Option 2

POC: Gregg

OVERVIEW NOTES

- Provisions are a combination of requirements and recommendations
    - This can include requirement and recommendation items within the same block
        - This brings the recommendation into sharper focus
    - Many can also include adjectival ratings above and below PASS
- Only PASS items aspects are required
- Assertions can be made against Adjectival ratings
- Conformance at bronze level does not involve assertions but they are part of silver and gold

Some things are not present in this Option or have been replaced with other things These include:

- Views (replaced by “pages, documents, and software)
- Being ok with only completing the parts “essential” to the task.
- Adaptive requirements (replaced by “conditional” and other)
- Extensible requirements
- Protocol-Based Testing
- Testing the Quality of and assertion that a protocol was followed
- User Testing

The rationale for each of these changes or omissions is included at the end of this document under Rationale for modification or omission of some approaches or terms” [OR JUST CLICK ON THE LINK IN THE LIST ABOVE]

How this Option stacks up against SURVEY criteria

(a quick guide to the detailed information in the full description below)

- **How does the approach expand the range of disabilities supported?**
    - (1) through the incorporation of new types of conformance (assertions) at the silver and gold levels -
      (2) through the inclusion recommendations mixed in with requirements -
      (3) by giving credit for working on recommendations (even non-testable ones)
- **How can the model be updated quickly and easily to support emerging technologies and interactions?**
    - it is technology agnostic and does not mix HOW TO in with WHAT MUST BE TRUE. so alternate technologies can be used for authoring and alternate accessibility strategies can be used as new technologies allow them.

- **How Is guidance expressed in generic terms so that they may apply to more than one platform or technology?**
    - there are no technology specific items used in this approach
- **How does the guidance motivate organizations to go beyond minimal accessibility requirements?**
    - (1) through the incorporation of new types of conformance (assertions) at the silver and gold levels -
      (2) through the inclusion recommendations mixed in with requirements -
      (3) by giving credit for working on recommendations (even non-testable ones)
- **How does the guidance motivate more organizations to work towards the minimum accessibility requirements?**
    - this option gives credit for moving TOWARD conformance - not just for meeting and exceeding. In fact it gives 2 levels of recognition
- **How does it provide guidance for people and organizations that produce digital assets and technology of varying size and complexity?**
    - Hard to know exactly what this means - but it does provide guidance in a way that is useable by large and small and single authors
- **Validity - How does the score at the outcome level reflect the accessibility of the product?**
    - While no level of conformance makes a product accessible to all people with disabilities - this option has clear objective criteria so you can't pass and not meet minimum standard. Can't game it.
- **Reliability - How is the score at the outcome level reproducible and consistent?**
    - There is no SCORE per se - but outcome level evaluations are all objective with high Inter-rater reliability
- **Complexity - How does the approach take a reasonable time to test? and consistent?**
    - Again tests are objective and therefore consistent. And they do not require complex judgment
    - That said -- there is another level that is looser and requires more judgment but is not objective so easier to rate
- **How does the proposal support internationalization?**
    - Terms are objective and technical so are more easily translated than adjectival - for conformance. Adjectival is allowed above and below conformance but that is for reinforcement so technical accuracy less critical
- **How Is the proposed approach and conformance model clear and easy to understand?**
    - this option has fewer types of tests and they are all clear and distinct from each other.
- **How does the proposal promote equity across all disability categories?**
    - this option is much better that WCAG2 in allowing incorporation of important but non-testible recommendations - that authors can get credit for if they follow
- **How is it realistic?**

- It is in many ways simpler than WCAG though it provides more flexibility for getting credit for going outside of the minimum testable criteria

How this Option stacks up against the original evaluation criteria

Does it use other ways of measuring and/or evaluating where appropriate so that more needs of people with disabilities may be included?

Option 2 offers both outcome and assertion methods to allow testability for broader coverage. It also includes recommendations mixed in with requirements to get things in front of designers eyes even when not “testable”. Some of these are stand alone but this model also allows the recommendations to be included as higher levels over Pass on a requirement (that is, where as the PASS level of a requirement must be testable - the adjectival levels above pass do not necessarily need to be testable)-- making sure that any filters that don't include recommendations would still include them.

