeser- vations	form answe r	total	attempte d in StroCard s	TP 1	TP 2	TP 3	TP 4	TP 5	TP 6	TP 7	TP 8	TP 9	TP1 0	area
General statement from the Tester	1	0	1	0					O						information
reasoning for a recommended change would be good	1	0	1	0									O		information
explain "ET" "(Entwicklungsbegleitender Test - Development accompanying test)	1	0	1	0			O								information
more levels than passed/not passed	1	0	1	0			O								structure
link to "allgemeine Anmerkungen"	0	1	1				F								structure

Comparison: Which report was preferred?

Preferred report	TP1	TP2	TP3	TP4	TP5	TP6	TP7	TP8	TP9	TP10
BITV report										
StroCards	F	F	F	F	F	F	F	F	F	F

Why was StroCards better?

Statement	Obeservations	form answer	total	page	TP1	TP2	TP3	TP4	TP5	TP6	TP7	TP8	TP9	TP10
use of colours	9	5	9	general	FO	O	FO		FO	O	FO	O	O	FO
customizable	9	3	9	general		O	FO	O	O	O	FO	O	O	FO
easier to understand	6	7	8	general		F	FO	FO	FO	FO	FO	O		F
diagrams (were demonstrative)	7	6	8	dashboard/page overview	FO	O	F	O			FO	FO	FO	FO
filtering (makes working with it easier)	6	5	7	general	O	FO	O			FO	FO	FO	F	
you get a good feeling for the status of the website very fast	6	1	6	general			O		O	O		O	O	FO
code is highlighted in the text	6	0	6	report	O		O		O	O	O		O	
color-coding	4	3	5	general	FO		FO		O			O		F
icons / symbols	5	0	5	general	O		O			O	O			O
linking to different parts in a page, interactive Table of Contents	5	0	5	report		O	O		O	O	O			
more appealing	3	2	4	general	O	FO	O	F						
hierarchies	3	1	4	general	O		O				O	F		

A7. Qualitative Evaluation

Statement	Observations	form answer	total	page	TP1	TP2	TP3	TP4	TP5	TP6	TP7	TP8	TP9	TP10
more motivated to work on accessibility of website	4	0	4	general	O			O				O		O
better structure	4	0	4	general	O		O			O				O
the same color coding is used throughout the application	3	0	3	general	O							O		O
not everything at once	2	2	3	general	F					O	FO			
easy to see who has to do what (roles)	1	2	3	dashboard			F		F	O				
user groups are interesting	1	2	3	dashboard		O	F		F					
more inviting to read	3	0	3	general	O		O				O			
so much more value is added with sorting SC by role or UAN	3	0	3	dashboard			O		O					O
proportions of the 3 cases	1	3	3	dashboard/page overview			FO		F			F		
report: link to reference is very clear	3	0	3	report			O			O				O
symbols make colours accessible	0	2	2	general			F							F
interaction	1	2	2	general									F	FO
larger Text	1	1	2	general	O							F		
it feels more positive: It doesn't say "no passed" everywhere	2	0	2	general	O			O						
color coding and meaning of symbols was easy to learn	2	0	2	general	O									O
layout feels more modern	2	0	2	general		O				O				
accessible through symbols	2	0	2	general			O							O
it's easier to compare pages	2	0	2	page overview					O	O				
default settings are well chosen	2	0	2	dashboard			O							O
report for a single page	2	1	2	report		FO	O							
Breadcrumbs	1	1	1	general	FO									
differentiating between not applicable and pass	1	1	1	general			FO							
table view	1	1	1	report			FO							
is presents everything necessary	1	0	1	general	O									
familiar layout	1	0	1	general		O								
create relevant reports	1	0	1	report		O								
appreciates note about the unusual number of SC	1	0	1	dashboard/page overview					O					

A7. Qualitative Evaluation

Statement	Obeservations	form answer	total	page	TP1	TP2	TP3	TP4	TP5	TP6	TP7	TP8	TP9	TP10
tester imagines it to be motivating to see red bars become smaller over time	1	0	1	dashboard/page overview										O
reports for a single role are helpful	1	0	1	report		O								

