Lab Testing & Evaluations Protocol

1.1 Purpose

Our goal is to evaluate products designed and developed for people with disabilities to provide an unbiased assessment of the products as well as to support the product developers by helping them improve their products through our evaluation. This protocol is designed to establish a structured methodology and evaluation metrics for accessibility and usability of products for individuals with disabilities using expert-based assessments. The main goals of the protocol are to:

1. Identify accessibility barriers that may affect users with disabilities
2. Assess compliance with recognized accessibility standards and guidelines
3. Evaluate compatibility with assistive technology
4. Determining the accessibility readiness of the product
5. Provide recommendations for improvement of the product

These protocols will be applied to: Web applications, Mobile applications, Desktop software, Digital platforms & services, and Hardware products with digital interfaces. A special focus is on emerging AI based tools for which there is a lack of prior evidence of success and established testing protocols.

2.1 Accessibility Experts (Participants)

The evaluation will be conducted by accessibility evaluators/experts with relevant skills, including:

1. Accessibility standards and regulations
2. Assistive technologies
3. Inclusive design principles
4. Disability-related accessibility requirements

Experts will be selected based on the qualifications needed for testing and evaluating the product, and a panel of caregivers, users, experts, and people with disabilities. Accessibility standards would be followed during testing, as applicable: WCAG 2.2 Level AA (World Wide Web Consortium, 2023), Section 508 Requirements (U.S. Access Board, 2017), EN 301 549 (European Telecommunications Standards Institute et al., 2021) and Organizational accessibility policies.

3.1 Evaluation Methods

Experts will use the following methods to test and evaluate the product:

3.1.1 Primary Evaluation Method

Evaluation metrics for product assessment are intended to support consistent, transparent, and evidence-based assessment of assistive technology products to ensure bias-free objective assessment of the products. A clear set of metrics (as appropriate) will be developed and agreed on by experts and shared with product vendors prior to the testing. Table 1 shows the proposed evaluation metrics that draws on the usability construct in ISO 9241-11(International Organization for Standardization, 2018), the software quality characteristics in ISO/IEC 25010 (ISO/IEC, 2011), and assistive-technology outcome measures from the rehabilitation literature (Day & Jutai, 1996; Demers et al., 2002)

Table 1. Proposed Evaluation Metrics for Product Assessments

Evaluation Metric	Definition	Key Question	Sources
Usability Findings	Measures how easy the product is to learn, use, and integrate into daily activities across different users and abilities.	*Can users effectively and comfortably use the product in everyday life?*	ISO 9241-11:2018
Independence Outcomes	Assesses whether the product increases autonomy, reduces reliance on assistance, and supports independent completion of tasks.	*Does the product increase meaningful independence?*	Demers et al., 2002; Day & Jutai, 1996
Environmental Performance	Evaluates how effectively the product functions across real-world settings including home, workplace, healthcare, and community environments.	*Does the product perform consistently across environments?*	ISO/IEC 25010:2011
Durability Observations	Examines product reliability, resilience, and consistency during repeated or prolonged use.	*Can the product withstand routine use over time?*	ISO/IEC 25010:2011
Connectivity Requirements	Assesses dependence on internet access, external devices, applications, or technical infrastructure needed for operation.	*What level of connectivity or setup is required for use?*	ETSI et al., 2021 (EN 301 549, Cl. 4)
Accessibility Considerations	Evaluates how well the product accommodates diverse abilities, needs, and potential barriers to use.	*Who may encounter barriers when using this product?*	W3C, 2023 (WCAG 2.2)
Best-Fit User Profiles	Identifies users, environments, or situations where the product may provide the greatest benefit.	*Who is this product most suitable for?*	Demers et al., 2002
Identified Limitations	Documents observed constraints, challenges, or conditions affecting performance.	*Under what circumstances might the product not meet user needs?*	ISO/IEC 25010:2011

Figure 1 shows the structure of the assessments we are using for product evaluation. Product evaluation proceeds through three parallel evaluation track that converges on a single accessibility-readiness. The first track universal accessibility applies the POUR principles (Perceivable, Operable, Understandable, and Robust) (World Wide Web Consortium, 2023). The second track Disability specific accessibility evaluates the product separately against cognitive, physical, sensory, communication and psychosocial needs, reflecting the disability specific metric in detail in section 3.1.3. The third track that is assistive technology testing, verifies compatibility with specific tools that users rely on such as screen readers, keyboard only navigation, switch access, voice

control and magnification software. Scoring and findings from all three tracks feed into an overall accessibility readiness score.