Does it have a maintenance and extensibility model for guidelines that can better meet the needs of people with disabilities using emerging technologies and interactions?

- This model does not have an “extensible” type - for the reasons cited below. But all items are designed to be extensible naturally (and without having to revise the standard) by being written in technology agnostic form -- and being based on outcome to the user in human terms. As new technologies are introduced, if they meet the guidelines - it will work for people with different disabilities

Is guidance expressed in generic terms so that they may apply to more than one platform or technology?

- That is a key feature - requirement - for items under this model

Does the guidance motivate organizations to go beyond minimal accessibility requirements?

- In three ways
    a. All provisions have adjectival (v poor, poor, pass, good, v good) -- (names can change) encouraging people who fall short to make progress toward pass - and rewarding those who do more than just pass
    b. It provides recommendations in addition to requirements to show how to go beyond
    c. It allows ‘non-testable’ language to be used at higher levels of testable criteria (experimentally to see how this works)

d. It provides, silver and gold recognition for those that go beyond minimum requirements

Does it provide guidance for people and organizations that produce digital assets and technology of varying size and complexity?

- This is orthogonal to the requirements and recommendations --

Validity - does the score at the outcome level reflect the accessibility of the product?

- there are two levels. One that can be used in regulations and one that recognized extra effort at the adjectival level for those that want to go beyond.

Reliability - is the score at the outcome level reproducible and consistent?

- outcome level requirements are testable with high inter-rater reliability.
- In addition - the assertion level also has high inter-rater reliability and lets people go beyond outcome requirements

Complexity - does the approach take a reasonable time to test?

- This approach is simpler to understand and to use. Testing includes both automatic (fast) and human-assisted (faster than pure human) and human testing depending on the item.
- The focus is on automated as much as possible with human as needed and to allow recognition of effort beyond “pass”

Does the proposal support internationalization?

- Terms are objective and technical so are more easily translated than adjectival - for conformance. Adjectival is allowed above and below conformance but that is for reinforcement so technical accuracy less critical

Does the proposal increase the testability?

- This approach assures / maintains or improves testability and acceptability to regulatory agencies - and also **expands** what is testable through including both **adjectival** and **assertions** (of protocols etc.)

Is the proposed conformance model easy to understand?

- It is much simpler

    - There are **requirements** - that you must do to conform
        - Some of these are **outcomes** (like WCAG 2.0)
        - Some are **assertions** (that one has done some process or protocol)

- **If you do less - or more** there are **adjectival** scores that you can use to recognize your efforts **toward** or **beyond** just passing

- There are recommendations - of things to do beyond basic requirements
  - All are included with the requirements
  - Some are included as part of requirements items (as adjectival) (this is experimental to see if it works

Does the proposal promote equity across all disability categories?

- this option is much better that WCAG2 in allowing incorporation of important but non-testible recommendations - that authors can get credit for if they follow

4. Testing

WCAG 3.0 consists of testable (requirements) and non-testable (recommendation) components.

The testable components include both ***outcomes*** and ***assertions***.

> **Outcomes** are written as testable criteria that allow testers to determine if the content they are evaluating satisfies the criteria.
>
> **Assertions** are only testable in that one can test that the **assertion** has been made - not that the **assertion** is true. As a result **assertions** are treated separately from outcomes

Testing outcomes use items, pages, user processes, and the aggregate to define what is being tested.

Items are the smallest testable unit. They may be interactive components such as a drop down menu, a link, or a media player. They may also be units of content such as a word, a phrase, an icon, or an image.