StroCards: Preferred and probably most viewed Diagrams/Pages

statement	Obeservations	form answer	total	TP1	TP2	TP3	TP4	TP5	TP6	TP7	TP8	TP9	TP10
diagrams / Dashboard	0	2	2	F			F						
diagram: executive summary	2	3	3					FO			F		FO
diagram: roles	4	6	7			FO	F	FO	O		F	F	FO
diagram: user groups	7	7	8	FO		FO	F	FO	FO		FO	O	FO
diagram: UANs rather than combined	1	2	2								F		FO
diagram: combined UANs rather than all	3	1	3	O		FO	O						
diagram: which UAN chart depends on scenario	1	0	1					O					
diagram: Scores	2	2	4				O			F		F	O
depends on situation	2	1	3					O	O	F			

StroCards: Wishes, Ideas and Recommendations

Statement	Obeservations	form answer	total	page	TP1	TP2	TP3	TP4	TP5	TP6	TP7	TP8	TP9	TP10
explanation for principles diagram	4	1	4	dashboard		FO				O	O			O
incorporate screenshots	2	2	3	general					FO		F			O
no farther ideas	0	3	3	general		F		F				F		
explain why bars are shorter in roles/group diagrams	2	1	2	dashboard						O	FO			
pin checkboxes (display passed/not passed/not applicable) and settings in a sidebar	2	1	2	dashboard/page overview									O	FO
toggle display button for each chart right next to it	2	1	2	dashboard	O									FO
Issue tracker	2	0	2	general	O	O								

A7. Qualitative Evaluation

Statement	Observations	form answer	total	page	TP1	TP2	TP3	TP4	TP5	TP6	TP7	TP8	TP9	TP10
anchor link to the "general notes" ("Allgemeine Anmerkungen")	2	0	2	report			O						O	
score bar chart: different colors (depending on score value)	2	0	2	dashboard					O			O		
calls to action, criticality flags, priorities	2	2	2	general	FO					FO				
bars: hover tooltip	1	1	1	dashboard/page overview	FO									
less overwhelming settings for first visit	0	1	1	dashboard		F								
print/export	0	1	1	general							F			
make order of charts on DB customizable	1	1	1	dashboard										FO
reports are on a new page, could have a hover-over preview	1	0	1	dashboard/page overview	O									
notes on improvement could be differentiated from the rest of the text	1	0	1	report	O									
advanced setting: emphasize the button when a change has been made	1	0	1	dashboard	O									
reload changes in advanced settings automatically	1	0	1	dashboard									O	
a link to the WCAG reference for the roles	1	0	1	dashboard	O									
report: order of categories should be different: not passed first because most important	1	0	1	report		O								
it is not obvious that a report for a single role is a report for all pages: should be explained	1	0	1	report		O								
if "not passed" SC are not checked, there might be a possibility of a user thinking that all SC are passed	1	0	1	general					O					
report: show number of SC with each heading	1	0	1	report									O	
more navigation elements at the end of a page (like "back to top" button or breadcrumbs)	1	0	1	general									O	
a Table with contents with all possible generated pages linked, showing which ones have been viewed already	1	0	1	general									O	
format code snippets more so they are differentiated more clearly from the rest of the text	1	0	1	report										O

A7. Qualitative Evaluation

Statement	Observations	form answer	total	page	TP1	TP2	TP3	TP4	TP5	TP6	TP7	TP8	TP9	TP10
button to reset to default settings	1	0	1	dashboard										0

StroCards: Functions and Views to leave out and negatively perceived aspects

statement	TP1	TP2	TP3	TP4	TP5	TP6	TP7	TP8	TP9	TP10	Observations	form answer	total	Page
nothing	F			F			F	F	F	F	0	6	6	
Scores		O			F	F					1	2	3	
Dashboard		F									0	1	1	
diagram: principles		O	F								1	1	2	
Legend			O	O		O					3	0	3	
List View	F										0	1	1	report
bars: more info on hover	F										0	1	1	dashboard
many bar charts		F									0	1	1	dashboard
diagram: principles		FO	FO				O	F			3	3	4	dashboard
none			F		F				F	F	0	4	4	general
Navigation				FO		O		O	O		4	1	4	general
Scores		O	O			F		FO			3	2	4	dashboard

StroCards Report: List or Table?