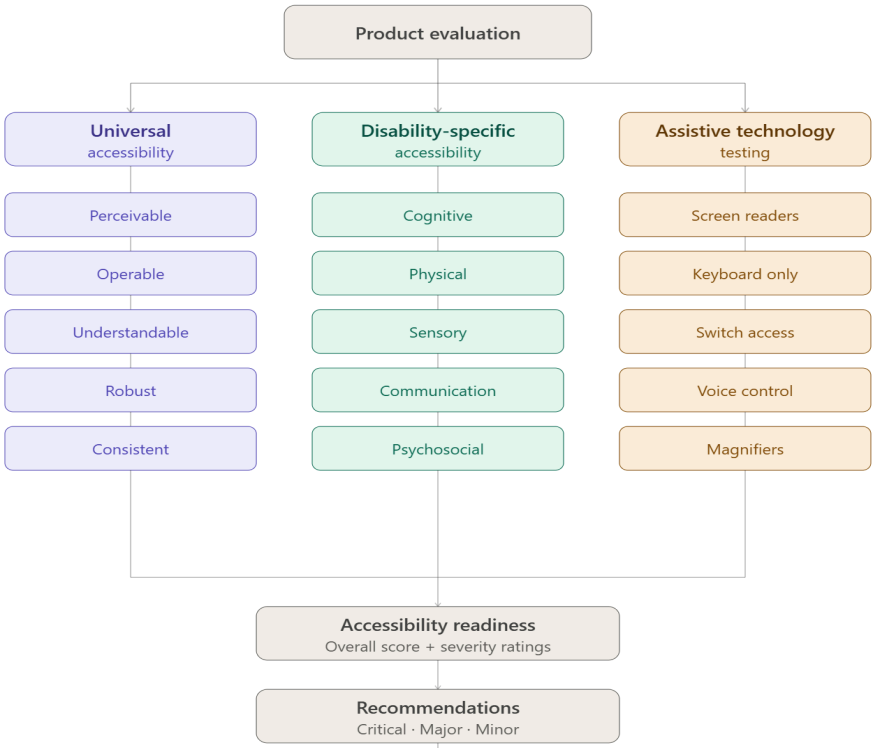

Figure 1. Structured Metrics for Product Evaluation

3.1.2 Expert heuristic review

The expert will use the product testing to establish accessibility heuristics and accessibility standards. Accessibility heuristics are expert based evaluation principles (Nielsen, 1994), they provide a setructured set of criteria that guide evaluators in assessing whether a product can be effectively used by people with diverse abilities. These are derived from established accessibility standards and human-computer interaction research, particularly the Web Content Accessibility Guidelines (WCAG) developed by the World Wide Web Consortium (W3C), universal design principles, and accessibility evaluation frameworks used in assistive technology and inclusive design research. In this study, accessibility heuristics are based on the four founding principles of the Web Content Accessibility Guidelines - Perceivable, Operable, Understandable, and Robust ("POUR"; World Wide Web Consortium, 2023). There are certain examples that come from established standards, such as, verifying that images contain alternative text, ensuring sufficient color contrast, confirming that all functions are keyboard accessible, checking that instructions and error messages are easy to understand, and verifying compatibility with assistive technologies such as screen readers. Below are the metrics:

Table 2. Accessibility Evaluation Metrics (Universal Accessibility)

Metric	Evaluation Question
Perceivability	Can users access and perceive all information through at least one sensory channel (e.g., text alternatives, captions, sufficient contrast)?

Metric	Evaluation Question
Operability	Can users effectively navigate and interact with the product using different input methods (e.g., keyboard, voice, switch devices)?
Understandability	Is the content, navigation, and interaction flow easy to understand and use?
Robustness	Does the product function reliably across assistive technologies and platforms?
Consistency & Predictability	Does the interface behave consistently and meet user expectations throughout the product?

Source: Adapted from the POUR principles, WCAG 2.2 (World Wide Web Consortium, 2023).

Each metric can be scored on a 5-point scale (1 = Poor, 5 = Excellent) to determine the product's overall accessibility performance.

3.1.3 Disability Specific Metrics

Metrics across the types of disabilities can differ, i.e., cognitive, sensory, physical, and communication. We will have distinct set of metrics across each of these categories as shown in tables 3-6.