Pages include all content visually and programmatically available without moving to a different **url.**. Conceptually, views correspond to the definition of a web page as used in WCAG 2.X. Outside of web-content the concept can be used to refer to a document or to a software program

NOTE: Several attempts to create something smaller than a web-app, mobile-app or software program but all of them fall apart on carful examination. In all of these “app” worlds there are an infinite number of views as things on a “page” (in and app or application) expand, as toolbars appear and disappear, as software moves through a process, as people move around and invoke different tools in a workspace or move through a virtual building or space. Only the simplest form of app that is “page-like” could be evaluated using views - and then only because each “page” is considered a stable view. Some SC will apply to views but many do not. And

those that apply to views most all work if you apply them to the application as a whole. Also, evaluation of an app with an indeterminate number of views is not practical or reproducible..

User processes are a series of user actions, and the distinct interactive pages and items that support the actions, where each action is required in order to complete an activity. A user process may include a subset of items in a page or a group of pages.

Examples of a process include:

- Logging into a site and being recognized as an authenticated user;
- Ordering an item, in which case the process includes the entire set of tasks from searching for the item, viewing all information about the item, viewing alternative items that might be offered, viewing any special deals that might be offered, adding it to the shopping cart, paying for it, and receiving confirmation;
- Submitting tax information, from start to end of the process; and
- Interacting with other users in a virtual reality environment.

A process is comprised of one or more pages or subsets of pages. Only the parts of the pages that support the broad user process are included in a test of user process.

Since an item might be inaccessible to some user (e.g. a movie to a deaf-blind user or a complex graph to a blind user) but covered by another accessible item (e.g. a movie script or long description), evaluations for conformance are only done at the page or process level. The aggregate is the combination of, pages and user processes that collectively comprise the site, set of web pages, web app, etc.

4.1 Types of tests

All WCAG 3.0 ***outcome*** and ***assertion*** tests are binary (true/false)

However - the true/false criteria only provides an evaluation of whether the subject content meets the minimum criteria.

Therefore other evaluative measures are proposed that

- allow there to be higher levels above ‘passing’ to acknowledge better performance/accessibility, and
- (to reward progress toward passing - for things that fall short -- mechanisms for showing progress toward meeting the criterial. For example - scores of **very poor** and **poor** below passing **(or two other adjectives that indicate progress but not sufficiency)** so that progress can be seen toward passing

.

Testing outcomes might involve a combination of automated evaluation, semi-automated evaluation, and human evaluation.

Although content may satisfy all outcomes the content almost assuredly will not be usable by all people given the wide variety, degrees and combinations of disabilities.

The three mechanisms are provided to address this gap by providing a range of ways to evaluate and impact the user experience.

4.1.1 Computational tests

Computational tests rely on measuring properties of the content based on nominal values. The test results are objectively verifiable, to avoid variation of test results between different testers. Values are quantitative, and could be boolean, for example to check the presence of titles, text alternatives, and accessible names. Other values could include numerical thresholds; for example, to check color luminosity ratios.

Each method using computational tests includes:

- the nominal values being tested; and
- an algorithm to measure the properties of the content based on the nominal values.

NOTE: Setting a threshold de facto turns a numeric test into a binary (true/false) test. However, for those criteria where there is a numeric threshold, there is both an opportunity to set higher thresholds for more accessible content - and an opportunity to use progress toward the threshold as an indicator of progress for content that is working toward conformance.

4.1.2 Conditional tests

These are the same as computational tests except that they include conditions under which the outcome must be true (or meet threshold). These take the form of

- If xxx then yyyy (is true) OR
- Where xxxx, yyyyy (is true)

These allow requirements to either (1) be constrained either to just those places where they are important, or (2) avoid having them apply in places where they are not possible or reasonable for some reason.

Example

- If content flashes more than 3 times within any one second period, then the area and nature of the flash is less than the general flash threshold and the red flash threshold. OR
- Where content flashes more than 3 times within any one second period, the area and nature of the flash are less than the general flash threshold and the red flash threshold.

4.1.3 Qualitative tests

Qualitative tests rely on evaluating content based on a set of defined expectations and exceptions. The set of expectations and exceptions limit the scope of decisions, to minimize variation of test results arrived at by different testers. Still, some level of qualitative assessment is required, therefore the accuracy of the test results also depends on the knowledge and context of the testers to some degree.

