# WCAG 3.0 Comments

Tero Pesonen and Kimmo Sääskilahti

## The proposal fundamentally alters the way in which WCAG is tested

According to the WCAG 3.0 draft, one has to test for a greater number of fine-grained, atomic outcomes, each of which will have its unique assessment formula (a method of determining the degree to which the outcome passes). Additionally, every one of these outcomes needs to be heeded when determining the score for a criterion, as there are no longer conformance levels (A, AA, AAA) that limit the testing scope. These factors pose a number of issues that fundamentally change the way in which WCAG conformance is tested. Our fear is that the new testing approach renders effective accessibility testing infeasible, and will leave service developers with only poor quality testing data produced mainly semi-automatically instead of manually observing and analyzing the service – in rank contrast to the intent of the new draft.

### Testing is difficult to carry out efficiently

In WCAG 3.0, one can no longer simply report on issues that fail WCAG. Instead, one has to report the degree to which all outcomes, including those part of successful criteria, are met. This is because the test result is based on an overall score of all criteria (with some caveats), not on finding out which criteria pass or fail or which are relevant in the testing context. Given that there are also no longer different levels of conformance (A, AA, AAA), one must score each criteria and all its outcomes for every tested process. What is more, the scoring formulae for these outcomes vary. Given, then, the vast number of the outcomes, these formulae are perforce very difficult to memorize reliably and must be consulted carefully during a test.

Conclusion: The testing procedure becomes exceedingly taxing, as each of the (hundreds?) of individual outcomes must be checked for their specific success formulae and the respective data collected and calculated for the test artefact. It goes without saying that the manual effort required is vast and altogether unsuitable for current manual testing approaches whereby the auditor can report only on criteria that fail, and by this means invest time and effort on explaining how and why the criterial fail and how the service owner can best amend the issues in order to improve the service accessibility.

**Now, however, the focus in WCAG 3.0 shifts from testing that serves the customer and the web service users to a testing that mechanically, yet unproductively, reports on passed atomic outcomes, with a lot of room for calculation and data assembly errors yet with little benefit for the service owner or its users.**

Today, a manual WCAG test report of a web service can comprise 20 to 30 pages. If the report expands manifold owing to the overhead discussed above, one will have to re-think the entire testing procedure to somehow accommodate for all the mandatory data.

This will likely preclude many manual auditors from offering their services in the future, as, on one hand, they cannot simply absorb this level of extra work without raising their prices substantially, and, on the other, the customers who here in Europe are mainly composed of the public sector cannot, in their turn, pay more for these services, which are already pricey due to the degree of manual work required. It is likely that the industry will thus shift towards increasingly automated and hence poorer quality testing procedures that conform to WCAG in name only. One gathers that is not a desirable outcome.

**This, one notes, is in stark contrast to the intent of WCAG 3.0, which is to augment a more qualitative and relevant testing approach. It will now likely do the opposite.**

### Conformance testing is difficult to carry out reliably and in a reproducible manner

Gregg Vanderheiden expounded in his review on the serious problems that the current draft imposes on conformance testing and test reproducibility. We agree with Vanderheiden’s observation: **Qualitative scoring cannot be used for conformance testing**; a test must comprise criteria that either fail or pass if said criteria are to be employed in any legislative manner. Otherwise, two testers may produce altogether different results, which is unacceptable for WCAG in its present role in European legilastion.

**There is also the problem of scope**, which in the current draft is left poorly delineated. Who determines what the processes that should be subjected to testing are? How can one ensure that two similar services or apps or documents when tested by different providers account for similar scope and type of processes or use cases?

Given the taxing nature of the WCAG 3.0 testing effort, as discussed above, one may be tempted to opt for simple processes with less complex user interface components or aspects, and/or limit the number of processes to a minimum – what ever the smallest acceptable number might be – in order to make the testing effort at least somewhat financially feasible. This, then, would further compound the issue of expected lower quality of testing, which we presaged above.

### WCAG 3.0 success levels are impractical

**We expect there to be little interest in WCAG conformance beyond the minimum level required by the law**. This, we must say, has been and remains the case with WCAG 2.x; and given the sheer effort of passing the qualitative and user-testing-based silver ang gold levels in 3.0, striving for a greater WCAG level amounts to a major undertaking in which only very few organizations are in the real world interested. We therefore suggest more emphasis is put on simplifying and clarifying the minimum WCAG level that in our experience caters to 99% of WCAG applications (as has now been successfully done in WCAG 2.x with the A and AA nomenclature).

**Moreover, the terminology of bronze, silver and gold levels fits poorly for the intended use of WCAG.** The terms are not professional and are unheard of in conformance specifications. That is, in the European cultural and linguistical context the terminology may even threaten to undermine WCAG 3.0 credibility amongst adopters. The draft should revert back to employing the WCAG 2.x A, AA and AAA level in its terminology for conformance.