Evaluation shown in Table 3 focuses on the internal mental effort required to complete a task. Even when a product is technically accessible, excessive cognitive demands can reduce usability for individuals with various neurodivergent conditions. This examines six complementary dimensions of cognitive accessibility: cognitive load, attention support, instruction clarity, error recovery, task sequencing, and memory demands. These metrics identify where cognitive burden occurs throughout the interaction, enabling evaluators to diagnose specific accessibility challenges rather than assigning a single overall score. Conditions include ADHD, Autism, Intellectual & Developmental Disabilities, Learning Disabilities and Executive Function Difficulties. Here, constructs are defined by cognitive load theory (Sweller, 1988; Sweller et al., 2011), the NASA-TLX workload framework (Hart & Staveland, 1988), and the W3C's cognitive accessibility guidance (W3C Cognitive and Learning Disabilities Accessibility Task Force, 2021).

Table 3. Metric for Cognitive & Neurodevelopmental Disabilities

Metric	**1 (Poor)**	**2 (Fair)**	**3 (Adequate)**	**4 (Good)**	**5 (Excellent)**
Cognitive Load	Information is overwhelming and difficult to process	High cognitive demands create frequent challenges	Moderate complexity with some support	Information is generally well-organized and manageable	Information is clear, organized, and easy to process
Attention Support	Frequent distractions interfere with task completion	Multiple distractions reduce focus and task completion	Some focus-supportive features available	Environment supports focus with few distractions	Strong support for sustained attention and task completion
Instruction Clarity	Instructions are confusing or unclear	Instructions require significant interpretation	Instructions are generally understandable	Instructions are clear with minor ambiguities	Instructions are simple, clear, and easy to follow

Error Recovery	Errors are difficult to identify or correct	Limited support for correcting mistakes	Basic error messages and recovery options provided	Clear error messages with helpful guidance	Errors are clearly identified with effective recovery guidance
Task Sequencing	Tasks are confusing and poorly structured	Task flow is difficult to follow	Tasks follow a reasonable sequence	Tasks are logically organized with clear progression	Tasks are logically organized into manageable steps
Memory Demands	Heavy reliance on memory and recall	Significant memory burden exists	Moderate memory requirements	Memory demands are reduced through via prompts and supports	Minimal memory burden via prompts and supports

Sources: Sweller (1988); Sweller et al. (2011); Hart & Staveland (1988); W3C Cognitive and Learning Disabilities Accessibility Task Force (2021).

Evaluation via Table 4 determines whether information remains accessible regardless of a user's visual or hearing abilities. It evaluates whether information is available through multiple sensory channels rather than relying on a single mode of presentation. Consistent with the Perceivable principle of the Web Content Accessibility Guidelines (WCAG), the evaluation assesses both technical accessibility and content accessibility. The metrics examine screen reader compatibility, visual accessibility, captioning and transcripts, contrast and visibility, and multimodal information access. These measures together determine whether users can effectively perceive and access information through alternative formats that accommodate diverse sensory needs.

Table 4. Metric for Sensory Disability (Visual & Hearing Disability)

Metric	**1 (Poor)**	**2 (Fair)**	**3 (Adequate)**	**4 (Good)**	**5 (Excellent)**
Screen Reader Compatibility	Content inaccessible to screen readers	Significant screen reader issues exist	Basic screen reader functionality supported	Good compatibility with common screen readers	Fully compatible and optimized for screen reader use
Visual Accessibility	Major barriers for blind or low-vision users	Limited visual accessibility support	Meets minimum accessibility requirements	Strong support for low-vision accessibility needs	Excellent support across visual disability needs
Captioning & Transcripts	No captions or transcripts available	Captions are incomplete or inaccurate	Basic captions or transcripts provided	Accurate captions and transcripts are available	Comprehensive and high-quality captioning support
Contrast & Visibility	Poor visibility and insufficient contrast	Some visual elements are difficult to distinguish	Meets minimum contrast requirements	Good visibility under most conditions	Excellent contrast and visual clarity

Multimodal Information Access	Information available through only one sensory channel	Limited alternative formats available	Some multimodal access is provided	Information is accessible through multiple formats	Fully supports multimodal access and user preferences

Sources: WCAG 2.2 Success Criteria 1.1.1 and 1.2.x (World Wide Web Consortium, 2023); Section 508, Chapter 5 (U.S. Access Board, 2017); WebAIM contrast guidance (WebAIM, 2026).

Table 5 metrics evaluates whether users can physically interact with and operate the product independently, regardless of their motor abilities. Unlike the previous evaluations, which focus on perception and cognition, this category examines the physical demands required to complete tasks. It is based on the Operable principle of WCAG, the evaluation assesses whether interactions require excessive motor precision, physical effort, or a specific input device. The metrics include alternative input support, keyboard accessibility, fine motor requirements, target size and spacing, time constraints, and fatigue reduction.