Each method using qualitative tests includes:

- the defined expectations being tested; and
- guidance on evaluating how well the content meets the defined expectations.

Quatlitative tests are not objective and cannot be use for requirements.

They do have a very important role for both **recommendations** and may also be usable for adjectival scoring either to show progress **toward** conformance or to talk about achievement beyond conformance. (They may also be useful in protocols - see Assertion-based testing)

4.2 Types of requirements and Recommendations

There are two types of requirements that can be tested: prescriptive-outcome, and assertion-based.

- **Prescriptive outcome requirements** include a predetermined set of options for determining how to conform, and how to measure conformance. These requirements are universal, meaning that how to meet them does not vary based on context.

- **Assertion-based requirements** are not based on **outcome** but rather on **assertions** of some sort that a protocol or process has been practiced in the development of the content/product.

4.2.1 Prescriptive-outcome requirements

Prescriptive-outcome requirements include a predetermined set of outcomes or options for determining how to conform, and how to measure conformance. These requirements are universal, meaning that how to meet them does not vary based on context or technology.

NOTE: these requirements should not be technology specific. That would be the role of techniques and they should not be normative - nor in the standard itself (or the standard will quickly become dated as other techologies are introduced or other techniques are discovered or developed to solve the problem.

Prescriptive technical requirements are to be used **only** where interoperability is required. In WCAG we don't have any at the present time. They more often occur in mechanical and electrical connections - or data standards like TCP/IP etc.

All other requirements should be **outcome** based so as to allow

- Authors to introduce new techniques that are better at doing something (e.g. ARIA was introduced to allow people to meet WCAG requirements in better ways
- New technologies can be introduced that require different approaches to meeting a requirement (e.g. non-markup language technologies could never have met WCAG if it had prescribed markup related requirements (even though that would have made WCAG much easier to understand and use with HTML content.

4.32.2 Assertion-based Requirements

- These requirements are not based on **outcome** but rather on **assertions** of some sort that a protocol or process has been practiced in the development of the content/product.
    - For Example, an **assertion** that a ***process*** has been carried out, or that ***user involvement*** was ensured or ***user testing*** has been carried out etc. These do not guarantee any *outcome*, but they can be tested simply by determining whether or not the author has made the ***assertion***. This is a simple true false test.
    - Since it is not outcome based it wouldn't fall in the same category as outcome based requirements, since there is no guarantee that a process will in fact improve the outcome at all (although in most cases it would improve it, and may even result in better outcomes than the outcome related provisions alone.) There is no way to independently check whether or not the process was carried out (other than that an **assertion** was made) and it is very easily gamed, and has been games in the past (i.e. a company carries out a process but it has no impact on the actual outcome or accessibility of the product/website).

The effectiveness of **Assertion** based requirements can be improved by requiring that the **assertion** include documentation on which protocol was used, why, and what the outcomes were. ach include documentation on which protocol was used, why, and what the outcomes were.

Assertion based requirements can be use for both ***protocols*** and for any other activity that might impact outcomes but is not itself an outcome.

Industry accepted protocols are used to evaluate whether a process was used that could improve accessibility. Examples include evaluation through user-centered design methods, plain language testing, and both user and expert usability testing. Since external evaluators are not allowed inside an organization to independently confirm whether any of these things were actually done (for obvious product secrecy and other reasons) **Assertion**-based testing involves an organization asserting that they have done something and the external evaluator simply

confirming that the **assertion** has been made. **Assertions** often apply to the aggregate or user process. Methods would provide guidance for different procedures that are proven to improve accessibility and would test the declarations or **assertions** about the protocols used. Requiring that documentation on the use of the process and any outcomes would increase likelihood of impact on outcomes. However in order to be objectively testable (e.g. the author can determine before product/website release that they have passed) these **assertion** based requirements would be limited to strict pass/fail (did they organization assert or not assert).. All other impact on outcomes would be measured by outcome oriented testing.