Statement	TP1	TP2	TP3	TP4	TP5	TP6	TP7	TP8	TP9	TP10	Observations	form answer	total
Table	FO	FO	FO	FO	FO	FO			FO	FO	8	8	8
Table: better to read, lines help to separate			FO	O	O		O	F	O	FO	6	3	7
both							FO	FO			2	2	2
List could be better on mobile							F				0	1	1
List feels like more								FO			1	1	1
List											0	0	0
Table: design fits better		O									1	0	1
Table: you can see in a single column which SC are concerned			O								1	0	1
Table: could be problematic with extremely long comments			O								1	0	1
change of settings wasn't visible right away						O			O	O	3	0	3

A7. Qualitative Evaluation

Statement	TP1	TP2	TP3	TP4	TP5	TP6	TP7	TP8	TP9	TP10	Obeservations	form answer	total
Table: feels longer							0				1	0	1

Bibliography

- BIK BITV-Test. (n.d.). *BIK BITV-Test / Abschließender BITV-Test*. Retrieved June 12, 2020, from https://www.bitvtest.de/bitv_test/einfuehrung/abschliessend.html
- BIK BITV-Test. (2021, January 8). *Über den BITV/WCAG-Test*. <https://testen.bitv-test.de/ueber-den-bitv-test.php>
- Bounford, T., & Campbell, A. (2001). *Digitale Diagramme: [Info-Grafiken professionell gestalten]* (p. 191 Seiten). Stiebner.
- Cooper, A., Reimann, R., & Cronin, D. (2012). *About Face 3: The Essentials of Interaction Design*. John Wiley & Sons.
- Fischer, D. (2020, April 29). *Vereinfachte Überwachung*. DIAS GmbH, virtual. 29.04.2020
- Henka, A., & Zimmermann, G. (2014). Persona Based Accessibility Testing. In C. Stephanidis (Ed.), *HCI International 2014—Posters' Extended Abstracts* (pp. 226–231). Springer International Publishing. http://link.springer.com/chapter/10.1007/978-3-319-07854-0_40
- Hudson, R. (2011, November 22). *Accessibility Barrier Scores / Web Usability*. <https://usability.com.au/2011/11/accessibility-barrier-scores-2/>
- Huntsman Security. (n.d.). *Huntsman Security: Defence-Grade SIEM Cyber Security Platform*. Huntsman. Retrieved January 8, 2021, from <https://www.huntsmansecurity.com/>
- IBM Accessibility. (2020, December 18). *IBM Accessibility: Design research*. www.ibm.com/able/toolkit/design/research
- IBM Accessibility Checklist 7.1*. (2019, June 23). https://www.ibm.com/able/guidelines/ci162/accessibility_checklist.html
- Jordan, P. W., Thomas, B., McClelland, I. L., & Weerdmeester, B. (1996). *Usability Evaluation In Industry*. CRC Press.
- Kaplan, R. S., & Norton, D. P. (1992, January 1). The Balanced Scorecard—Measures that Drive Performance. *Harvard Business Review*, January-February 1992. <https://hbr.org/1992/01/the-balanced-scorecard-measures-that-drive-performance-2>
- Kaplan, R. S., & Norton, D. P. (1993, September 1). Putting the Balanced Scorecard to Work. *Harvard Business Review*. <https://hbr.org/1993/09/putting-the-balanced-scorecard-to-work>
- Kaplan, R. S., & Norton, D. P. (1996a). *The balanced scorecard: Translating strategy into action*. Boston, Mass.: Harvard Business School Press. <http://archive.org/details/balancedscorecar00kapl>