Table 5. Metric for Physical Disability

Metric	**1 (Poor)**	**2 (Fair)**	**3 (Adequate)**	**4 (Good)**	**5 (Excellent)**
Alternative Input Support	No support for alternative input methods	Very limited support for assistive inputs	Supports some assistive input methods	Supports most common assistive input technologies	Fully compatible with diverse assistive input methods
Keyboard Accessibility	Major functions inaccessible via keyboard	Many interactions require mouse input	Most functions accessible via keyboard	Nearly all functions accessible via keyboard	Complete keyboard accessibility with logical navigation
Fine Motor Requirements	Requires precise motor control to operate	High motor precision required for many tasks	Moderate motor demands	Minimal precision required	Easily usable with limited dexterity or motor control
Target Size & Spacing	Controls are too small and difficult to activate	Many controls present accessibility challenges	Controls meet minimum usability expectations	Controls are comfortable for most users	Controls are large, well-spaced, and highly accessible
Time Constraints	Strict timing requirements create barriers	Time limits create significant usability issues	Some flexibility is provided	Most timing requirements can be adjusted	Users have complete control over timing requirements
Fatigue Reduction	Use causes significant physical fatigue	Extended use causes discomfort	Moderate physical effort required	Product minimizes unnecessary physical effort	Designed to minimize physical effort and strain

Sources: WCAG 2.2 Success Criteria 2.1.1–2.1.4 and 2.5.8 (World Wide Web Consortium, 2023); ISO 9241-171 (ISO, 2008).

Table 6 centers on whether users with diverse communication needs can interact with the product independently and effectively. This focuses on expressive communication rather than information access or physical interaction. It evaluates whether the product supports alternative methods of communication instead of assuming spoken or typed communication as the only means of interaction. The metrics assess compatibility with augmentative and alternative communication (AAC) technologies, support alternative communication methods, communication independence, and response time flexibility. Therefore, these measures determine whether the product accommodates diverse communication preferences while enabling users to communicate effectively and at their own pace.

Table 6. Metric for Communication Disability

Metric	1 (Poor)	2 (Fair)	3 (Adequate)	4 (Good)	5 (Excellent)
AAC Compatibility	Incompatible with AAC technologies	Significant barriers remain for AAC users	Basic AAC compatibility exists	Good support for AAC technologies	Fully compatible with AAC tools and workflows
Alternative Communication Methods	Speech is required for communication	Very limited alternatives available	Some alternative communication options exist	Multiple communication methods are supported	Extensive communication flexibility is provided
Communication Independence	Users require assistance to communicate	Significant support is required	Moderate independence is possible	Most users can communicate independently	Fully supports independent communication
Response Time Flexibility	Strict response expectations create barriers	Limited response flexibility is available	Moderate flexibility is provided	Users can respond comfortably at their own pace	Complete flexibility supports diverse communication needs

Sources: ISO 9999 assistive products classification (ISO, 2016); American Speech-Language-Hearing Association (July 15, 2026) AAC guidance.

4.1 Automated Testing Tools

Automated accessibility testing identifies technical accessibility issues and measure compliance with established accessibility standards. Automated tools efficiently detect code-level errors and WCAG violations that may be difficult to identify manually; however, they cannot fully evaluate usability, cognitive accessibility, or user experience. Therefore, this evaluation combines multiple automated tools to assess technical accessibility across WCAG compliance, assistive technology compatibility, keyboard accessibility, code quality, document accessibility, readability, and overall accessibility coverage. Table 7 shows metrics that provide an objective baseline that complements expert heuristic evaluation and user testing.

Table 7. Automated Accessibility Testing Metrics

Metric	Tool(s) Used	Measurement Criteria
Accessibility Violations	Axe, WAVE, Lighthouse	Number and severity of WCAG violations detected.

WCAG Compliance Score	Axe, Lighthouse	Percentage of criteria satisfied (A, AA, AAA levels).
Colour Contrast Compliance	WebAIM Contrast Checker, Colour Contrast Analyser	Percentage of text and UI elements meeting WCAG contrast requirements (4.5:1 normal text, 3:1 large text).
HTML Validity	W3C Markup Validator	Number of markup errors affecting accessibility and assistive technology compatibility.
ARIA Compliance	Accessibility Insights, Axe	Correct use of ARIA roles, labels, states, and properties.
Keyboard Accessibility	Accessibility Insights, Axe	Percentage of interactive elements accessible via keyboard navigation.
Form Accessibility	Axe, WAVE	Presence of accessible labels, error messages, and instructions for form controls.
PDF Accessibility	Adobe Acrobat Accessibility Checker	Percentage of PDF elements correctly tagged and reading order preserved.
Readability Score	Hemingway Editor	Reading level and complexity score of content.
Accessibility Coverage	Combined Tools	Percentage of pages/screens passing automated accessibility checks.