4.2.3 Recommendations

Along with, and mixed right in with, the testable requirements would be recommendations. These would include things that are very beneficial for accessibility but that cannot be required for some reason-- usually because there is no (current way to) test their outcome objectively.

- Some examples
    - ***Write as clearly and simply as can - avoiding jargon (that is not accompanied by a plain language explanation) and erudite terms (learned, scholarly terms with emphasis on knowledge gained from books or advance education) when simpler terms could be used.***
    - (for more and better examples see Cognitive Accessibility Guidance and Making Content Usable for People with Cognitive and Learning Disabilities
- These ***recommendations*** do not have a pass fail but could have adjectival scores (very poor, poor, adequate, good very good) with descriptions of what each means.

These recommendations are mixed in with requirements to get things in front of designers eyes even when not “testable”.

- Some of these are stand alone recommendations
- But this model also allows the recommendations to be included as higher levels over Pass on a requirement (that is, where as the PASS level of a requirement must be testable - the adjectival levels above pass do not necessarily need to be testable)-- making sure that any filters that don't include recommendations would still include them.

5. Scoring

Scoring has no role in testing of pass fail of an outcome oriented requirement. (if it did - it quickly becomes subjective and cannot be used in regulation).

However, there is a strong role for it in

- Getting credit for going beyond minimum requirements
- Noting progress toward conformance for sites/products that do not meet the minimum requirements.

Each outcome should have methods associated with different technologies. These would be documented in a separate informative document (or appendix) to the standard -- so that these can be updated regularly as technologies evolve (authoring and testing technologies) and new methods added as new technologies are released into the field.
Each method contains tests and techniques for meeting that outcome.

5.1 Scoring tests

The basic test for each requirement is a simple true false test. Did it pass or not.

In addition to the pass/fail used to determine conformance however it would be useful for there to be a scale and adjectives added both for content below and above passing. This can be use both to reward those who go beyond the minimum and acknowledge and encourage progress by those who fall below the minimum requirements.

For example

-
- [to provide minimum required readability for those with low vision and color vision differences]
- REQUIREMENT: The contrast for blocks of text is greater than 4.5 for standard print and 3 for large print
    - Adjectival measures
        - 1 and .5 to 2 and 1 are very poor
        - 2 and 1 to 3 and 2 are poor
        - 3 and 2 to 4.5 and 3 are Pass
        - 4.5 and 3 to 5.5 and 4.5 are Good
        - 5.5 and 4.5 to 7 and 5 are Very good

5.2 Scoring

Scoring could occur in two forms

1. Rank order scoring
    a. Very poor, poor, pass good and very good are each given a numeric value
    b. Scoring is done by adding them up
    c. This is highly discouraged - since it is very prone to gaming and very-good can be used to offset very poors and give the impression of passing
2. Adjectival reporting
    a. Better method is to report by category.

b. This site has
    i. 4 % very poor
    ii. 8 % poor
    iii. 60 % pass
    iv. 25% Good
    v. 3% Very good
c. This gives a clear indication of a site that has some serious and moderate problems that need to be fixed. So it clearly does not conform and that needs to be resolved. However it also indicates that they have gone beyond what is required in some areas.

3. Adjectival reporting by category
    a. This adjectival reporting could be even more effective if it were reported in different dimensions.
        i. By function -
            1. Perceivable, Operable, Understandable, Robust
            2. This works if the items can be UNIQUELY tagged with these for characteristics
        ii. By disability
            1. This would be very problematic since a majority of the provisions cannot be UNIQUELY tagge with one disability or functional impairment (many apply to many disabilities and some to all). Even if you skip disabilities and try functional impairments - you encounter the same problem
        iii. By something else very tangible

6. Conformance

6.1.1 Conformance levels

WCAG 3.0 defines three levels of conformance: bronze, silver, and gold.