A7. Qualitative Evaluation

- Kaplan, R. S., & Norton, D. P. (1996b). Linking the Balanced Scorecard to Strategy. *California Management Review*, 39(1), 53–79. <https://doi.org/10.2307/41165876>
- Laugwitz, B., Held, T., & Schrepp, M. (2008). *Construction and Evaluation of a User Experience Questionnaire*. 5298, 63–76. https://doi.org/10.1007/978-3-540-89350-9_6
- Les directives AnySurfer—AnySurfer*. (2020, July 10). <https://www.anysurfer.be/fr/en-pratique/directives>
- Lopes, R., Votis, K., Carriço, L., Tzovaras, D., & Likiothanassis, S. (2009). Towards the universal semantic assessment of accessibility. *Proceedings of the 2009 ACM Symposium on Applied Computing*, 147–151. <https://doi.org/10.1145/1529282.1529311>
- Milne, S., Dickinson, A., Carmichael, A., Sloan, D., Eisma, R., & Gregor, P. (2005). Are guidelines enough? An introduction to designing Web sites accessible to older people. *IBM Systems Journal*, 44(3), 557–571. <https://doi.org/10.1147/sj.443.0557>
- Nielsen, J. (1994, April 24). *10 Usability Heuristics for User Interface Design*. Nielsen Norman Group. <https://www.nngroup.com/articles/ten-usability-heuristics/>
- Nielsen, J. (1996, January 1). *Focus Groups in UX Research: Article by Jakob Nielsen*. Nielsen Norman Group. <https://www.nngroup.com/articles/focus-groups/>
- Nielsen, J. (2004, February 29). *Risks of Quantitative Studies*. Nielsen Norman Group. <https://www.nngroup.com/articles/risks-of-quantitative-studies/>
- Prevos, P. (2015). *Visualising water quality: A graphical index for drinking water system performance*.
- Ripley, D. L., & Politzer, T. (2010). Vision Disturbance after TBI. *NeuroRehabilitation*, 27(3), 215–216. <https://doi.org/10.3233/NRE-2010-0599>
- Roles*. (2018, May 22). A11yCLE. <https://a11ycle.com/digital-accessibility/roles/>
- Ryan, S. T. (2020). *Get in Gear: The Seven Gears that Drive Strategy to Results*. CRC Press.
- Schneiderman, A. M. (2006, August 13). *History of the First Balanced Scorecard*. http://www.schneiderman.com/Concepts/The_First_Balanced_Scorecard/BSC_INTRO_AND_CONTENTS.htm
- SecurityScorecard. (n.d.). *Security Ratings & Cybersecurity Risk Management / SecurityScorecard*. Retrieved June 18, 2020, from <https://securityscorecard.com/>
- Swallow, D. M., Petrie, H., & Power, C. D. (2016, August 21). *Understanding and supporting web developers: : Design and evaluation of a web accessibility information resource (WebAIR)* [Proceedings Paper]. Universal Design 2016: Learning from the Past, Designing for the Future (Proceedings of the 3rd International Conference on Universal Design, UD2016); IOS Press. https://doi.org/10.1/Swallow_WebAIR_UD2016.pdf

A7. Qualitative Evaluation

- Swallow, D. M., Power, C., Petrie, H., Bramwell-Dicks, A., Buykx, L., Velasco, C. A., Parr, A., & Connor, J. O. (2014). Speaking the Language of Web Developers: Evaluation of a Web Accessibility Information Resource (WebAIR). In K. Miesenberger, D. Fels, D. Archambault, P. Peñáz, & W. Zagler (Eds.), *Computers Helping People with Special Needs* (pp. 348–355). Springer International Publishing. https://doi.org/10.1007/978-3-319-08596-8_54
- The British Standards Institution. (2020). *Accessibility requirements for ICT products and services*: BSI British Standards. <https://doi.org/10.3403/30401206>
- T-Systems Multimedia Solutions GmbH. (2019). *Beispielbericht Barrierefreiheit*.
- Tufte, E. R. (2007). *The visual display of quantitative information* (2. ed., 5. print., p. 197 Seiten). Graphics Press.
- University of Cambridge. (n.d.). *Exclusion Calculator Lite v2.1*. Retrieved June 12, 2020, from <http://calc.inclusivedesign toolkit.com/Vigetlabs/interactive-wcag>
- Vigetlabs/interactive-wcag. (2020, December 30). GitHub. <https://github.com/vigetlabs/interactive-wcag>
- W3C. (n.d.). *How to Meet WCAG (Quick Reference)*. Retrieved July 10, 2020, from https://www.w3.org/WAI/WCAG21/quickref/?versions=2.0¤tsidebar=%23col_customize
- W3C. (2018, June 5). *Web Content Accessibility Guidelines (WCAG) 2.1*. <https://www.w3.org/TR/WCAG21/>
- W3C. (2020, September 25). *Results of Questionnaire Accessibility Icon Set 3—Review (Approval?)—Web-Based Straw-poll and Balloting System*. <https://www.w3.org/2002/09/wbs/1/wai-icons-3/results#x2speech>
- W3C WAI. (2018, April 5). *Web Content Accessibility Guidelines (WCAG) 2.1*. <https://www.w3.org/TR/WCAG21/>
- W3C WAI. (2020, November 24). *Web Accessibility Perspectives: Explore the Impact and Benefits for Everyone*. Web Accessibility Initiative (WAI). <https://www.w3.org/WAI/perspective-videos/>
- W3C WAI, White, K., Abou Zahra, S., & Lawton Henry, S. (2020, December 1). *Writing for Web Accessibility – Tips for Getting Started*. Web Accessibility Initiative (WAI). <https://www.w3.org/WAI/tips/writing/>
- W3C Web Accessibility Initiative (WAI). (2020, December 24). *Get resources for....* Web Accessibility Initiative (WAI). <https://www.w3.org/WAI/roles/>