5.1 Test Scenarios

5.1.1 Barrier Walkthrough

In this phase, the expert uses their knowledge to go through the product step by step, pretending to be a person with a specific disability, and documents every obstacle they hit along the way. The scenarios listed are the typical journeys a real user would take through any product, framed against the relevant WCAG 2.2 success criteria (World Wide Web Consortium, 2023):

The scenarios listed are the typical journeys a real user would take through any product:

1. User registration: Can someone with a screen reader fill out a sign-up form? Are the fields properly labeled?
2. Authentication: Can a keyboard-only user log in? Does a CAPTCHA block someone who can't see images?
3. Navigation: Can someone using only Tab and arrow keys move around the site without getting stuck?
4. Content consumption: Can a deaf user still get value from a video without sound? Is text readable at high zoom?
5. Form completion: Are error messages clear enough for someone with a cognitive disability to understand and fix their mistake?
6. Purchasing or transaction workflows: Can a user with a tremor complete a checkout without accidentally triggering the wrong button? Are time-out warnings accessible?
7. Customer support access: Can someone using a screen reader find and use the help chat or contact form?

6.1 Data Collection Forms

Data collection forms will be created and customized over time for various product categories.

References

American Speech-Language-Hearing Association. (July 15, 2026). Augmentative and alternative communication (AAC). https://www.asha.org/public/speech/disorders/aac/

Day, H., & Jutai, J. (1996). Measuring the psychosocial impact of assistive devices: The PIADS. Canadian Journal of Rehabilitation, 9(2), 159–168.

Demers, L., Weiss-Lambrou, R., & Ska, B. (2002). The Quebec User Evaluation of Satisfaction with Assistive Technology (QUEST 2.0): An overview and recent progress. Technology and Disability, 14(3), 101–105. https://doi.org/10.3233/TAD-2002-14304

European Telecommunications Standards Institute, European Committee for Standardization, & European Committee for Electrotechnical Standardization. (2021). EN 301 549 V3.2.1 — Accessibility requirements for ICT products and services. ETSI. https://www.etsi.org/deliver/etsi_en/301500_301599/301549/03.02.01_60/en_301549v030201p.pdf

European Union. (2016). Directive (EU) 2016/2102 of the European Parliament and of the Council on the accessibility of the websites and mobile applications of public sector bodies. Official Journal of the European Union.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), Human mental workload (pp. 139–183). North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9

International Organization for Standardization. (2008). ISO 9241-171:2008, Ergonomics of human-system interaction, Part 171: Guidance on software accessibility. ISO.

International Organization for Standardization. (2016). ISO 9999:2016, Assistive products for persons with disability, Classification and terminology. ISO.

International Organization for Standardization. (2018). ISO 9241-11:2018 — Ergonomics of human-system interaction, Part 11: Usability: Definitions and concepts. ISO.

International Organization for Standardization/International Electrotechnical Commission. (2011). ISO/IEC 25010:2011, Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE), System and software quality models. ISO/IEC.

International Organization for Standardization/International Electrotechnical Commission. (2012). ISO/IEC 14289-1:2014, Document management applications, Electronic document file format enhancement for accessibility (PDF/UA-1). ISO/IEC.

Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen & R. L. Mack (Eds.), Usability inspection methods (pp. 25–62). John Wiley & Sons.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Cognitive load theory. Springer. https://doi.org/10.1007/978-1-4419-8126-4

U.S. Access Board. (2017). Information and communication technology (ICT) standards and guidelines (36 C.F.R. Part 1194, Section 508 refresh). https://www.access-board.gov/ict/

WebAIM. (2026). Contrast checker. Utah State University. https://webaim.org/resources/contrastchecker/

World Wide Web Consortium. (2021). Web Content Accessibility Guidelines (WCAG) 2.2, Accessible Rich Internet Applications (WAI-ARIA) 1.2. https://www.w3.org/TR/wai-aria-1.2/

World Wide Web Consortium. (2023). Web Content Accessibility Guidelines (WCAG) 2.2 (W3C Recommendation, October 5, 2023; edited December 12, 2024). https://www.w3.org/TR/WCAG22/

World Wide Web Consortium, Cognitive and Learning Disabilities Accessibility Task Force. (2021). Making content usable for people with cognitive and learning disabilities (W3C Working Group Note). https://www.w3.org/TR/coga-usable/