6.1.1.1 Bronze

- Bronze is the **minimum conformance level**.
- To meet WCAG at Bronze level content must meet all of the **Outcome Requirements** at PASS level.
- (Content that does not meet the **outcome Requirements** of the bronze level does not conform to WCAG 3.0 at any level.)
- NOTE

NOTES:

- While there is a lot of overlap between WCAG 2.X and WCAG 3.0, WCAG 3 includes additional measures and different scoring mechanics. As a result, some level of conformance to WCAG 3.0 may or may not be backwards compatible with WCAG 2.X. If there is one - bronze is the most likely level though Bronze might be easier to pass than level AA of WCAG 2.x and fall between Level A and Level AA

6.1.1.2 Silver

- Silver is a **higher conformance level.**
- To achieve Silver you must first pass Bronze
- In addition you must pass the adjectival level of **Good** for at least 50% of BOTH the Requirements (that are not binary*) AND the Recommendations
- *Binary means there are no adjectival scores above pass).

6.1.1.3 Gold

- Gold is the **highest conformance level**.
- To achieve Gold you must first pass Silver
- In addition you must
    - pass 75% of the provisions that are not binary at adjectival level of **Good** and 50% at **Very good**).
    - And
    - Must provide **assertions** that 50% or more of the protocols for improving accessibility were followed in creating the content/product
        - (here requiring documentation and qualitative discussion of impact on outcome could be helpful but also problematic.
            - Some companies might be barred by their lawyers from making **assertions** but even more would be barred from doing reports
            - Companies with really good accessibility could find it harder to show how protocols had impact on their product designs (since they were already making accessible products -- than would companies whose products are terrible at accessibility.

Rationale for modification or omission of some approaches or terms

Views

When working on WCAG2ICT a lot of effort was put into trying to make Views work. Two leading proponents of this were Peter Korn and Gregg Vanderheiden (and some others)

but especially Peter. This was felt to be much better than just treating software as a whole - and there was not such thing as pages.

All of the attempts to create something smaller than a web-app, mobile-app or software program fell apart on carful examination. In all of these “app” worlds there are an infinite number of views as things on a “page” (in and app or application) expand, as toolbars appear and disappear, as software moves through a process, as people move around and invoke different tools in a workspace or move through a virtual building or space. Only the simplest form of app that is “page-like” could be evaluated using views - and then only because each “page” is considered a stable view. Some SC will apply to views but many do not. And those that apply to views most all work if you apply them to the application as a whole. Also, evaluation of an app with an indeterminate number of views is not practical or reproducible.

And games only made the problem more evident.

Being ok with only completing the parts “essential” to the task.

This approach often leads to missing things that are not thought to be important but are very important to some users. (And important enough to the author to include them)

For example - when ordering - a list often looks like a process that includes

- searching for the item,
- viewing all information about the item,
- adding it to the shopping cart,
- paying for it, and
- receiving confirmation

(In fact this is the type of list I used to cite - until I got push back from people with disabilities that there wer other things on the page that were equally as important or more.

For actual shoppers however two additional items may be as critical to them

- viewing alternative items that might be offered,
- viewing any special deals that might be offered
- viewing the ads on the page

In fact some shoppers search for a key item only to use it to find all the related items where what they are looking for can be found. And users have also reported that the ads on pages were their primary source of information on new things -- and also offers were key to their being able to make their precious funds go further.

Adaptive requirements

I did not include Adaptive requirements - that were described as

- **“Adaptive requirements** includes an additional checkpoint to determine user-specific context that may require varying which computational and qualitative tests will be used, and potential variations in the desired outcome.This means that the same computational and qualitative tests may need to be applied multiple times in different ways, and some variations may pass or fail while others have a different result.”

I don't understand how there can be any user-specific contexts unless W3C is going to specify in advance what those contexts are and identify the users so that testing on those users- in that context can be done before produce or website release. Otherwise it cannt be a requirement that is enforced.

Extensible requirements

I did not include **Extensible requirements** - that were described as

- **Extensible requirements** allow for computational and qualitative tests to be applied using an appropriately selected specification out of many when multiple specifications or standards are available for the same test (for example, luminance). The testers are responsible for documenting which specification is applied and why.
- Testing may occasionally require determining and referencing which specifications are being tested against. These variances can be due to regional laws or different means for measuring. For example, luminance may be measured based on APCA or IEC-4WD. In these cases, an extensible method of testing is being used, in which the selected specification used for evaluation is documented, along with why that specification was chosen.