A7. Qualitative Evaluation

- W3C Web Accessibility Initiative (WAI), Abou Zahra, S., & Brewer, J. (2017, May 15). *Diverse Abilities and Barriers*. Web Accessibility Initiative (WAI). <https://www.w3.org/WAI/people-use-web/abilities-barriers/>
- W3C Web Accessibility Initiative (WAI), & Lawton Henry, S. (2019, September 20). *Design and Develop Overview*. Web Accessibility Initiative (WAI). <https://www.w3.org/WAI/design-develop/>
- W3C World Wide Web Consortium. (2018). *Web Content Accessibility Guidelines (WCAG) 2.1 Recommendation 05*. <https://www.w3.org/TR/WCAG21/#background-on-wcag-2>
- WAI-Engage: Web Accessibility Community Group. (2012, May 10). *Accessibility Responsibility Breakdown—WAI-Engage: Web Accessibility Community Group*. https://www.w3.org/community/wai-engage/wiki/Accessibility_Responsibility_Breakdown
- WHO. (2010). *Visual Impairment and Blindness 2010*. https://www.who.int/blindness/data_maps/VIFACTSHEETGLODAT2010full.pdf?ua=1

List of Figures

Figure 1 distribution of skill levels of users (Cooper et al., 2012, p. 42)	15
Figure 2 "The Balanced Scorecard Links Performance Measures" as depicted in the first article by Kaplan and Norton (Kaplan & Norton, 1992)	18
Figure 3 example scorecard : industry overview (https://csipbl.com/software/scorecard-examples/industry-overview/)	20
Figure 4 example scorecard for employee performance (https://www.datapine.com/blog/kpi-scorecard-examples-templates-to-track-performance/)	20
Figure 5 example scorecard: strategy map (https://balancedscorecard.org/bsc-basics/what-is-a-strategy-map/)	21
Figure 6 Essential 8 Security Scorecard (https://www.huntsmansecurity.com/products/essential-8-security-scorecard/#:~:text=The%20Essential%20Scorecard%20is,mitigate%2085%25%20of%20cyber%20threats)	22
Figure 7 possible variations of the exemplary vision icon to differentiate between no vision and limited vision	44
Figure 8 first pie chart by the four principles created with pygal	46
Figure 9 bar chart created with pygal showing success criteria by role	46
Figure 10 a stacked bar chart created with pygal	47
Figure 11 development of the custom bar chart design including icons	47
Figure 12 bar chart with scores only, created with pygal	48
Figure 13 Page overview initial design	51
Figure 14 possible design for the data for the role of writer in the dashboard with a line chart instead of a bar chart	52
Figure 15 possible design for the data for impairments related to vision in the dashboard with two line charts	53
Figure 16 two variations for a chart by principle. Left: vertical stacked bar chart with a tool tip. Right: pie chart	53
Figure 17 horizontal bar chart with separate bars for passed/failed success criteria and a separate bar for the sum of success criteria per role	54
Figure 18 horizontal bar chart for scores for each role	54

