Silver Measurements and their Rubrics

September 2019

This document examines the various proposals of measurements for Silver and how to set up a rubric (or classification) of how the existing WCAG success criteria and the proposed test guidance for new Silver guidance could be scored. The purpose is to do feasibility testing of what measures in a conformance ranking best perform to give results that match the SIlver conformance requirements and goals.

From [Proposal to Define Scoring Parameters (June 2019)](#) and the accompanying [Scoring System Proposal](#) spreadsheet

Impact
- Don't rank user needs
- Spirit of Accessibility - this could fit better in another area than the measure of the point system. These could be Methods that get points for an organization. Others thought it should at least wait until we have more measurement numbers.
    - Consider usability as core to accessibility for PWD;
    - Report their mistakes transparently and ask for help from PWD to improve;
    - Use complaints and recruitment of PWD to change their products and innovate;
    - Evaluate how their function may increase the exclusion or exploitation of PWD, and work to solve that;
    - Look to create better impact for PWD and increase the diversity of disabilities they reach.
- Effectiveness (this is by Technique instead of by Success Criterion, so it won't fit the spreadsheet. The process of measuring techniques is troubling, because of the needs of retrofitting from WCAG. If there are bad Techniques, we should get rid of them or re-write them.)
    - Effectiveness is how well this method or test meets all of the needs of the different user groups. This refers to the Effectiveness of the Technique or Method, not how well the organization implemented it.
- Customization - This should be discussed with AOM and CSS at TPAC,

Reach[1] and Process Inclusion of PwD
- Usability Testing
- Meaningful Involvement

Measures of investment in accessibility:
- Ease of Implementation (how actionable it is)

---

[1] Reach: the diversity of people who are included, a measure of reaching a wider range

- ○ Earliness of Action (more points for prompt start)
- ○ Time Required (more points for taking on big tasks)
- ○ Degree of Specialization Required (more points for harder tasks)
- ○ Easy Checks - Bonus Points for Degree of List Completion
- Difficulty or Complexity of Testing.
  - ○ Auto-Testing vs Manual Testing
  - ○ Update Frequency

From [Scoring and Reporting Proposal - JF](#)

- Effort (to create/to test)
- Impact on users
  - ○ For each user group that is specifically impacted by the requirement = 1 multiplyer. (For example, Headings benefit both non-sighted users and users with cognition issues, thus multiplyer there would be X2. For futher discussion. This also presumes 4 basic groups: visual, auditory, mobility and cognition. We may also want to consider other scenarios, such as input modes - voice input for example)
- Overall Severity of Existing SC
  - ○ This starts with the assumption that not all SC are equal in severity; for example "Language" of page is far less severe than providing captions on videos, or ensuring every target on the page is keyboard accessible.
    I'm thinking here that we assign specific point values to each existing SC, which are then calculated with the multiplyers. This "value assignment" will require some discussion and thinking, and is (I will suggest) a critical part of our moving existing SC into the new format.
- Prevelance and Quantity/Quality
  - ○ Some 'failures' can be of one item or requirement failing mulitple times, or a percentage of the time. Examples include number of images on a page, or number of duplicate ID values, etc. This will actually be a subtractive multiplyer: for example if 95% of the images have a good alt text, multiply score by .95, or for each duplicate ID value = minus 10%. More discussion required.)
- Methods Ranking
  - ○ Another aspect to consider is when there are multiple methods to achieve the goal - do we want to apply a ranking that rewards better methods over less better methods? For example, alt text can be supplied using aria-label, aria-labeledby, alt, caption - with using @alt being the "best" method, followed by either aria method, and then trailing last is caption, which meets the functional need [sorta], but is the less preferred method. Again, for discussion, but for illustration, best method = 10, secondary method = 7, and subsequent methods = 5)