I don't see how these are really extensible requirements. I only see things being converted or conditioned (if this then that)

I would call those with conditions **conditional requirements** (and I added that type above to cover this). For example - if you are using technologies that use a different colorspace - then you use the requirements for that colorspace.

If you are just using a different unit of measure (e.g. mm vs inches) you simply convert your units into the other units.

You are extending the requirement (making it longer in any way) - you are just either converting it - or if conversion is not possible - you are applying the measure appropriate for that technology - and it is not an extension but an application appropriate for that technology.

If you arent doing one or the other - then what you are doing is not pre-defined in the standard and there can't be uniformity of judgment - and I cannot see how it could be required and enforced.

Protocol-Based Testing

This is a very valuable concept but not well titled. You can't actually test the protocol unless you get inside the company and view their processes. And you can't asses it for a product or web page or document -- unless you were there when that product or web page or document was developed to determine if the protocol was followed for the product. (and it is against products, pages or documents that you assign pass fail - not the company).

Companies do not allow people inside while new things are being produced and after the fact is too late.

What we CAN test is the **assertion** or **claim** they make that prototocl or process was followed. We can easily see if they made the ‘assertion”. (and let ISO 9000 etc. help ensure that they don't make false assertions.

So I renamed this “Assertion-Based Testing” since that is in fact all that we can test.
But I feel this is potentially a valuable or even critical arrow in our quiver going forward.

Testing the Quality of and assertion that a protocol was followed

- The quality of the implementation and outcome can also be assessed to ensure adhering to the protocol created a more accessible outcome.

This is essentially **outcome** testing - and that is covered elsewhere. The thing about protocols is that it only tests the process - not the outcome. If one is going to try to test the outcome without using outcome measures - (e.g. with user testing) then it is no

longer objective. You can get very wildly differing results depending on the users chosen.

User Testing

User testing is wonderful - and should be one of the protocols that people attest to using. It is fantastic for improving the development process. But it cannot be used as an outcome test. The results are highly dependent on the users chosen. So a pass/fail cannot be determined -- or even sem-reliable adjectival. We would have to specify exactly which users and their exact skill levels in order for an author/developer to determine if their product passed or failed before they released it.

Examples

[The instructions asked us to pick one or more SC and create examples of how this would work]

I think this is very dangerous - since providing one example doesn't mean it would work for all the others but …with that warning - here is an example to illustrate the gist - though without testing against all the SC I don't know if this approach will hold up without modification - or in the case of major failure - going back to the drawing board.

Here is one for the contrast provision

- **REQUIREMENT: The contrast for blocks of text is greater than 4.5 for standard print and 3 for large print - and there is nothing preventing viewing with lower contrast.**
    - Adjectival measures
        - SP = 2.0 to 4.4 and LP = 1.0 to 2.9 are poor
        - SP = 4.5 to 5.4 and LP = 3.0 to 3.9 are Pass
        - SP = 5.5 to 6.9 and LP = 4.0 to 4.4 are Good
        - SP = 7.0 and up and LP = 4.5 and up are Very good
    - NOTE 1: these scores ONLY apply to this example.
    - NOTE 2: Poor, Pass, Good and Very Good are stand ins - may change
    - NOTE 3: LP = large print SP = standard print

- PASSING it (**Bronze**)
    - would <u>for this provision (but not necessarily for all provisions)</u> be the equivalent of level AA
- Getting to **Silver**
    - would mean passing 50% of provisions like this at GOOD (which <u>for this provision but not necessarily for all provisions</u> is better than Level AA but less than Level AAA

- Getting to **GOLD**
    - would mean passing 75% of them at GOOD
    - and 50% of them at VERY GOOD (currently level AAA for this provision but not necessarily for all provisions)
    - and asserting 50% of protocols were followed like “plain language review” “testing with people with disabilities” etc.