A7. Qualitative Evaluation

Figure 19 example of a tool tip in a custom bar chart	54
Figure 20 design for a legend	54
Figure 21 summary with the highest and lowest scores for the UAN dimension	55
Figure 22 chart presenting success criteria by level of conformance. The typical order (A, AA, AAA) is changed and sorted by value. Keeping the natural order would be in accordance with Nielsen's Usability Heuristic #2 (Nielsen, 1994)	55
Figure 23 example for a chart using combined UAN	55
Figure 24 example of generating icon variations in order to differentiate between limited vision and no vision	56
Figure 25 two variations of a chart showing success criteria by UAN. Left: passed and not applicable success criteria are treated separately. Right: not applicable success criteria are summed with the passed ones	56
Figure 26 Distribution of Answers: How experienced are you with the use of computers?	68
Figure 27 Distribution of Answers: How experienced are you with Web Accessibility?	68
Figure 28 Distribution of Answers: Have you ever conducted or commissioned a Web Accessibility test?	68
Figure 29 UEQ Results BITV report: Distribution of Answers per Item	72
Figure 30 UEQ Results StroCards: Distribution of Answers per Item	72
Figure 31 UEQ Results BITV report: Mean value per Item	73
Figure 32 UEQ Results StroCards: Mean value per Item	73
Figure 33 UEQ Results BITV report: Qualities	74
Figure 34 UEQ Results StroCards: Qualities	74
Figure 35 UEQ Results BITV report: Scales	74
Figure 36 UEQ Results StroCards: Scales	74
Figure 37 UEQ Results BITV report: Benchmark Comparison	74
Figure 38 UEQ Results StroCards: Benchmark Comparison	74

List of Tables

Table 1 UANs listed in EN 301 549 compared to the user groups used by T-Systems	29
Table 2 responsibilities defined by WAI-Engage and how A11YCLE grouped them further.....	34
Table 3 comparison of icons used in different resources for the presented User Accessibility Needs and the icons developed for StroCards	43
Table 4 icon for each role	45
Table 5 Statement count: what did testers like about the BITV report?	76
Table 6 Statement count: What testers did not like about the BITV report	77
Table 7 Statement count: What additional features testers would like to see in the report and whether it was attempted in StroCards.....	79
Table 8 Why testers preferred StroCards and their positive comments	83
Table 9 Diagrams that testers liked especially or would use the most.....	85
Table 10 Questionnaire answers and observations which charts testers would leave out	85
Table 11 observed negative comments and misunderstandings	90
Table 12 improvement ideas from testers (through questionnaire and observed).....	91

List of Screenshots

Screenshot 1 Shortened example of a BITV report with all success criteria passed. https://testen.bitv-test.de/index.php?a=ti&sid=3088	12
Screenshot 2 Filter options in the WCAG Quick Reference: the selectable tags change with the settings of the checkboxes	35
Screenshot 3 an excerpt from a StroCards report showing some success criteria in list view	49
Screenshot 4 an excerpt from a StroCards report showing some success criteria in table view	49
Screenshot 5 Dashboard: Header and settings	86
Screenshot 6 advanced settings expanded, default charts are checked	86
Screenshot 7 Dashboard: executive summary	87
Screenshot 8 Dashboard: success criteria by Role.....	87
Screenshot 9 Dashboard: combined User accessibility Needs, example for placement of legend...	88
Screenshot 10 Dashboard: all User Accessibility Needs (without a legend).....	88
Screenshot 11 Dashboard: Scores for each Role.....	89
Screenshot 12 Dashboard: success criteria by Principle.....	89
Screenshot 13 Report: Display Settings and table of Contents	89
Screenshot 14 Report: Table with tested pages and links to each dashboard and report.....	90

Acronyms

BITV. *Barrierefreie-Informationstechnik-Verordnung (Accessible Information Technology Regulation)*

HTML. Hypertext Markup Language

IAAP. *International Association of Accessibility Professionals*

JSON. JavaScript Object Notation

SQL. *Structured Query Language*

SUS. System Usability Scale

UAN. *User Accessibility Need*

UEQ. User Experience Questionnaire

WAI. *Web Accessibility Initiative*

WCAG. *Web Content Accessibility Guidelines*

WHO. *World Health Organization*